Full-Band Simulations of Single-Particle Resonant Tunneling in Transition Metal Dichalcogenide-Based Interlayer Tunneling Field-Effect Transistors

Xian Wu, Xuehao Mou, Leonard F. Register, and Sanjay K. Banerjee
Department of Electrical and Computer Engineering and Microelectronics Research Center
The University of Texas at Austin, Austin, Texas, USA
xianwu@utexas.edu

Abstract—We model and simulate the resonant tunneling and I-V characteristics of Interlayer Tunneling Field-Effect Transistors (ITFETs) based on transition metal dichalcogenide monolayers, MoS2 layers here, using quantum transport simulations with a full-band model. Gate-controllable resonant peaks are demonstrated and the short channel effects on resonance broadening are studied.

Keywords—resonant tunneling; tunneling transistor; transition metal dichalcogenide; quantum transport

I. INTRODUCTION

The Interlayer Tunneling Field-Effect Transistor (ITFET) (originally referred to as a “Double Electron Layer Tunneling Transistor” in [1]) is an attractive, non-conventional device intended for high-speed electronic applications. Based on conduction band (CB)-to-CB or valence band (VB)-to-VB resonant tunneling [2] between two two-dimensional (2D) electron gases separated by a tunnel barrier, ITFETs show gate-tunable negative differential resistance (NDR) characteristics [3,4], which can be applied to digital logic circuits [5].

Adjacent semiconductor quantum wells were the initial host proposed in [1] to achieve resonant tunneling for ITFETs. Recently, such resonant tunneling between the natural near-perfect 2D graphene layers has been studied both in theory and experiments [6-8]. Furthermore, quasi-2D monolayers of the transition metal dichalcogenides (TMDs) composed of a layer of transition metal atoms (Mo, W, etc.) sandwiched by two layers of chalcogen atoms (S, Se, etc.) [9], also provide a promising platform for ITFETs. Use of TMDs may reduce short-channel resonance broadening associated Heisenberg uncertainty.

Due to the complex band structures, the modeling of TMD materials in electronic devices commonly employs simple effective-mass-based parabolic band approximations [10], which neglects their nonparabolicity and higher-lying energy bands. In this paper, we apply a more accurate multi-band model for, for specificity here, MoS2 monolayers to study the resonant tunneling effect and I-V characteristics of short channel ITFETs based on these less explored materials.

II. SIMULATION METHOD

A. Device concept

Fig. 1(a) shows the structure of a MoS2-based ITFET schematically, where two MoS2 monolayers are partially overlapped with a tunnel barrier layer between. The left end of the top (T) MoS2 layer and the right end of the bottom (B) MoS2 layer extending beyond the overlap region serve as the electrical leads and are separately biased. In this paper, the bottom right or “source” (S) lead is always grounded and the top left or “drain” (D) lead is biased with voltage $V_{DS}$, which creates a Fermi level difference between two leads $E_{F,D} - E_{F,S} = -eV_{DS}$ where $e$ is the magnitude of the electron charge. Electrons in the overlap region are controlled by a back gate with voltage $V_{BG}$ and a top gate with voltage $V_{TG}$. It is possible for electrons or holes to tunnel from one layer to the other in this region, forming an interlayer tunneling current. The magnitude of the tunneling current depends on the alignment of the band structures in two layers and the Fermi level difference between them, as illustrated by Fig. 1(b). If carrier energies and in-plane crystal momenta are well defined, then when the band structures are aligned (when the difference $\Delta \phi$ between the electrostatic potentials in the top layer ($\phi_T$) and the bottom layer ($\phi_B$) becomes zero) a non-zero $V_{DS}$ can lead to a large resonant interlayer tunneling current, but when $\Delta \phi$ is nonzero, energy and momentum cannot be conserved simultaneously and the current would be eliminated. However, in practice weaker but nonzero currents still can flow off-resonance due to energy or momentum uncertainty in the tunneling process, such as due to scattering and, as to be considered here, short channel effects.

This work was supported by the Semiconductor Research Corporation’s Nanoelectronics Research Initiative (SRC-NRI) via the South West Academy of Nanoelectronics (SWAN). Supercomputing resources were provided by the Texas Advanced Computing Center (TACC).

![Fig. 1. (a) Schematic structure and biasing scheme of a MoS2-based ITFET. (b) Band diagrams illustrating the operation of ITFETs in (left panel) and out of (right panel) resonance. $\Delta \phi$ is the electrostatic potential difference between the layers, which is a function of source (S), drain (D), and top (T) and bottom (B) gate voltages.](image-url)
B. Electrostics modeling

Rigorously, $\varphi_T$ and $\varphi_B$ are functions of position along the tunneling region and should be calculated self-consistently using the Poisson’s equation along with the transport calculations discussed in the upcoming Section II-C. For the purpose of capturing the underlying principles of ITFETs here, we approximate $\varphi_T$ and $\varphi_B$, as well as the Fermi levels $E_{FT} = E_{FD}$ and $E_B = E_F$, as uniform within the respective layers. Therefore, the quasi-one-dimensional capacitive relations give

\[
e \left( p_T - N_{n=2}^{(T)} - N_{A,T}^{(c)} \right) = C_{G,T} \varphi_{OXT} + C_I (\varphi_T - \varphi_B)
\]

\[
e \left( p_B - N_{n=2}^{(B)} + N_{A,B}^{(c)} \right) = C_{G,B} \varphi_{OXB} + C_I (\varphi_B - \varphi_T)
\]

where $n_T$ and $p_T$ ( $n_B$ and $p_B$ ) are the electron and hole concentrations as functions of $\varphi_T$ ($\varphi_B$) and the Fermi level $E_{FT}$ ($E_{FB}$) in the top (bottom) layer; $N_A^{(c)}$ and $N_{A,T}^{(c)}$ are the similarly defined ionized donor or acceptor concentrations; $C_{G,T}$ ($C_{G,B}$) is the top (bottom) gate oxide capacitance; $C_I$ is the tunnel barrier capacitance; $\Delta \varphi_{OXT}$ ($\Delta \varphi_{OXB}$) is the electrostatic potential drop across the top (bottom) gate oxide determined by $\varphi_T$ ($\varphi_B$), gate voltage $V_{GC}$ ($V_{BG}$) and the gate material work function and electron affinity of the MoS$_2$ monolayer [10]. Thus, $\varphi_T$ or $\varphi_B$ are ultimately controlled by all four-terminal voltages.

Illustratively, here we use a capacitance equivalent to that of 1 nm SiO$_2$ for $C_I$, $C_{G,T}$ and $C_{G,B}$, ideal undoped channels, 4.3 eV effective electron affinity for MoS$_2$, and 4.1 eV work function for both gates. Under these conditions, the CB edge in each layer is located near the Fermi level under small voltages.

C. Transport simulation approach

The bulk crystal structure of a MoS$_2$ monolayer has a hexagonal primitive lattice. In order to model the transport properties, we use a rectangular unit cell in the x-y plane with twice the area of a primitive unit cell, as shown in Fig. 2(a). The lattice vectors of this unit cell arrangement are $\mathbf{a}$, $\mathbf{a'}$. We take the direction of current flow to be along the x axis and number the unit cells as shown in Fig. (b). The $n=1$ and $m=N$ cells are the left and right boundaries of the tunneling region, respectively.

To describe the electronic structure of the MoS$_2$ layers, here as in [11] we use a tight-binding model with a basis set of maximally-localized Wannier functions (MLWFs) [12] $|nma\rangle$ with hopping matrices $(H_{nma,n'm'a'})_{\alpha \alpha'} = (nma|H|n'm'a')$ calculated from density functional theory (DFT) using OpenMX [13]. Here $|nma\rangle$ denotes the $\alpha$th basis function in the unit cell denoted by $n\mathbf{a} + m\mathbf{a'}$. In this study we use the Mo atom’s $d_{xz}$, $d_{yz}$ and $d_{x^2-y^2}$ orbitals, including separate spin up and down orbitals to allow for spin-splitting, from which to construct the MLWFs and obtain $M = 12$ basis functions in a bilayer rectangular unit cell. The electron states $|\psi\rangle$ in this system can be written as a linear combination of all orbitals $|nma\rangle$ with coefficients $\psi_{nma}$. We write $|\psi_{nm}\rangle = (\psi_{nm1}, \ldots, \psi_{nmM})^T$ to denote the tight-binding wavefunction within each cell.

We treat our device as sufficiently wide so that translation invariance in y direction is retained and Bloch’s theorem can be applied. For the state with wavevector $k_y$ in the y direction, $\psi_{nm}(k_y) = e^{ik_y m a} \psi_{n0}(k_y)$, we write $\psi_n(k_y) = \psi_{n0}(k_y)$ and $H_{nn'}^{(hop)}(k_y) = \sum_m e^{ik_y m a} H_{n0n'm}^{(hop)}$. For the following analysis,

$k_y$ is included implicitly in $\psi_n$, $H_{nn'}^{(hop)}$ and related terms.

Similar to the work of [11], the Hamiltonian of this system has the form

\[
H_{nn'} = \begin{bmatrix} H_{nn',T} + e\varphi_T I & H_{nn',B} \\ H_{n'B,n'T} & -e\varphi_B I & H_{nn',B} \end{bmatrix}
\]

(2)

where $I$ is the unitary matrix; $H_{nn',T}^{(hop)}$ and $H_{nn',B}^{(hop)}$ are the intralayer hopping matrices of each MoS$_2$ monolayer obtained from the DFT calculations; $H_{n'B,n'T}^{(hop)}$ and $H_{n'B,n'T}^{(hop)}$ describe the interlayer interactions. In this study we simply take $H_{n,T,n'B}$ and $H_{n,B,n'T}$ equal to $0$ when $1 < n = n' < N$ and equal to 0 otherwise, such that interlayer hopping occurs only between the same basis functions in nearest interlayer cells in the tunneling region. We then use the scalar parameter $t$ to describe the interlayer tunnel strength. In further studies, these matrices can be replaced by more accurate hopping matrices through specific interlayer barriers from DFT calculations, or even by matrices including non-local interactions, e.g., the Fock exchange interaction as in [14].

In the transport direction, we assume that the leads are semi-infinite and perfectly absorbing. Thus, for each energy $E$ and wave-vector $k_y$, the wavefunction of the ITFET in the leads ($n \leq 1$ or $n \geq N$) can be written as $\psi_n = \psi_{n}^+ + \psi_{n}^-$, where the plus (minus) signed term indicates the components propagating or evanescent in the positive (negative) $x$ direction. In turn, $\phi_n^+$ ($\phi_n^-$) can be written as a linear combination of the multiple independent right (left) propagating or evanescent modes of these leads, $\phi_n^+ (\phi_n^-)$, with complex phase factors $\lambda_n^+$ ($\lambda_n^-$),

\[
\psi_{n}^\pm = \sum_i c_{i\pm}^n (\lambda_{i\pm}^\mp)^n \phi_i^n.
\]

(3)

All of these lead modes can be obtained by solving the quadratic eigenvalue equations,

\[
H_{10} \phi + \lambda H_{00} \phi + \lambda^2 H_{01} \phi = \lambda \Phi \phi.
\]

(4-a)

\[
H_{N(N+1)} \phi + \lambda H_{(N+1)(N+2)} \phi + \lambda^2 H_{(N+1)(N+2)} \phi = \lambda \Phi \phi.
\]

(4-b)

considering nearest neighbor rectangular unit cells in the x direction (four or more nearest neighbor atomic planes) where (4-a) is for the D lead and (4-b) for the S lead. Each equation,
(4-a) and (4-b), has 2M non-zero solutions, where, again, M is the number of basis functions in a unit cell. M of these solutions, $\Phi_1^±$ and $\lambda_1^±$, are the modes that either propagate ($|\lambda_1^±| = 1$) or evanesce ($|\lambda_1^±| < 1$) in the positive x direction, with only the propagating modes carrying current. The remaining M solutions, $\Phi_i^±$ and $\lambda_i^±$, are the modes that either propagate ($|\lambda_i^±| = 1$) or evanesce ($|\lambda_i^±| > 1$) in the negative x direction.

Use of (3) and (4) allows us to relate the wave function components of an electron state of the ITTFET system at two adjacent cells in the leads, e.g., $\psi_1^+$ and $\psi_0^-$ ($\psi_1^+$ and $\psi_0^-$), by

$$\psi_1^± = T_D^±\psi_0^± = \mathbf{Q}_±^±\mathbf{P}_±^±\mathbf{Q}_±^±\psi_0^±,$$

(5)

where

$$\mathbf{Q}_± = (\Phi_1^±, \Phi_2^±, \ldots, \Phi_M^±), \quad \mathbf{P}_± = \text{diag}(\lambda_1^±, \lambda_2^±, \ldots, \lambda_M^±)$$

(6)

with $\Phi_1^±$ and $\lambda_1^±$ solved from (4-a). Similar relations can be obtained between $\psi_{N+1}^±$ and $\psi_N^±$ via $T_{SL}$ constructed using the solutions from (4-b).

With these relations, we can use the approach as in [14] to set up a transport equation for the channel region

$$(\mathbf{E} - \mathbf{H} - \Sigma)\mathbf{\Psi} = \mathbf{S},$$

(7)

where $\mathbf{H} = [H_{ij}]_{N\times N}$, $\Sigma = [\Sigma_{ij}]_{N\times N}$, $\mathbf{\Psi} = [\psi_i]_{N\times 1}$, $\mathbf{S} = [s_i]_{N\times 1}$ with $i,j = 1, \ldots, N$. $\Sigma_1 = H_{10}T_{D}^{-1} - s_1 = H_{10}(1 - T_o^{-1}T_{D}^{-1})\psi_0^±$, $\Sigma_{N+N} = H_{N+1}T_{S}^{-1} - s_N = H_{N+1}(1 - T_{S}^{-1}T_{D}^{-1})\psi_{N+1}^±$, and all remaining $\Sigma_i$ and $s_i$ are zero.

For any given energy E and $k_y$, we first solve (4) to obtain all transport modes in each lead. Each propagating mode $\Phi_i^±$ gets $k_x$-dependent propagation direction injected into the channels, producing an associated $\psi_0^±(E, k_y)$ or $\psi_0^±(E, k_y)$ as required for the source term $\mathbf{S}$ in (7). With $\mathbf{S}$ the ITTFET state $\mathbf{\Psi}$ for that incident lead mode be obtained from (7). The tight-binding probability current flow carried by this ITTFET state $\mathbf{\Psi}$ is [15]

$$j_{D}^{±}(E) = 2\text{Im}(\psi_1^±H_{10}\psi_0^±)/\hbar$$

$$j_{S}^{±}(E) = 2\text{Im}(\psi_1^±H_{N+1}\psi_0^±)/\hbar$$

(10)

The total probability current flow is thus

$$j_e = j_{D}^{+} + j_{S}^{+} = j_{D}^{+}(E) + j_{S}^{+}(E)$$

(11)

The transmission coefficient $T$ for the state injected into the D (S) lead can be defined as

$$T_{D(S)}^{±} = j_{E,D(S)}^{±}(E) / j_{E,D(S)}^{±}(E)$$

(12)

Note that for any injection state labeled by energy $E$, wave vector $k_y$, injection lead $i$, and index $n$ from solving (4), (7) can be set up and solved independently. Therefore parallelized computing is readily used to reduce computing time.

Finally, the total macroscopic charge current is

$$j_{DS} = -e\sum_{E, k_y, n} f(E - E_F)j_e(E, k_y, i, n)$$

(13)

where $f(E)$ is the Fermi distribution function.

III. RESULTS

Fig. 3 shows the transmission coefficient $T$ of a randomly-chosen injection state near the CB edge as a function of the interlayer potential difference $\Delta \phi$, with different channel lengths $L$ and different hopping strengths $t$. As expected, this state has a strong transmission peak near $\Delta \phi = 0$ where resonant tunneling occurs. With a shorter channel, this peak is weakened and broadened, which is undesired and needs to be suppressed for the logic gates based on ITTFETs [1]. This broadening is due to Heisenberg momentum-position uncertainty for the finite channel length, translated to energy uncertainty via the charge carrier incident group velocities $v = \partial E / \hbar \partial k$. Thus, slower carriers produce less broadening, turning the heavy mass carriers in TMDs into an advantage. A moderate interlayer hopping strength doesn’t affect the shape of the transmission curves. However, a large $t$, e.g. 10 meV in Fig. 3(b), leads to the peak transmission probability saturating toward unity near $\Delta \phi = 0$. This saturation also effectively broadens the resonant tunneling by increasing the relative importance of the off-resonance tails which are not in saturation. Therefore, the tunneling barrier should be designed to allow sufficient tunneling current without excessively broadening the resonant tunneling.

The I-V curves in Fig. 5(a) present gate-controllable resonant tunneling in ITTFETs for non-saturating and in saturating current. With the back gate grounded ($V_{BG} = 0$), the channel potentials in two layers shift with respect to each other under different $V_{TG}$, which, in turn, leads to resonant tunneling at different $V_{DS}$ providing the mechanism of gate control. With our simulation parameters, a 0.2 V $V_{TG}$ is sufficient to shift the resonant $V_{DS}$ by only 25 mV, but this change still should be sufficient for switching in bi-stable logic gates in [1] in the absence of current saturation.

The degree of off-resonance current in the I-V tails at positive $V_{DS}$ and the asymmetries with respect to both $V_{TG}$ and $V_{DS}$ in Fig. 5(a) are associated with the filling of the channel states consistent with the density of states. This behavior can be effectively described via a quantum capacitance $C_Q$ connected in series between the lead and channel [16]. As illustrated in Fig. 5(b), when a positive (negative) $V_{TG}$ is applied, the Fermi level
shifts to a higher (lower) $C_Q$ regime and, thus, the drain has better (worse) control over the channel, while the gate has less (more) control. Note that the 300K thermal carrier distribution couples to the non-parabolic higher energy band structure producing the shown $C_Q$. All the above simulations were performed with 15 nm channel and $V_{BG} = 0$.

### IV. Conclusion

We demonstrate the theoretical feasibility of achieving gate-controllable resonant tunneling of an ITFET based on bilayer MoS$_2$ system. Tradeoffs of improving the ITFET performances are addressed. A sharp resonant peak is desired for ITFET-based logic gates. However, device scaling broadens the peak. A larger interlayer tunneling strength required for a sufficient current also broadens the resonance peak. Low resistivity is also desired for the MoS$_2$ layers to carry sufficient current, which requires higher carrier concentrations. However, as illustrated in Fig. 5, this will lead to a high $C_Q$, which weakens the gate control over the resonance.

In terms of the simulation method, we incorporated a multi-band model for MoS$_2$ monolayers into quantum transport simulations. Some of the underlying physics, e.g., the effect of quantum capacitances, can be addressed here more accurately than by a parabolic model. Furthermore, with the inclusion of non-local exchange effects, this method should allow consideration of current flow in the presence of possible interlayer exciton formation [14] in perhaps more compatible TMD systems, for exotic devices applications such as the Bilayer pseudoSpin Field-Effect Transistor (BiSFET) in [17].

### REFERENCES


