

Approaching Ballistic Transport in Monolayer MoS₂ Transistors with Self-Aligned 10 nm Top Gates

Chris D. English, Kirby K.H. Smithe, Runjie (Lily) Xu, and Eric Pop
Stanford University, Electrical Engineering, Stanford CA 94305, U.S.A.; Contact: epop@stanford.edu

Abstract—We present the first study of 10 nm self-aligned top-gated field-effect transistors (SATFETs) based on monolayer MoS₂. Using a novel fabrication process, we achieve record saturation current, \( I_{\text{Dsat}} > 400 \mu \text{A/\mu m} \), sub-threshold slope down to 80 mV/dec and equivalent oxide thickness (EOT) \( \approx 2.5 \text{ nm} \). Combining transistor modeling with careful gate capacitance and contact resistance measurements, we provide the first analysis of diffusive vs. ballistic transport in monolayer MoS₂ FETs. Results indicate the onset of ballistic transport with transmission up to 0.25 at low temperature. We also suggest a feasible route to advance MoS₂ transistors further to the ballistic limit.

I. INTRODUCTION

Scaling field-effect transistors (FET) below 10 nm requires channels \(< 3 \text{ nm} \) thin for well-behaved device electrostatics [1, 2]. In this regime, sub-nm thick monolayer semiconductors like MoS₂ do not suffer from surface roughness scattering, unlike Si, Ge, III-Vs bulk materials [3-5]. MoS₂ is also appealing like Si, Ge, III-Vs bulk materials [3-5]. MoS₂ is also appealing for highly scaled n-type transistors due to: 1) high-quality wafer-scale monolayer growth [6], and 2) Fermi level pinning near the conduction band [7, 8]. In addition, theoretical studies of MoS₂ FETs in the ballistic limit (\( L < 10 \text{ nm} \)) project advantages due to lower leakage currents enabled by the larger band gap (>1.8 eV) [9-11]. Some experimental studies have examined short channel behavior [7, 12-14], but until now, no one has demonstrated quasi-ballistic transport.

In this work, we demonstrate for the first time a self-aligned top-gate FET (SATFET) fabrication that allows ~10 nm gate lengths near the scaling limit, yielding good contact resistance (\( R_C \)) and record high saturation current in a monolayer MoS₂ FET (\( I_D > 400 \mu \text{A/\mu m} \) at room temperature). Furthermore, we demonstrate the onset of ballistic transport and provide insights to achieving fully scaled, ballistic devices. We also include a complete thermal analysis with analytic and numerical modeling to demonstrate the importance of accounting for Joule heating in the context of quasi-ballistic transport.

II. DEVICE FABRICATION

Devices are fabricated on monolayer MoS₂ (thickness \( d \approx 0.615 \text{ nm} \)) grown by chemical vapor deposition (CVD) on SiO₂(30 nm)/Si substrates [6]. After defining the channel width (\( W = 1 \mu \text{m} \)) by XeF₂ etching, a narrow electrode is patterned over the MoS₂ by e-beam lithography. Initially, we deposit a 2 nm seed layer of Al into the pattern, which fully oxidizes after exposure to air. Subsequently, we deposit thicker Al to form the top-gate (Fig. 1a). In air, the gate forms a 5 nm self-passivated surface oxidation layer around the entire electrode (Fig. 1b) which completes the gate dielectric. Next, we deposit 10 nm Au in high vacuum (10⁻⁸ Torr) [6] over the entire structure, forming self-aligned contacts with a small underlap (\( L_{\text{sp}} \approx 10 \text{ nm} \), Fig. 1c). Finally, we deposit thicker Ti/Au electrodes further from the channel (Fig. 1d). Cross-sectional TEM (Fig. 1e) reveals top-gate length \( L \approx 10 \text{ nm} \) with uniform surrounding Al₂O₃ thickness \( t_{\text{ox}} \approx 5-6 \text{ nm} \).

III. DEVICE MEASUREMENTS

All devices were measured in a vacuum probe station (~10⁻⁵ Torr), with the heavily doped p⁺ Si substrate serving as a back-gate (BG) if needed. We first examine the top-gate capacitance (\( C_{\text{ox}} \)), which is required to interpret our subsequent measurements. We directly measure \( C_{\text{ox}} \) on a long channel device (\( L = 2 \mu \text{m} \), Figs. 2a-b) fabricated with our self-aligned process, finding \( C_{\text{ox}} \approx 1.36 \mu \text{F/cm}² \) and EOT \( \approx 2.5 \text{ nm} \). However, for \( L = 10 \text{ nm} \) devices, we estimate EOT \( \approx 5 \text{ nm} \) by examining the shift in threshold voltage (\( V_T \)) of the top-gate vs. the back-gate voltage (\( V_{\text{BG}} \)) [15] in Fig. 2c. This apparent change in EOT arises from fringing field effects in the short channel devices, when the assumption of a parallel plate capacitance (\( C_{\text{air}}/t_{\text{ox}} \)) is no longer valid. Field lines from the Al gate to the MoS₂ channel fringe laterally (Fig. 2d, inset), reducing the peak field at the oxide-MoS₂ interface. Analytic approximations [16] for the electric field (with \( t_{\text{ox}} = 5 \text{ nm} \)) reveal the EOT rises for \( L < 50 \text{ nm} \) (Fig. 2d), highlighting the need for thinner oxides and double gates at the scaling limit.

To record \( I-V \) characteristics of our SATFETs, we used pulsed voltage measurements with 125 µs duration and 0.1 s period [17]. This was needed because most SATFETs exhibited slightly hysteretic DC behavior, indicating that our top gate oxides could be further improved. Pulsed measurements eliminate the hysteresis, allowing extraction of \( V_T \), carrier density \( n \), and the intrinsic device behavior. (However, a drawback is that we cannot probe the lowest sub-threshold currents.)

\( I-V \) characteristics of a \( L = 10 \text{ nm} \) SATFET are shown in Figs. 3a-b for the ambient temperature \( T_0 = 225 \text{ K} \) (the actual device temperature, \( T \), is much higher, as discussed later). For \( V_{\text{TG}} = 5 \text{ V} \) and \( V_{\text{BG}} = 0 \text{ V} \), we achieve \( I_D \approx 260 \mu \text{A/\mu m} \) at \( V_{DS} \approx 1 \text{ V} \), which saturates to over 400 \( \mu \text{A/\mu m} \) at \( V_{DS} > 2 \text{ V} \). In this regime, we also observe slight negative differential conductance, a sign of self-heating. Driving \( V_{DS} \), \( V_{\text{TG}} \), as well as \( V_{\text{BG}} \) to high bias we can obtain \( I_D \) up to 425 \( \mu \text{A/\mu m} \) for \( T_0 = 300 \text{ K} \).

Sub-threshold characteristics (Figs. 3c-d) reveal good device behavior down to \( L = 30 \text{ nm} \), with sub-threshold slope \( SS \approx 80 \text{ mV/dec} \). For \( L = 30 \text{ nm} \), the fringing fields lead to a lower \( V_T \), preventing access to the deep sub-threshold regime without breaking the thin Al₂O₃. The minimum SS for \( L = 10 \text{ nm} \) is 250 mV/dec, which could be steeper below our measurement limit. The resolution of our pulsed measurements also limits probing the \( I_{\text{OFF}} \) floor, underestimating the \( I_{\text{ON}}/I_{\text{OFF}} \) ratio.
In Fig. 4, we extract the effective mobility ($\mu$) and contact resistance ($R_C$) using the transfer length method (TLM) with varying $L$ (Fig. 1f). We see minor $\mu$ degradation between FETs without top-gates (34 cm²V⁻¹s⁻¹) and FETs with top-gates (30 cm²V⁻¹s⁻¹), indicating a relatively good top-oxide interface. Interestingly, contrary to FETs with bulk channel materials, $\mu$ appears to be largely independent of $V_{TG}$ (Fig. 4b). The $R_C$, which is limited mostly by access resistance [7], is lowest at 80 K due to increased $\lambda$. By adjusting $V_{TG}$, we can decrease this access resistance to obtain $R_C \approx 1.7 \text{ k}\Omega \text{um}$ at $T_0 = 300 \text{ K}$.

**IV. DIFFUSIVE VS BALLISTIC TRANSPORT**

Taking inspiration from previous work [18, 19], we analyze our data with a virtual source (VS) model that assumes $I_D = WQ_n(0)V_{IN}F_{SAT}$, where $Q_n(0)$ is the charge at the top of the barrier (ToB), $V_{IN}$ is the injection voltage, and

$$F_{SAT} = \left( \frac{Tr}{2-Tr} \right)^{1-\exp\left(\frac{\phi_{inj}}{k_BT}\right)} 1+\frac{Tr}{2-Tr} \exp\left(-\frac{\phi_{inj}}{k_BT}\right)$$

(1)

where $Tr$ is the transmission across the ToB, $V_{DS}$ is the intrinsic voltage, and $V_{th} = k_BT$ is the thermal voltage. This modified VS emission-diffusion (VSED) model connects the diffusive and ballistic regimes through $Tr$, which varies from near zero at the diffusive regime to near unity at the ballistic limit. $v_{inj} = (2k_BT/\pi m)^{1/2}$ where the effective mass $m^* = 0.45m_e$ and $m_e$ is the electron mass. $Tr$ is calculated from the effective channel length ($L_{EFF}$) and the mean free path for back-scattering ($\lambda$):

$$Tr = \frac{\lambda}{L_{EFF}} \quad \lambda = 2\left(\frac{v_{th}}{v_{inj}}\right)^2 \mu$$

(2)

Since $v_{inj}$ is a function of temperature, we self-consistently account for device self-heating, following ref. [20].

Fig. 5a shows the results of this VSED model fitted to our $L = 10 \text{ nm}$ device, by adjusting the $L_{EFF}$ parameter. At $V_{DS} < 1 \text{ V}$ the device is dominated by $R_C$, which decreases exponentially as the Schottky barrier is lowered at the drain. At $V_{DS} > 1 \text{ V}$, the contribution from $R_C$ is significantly lower than the channel resistance, such that $V_{THS}$ approaches $V_{DS}$. Without self-heating (dashed lines in Fig. 5a) the VSED model cannot match the experimental data (symbols). However, the measured $I_D$ continues to increase at high-bias, consistent with the positive $T$ dependence of $v_{inj}$. The calculated $T$ rise of the device is shown in Fig. 5b.

As a result, we obtain transmission $Tr \approx 0.15$ for $T = 300–400 \text{ K}$, indicating only the onset of ballistic transport. This is not unexpected, since $\lambda \approx 2 \text{ nm}$ for monolayer MoS$_2$ on SiO$_2$ at these carrier mobilities (Fig. 5c). However, at $T_0 = 80 \text{ K}$ (self-heating to $T \approx 175 \text{ K}$), we observe no increase in drive current despite an increase in $\mu$ and decrease in $R_C$ (Fig. 4). This is consistent with quasi-ballistic transport, where the decrease in $v_{inj}$ with smaller $T$ becomes apparent. Fig. 5d shows that we extract an effective electron velocity $v_{EFF} \approx 9 \times 10^6 \text{ cm/s}$ and $Tr = 0.25$ at $T_0 = 80 \text{ K}$ ($T \approx 175 \text{ K}$), indicating quasi-ballistic transport behavior. This improved $Tr$ does not result from an increase in $\lambda$, which is approximately constant since the $T$-dependence of $\mu$ is balanced by the $T$-dependence of $v_{inj}$. Rather, the improvement in $Tr$ results from a smaller $L_{EFF}$, where $L_{EFF} = 6.5 \text{ nm}$ for $T_0 = 80 \text{ K}$. At lower $T$, electrons injected over the ToB are more readily swept to the drain without scattering (by thermal energy $\approx k_BT$) back to the source, producing a shorter $L_{EFF}$ (Fig. 5c inset).

At high $\mu_{DS}$, $v_{EFF} = \mu_{DS}F_{SAT} \approx v_{inj}Tr / (2-Tr)$. Thus, to realize the full potential of MoS$_2$ FETs and obtain $v_{EFF} \approx v_{inj}$, a higher $Tr$, higher $\mu$ (for higher $\lambda$) and lower $L_{EFF}$ must be a priority. To achieve lower $L_{EFF}$ in our devices, doped contact underlap regions and improved EOT are necessary to produce a sharper ToB profile. $\mu$ must also be increased by reducing disorder scattering and improving the MoS$_2$-oxide interface [21].

In Fig. 6, we project how improvements to $L_{EFF}$ and $\mu$ ($\lambda$) could lead to increased $I_D$ for sub-8 nm MoS$_2$ FETs. With our VS model, we choose the smallest $L_{EFF}$ possible while maintaining electrostatic integrity against short channel effects, i.e. $L \geq \beta(k_{TB}F_{GS}C_{ox}/C_{TUB})^{1/2}$ where $\epsilon_{ox}$ and $C_{ox}$ are dielectric constants of the MoS$_2$ channel and top-oxide, and $\beta \geq 2.5 [1, 10]$. Assuming $EOT = 0.8 \text{ nm}$, this yields a minimum $L_{EFF} \approx 3 \text{ nm}$ for monolayer MoS$_2$. Fig. 6a projects the performance of MoS$_2$ FETs assuming $L_{EFF} = 8 \text{ nm}$ and $L_{EFF} = 3 \text{ nm}$, with $\mu = 50 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ from this work and $\mu = 100 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ which may be possible by repairing S vacancies [21]. We also assume drain induced barrier lowering (DIBL) $\approx 70 \text{ mV/V}$ and $SS = 100 \text{ mV/Dec}$, similar to ref. [11]. In general, scaling from $L_{EFF} = 10 \text{ nm}$ to $3 \text{ nm}$ yields a 40–50% boost in $I_D$. Increasing $\mu$ from 30 to 100 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ also results in a 50% increase ($\lambda \approx 2 \text{ to } 6.5 \text{ nm}$). If both an ideal $L_{EFF} \approx 3 \text{ nm}$ and $\mu \geq 100 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ are obtained, $I_D$ could be increased from $\sim 300 \text{ mA/\mu m}$ to $\sim 12 \text{ mA/\mu m}$ ($V_{TH} = V_{DS} = 1 \text{ V}$). This corresponds to an increased $Tr$ from $0.15$ to $0.68$ (Fig. 6b), marking near ballistic transport.

We note that bulk materials like Si, Ge, and III-Vs have difficulty in this scaling limit because 1) their $\mu$ and $\lambda$ decrease at channel thickness $d \leq 3 \text{ nm}$ [3-5, 22], and 2) the smaller effective masses of these materials increase the band-to-band tunneling leakage in the sub-threshold regime [22], resulting in larger $I_{OFF}$. The larger $m^*$ (and bandgap) of MoS$_2$, while detrimental to $v_{inj}$, allows for significantly lower $I_{OFF}$ compared to bulk materials at very small $L$ [9-11]. Therefore, after taking into account the effects of channel length and thickness scaling on both $I_{ON}$ and $I_{OFF}$, monolayer semiconductors like MoS$_2$ near the ballistic limit could emerge as promising candidates for low-power computing.

**CONCLUSION**

We described a self-aligned fabrication process enabling MoS$_2$ FETs with $L \approx 10 \text{ nm}$. We demonstrate record saturation current for a monolayer MoS$_2$ FET (> 400 $\mu A/\mu m$) as well as low $R_C$ and good $\mu$, among the best known to date for a monolayer semiconductor. We assessed the quasi-ballistic transport behavior of our devices including self-heating, demonstrating transmission up to 0.25 below room $T$. Our model also indicates that realistic improvements to $L_{EFF}$ and $\mu$ could move MoS$_2$ FETs closer to the ballistic limit, allowing them to compete with or surpass existing CMOS technologies.

**ACKNOWLEDGEMENT**

This work was supported in part by the AFOSR, the NSF EFRI 2-DARE program, and the Stanford Graduate Fellowship (SGF) to C.D.E. and K.K.H.S. We thank Prof. K. C. Saraswat for helpful discussions.
Fig. 1. Self-aligned top-gate FET (SATFET) fabrication: (a) deposition of Al gate electrode with seed layer, (b) formation of self-passivated Al₂O₃ gate dielectric, (c) self-aligned Au source and drain. (d) SEM image (top view) of SATFET. (e) Colorized cross-sectional TEM of SATFET with \( L \approx 10 \) nm. The monolayer MoS₂ cross-section is difficult to image, but it appears faintly visible at the contacts. (f) TLM structure with \( L = 10-100 \) nm used to extract mobility and contact resistance, required for the in-depth device analysis.

Fig. 2. (a) \( I_D \) vs \( V_{TG} \) curve for a long channel (\( L = 2 \) μm) SATFET on MoS₂. Arrows indicate forward and backward sweeps with minimal hysteresis. Inset: Optical image of device. (b) Measured top gate oxide capacitance vs. \( V_{TG} \) for the same device, with the source and drain grounded. (c) Top gate threshold voltage (\( V_T \)) vs. back-gate \( V_{BG} \) (symbols) for short channel SATFET (\( L \approx 10 \) nm). Slope of the fit (red line) yields ratio of top and bottom oxide capacitance, allowing the extraction of \( C_{BOX} \) because \( C_{BOX} \) is well-known [15]. (d) Simulated equivalent oxide thickness (EOT) vs. \( L \) (solid line) for Al₂O₃ physical oxide thickness \( t_{ox} = 5 \) nm. EOT approaches the parallel plate approximation (dashed) for long \( L \). The star symbol represents the \( \approx 10 \) nm device measured in this work. Inset: Fringing fields at the oxide-MoS₂ interface.
Fig. 3. Measurements of short-channel SATFETs. (a) $I_D$ vs. $V_{TG}$ and (b) $I_D$ vs. $V_{DS}$ at $T_0 = 225$ K for SATFET with $L = 10$ nm. (c) Sub-threshold behavior at $T_0 = 300$ K for SATFET with $L = 10$ nm, and (d) with $L = 30$ nm.

Fig. 4. (a) Measured effective mobility ($\mu$, red, left axis) and contact resistance ($R_C$, blue, right axis) vs. temperature. (b) $\mu$ (red, left axis) and $R_C$ (blue, right axis) vs. top-gate voltage ($V_{TG}$). Results obtained from TLM structure shown in Fig. 1f. In-depth analysis of such TLM structures is described in [7].

Fig. 5. (a) Measured $I_D$ vs $V_{DS}$ (symbols) at $T_0 = 225$ K for $V_{TG} = 0, 1$ V compared to the VSED model with (solid lines) and without (dashed) self-heating. Symbol colors correspond to channel temperature. (b) Average device temperature ($T$) vs. $V_{DS}$ from self-heating corresponding to the output curves in a). Inset: Finite element thermal simulation of a SATFET with $L = 10$ nm showing heat sinking at the contacts. (c) Transmission ($Tr$) vs. $V_{DS}$ for a 10 nm SATFET at $T_0 = 225$ K. Inset: conduction band diagram of transport across the barrier. (d) $Tr$ (green, left axis) and effective velocity ($v_{EFF}$, orange, right axis) vs. $T$.

Fig. 6. (a) Model projections of $I_D$-$V_{TG}$ assuming $EOT = 0.8$ nm and $I_{OFF} = 100$ nA/µm for $V_{DS} = 1$ V. The VSED model assumes $DIBL = 70$ mV/V and $SS = 100$ mV/Dec. For comparison, simulations use two values of $\mu = 30$ cm²/V·s (red) and 100 cm²/V·s (blue), and two values of $L_{EFF} = 3$ nm (solid lines) and 8.0 nm (dashed). (b) Estimated transmission vs. long-channel mobility of the material, for $L_{EFF} = 3$ nm (solid) and 8.0 nm (dashed).