Effect of phonon-assisted tunneling on the subthreshold swing of tunnel field-effect transistor

Yao Xiao,[1,*] Xiangwei Jiang,[2,†] and Lin-Wang Wang[1,‡]

[1] *State Key Laboratory of Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China*

[2] *Department of Mathematical and Physical Sciences, National Natural Science Foundation of China, Beijing 100085, China*

(Received 20 November 2023; revised 28 March 2024; accepted 15 May 2024; published 20 June 2024)

The tunnel field-effect transistor (TFET) is considered to be one of the best hopes for achieving a subthermal switch with a subthreshold swing (SS) smaller than the Boltzmann limit, $(k_B T/q) \times \ln(10)$, which is 60 mV/dec at room temperature. However, many experimental studies show that the realistic SSs of TFETs are far inferior to the idealized simulation results. To explain the discrepancy between experiments and simulations, we developed a first-principles model of multiphonon-assisted tunneling and calculated the parasitic leakage current induced by phonon-assisted tunneling. The purpose of this work is to show the importance of phonon-assisted transport in the so-called cold-source device designs. The results show that this is an unavoidable intrinsic mechanism and that phonon-assistance effects impose a fundamental limit on the TFET performance.

DOI: 10.1103/PhysRevApplied.21.064046

I. INTRODUCTION

In the post-Moore era, different nanosized microelectronic devices have been proposed. Besides further shrinking of the device dimensions, another urgent need is to reduce the power consumption of these devices. Most power consumption comes from the interconnect due to their wire capacitances, with the energy per switch proportional to the square of the supply voltage, i.e., V^2. Thus, a major goal is to reduce the supply voltage, V. However, due to the Boltzmann distribution of carrier occupation in the source electrode, V is limited by the subthreshold swing (SS), i.e., the inverse slope of the I-V curve of the traditional transistor, which has a fundamental limit of 60 mV/decade at room temperature. As a result, the traditional transistor operates within the range of 0.5–1.0 V for the supply voltage. Thus, one design is to have a "cold source," where the source electrode's electron density of states (DOS) is significantly reduced above its Fermi energy, thus there is no electron occupation, despite the Boltzmann distribution among the electronic states.

The cold-source design includes carbon nanotubes [1], a Dirac source [2], as well as broken-gap-like band structures [3–5]. The tunnel field-effect transistor (TFET) is also proposed to have a steep-slope I-V curve to overcome the 60 mV/decade limit [6]. In a sense, it is an extreme case of a cold source. In a TFET, when the channel potential is lowered by the gate voltage to a position where its conduction-band minimum (CBM) becomes lower than the source's valence-band maximum (VBM), the electron from the valence band can tunnel to the channel's conduction band via band-to-band tunneling (or Zener tunneling [7]), and thus, has current I. However, when the channel's CBM is higher than the source's VBM, there is no overlap between the source DOS and channel DOS; thus, there is no current. In this argument, the SS can be zero. In terms of the cold-source design mechanism, the DOS of the source electrode above the VBM is zero in the TFET (if we ignore its conduction band, which is high above); thus, it is a perfect case of cold-source design.

The above argument, however, has ignored the effect of electron-phonon coupling. Notably, it is the electron-phonon coupling that causes a quasiequilibrium in the source electrode, and hence, it has a Boltzmann distribution. The electron-phonon coupling can also pump electrons from the source's valence band to the channel's conduction band, taking the phonon energy before the DOS overlap, as shown in Fig. 1. Hence, one fundamental question is how large is the effect of electron-phonon coupling in the cold-source microelectronic device design? Would such an effect significantly spoil the design and make the cold-source idea an unviable design principle for overcoming the 60 mV/decade SS? However, this is a dynamic process rather than an equilibrium or quasiequilibrium

*Corresponding author: yaoxiao216@semi.ac.cn
†Corresponding author: xwjiang@semi.ac.cn
‡Corresponding author: lwwang@semi.ac.cn

 © 2024 American Physical Society

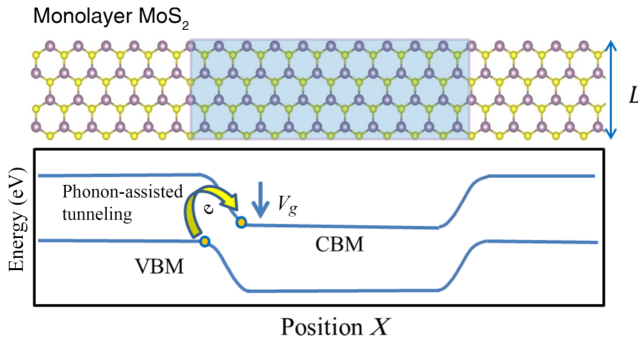

FIG. 1. Schematic view of the calculation structure of the MoS_2 TFET. Supercell length is 43.75 Å, and width L is 10.82 Å. Shaded region is the channel region, and two sides are the electrode region. Schematic diagram shows the possible phonon-assisted tunneling process.

situation. To answer the above questions, we need to actually calculate the phonon-assisted current and compare that to the zero-phonon current (which is assumed in the cold-source design).

Most current device simulations ignore the effect of electron-phonon coupling (EPC). Even if the EPC is included, it is approximated in the electron-phonon-coupling-induced self-energy term [8,9]; thus, it cannot be used to describe the transfer of the electron from one state to another. In the current study, we explicitly calculate the phonon-caused electron transfer and compare that to zero-phonon transport. First, for the zero-phonon mode-transition calculation, we provide a straightforward Fermi golden rule method, while the state-to-state coupling constant is calculated from the anticrossing gap. To cross validate the results, we also use our plane-wave nonlocal pseudopotential-based scattering-state method [10], which is different from the popular nonequilibrium Green's function (NEGF) method, to calculate the elastic quantum transport. This allows us to cross-check the methods with the same plane-wave basis. Second, for the calculation of phonon-assisted transfer, we have employed two different methods, perturbation theory and the nonadiabatic molecular dynamics method following the time-dependent Schrödinger equation. In perturbation theory, we have developed a time-integration approach based on molecular dynamics (MD), which avoids the need to compute explicitly the electron-phonon coupling constants. However, this method only considers the assistance of a single phonon. For a multiphonon effect, we use our density-based nonadiabatic molecule dynamics (NAMD) formalism [11]. This density-based formalism only needs one calculation to get all the stochastic probabilities, which makes the calculation much faster. The use of different methods allows us to cross-check the results, enhancing the overall reliability of the results.

The TFET is an ideal system to investigate the electron-phonon coupling effect since the zero-phonon electron transport and phonon-assisted electron transport can be treated in similar fashion, thus simplifying their comparison. In a realistic TFET device, the band profile along the transport direction is in a range of tens of nanometers, with the tunneling junction across several nanometers. As a result, the TFET usually suffers from a small ON-state current. In our simulation, we have used a relatively sharp potential profile to reduce the size of the device system, hence reducing the computational cost. Nevertheless, since both the zero-phonon tunneling and the phonon-assisted transition depend on the same potential profile (the spatial overlap of the VBM and CBM wave functions), the ratio between the phonon-assisted and zero-phonon transition amplitudes should be relatively independent of the potential profile. We have used the MoS_2 two-dimensional (2D) material as our TFET channel material. This system has been studied extensively for TFETs due to its cleanness for passivation and the absence of defect states [12,13].

Experimentally, the TFET is only observed with a SS value less than 60 mV/decade with current densities below 10 nA/μm [14–17]. Achieving a SS value below 60 mV/decade throughout the working voltage interval is still impossible. Researchers in previous studies believed that the nonideal characteristics of devices, such as a Shockley-Read-Hall process, gate deficiency, and trap-assisted tunneling, caused the increase of SS [18–22]. Recently, Teherani *et al.* [23] suggested that intrinsic phenomena, such as Auger, phonon assistance, and radiation generation, may limit the performance of the TFET. Teherani *et al.* [24] proposed an intrinsic link between the Auger and band-to-band tunneling rates. However, the effect of electron-phonon coupling has not been fully studied. Here, we point out that electron-phonon coupling can significantly affect the cold-source design and make the idea of a zero-phonon I-V curve impossible. In the case of MoS_2, phonon-assisted charge transfer can be as large as zero-phonon transfer, making phonon-assisted transport prominent. Our work indicates that choosing a material with a small electron-phonon coupling is an important factor to consider in the design of a cold-source device.

II. THE CALCULATION MODEL AND APPROACH

A. The calculation model

In this work, we use an ideal supercell of monolayer MoS_2 with 168 atoms to construct a TFET device model, as shown in Fig. 1. The supercell length is 43.75 Å, and width is 10.82 Å. The shaded region in Fig. 1 is the channel region and the two sides are the electrode region. All the calculations are based on the PWmat code [25,26] at the level of local-density approximation (LDA), with a plane-wave cutoff energy up to 60 Rdy. The SG15

norm-conserving pseudopotential [27] is used. We use a $1 \times 5 \times 1$ k-point grid to integrate the Brillouin zone. To relax the atomic structure, we set a residual force tolerance of 0.005 eV/Å. The calculated band-gap value is fortuitously 1.90 eV due to error cancellation (the lowering of the LDA method, and the increase due to the lack of spin-orbit coupling); this is close to the experimental value of 1.89 eV [28]. The LDA relaxed equilibrium lattice constant is $a = 3.16$ Å, which is in good agreement with the experimental value of 3.1 Å [29]. To simulate the working state of TFET devices, we add an external potential, as shown in Fig. 1. The concrete expression can be seen in the Supplemental Material [30]. Notably, the potential in the x direction has a relatively sharp bending shape within a region of about 13.0 Å widths. This is much sharper than what can be realized experimentally by a realistic gate electrode. This can lead to an overestimation of the tunneling current. But as we discussed above, this should not affect the relative ratio between the zero-phonon current and phonon-assisted currents, since they both depend on spatial overlaps between the electronic states in the left-electrode region and the electronic states in the central channel region in a similar fashion. Also note, to carry out the perturbation treatment through the MD simulation, we have used a potential profile with a potential well in the middle of the system, instead of an open-boundary-condition system. The tunneling or phonon-assisted transition is approximated by the transition from the VB eigenstates, which localize in the electrode, to the CB eigenstates in the middle well (channel) of the system. The wave functions of corresponding energy levels at the gate voltage are shown in Fig. 2. The modulation effect of the gradient electric field on the wave function can be observed. Due to the discrete nature of the finite numbers of states in the potential wells, we use an approximation to smear the eigenstate energies, as described in the next section.

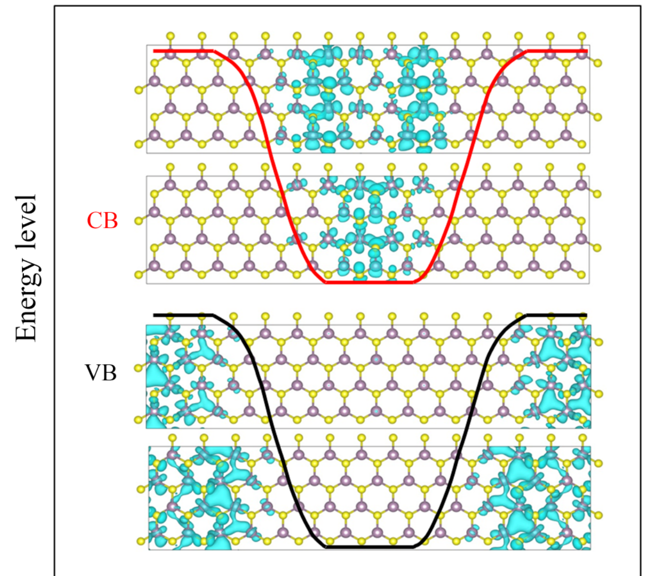

FIG. 2. Wave functions of different energy levels under gate voltage. Modulation effect of gradient electric field on wave function can be observed.

B. Theoretical model

The calculation for the zero-phonon transition rate is based on the Fermi golden rule:

$$W = \sum_{j,i} \frac{2\pi}{\hbar} |H_{ji}|^2 \delta(\varepsilon_j^c - \varepsilon_i^v - V_g), \tag{1}$$

where ε_i^v and ε_j^c are eigenenergies of the initial VB states in the electrode and final CB states in the channel, respectively. The effect of the gate voltage, V_g, is to change the channel-region potential well depth. One major issue is how to calculate the coupling constant, H_{ji}. A typical way to calculate the coupling is to use a perturbation on top of the original Hamiltonian. But here, such a perturbation for zero phonons is difficult to define. This is a tunneling coupling, representing the connection between the VB and CB states. Such coupling is like coupling in the Marcus theory [31,32] for a state transition. One approach is to change the Hamiltonian to cause the energy crossing between the i and k states, then to use the minimum avoid-crossing (anticrossing) gap of the eigenenergy to determine the coupling constant (the coupling constant is half of the anticrossing minimum gap). We have varied a small constant electric field on top of a fixed potential profile to cause an anticrossing of the VB and CB state energies, and to yield the anticrossing gap between all the confined VB states and CB states, as shown in Fig. 3.

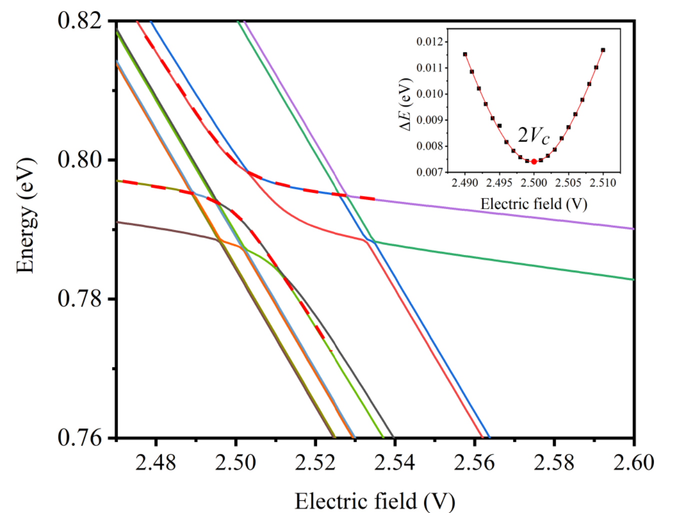

FIG. 3. Coupling process of E_{CB} and E_{VB} driven by an electric field. Inset is the fitting of the anticrossing energy gap. Here, V_c is the coupling constant.

In using Eq. (1), we have also smeared the VB energy, ε_i^v, and CB energy, ε_j^c. Considering that in an open-boundary-condition system the source electrode is infinitely long towards the left, and channel is also infinitely long towards right, both ε_i^v and ε_j^c are continuous. To represent this real situation, in the energy-conservation delta function of Eq. (1), we have smeared ε_i^v into a uniform density of state within the range of $[0.5(\varepsilon_{i-1}^v + \varepsilon_i^v), 0.5(\varepsilon_i^v + \varepsilon_{i+1}^v)]$. The same is true for ε_j^c. In the calculation, multiple perpendicular k points are also used, and tunneling happens within the same perpendicular k points. Besides, since electron transfer can come from both sides (left and right) of the potential bending and coupling, we divide the calculated W by a factor of 2. The same is true for the perturbation and NAMD results, as shown in the following.

Following Eq. (1), the 2D MoS_2 TFET drain current per unit width length is given by

$$I = qW/L, \tag{2}$$

where q is the elementary charge. L represents the linewidth of the MoS_2 TFET device.

The resulting tunneling current as a function of the gate voltage is shown in Fig. 4 as the blue line. As we can see, the tunneling $I(V)$ curve has a steep slope at $V_g = 0$ (where the source VBM energy coincides with the channel CBM energy). This is in accordance with the TFET design, where the SS is much smaller than the 60 mV/decade. The big oscillation in the $V_g > 0$ region, where there is a large overlap between the source VB states and channel CB states, is caused by the discrete nature of the confined states due to the finite potential well length in Fig. 1 (the energy-smearing approximations used above do not completely remove the discreteness feature).

The above approach provides a unique way of using the Fermi golden rule to calculate the tunneling current. We have yet to find similar way in the literature for such calculations. However, a more common way to calculate such a tunneling transition is to calculate its quantum elastic transport. This is commonly done using the NEGF method [33]. We have also developed a direct way to calculate the scattering states using the plane-wave basis set and nonlocal pseudopotential Hamiltonian [9]. This approach was used in our previous study to simulate a 2D material TFET [34]. Here, we used the same method to calculate the $I(V)$ curve, with the renewed implementation in the PWmat code. We used the exact same potential profile at one side of Fig. 1 (with open-boundary conditions at the other side), and the same pseudopotential Hamiltonian as in the calculation above. The results are shown as the pink line, and more details are presented in the Supplemental Material [30]. As expected, it does not have a big oscillation, as in the above calculation

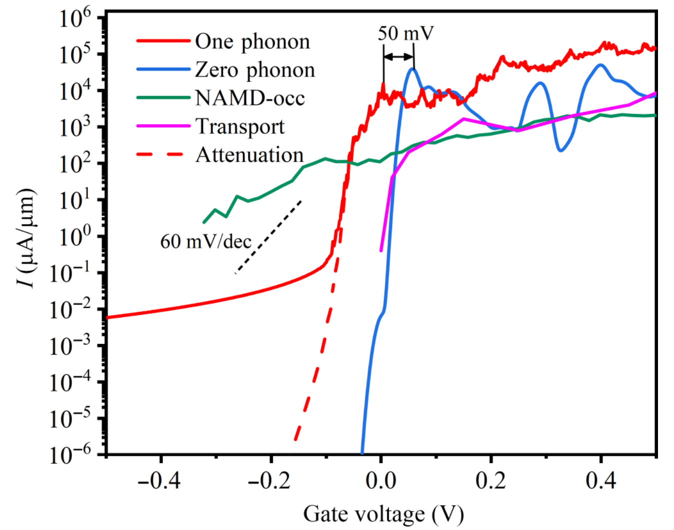

FIG. 4. I-V of MoS_2 TFET for different physical mode results. Gate-voltage origin is set as VBM, which energetically aligns with the CBM.

(since there is no usage of the confined discrete quantum well states). Its amplitude is in the middle of the oscillation for the Fermi golden rule method. The slope near $V_g = 0$ is similar to the Fermi golden rule result. This confirms that the Fermi golden rule result is reliable. This is important as the phonon-assisted transition calculation is similar to the Fermi golden rule calculation; both are based on the quantum well, and thus, are affected by the quantum well finite number of states in the same way. On the other hand, scattering-state calculations are based on the open-boundary condition, without the use of a quantum well. Thus, by comparing the quantum well Fermi golden rule calculation and quantum transport calculation, we can estimate the effect of the finite quantum well.

After finishing the zero-phonon transport, we now turn to the phonon-assisted transport calculation. It is desirable to calculate the phonon-assisted effect with open-boundary methods like in the quantum transport calculation. However, currently, no such method is available. Although quantum transport calculations can compute zero-phonon transport, as shown in Fig. 4, there is currently no equivalent approach to incorporate the phonon-assisted effect. Researchers have used an electron-phonon coupling-induced self-energy term in the NEGF formalism [8,9]. Such methods cause energy smearing and broadening but do not explicitly capture the effect of transferring one electron from one scattering state to another scattering state with an different energy, as discussed herein. The Boltzmann equation approach commonly used to describe this phenomenon is inadequate, as it does not accurately account for quantum interference and dephasing. Due to the lack of open-boundary-condition methods, we use the

quantum well approximation, as shown in Fig. 2, to study the phonon-assisted transition. The effects of the finite number of states in the quantum well have been shown in the comparison between the zero-phonon Fermi golden rule calculation and the quantum transport calculation. Thus, here, we can compare our phonon-assisted result with the zero-phonon Fermi golden rule result; both of them use the same quantum well setting. One can still use the Fermi golden rule, with the phonon-induced coupling between VB state i and CB state j, for all the phonon modes ν. However, this requires us to calculate all the electron-phonon coupling constants with all the phonon modes. More importantly, one can only use the so-called static approximation for the coupling constant. As we discuss below, in our situation, that tends to overestimate the coupling constant. Instead, we use the adiabatic state coupling. Here, we use an integration of the MD simulation to calculate the transition rate. More specifically, we carry out *ab initio* MD calculations, at a given average room temperature (but using the *NVE* ensemble, so the trajectory is a true Newton's law trajectory). During MD, we first calculate the coupling constant $D_{ji}(t) = \left\langle \psi_j^{c0} | H(t) | \psi_i^{v0} \right\rangle$; here, time-independent ψ_i^{v0} and ψ_j^{c0} are the source VB state and channel CB state at the zero-temperature relaxed atomic structure (no phonon excitations), respectively. After MD to $t = T$, we can carry out the following integration using $H_{ji}(t) = D_{ji}(t)$:

$$C_{ji} = \int_0^T H_{ji}(t) e^{i(\varepsilon_j^{c0} - \varepsilon_i^{v0} - V_g)t} \cos(\pi t/2T)^2 dt. \tag{3}$$

Here, the attenuation factor $\cos(\pi t/2T)^2$ is used to avoid high-frequency oscillation in the final result. This is actually how the first-order perturbation theory is derived: C_{ji} is the wave-function coefficient projected to CB state j provided its coefficient at VB state i is one at $t = 0$ and the wave function is evolved following the time-dependent Schrodinger equation. In Eq. (3), we use the gate voltage, V_g, to indicate the fact that the gate voltage can shift all the conduction-band energies in the channel region. If $H_{ji}(t)$ only has one phonon mode, Eq. (3) will lead to the Fermi golden rule with one-phonon energy. The tunneling transition rate can then be calculated as

$$W = \sum_{j,i} |C_{ji}(T)|^2 / T. \tag{4}$$

This is analogous to the zero-phonon Fermi golden rule of Eq. (1).

One subtle issue is the calculation of $H_{ji}(t)$ as $D_{ji}(t) = \left\langle \psi_j^{c0} | H(t) | \psi_i^{v0} \right\rangle$. Notably, by definition, using zero-temperature ground state H_0, this D_{ji} is zero. Using $D_{ji}(t)$ as $H_{ji}(t)$ is called the static approximation [35]. In practice, however, we find that such a static approximation often leads to unreasonably large coupling constants (e.g., a few eV). This is mostly caused by some soft acoustic phonon modes, where $H(t)$ and H_0 might have large relative shifts for their atomic positions, which leads to large $\left\langle \psi_j^{c0} | (H(t) - H_0) | \psi_i^{v0} \right\rangle$. One reason for this is that the numerical MD trajectory is far beyond the linear approximations of the atomic displacement used in the electron-phonon coupling formalism. To solve this problem, we use $H_{ji}(t) = (dD_{ji}(t)/dt)/(\varepsilon_j^{c0} - \varepsilon_i^{v0})$. Within the linear-approximation formalism of the electron-phonon coupling, with harmonic phonon modes and a strict Fermi golden rule, this should yield the same results as those obtained using the original $D_{ji}(t)$. But, in practice, it removes the large $D_{ji}(t)$. Furthermore, it relies on the change of the Hamiltonian with time, instead of the absolute Hamiltonian under a fixed basis set; thus, in a sense, it is more like the adiabatic approximation, where the simulated wave function changes with time along the adiabatic states and only has transitions when the adiabatic states change into each other (e.g., like the anticrossing). Thus, this is also similar to the zero-phonon tunneling calculation and its coupling, as represented in Fig. 3. Only this time, anticrossing is driven by time-varying $H(t)$. This makes the comparison with the zero-phonon transition more natural. The calculated one-phonon transition rate is shown as the red curve in Fig. 4 with $T = 4000$ fs.

Finally, the above calculations only account for first-order perturbation theory; hence, it only has the effect of one phonon. To include all the electron-phonon coupling effects, we use the NAMD simulation [10]. In NAMD, the time-dependent Schrodinger equation is used to track the evolution of $\psi(t)$ under the *ab initio* MD Hamiltonian, $H(t)$. Besides, a dephasing time, tau (20 fs), and a Boltzmann factor are used to introduce surface hopping [11]. We used the PWmat implementation of the NAMD algorithm [36]. In a sense, perturbation theory also tracks the time-dependent Schrodinger equation of the wave function, albeit under the first-order approximation. The NAMD includes the electron-phonon coupling for all orders. Besides, it also depends on the change of Hamiltonian $H(t)$ with time, instead of the absolute $H(t)$ values; thus, it is similar to the adiabatic approximation in our perturbation-theory treatment. In the NAMD simulation, the system has a closed-boundary condition. The eigenstate wave-function overlap (hence, the change of the eigenstates) at time steps t and $t + \Delta t$ allow us to evaluate the wave-function evolution following the time-dependent Schrodinger equation. This simulation technique assumes that carrier dynamics do not have a back-reaction effect on nuclear dynamics, which is known as the classical path approximation. This approximation simplifies the simulation process by neglecting the influence of carrier dynamics on nuclear dynamics. As the NAMD

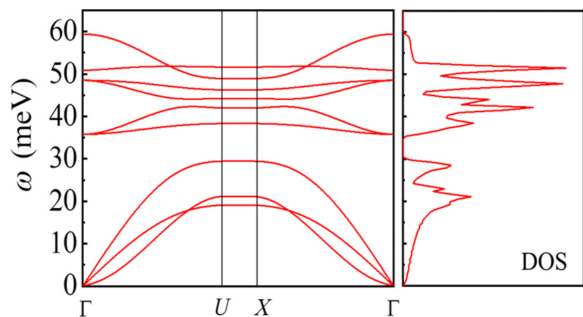

FIG. 5. Single-layer phonon-dispersion curves and density of states.

simulation does not use an open-boundary condition, the controls of the Fermi energy are not applied on the left and right sides. To calculate the transition rate in NAMD, an initial condition is set where all VB states are occupied, and all CB states are empty at $t = 0$, with a given V_g (gate voltage). This sets the condition for studying the occupation of CB states over time. By conducting the NAMD simulation, we can determine the rate at which some of the CB states become occupied over time. This analysis provides valuable insights into carrier dynamics and the transition of electrons from the VB to the CB states. Thus, once again, the transition rate, W, equals the amplitude in the channel region divided by the simulation time, T. The NAMD result is shown in Fig. 4 as the green curve, and more details of NAMD are given in the Supplemental Material [30].

III. RESULTS AND DISCUSSION

All our results are summarized in Fig. 4. In all of our calculations, at each k point, the highest 10 energy levels in the source VB and the lowest 26 energy levels in the channel CB are used. For the perturbation treatment, at the very negative region of V_g, there is a long tail. We find that this tail is caused by the numerical truncation of our time integration in Eq. (3). In the Supplemental Material [30], we show that, when we increase the total simulation duration time T, this tail gets smaller in amplitude. We thus can draw an extrapolation line along the steepest decay region, and ignore the tail, as indicated by the dashed line in Fig. 4. We see that the onset of this dashed line is indeed a left shift from the zero-phonon line by 0.05 eV, corresponding roughly to the average phonon modes. The calculated phonon density of state is shown in Fig. 5, with a phonon-energy range from 0 to 60 meV. The early onset of the $I(V)$ curve is likely to spoil the design of the cold-source device. Surprisingly, the one-phonon curve in the $V_g > 0$ region (where the source VB and channel CB DOS overlap) has a current even larger than the zero-phonon counterpart. This means that the electron-phonon coupling effect is predominant. Thus, without considering it, the predicted $I(V)$ curve can be really wrong.

Since the first-order one-phonon curve does not include the multiple-phonon effect, it is more revealing to consider the NAMD curve. This time, in the $V_g > 0$ region, it is rather close to the quantum-transport-calculated results. It is smaller than the one-phonon-perturbation result. This might indicate that, despite the use of the adiabatic coupling constant $(dD_{ji}(t)/dt)/(\varepsilon_j^{c0} - \varepsilon_i^{v0})$, the perturbation theory still overestimates the transition. One possibility is the use of ψ_i^{v0} and ψ_j^{c0} in the perturbation formalism, instead of the use of the true adiabatic states $\psi_i^v(t)$ and $\psi_j^c(t)$. In the NAMD, the phonon-assisted transition is based on the adiabatic states. But even if we scale down the one-phonon curve to be consistent with the NAMD curve in the $V_g > 0$ region, the early onset of the curve at V_g near -0.05 eV will still not change. Besides, the use of the same ψ_i^{v0} and ψ_j^{c0} for the transition-state calculation, while the gate voltage changes, could also lead to some approximations. But we believe the error introduced should be small, since the confinement of the VB and CB states is controlled by the shape of the band edges of the VBM and CBM (Fig. 2). While the gate-voltage change does introduce some shape changes, overall the shape change should not be large, especially near the region of $V_g = 0$ V. Once again, we are interested in a comparative study of without phonons and with phonons, and in our calculation, they are affected by such approximations to the same degree, just like the sharpness of the potential.

A prominent feature of the NAMD curve is slow decay in the $V_g < 0$ region. We carefully checked this result by using different simulation times T, as shown in the Supplemental Material [30]. We found that the result was unchanged, regardless of the simulation time used. As a matter of fact, in most cases, the electron changes from the source VB state to the channel CB state within 1–1000 fs [37], indicating a very fast process.

Finally, we see that, if we use the NAMD curve as a guide, its $I(V)$ curve slope is even smaller than the standard 60-mV/decade curve. One possible reason for this is the eigenenergy fluctuation of the electronic states for the CB states and VB states. Because of this fluctuation, they have an overlap before the zero-temperature CB and VB local density of states have an overlap. Such a fluctuation has a slow decay as a function of the CB and VB energy separation. Such fluctuation can be enhanced by the localization of the wave functions. It would be interesting to study the effects of such fluctuations in real devices in the future. Overall, our results indicate that it is difficult to use the TFET to realize a very small SS due to electron-phonon coupling.

IV. CONCLUSION

We developed methods to calculate the zero-phonon and one-phonon-assisted electron transitions in a TFET device.

We also deployed a quantum transport method to calculate the elastic tunneling transport and the NAMD method to calculate electron-phonon-coupling-induced transport, including all the multiphonon processes. The use of different methods allows us to cross-check the results, thus increasing the reliability of these calculations. Our goal is to study the effect of electron-phonon coupling in the cold-source design to reduce the SS. We used the MoS_2 2D system TFET as one example. Combining all our calculations, we found that the electron-phonon coupling played a very important role. It is likely that, with the electron-phonon coupling, it will be difficult for the TFET to significantly reduce the SS from the conventional 60-mV/decade limit for the MoS_2 device. Our study reveals the importance of electron-phonon coupling in the design of post-Moore's law devices. For example, it might be helpful to choose a small electron-phonon-coupling material as the channel material to surpass the conventional 60-mV/decade limit.

ACKNOWLEDGMENT

The work of Y.X. and L.W.W. was supported by the National Natural Science Foundation of China (Grants No. T2293700 and No. T2293702).

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

[2] J. Lyu, J. Pei, Y. Guo, J. Gong, and H. Li, A new opportunity for 2D van der Waals heterostructures: Making steep–slope transistors, Adv. Mater. **32**, 1906000 (2020).

[3] F. Liu, C. G. Qiu, Z. Y. Zhang, L. M. Peng, J. Wang, Z. H. Wu, and H. Guo, in *IEDM Tech. Dig.* (2018), pp. 33.2.1–33.2.4.

[4] J. T. Smith, S. Das, and J. Appenzeller, Broken-gap tunnel MOSFET: A constant-slope sub-60-mV/decade transistor, IEEE Electron Device Lett. **32**, 1367 (2011).

[5] W. Z. Gan, R. J. Prentki, F. Liu, J. H. Bu, K. Luo, Q. Z. Zhang, H. L. Zhu, W. W. Wang, T. C. Ye, H. X. Yin, Z. H. Wu, and H. Guo, Design and simulation of steep-slope silicon cold source FETs with effective carrier distribution model, IEEE Trans. Electron Devices **67**, 2243 (2020).

[6] H. Lu and A. Seabaugh, Tunnel field-effect transistors: State-of-the-art, IEEE J. Electron Devices Soc. **2**, 44 (2014).

[7] E. Kane, Zener tunneling in semiconductors, J. Phys. Chem. Solids **12**, 181 (1959).

[8] A. Afzalian, E. Akhoundi, G. Gaddemane, R. Duflou, and M. Houssa, Advanced DFT-NEGF transport techniques for novel 2-D material and device exploration including HfS_2/WSe_2 van der Waals heterojunction TFET and WTe_2/WS_2 metal/semiconductor contact, IEEE Trans. Electron Devices **68**, 5372 (2021).

[9] A. M'foukh, J. Saint-Martin, P. Dollfus, and M. Pala, *Ab-initio* simulation of dissipative transport in tunnel devices based on heterostructures of 2D materials, J. Comput. Electron. **22**, 1257 (2023).

[10] A. Garcia-Lekue and L. W. Wang, Elastic quantum transport calculations for molecular nanodevices using plane waves, Phys. Rev. B **74**, 245404 (2006).

[11] J. Kang and L. W. Wang, Nonadiabatic molecular dynamics with decoherence and detailed balance under a density matrix ensemble formalism, Phys. Rev. B **99**, 224303 (2019).

[12] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, Single-layer MoS_2 transistors, Nat. Nanotechnol. **68**, 147 (2011).

[13] D. Logoteta, M. G. Pala, J. Choukroun, P. Dollfus, and G. Iannaccone, A steep-slope MoS_2-nanoribbon MOSFET based on an intrinsic cold-contact effect, IEEE Electron Device Lett. **40**, 1550 (2019).

[14] S. H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, in *2009 Symposium on VLSI Technology* (2009), pp. 178–179.

[15] R. Gandhi, Z. Chen, N. Singh, K. Banerjee, and S. Lee, Vertical Si-nanowire *n*-type tunneling FETs with low subthreshold swing ($\leq$50 mV/decade) at room temperature, IEEE Electron Device Lett. **32**, 437 (2011).

[16] K. Tomioka, M. Yoshimura, and T. Fukui, in *2012 Symposium on VLSI Technology* (2012), pp. 47–48.

[17] S. Kanungo, G. Ahmad, P. Sahatiya, A. Mukhopadhyay, and S. Chattopadhyay, 2D materials-based nanoscale tunneling field effect transistors: Current developments and future prospects, npj 2D Mater. Appl. **6**, 83 (2022).

[18] S. Mookerjea, D. Mohata, T. Mayer, V. Narayanan, and S. Datta, Temperature-dependent *I–V* characteristics of a vertical $In_{0.53}Ga_{0.47}As$ tunnel FET, IEEE Electron Device Lett. **31**, 564 (2010).

[19] T. Yu, U. Radhakrishna, J. L. Hoyt, and D. A. Antoniadis, in *IEEE Int. Electron Devices Meet* (2015).

[20] A. M. Walke, A. Vandooren, R. Rooyackers, D. Leonelli, A. Hikavyy, R. Loo, A. S. Verhulst, K. H. Kao, C. Huyghebaert, G. Groeseneken, V. R. Rao, K. K. Bhuwalka, M. M. Heyns, N. Collaert, and A. V. Thean, Fabrication and analysis of a $Si/Si_{0.55}Ge_{0.45}$ heterojunction line tunnel FET, IEEE Trans. Electron Devices **61**, 707 (2014).

[21] U. E. Avci, B. Chu-Kung, A. Agrawal, G. Dewey, V. Le, R. Rios, D. H. Morris, S. Hasan, R. Kotlyar, J. Kavalieros, and I. A. Young, in *IEEE Int. Electron Devices Meet* (2015).

[22] S. Sant and A. Schenk, The effect of density-of-state tails on band-to-band tunneling: Theory and application to tunnel field effect transistors, J. Appl. Phys. **122**, 135702 (2017).

[23] J. T. Teherani, W. Chern, S. Agarwal, J. L. Hoyt, and D. A. Antoniadis, in *2015 Fourth Berkeley Symposium on Energy Efficient Electronic Systems E3S* (2015), pp. 1–3.

[24] J. T. Teherani, S. Agarwal, W. Chern, P. M. Solomon, E. Yablonovitch, and D. A. Antoniadis, Auger generation as an intrinsic limit to tunneling field-effect transistor performance, J. Appl. Phys. **120**, 084507 (2016).

[25] W. Jia, Z. Cao, L. Wang, J. Fu, X. Chi, W. Gao, and L. W. Wang, The analysis of a plane wave pseudopotential density functional theory code on a GPU machine, Comput. Phys. Commun. **184**, 9 (2013).

[26] W. Jia, J. Fu, Z. Cao, L. Wang, X. Chi, W. Gao, and L. W. Wang, Fast plane wave density functional theory molecular dynamics calculations on multi-GPU machines, J. Comput Phys. **251**, 102 (2013).

[27] D. R. Hamann, Optimized norm-conserving Vanderbilt pseudopotentials, Phys. Rev. B **88**, 085117 (2013).

[28] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Atomically thin MoS_2: A new direct-gap semiconductor, Phys. Rev. Lett. **105**, 136805 (2010).

[29] K. K. Liu, W. J. Zhang, Y. H. Lee, Y. C. Lin, M. T. Chang, C. Y. Su, C. S. Chang, H. Li, H. Y. M. Shi, H. Zhang, C. S. Lai, and L. J. Li, Growth of large-area and highly crystalline MoS_2 thin layers on insulating substrates, Nano Lett. **12**, 1538 (2012).

[30] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevApplied.21.064046 for the influence of time integration, quantum well length, and pseudopotential. In addition, the results of the transport model and the NAMD calculation model are included.

[31] R. A. Marcus, On the theory of oxidation-reduction reactions involving electron transfer. I, J. Chem. Phys. **24**, 979 (1956).

[32] R. A. Marcus, On the theory of electron-transfer reactions.VI. Unified treatment for homogeneous and electrode reactions, J. Chem. Phys. **43**, 679 (1965).

[33] H. Ilatikhameneh, Y. Tan, B. Novakovic, G. Klimeck, R. Rahman, and J. Appenzeller, Tunnel field-effect transistors in 2-D transition metal dichalcogenide materials, IEEE J. Explor. Solid-State Comput. **1**, 12 (2015).

[34] X. W. Jiang, J. W. Luo, S. S. Li, and L. W. Wang, in *IEEE International Electron Devices Meeting* (2015).

[35] L. Hkdin, New method for calculating the one-particle Green's function with application to the electron-gas problem, Phys. Rev. **139**, 663 (1965).

[36] F. Zheng and L. W. Wang, Ultrafast hot carrier injection in Au/GaN: The role of band bending and the interface band structure, J. Phys. Chem. Lett. **10**, 6174 (2019).

[37] E. Hauge and J. Stovneng, Tunneling times: A critical review, Rev. Mod. Phys. **61**, 917 (1989).