How Do Quantum Effects Influence the Capacitance and Carrier Density of Monolayer MoS$_2$ Transistors?

Robert K. A. Bennett and Eric Pop*

ABSTRACT: When transistor gate insulators have nanometer-scale equivalent oxide thickness (EOT), the gate capacitance ($C_G$) becomes smaller than the oxide capacitance ($C_{ox}$) due to the quantum capacitance and charge centroid capacitance of the channel. Here, we study the capacitance of monolayer MoS$_2$ as a prototypical two-dimensional (2D) channel while considering spatial variations in the potential, charge density, and density of states. At 0.5 nm EOT, the monolayer MoS$_2$ capacitance is smaller than its quantum capacitance, limiting the single-gated $C_G$ of an $n$-type channel to between 63% and 78% of $C_{ox}$, for gate overdrive voltages between 0.5 and 1 V. Despite these limitations, for dual-gated devices, the on-state $C_G$ of monolayer MoS$_2$ is 50% greater than that of silicon at 0.5 nm EOT and more than three times that of InGaAs at 1 nm EOT, indicating that such 2D semiconductors are promising for improved gate control of nanoscale transistors at future technology nodes.

KEYWORDS: semiconductor capacitance, quantum capacitance, centroid capacitance, electrostatics, field-effect transistor, 2D semiconductor

Two-dimensional (2D) semiconductors have emerged over the past decade as promising candidates for sub-10 nm metal-oxide-semiconductor field-effect transistors (MOSFETs). Using 2D monolayer semiconductors in such transistors is appealing from an electrostatic perspective because their ultrathin channel (<1 nm) reduces the impact of lateral fringing fields while increasing the 2D semiconductor’s out-of-plane capacitance $C_{ic}$ (sometimes called the inversion layer capacitance for bulk semiconductor transistors in the on-state). Although conventional bulk semiconductors (like silicon) suffer from mobility degradation as their channel thickness is reduced to a few nanometers, 2D semiconductors maintain good electron and hole mobilities even in their monolayer limit, thus simultaneously offering excellent electrostatic control and good on-state conductance.

In a field-effect transistor, the total gate capacitance $C_G = qdn_{sh}/dV_G$ (where $q$ is the elementary charge, $n_{sh}$ is the number of charge carriers (electrons or holes) per unit area, and $V_G$ is the gate voltage) is given by the series capacitance of $C_{ic}$ with the gate insulator capacitance, denoted here as $C_{ox}$ (acknowledging that gate insulators may have nitrides or other components), as shown in Figure 1a and eq 1:

$$\frac{1}{C_G} = \frac{1}{C_{ic}} + \frac{1}{C_{ox}}$$

In general, $C_G$ must be maximized to achieve the highest transconductance and current drive of a transistor at the lowest $V_G$. The semiconductor channel’s contribution to $C_G$ is negligible when $C_{ic} \gg C_{ox}$ at which point $C_G \approx C_{ox} = \epsilon_{ox}/t_{ox}$ where $\epsilon_{ox}$ and $t_{ox}$ are the insulator’s permittivity and thickness, respectively. If $C_{ic}$ is comparable to $C_{ox}$, however, then $C_G$ can limit $C_G$, thereby limiting the maximum carrier densities, current, and transconductance achievable in the FET on-state. For example, we demonstrate in this work that for monolayer MoS$_2$, $C_{ic}$ is negligible when the gate insulator’s equivalent oxide thickness (EOT) is $\geq 2.5$ nm, although the precise EOT at which $C_{ic}$ becomes negligible varies between semiconductors. $C_{ic}$ has two main components: the centroid capacitance (due to the penetration of the charge centroid into the semiconductor channel) and the quantum capacitance $C_q$ (due to Fermi level movement with respect to the energy bands in a semiconductor channel with finite density of states). For 2D semiconductors, the centroid capacitance has often been taken as infinite (implicitly assuming $C_{ic} \approx C_q$), where $C_q$ is evaluated as:

Received: October 6, 2022
Revised: December 29, 2022
Published: February 14, 2023

https://doi.org/10.1021/acs.nanolett.2c03913
Nano Lett. 2023, 23, 1666−1672
\[ C_q = 2\pi \frac{g_s \epsilon_{s} m^*}{h^2} \left( 1 + \frac{\exp[E_C/(k_B T)]}{2 \cosh[qV_{ch}/(k_B T)]} \right)^{-1} \]  

(2)

where \( g_s = 2 \) is the spin degeneracy, \( g_v \) is the valley degeneracy (\( \approx 2 \) and 6 for the lower K-valleys and the higher Q-valleys, respectively, of the conduction band in monolayer MoS\(_2\)), \( m^* \) is the effective mass, \( h \) is Planck’s constant, \( E_C \) is the electronic band gap (\( \approx 2.2 \) eV for monolayer MoS\(_2\), depending on its dielectric environment\(^{23} \)), \( k_B \) is the Boltzmann constant, and \( T \) is the absolute temperature (here, \( \approx 300 \) K). Above, \( \psi_{ch} \) is the channel potential, typically considered without regard to its variation in the channel of 2D semiconductors (i.e., infinite centroid capacitance) but self-consistently treated in this work with spatial variation of charge density. Although \( C_q \) is tiny in the off-state, in the on-state a large \( |\psi_{ch}| \) pushes the Fermi energy into the channel conduction or valence bands, causing the bracketed term in eq 2 to approach unity and saturating \( C_q \) to its degenerate value \( C_{dq} \) given by\(^{12,15} \)

\[ C_{dq} = 2\pi \frac{g_s \epsilon_{s} m^*}{h^2} \]  

(3)

Considering only the lowest energy conduction and valence bands, \( C_{dq} \approx 70 \) and 200 \( \mu \)F/cm\(^2\) for \( n \) - and \( p \) -type monolayer MoS\(_2\), respectively. Although including higher energy bands (e.g., the Q-valley along the T-line\(^{16} \) in the monolayer MoS\(_2\) conduction bands) would enable larger \( C_{dq} \) even these lower bound estimates of \( C_{dq} \) greatly exceed \( C_{ox} \) for any realistic EOT, leading most studies to neglect \( C_{ch} \).

However, previous experimental studies on monolayer semiconductors, including MoS\(_2\), MoSe\(_2\), WSe\(_2\), and black phosphorus, have reported values of \( C_{ch} \) and \( C_{sc} \) that are much smaller than their respective \( C_{eq} \) when the Fermi energy \( E_F \) is pushed beyond the band edges\(^{17,19} \). Although these smaller-than-anticipated capacitances could be attributed to extrinsic contributions (like defects), recent theoretical work has shown that for monolayer MoS\(_2\), other components of \( C_{sc} \) could limit it to values much smaller than \( C_{dq} \).\(^{20} \) However, the contribution of nonuniform carrier distributions across a 2D semiconductor’s thickness, as well as the impact that this reduced \( C_{eq} \) will have on \( C_{ch} \), remain largely unexplored.

In this work, we address these gaps in knowledge by self-consistently solving carrier statistics equations with the electrostatic potential distribution across a monolayer MoS\(_2\)-based MOS capacitor, as shown in Figure 1b. We consider the spatial variation of electrostatic potential \( V(z) \) [where \( z \) is the cross-plane coordinate labeled in Figure 1b], the volumetric charge density \( \rho(z) \), and the local density of states (LDOS)\(^{21,22} \) across the monolayer thickness, and we write

\[ \rho(z) = \mp q \int \text{LDOS}(E, z) \left[ \frac{1}{2} \mp \frac{1}{2} \pm f(E) \right] \, dE \]  

(4)

where \( E \) is the energy, \( f(E) \) is the Fermi–Dirac distribution, upper (lower) signs are for electrons (holes), and the channel carrier density \( n_{ch} \) is obtained by integrating \( |\rho(z)|/q \). Applying a gate voltage changes \( \rho(z) \) and \( n_{ch} \) by modulating the local electrostatic potential \( V(z) \), pushing \( E_F \) from midgap toward the conduction (valence) bands and populating the channel with electrons (holes).

Here, we assume that the intrinsic semiconductor \( E_F \) is at the midgap energy, allowing us to equate \( E_F \) with \( qV(z) \) when computing \( f(E) \) in eq 4, where \( qV(z) \) is also referenced to the midgap. As both \( \rho(z) \) and \( V(z) \) are unknowns, we solve eq 4 self-consistently with Poisson’s equation,

\[ \frac{\partial}{\partial z} \left( \varepsilon(z) \frac{\partial V}{\partial z} \right) = -\rho(z) \]  

(5)

where \( \varepsilon(z) \) is the permittivity. We discretize \( V(z) \) and \( \rho(z) \) along a one-dimensional grid, including the gate voltage boundary condition \( [V(z) = V_G] \) at the top of the gate insulator and a Neumann boundary condition \( [\partial V(z)/\partial z = 0] \) at the opposite side of the bottom insulator in the structure shown in Figure 1b. The dual-gate geometry is treated in Supporting Information Section S1.

We model \( \varepsilon(z) \) as a step-like function that transitions from \( \varepsilon_{SiO_2} \) to the MoS\(_2\) permittivity\(^{17} \) \( \varepsilon_{MoS_2} \) (where \( \varepsilon_0 \) is the permittivity of vacuum) at \( z = -t_{SiO_2}/2 \), and then to the permittivity of SiO\(_2\) \( \varepsilon_{SiO_2} \) at \( z = -t_{SiO_2}/2 \), where \( t_{SiO_2} = 0.615 \) nm is the MoS\(_2\) monolayer thickness. We note that this dielectric profile is approximate; many different estimates for the out-of-plane dielectric constant of monolayer MoS\(_2\) have been reported\(^{14,17,23,24} \) and it is unclear how \( \varepsilon(z) \) varies spatially at the insulator/MoS\(_2\) interface. Furthermore, it is uncertain if \( \varepsilon(z) \) is mostly constant across the thickness of the MoS\(_2\) monolayer or if, like graphene,\(^{25} \) \( \varepsilon(z) \) is a function of position within the monolayer. Once these factors are known, they can be incorporated into our model by substituting the appropriate dielectric profile into eq 5.

We extract the LDOS of monolayer MoS\(_2\) using density functional theory (DFT) with Quantum ESPRESSO software.\(^{26} \) All calculations are performed on a \( 151 \times 151 \times 1 \) k-point grid using projector-augmented wave pseudopotentials with kinetic energy cutoffs of 60 Ry for wave functions and 480 Ry for charge densities and potentials. After computing the LDOS for a primitive cell in three dimensions, we average the LDOS across the in-plane directions to represent it only as functions of \( E \) and \( z \). Then, to ensure that the LDOS at a specific energy will always sum to the magnitude of the DOS at that same energy, we express it as

\[ \text{LDOS}(E, z) = L(E, z) \text{DOS}(E) \]  

(6)
where \( L(E, z) \) is the spatial distribution of the LDOS,\(^{21} \) which we normalize at each energy level \( E_n \):

\[
L(E_n, z) = \frac{\text{LDOS}(E = E_n, z)}{\int_{-\infty}^{\infty} \text{LDOS}(E = E_n, z) \, dz}
\] (7)

Figure 2 panels a and b show \( L(E, z) \) in the conduction and valence bands, respectively. At \( E \leq E_C + 0.25 \text{ eV} \) (where \( E_C \) is the conduction band minimum), the LDOS is confined close to the center of the MoS\(_2\) monolayer (i.e., near the Mo atoms) with sharp, narrow peaks appearing just to the left and right of the main central peak. These sharp satellite peaks arise from the spatial distribution of the 4d\(^{z^2}\) orbital of Mo, which has been shown to dominate the DOS at these energies in previous studies.\(^{27,28} \) We also find that broad peaks centered close to the S atoms appear in the LDOS at \( E > E_C + 0.25 \text{ eV} \). As shown in the projected DOS (pDOS) in Figure 2c, the S atoms begin to contribute to the DOS in the conduction bands at \( \sim 0.26 \text{ eV} \) above the conduction band minimum, corresponding to the Q-K valley separation \( \Delta E_{QK} \) from our DFT simulations. Projected band structures have also shown that S atoms contribute weakly to the electronic structure of monolayer MoS\(_2\) at the conduction band minimum, but they contribute noticeably to the Q-valley.\(^{27,29} \)

We note that the exact \( \Delta E_{QK} \) for monolayer MoS\(_2\) in a vacuum is not precisely known\(^{30} \) and that the band structure of monolayer MoS\(_2\) can vary depending on strain\(^{31} \) or its surrounding dielectric environment.\(^{15} \) For example, the experimental \( \Delta E_{QK} \) for monolayer MoS\(_2\) en cascaded in quartz and WS\(_2\) is \( \Delta E_{QK} \approx 0.11 \text{ eV} \) and simulated values range...
between 0.071 and 0.270 eV, depending on the approach used. To accommodate this uncertainty in the value of \( \Delta E_{30} \), we investigate its effect on \( C_{sc} \) and \( n_{ch} \) in Section S2 of the Supporting Information. We also note from Figure 2a,b that the LDOS extends slightly beyond \( t_{ed} = 0.615 \) nm, which occurs because DFT simulations of 2D MoS\(_2\) assume that the semiconductor is surrounded by a vacuum; in reality, a semiconductor’s LDOS cannot so easily penetrate into an insulator. However, we shortly demonstrate that this nonideality should not significantly affect our results.

In Figure 3a,b, we calculate and plot \( C_{sc} \) of monolayer MoS\(_2\) on linear and logarithmic \( y \)-axes, respectively, by self-consistently solving eqs 4 and 5 under an applied gate bias, with the corresponding \( n_{ch} \) values plotted in Figure 3c. For now, we set the permittivity of the gate insulator to an extremely large value (\( \epsilon_{ox} \rightarrow \infty \)), so that \( \psi_{surf} = V_{G} \) allowing us to study the intrinsic capacitance of monolayer MoS\(_2\) by neglecting the potential drop across the gate insulator. We will shortly relax this assumption and study monolayer MoS\(_2\)-based capacitors with finite EOTs.

At \( \psi_{surf} - E_G/(2q) \approx 0.3 \) V, the capacitance of p-type MoS\(_2\) exceeds that of n-type MoS\(_2\), which is due to the DOS near the valence band edge being larger than the DOS near the conduction band edge (Figure 2c,d). However, as \( \psi_{surf} \) is pushed farther into the conduction bands, the slopes of both \( C_{sc} \) and \( n_{ch} \) increase sharply for n-type MoS\(_2\). This increase is due to the step-like increase in the DOS at the Q-valley, noting that thermal broadening in eq 4 allows the DOS from the Q-valley to also contribute when \( E_G \) is a few \( k_B T \) below the edge of the Q-valley. This effect has been experimentally observed in MoSe\(_2\) and WSe\(_2\) monolayers, which have electronic structures similar to that of MoS\(_2\) (including a Q-valley above the conduction band edge). Thus, the shape of the electron density in our Figure 3c resembles similar experimental curves for MoSe\(_2\) and WSe\(_2\), monolayers.

We also compare our computed \( C_{sc} \) values to the conventional \( C_q \) for both n- and p-type monolayer MoS\(_2\), which we have plotted alongside \( C_{sc} \) in Figure 3b. Although our calculated \( C_{sc} \) closely matches \( C_q \) at small gate voltages (i.e., nondegenerate surface potentials), we find that at high gate voltages (i.e., degenerate potentials), our computed \( C_{sc} \) values are substantially lower than the traditionally calculated \( C_q = C_{ed} = 70 \) \( \mu \)F/cm\(^2\) for n-type and 200 \( \mu \)F/cm\(^2\) for p-type monolayer MoS\(_2\), respectively.

To understand why \( C_q = C_{ed} \) only for nondegenerate potentials, we first plot the charge distributions and potential across the thickness of n-type (p-type) MoS\(_2\) when \( \psi_{surf} \) is 0.3 V below (above) the conduction (valence) band edge in Figure 4a,b. For this nondegenerate case, the charge distributions are nearly symmetric across the channel, closely matching the LDOS distributions shown in Figure 2a,b. This result is consistent with the potential profile shown in Figure 4b: there is nearly no potential drop across the monolayer MoS\(_2\) thickness, allowing electronic states to contribute to the carrier density equally efficiently, regardless of their location in the channel. Hence, \( C_{sc} \approx C_q \) in this regime.

Next, to understand why \( C_q < C_{sc} \) at degenerate potentials, we plot the charge distributions and potential across the thickness of monolayer n-type (p-type) MoS\(_2\) when \( \psi_{surf} \) is 0.3 V above (below) the conduction (valence) band edge in Figure 4c,d. We find that for this degenerate case, the charge distributions are heavily asymmetric and skewed toward the gate electrode. This asymmetry can be explained from the potential drops in...
cm$^2$ at 0.5 V to 5.35 $\mu$F/cm$^2$ at 1 V, remaining between 63% and 78% of $C_{ox}$. However, we note that $C_G$ is sensitive to small variations in $V_G$ and to the position of the Q-valley for n-type MoS$_2$, as captured in Figure 5a and discussed in Supporting Information Section S2.

Similarly, we find that the classical equation overestimates $n_{th}$ in 2D MOS capacitors with small EOTs, as shown in Figure 5c,d for n- and p-type MoS$_2$, respectively. At EOTs of 0.5 and 1 nm, the channel carrier density $n_{th}$ can be as small as 65% and 79% (79% and 89%) of $n_{th}^{\text{classical}}$ for n-type (p-type) MoS$_2$ in the $V_G$ range considered here. Contributions from the Q-valley are visible for the n-type device with EOT = 0.5 nm, causing an increase in $C_G$ and $n_{th}$ when the voltage is sufficiently high. This effect has also been observed experimentally in ionic-liquid-gated MoSe$_2$ and WSe$_2$ monolayers, with band structures similar to that of monolayer MoS$_2$.

We note that $n_{th}^{\text{classical}} = C_{ox}(V_G - V_T)/q$ should not be applied near or below $V_T$ because this expression neglects subthreshold effects. Instead, $n_{th}$ may be approximated in both the off- and on-states by taking $C_G$ as the series combination of $C_C$ and $C_{ox}$ and then integrating the result (up to the relevant $V_G$) to find the carrier density. However, as we demonstrate in Section S3 of the Supporting Information, correcting for $C_C$ in this manner still significantly overestimates both $C_G$ and $n_{th}$ in the on-state for 2D channels with low EOT, highlighting the importance of including both quantum and charge centroid effects when modeling these devices.

Finally, we assess how $C_G$ limits $C_G$ for monolayer MoS$_2$ compared to other semiconductors, including silicon. Although decreasing $t_{th}$ improves the channel electrostatics, $t_{th} < 5$ nm causes surface scattering to limit silicon carrier mobilities. For $t_{th} \approx 5$ nm, a previous study has shown that the $C_G$ of silicon limits a dual-gated silicon FET with an EOT of 0.5 nm to $C_G \approx 7 \mu$F/cm$^2$ at an overdrive $V_{OV} = V_G - V_T = 1$ V. In direct comparison, our calculations show that a similar dual-gated structure with n-type monolayer MoS$_2$ offers $C_G \approx 10.9 \mu$F/cm$^2$ (10.6 $\mu$F/cm$^2$ for p-type), over 50% greater than that of silicon. The MoS$_2$ advantage persists even when the silicon thickness is reduced to 2.5 nm, which yields $C_G \approx 8.1 \mu$F/cm$^2$. We refer the reader to Section S1 of the Supporting Information for a description of how we calculated $C_G$ and $n_{th}$ for dual-gated devices and for full $C_G$ and $n_{th}$ vs $V_G$ curves.

The $C_G$ of monolayer MoS$_2$ compares even more favorably to III–V semiconductors, whose low density of states are known to limit their $C_G$. A previous study has shown that $C_G$ limits dual-gated InGaAs MOS capacitors with EOT = 1 nm to $C_G < 1.6 \mu$F/cm$^2$ at channel thickness $t_{ch} = 25$ nm; as $t_{ch}$ decreases, this $C_G$ worsens because the DOS shrinks due to quantum confinement effects. Using the same approach as above, we find that a dual-gated monolayer n-type MoS$_2$ capacitor with EOT = 1 nm offers $C_G \approx 5.55 \mu$F/cm$^2$ (or 6.00 $\mu$F/cm$^2$ for p-type) at $V_{OV} = 1$ V, over three times higher than InGaAs. The results from Figure 5a also indicate that single-gated monolayer MoS$_2$ capacitors with EOT of 2.5 nm offer higher $C_G$ than those reported for single-gated In$_n$Ga$_{1-n}$As and InAs capacitors with similar or lower EOTs.

In conclusion, we have shown that variations in carrier density (i.e., the centroid capacitance), potential, and density of states across the thickness of monolayer MoS$_2$ limit its on-state $C_G$ to values well below its degenerate quantum capacitance. As a result, gate capacitance, current, and transconductance estimates made by classical equations must be corrected when evaluating the carrier density or estimating the mobility of devices with EOTs below ~2 nm. Nevertheless, we find that in strong inversion, the $C_G$ of dual-gated n-type monolayer MoS$_2$ capacitors is over 50% higher than dual-gated silicon MOS capacitors at EOT = 0.5 nm and over three times higher than InGaAs capacitors at EOT = 1 nm. The monolayer MoS$_2$ capacitance advantage is higher at lower EOTs, ultimately indicating that good current and transconductance may be achieved in such 2D transistors if their channel mobility and contact resistance continue to be improved.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.nanolett.2c03913.

Dual-gated capacitor calculations; description and variation of the Q-K energy valley separation; comparison to quantum capacitance approximation (PDF)
Corresponding Author

Eric Pop — Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0003-0436-8534; Email: epop@stanford.edu

Author

Robert K. A. Bennett — Department of Electrical Engineering, Stanford University, Stanford, California 94305, United States; orcid.org/0000-0001-7427-8724

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.nanolett.2c03913

Author Contributions

R.K.A.B. and E.P. conceived the idea and wrote the manuscript. R.K.A.B. carried out all calculations.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

R.K.A.B. acknowledges support from the Stanford Graduate Fellowship (SGF) and the NSERC PGS-D programs. The authors also acknowledge partial support from the Stanford SystemX Alliance and from ASCENT, one of six centers in JUMP, an SRC program sponsored by DARPA.

REFERENCES

(4) Klinkert, C.; et al. 2-D Materials for Ultrascaled Field-Effect Transistors: One Hundred Candidates under the Ab Initio Microscope. ACS Nano 2020, 14, 8605–8615.
(12) Ma, N.; Jena, D. Carrier statistics and quantum capacitance effects on mobility extraction in two-dimensional crystal semiconductor field-effect transistors. 2D Mater. 2015, 2, 015003.
(20) Fang, N.; Nagashio, K. Quantum-mechanical effect in atomically thin MoS2 FET. 2D Mater. 2020, 7, 014001.
(29) Kormányos, A.; et al. k·p theory for two-dimensional transition metal dichalcogenide semiconductors. 2D Mater. 2015, 2, 022001.
(34) In a physical device, the remainder of the potential would be dropped across the bottom insulator, which is not observed here because we apply a Neumann boundary condition at the far side of the bottom insulator. However, we have verified that the potentials across the channel of Figure 4b,d become nearly identical to those
obtained when we instead ground the far side of the bottom insulator (note that electrical ground corresponds to the midgap voltage) and allow the bottom insulator thickness to approach infinity.


How do Quantum Effects Influence the Capacitance and Carrier Density of Monolayer MoS$_2$ Transistors?

Robert K. A. Bennett$^1$ and Eric Pop$^{1,*}$

$^1$Department of Electrical Engineering, Stanford University, Stanford, California 94305, U.S.A.

$^*$Contact: epop@stanford.edu

S1. Monolayer MoS$_2$ Dual-Gated Capacitors

We calculate the charge density $n_{ch}$ and gate capacitance $C_G$ of dual-gated monolayer MoS$_2$ capacitors by solving equations (4) and (5) in the main text self-consistently, just as we do when computing these quantities for single-gated devices. Here we use the device schematic shown in Figure S1a, which is similar to the single-gated device in Figure 1b in the main text except that the relative permittivities of both the top and bottom insulator are set to 20 (approximately that of HfO$_2$) and the thicknesses of the top and bottom insulators are identical. We update the bottom boundary condition [previously $\partial V(z)/\partial z = 0$ at this boundary for single-gated devices] to $V(z) = V_G$, where $V_G$ is the gate voltage applied at both the top and bottom electrodes.

We plot $C_G$ and $n_{ch}$ at equivalent oxide thickness (EOT) = 0.5, 1, and 2.5 nm for both $n$-type and $p$-type monolayer MoS$_2$ in Figures S1b,c. At an overdrive voltage of 1 V, $C_G$ is almost 11 μF/cm$^2$ for an $n$-type dual-gated monolayer MoS$_2$ FET with 0.5 nm EOT, over 50% greater than that of a dual-gated silicon FET with the same EOT (considering a silicon channel with thickness $t_{ch} = 5$ nm; note that $C_G$ for monolayer MoS$_2$ is still higher than for dual-gated silicon with $t_{ch} = 2.5$ nm).$^1$ At an EOT of 1 nm, the $C_G$ of dual-gated MoS$_2$ is over three times greater than that of dual-gated InGaAs with the same EOT and $t_{ch} = 25$ nm. (Further decreasing the InGaAs thickness reduces the InGaAs $C_G$ because its DOS shrinks due to quantum confinement effects.)$^2$

![Figure S1.](image-url)

Figure S1. (a) Schematic of a monolayer MoS$_2$ dual-gated MOS capacitor with boundary conditions applied when solving equations (4) and (5) in the main text. (b) Gate capacitance $C_G$ and (c) carrier density $n_{ch}$ for these dual-gated devices. Solid red (blue) lines represent $n$-type ($p$-type) MoS$_2$. Dashed lines mark the total dual-gated oxide capacitance $C_{ox} = 2 \times (3.9\epsilon_0/EOT)$ (where the prefactor of 2 accounts for both gates) and the conventionally calculated carrier density $n_{ch}^{\text{classical}} = C_{ox}(V_G - V_T)/q$, where the threshold voltage is $V_T = \pm E_{d}/(2q)$ for $n$- and $p$-type devices, assuming the same gate metal.
S2. Effect of Q-K Energy Valley Separation

Our density functional theory (DFT) calculations yield an energy difference $\Delta E_{QK} \approx 0.26$ eV between the Q- and K-valleys in the conduction bands of monolayer MoS$_2$, as labeled in Figure S2a (we note this Q-valley is sometimes called T or Λ). However, the computed value of $\Delta E_{QK}$ for monolayer MoS$_2$ is highly dependent on the input parameters used in DFT (e.g., exchange-correlation functionals and pseudopotentials), where the settings that yield the actual $\Delta E_{QK}$ of monolayer MoS$_2$ in vacuum are presently unclear. The band structure of monolayer MoS$_2$ also varies with the surrounding dielectric environment, further complicating the question of which $\Delta E_{QK}$ is relevant for a given physical system.

To accommodate this uncertainty in the “correct” $\Delta E_{QK}$, we repeat calculations of the semiconductor capacitance $C_{sc}$ and carrier density $n_{ch}$ for $n$-type MoS$_2$ with $\Delta E_{QK} = 0.13$ eV and compare these results to those obtained using $\Delta E_{QK} = 0.26$ eV in the main text. As $p$-type MoS$_2$ does not have an associated $\Delta E_{QK}$ or similar low-energy peaks that contribute to its density of states (DOS) in the range of energies of interest, the $p$-type results would be unchanged and are not repeated here.

![Figure S2](image)

Figure S2. (a) Band structure of monolayer MoS$_2$ near the conduction and valence band edges obtained from DFT, where the energy separation $\Delta E_{QK}$ is labeled between the Q- and K-valleys. Energies ($E$) are not to scale (e.g., the band gap is reduced to highlight details of the conduction band). Note that we neglect spin-orbit coupling in this work, although for monolayer MoS$_2$, including spin-orbit coupling only negligibly influences the value of $\Delta E_{QK}$ obtained from DFT. (b) Computed semiconductor capacitance $C_{sc}$ and (c) carrier density $n_{ch}$ for $n$-type monolayer MoS$_2$ capacitors with Q-K energy separations $\Delta E_{QK} = 0.13$ and 0.26 eV.

To obtain a local DOS (LDOS) profile with $\Delta E_{QK} = 0.13$ eV, we take the LDOS used in the main text with $\Delta E_{QK} = 0.26$ eV and splice together the LDOS at energies $E < E_C + 0.13$ eV and $E > E_C + 0.26$ eV to create a continuous LDOS profile with $\Delta E_{QK} = 0.13$ eV. We then compute $C_{sc}$ and $n_{ch}$ with this LDOS using the same methodology as described in the main text.

As shown in Figures S2b and S2c, $C_{sc}$ and $n_{ch}$ are the same for both values of $\Delta E_{QK}$ we consider at low $\psi_{surf}$. However, as $\psi_{surf}$ increases, the states near the Q-valley contribute at lower energies for the LDOS profile with $\Delta E_{QK} = 0.13$ eV, resulting in an earlier onset for the second linear region of $C_{sc}$, thereby increasing $n_{ch}$. Although this lower $\Delta E_{QK}$ shifts this linear region of the $C_{sc}$ curve...
to the left, it does not significantly affect the maximum value of $C_{sc} \approx 40 \ \mu\text{F/cm}^2$ in the range of $\psi_{surf}$ we consider.

Next, we repeat the calculations of $C_G$ and $n_{ch}$ presented in Figures 5a,b of the main text at EOTs of 0.5, 1, and 2.5 nm using $\Delta E_{QK} = 0.13$ and 0.26 eV. As shown in Figures S3a and S3b, the value of $\Delta E_{QK}$ used affects neither $C_{sc}$ nor $n_{ch}$ at EOT = 2.5 nm since $C_G$ is dominated by the oxide capacitance $C_{ox}$ at sufficiently large EOTs. At EOT = 0.5 and 1 nm, however, we find that the higher $C_{sc}$ at $\Delta E_{QK} = 0.13$ eV allows $C_G$ and $n_{ch}$ to grow closer to the classical limit compared to $\Delta E_{QK} = 0.26$ eV. This result signifies that smaller $\Delta E_{QK}$ provides further advantage of monolayer MoS$_2$ over silicon and III-V channels in terms of capacitance and channel carrier density at a given overdrive voltage $V_{OV} = V_G - V_T$. In practice, note that $\Delta E_{QK}$ is controlled by the strain applied$^{5,6}$ and may be controlled by the environmental dielectric as well.

![Figure S3](image_url)

**Figure S3.** (a) Single-gate capacitance $C_G$ and (b) carrier density $n_{ch}$ for n-type MoS$_2$ with Q-K energy separations $\Delta E_{QK} = 0.13$ and 0.26 eV at EOTs of 0.5, 1, and 2.5 nm. Dotted green (solid red) lines represent the LDOS profile obtained using $\Delta E_{QK} = 0.13$ (0.26) eV. Dashed lines mark the oxide capacitance $C_{ox} = 3.9e_0$/EOT and the conventionally calculated carrier density $n_{ch}^{\text{classical}} = C_{ox}(V_G - V_T)/q$, where the threshold voltage is $V_T = E_G/(2q)$.

### S3. Comparison to Quantum Capacitance Approximation

As discussed in the main text, the classical approximation of gate capacitance, $C_G = C_{ox}$ (where $C_{ox}$ is the oxide capacitance) and the classical approximation for charge carrier density, $n_{ch}^{\text{classical}} = C_{ox}(V_G - V_T)/q$, are well-known to be inaccurate near or below the threshold voltage $V_T$. Instead, $C_G$ is typically modeled in 2D semiconductors as the series combination of the quantum capacitance $C_q$ and the oxide capacitance $C_{ox}$, yielding $C_G^{-1} \approx C_{ox}^{-1} + C_q^{-1}$. Since $C_q$ is a function of the semiconductor’s surface potential, when using this equation, $C_G$ must be solved iteratively such that the voltage drop across the oxide and semiconductor are self-consistent with $C_q$. Then, $n_{ch}$ may be approximated at any $V_G$ by integrating this result to obtain

$$n_{ch}^{\text{Q-corrected}} \approx 1 \int_{E_G/2q}^{V_G} \left[ C_q^{-1}(V'_G) + C_{ox}^{-1} \right] dV'_G. \quad (S1)$$

As shown in Figure S4 below, the approximation $C_G^{-1} \approx C_{ox}^{-1} + C_q^{-1}$ matches the rigorously calculated $C_G$ presented in the main text in the subthreshold region (here $|V_G| < E_G/2q$). However, this approximation considerably overestimates the $C_G$ with EOT $\leq$ 1 nm at larger overdrive voltages. This finding is consistent with our previous result from Figure 3b in the main text, which shows that $C_q$ similarly overestimates the semiconductor’s capacitance in the on-state.
Figure S4: Gate capacitance $C_G$ for $n$-type and $p$-type monolayer MoS$_2$ as functions of $V_G$ at EOT = 0.5, 1, and 2.5 nm. Solid red (blue) lines represent the $C_G$ of $n$-type ($p$-type) MoS$_2$ calculated using the full \textit{ab initio} approach described in the main text that includes charge centroid effects. Dash-dotted red (blue) lines are the approximation $C_G^\text{approx} \approx C_{\text{ox}}^\text{approx} + C_q^\text{approx}$ (where $C_q$ is calculated using the self-consistent approach described above) which does not include centroid effects, and black dashed lines mark the oxide capacitance $C_{\text{ox}}$. The discrepancies between solid lines and approximations highlight the importance of quantum and charge centroid effects in the on-state, which are most important at small EOTs.

Similarly, to compare our rigorously computed $n_{ch}$ in the main text to the carrier density corrected with only the quantum (not centroid) capacitance, we plot our calculated $n_{ch}$ from the main text alongside Equation S1 in Figure S5 below. For a more thorough comparison, we also include our original $n_{ch}^{\text{classical}}$ on the same plot. At an EOT of 2.5 nm, Equation S1 accurately approximates our rigorously calculated $n_{ch}$ in both the off- and on-states. However, at EOT = 0.5 and 1 nm, Equation S1 does not correctly predict the charge in the on-state significantly better than the classical approximation $n_{ch}^{\text{classical}}$. Again, this result can be understood based on Figure 3b in the main text, where we show that our rigorously calculated semiconductor capacitance $C_{\text{sc}} < C_q$ in the on-state. From these results, we conclude that although including corrections for $C_q$ enables good approximations of $C_G$ and $n_{ch}$ in and near the off-state, the more rigorous approach for calculating these quantities presented in the main text (i.e., including spatial variations in the density of states, potential, and volumetric charge density) should be used to understand 2D semiconductor devices with sub-1 nm EOT in the on-state.
**Figure S5**: Charge carrier density $n_{ch}$ for $n$- and $p$-type monolayer MoS$_2$ as functions of $V_G$ at (a) EOT = 2.5 nm, (b) EOT = 1 nm, and (c) EOT = 0.5 nm. Solid red (blue) lines represent $n_{ch}$ of $n$- ($p$-) type MoS$_2$ calculated using the full *ab initio* approach described in the main text, which includes charge centroid effects. Dashed red (blue) lines represent the quantum capacitance-corrected carrier density $n_{ch}^{Cq\text{-corrected}}$ (Equation S1, using the self-consistent approach described above) that does not include centroid effects, and black dashed lines mark the classically calculated charge carrier density, $n_{ch}^{\text{classical}} = \frac{C_\text{ox}(V_G - V_T)}{q}$. The discrepancies between solid lines and approximations highlight the importance of quantum *and* charge centroid effects in the on-state, which are most important at small EOTs.

**SUPPORTING REFERENCES**


