Impact of device scaling on the electrical properties of MoS$_2$ field-effect transistors

Goutham Arutchelvan$^{1,2,5,}\ast$, Quentin Smets$^1$, Devin Verreck$^1$, Zubair Ahmed$^1$, Abhinav Gaur$^2$, Surajit Sutar$^1$, Julien Jussot$^1$, Benjamin Groven$^1$, Marc Heyns$^{1,2}$, Dennis Lin$^1$, Inge Asselberghs$^1$ & Iuliana Radu$^1$

Two-dimensional semiconducting materials are considered as ideal candidates for ultimate device scaling. However, a systematic study on the performance and variability impact of scaling the different device dimensions is still lacking. Here we investigate the scaling behavior across 1300 devices fabricated on large-area grown MoS$_2$ material with channel length down to 30 nm, contact length down to 13 nm and capacitive effective oxide thickness (CET) down to 1.9 nm. These devices show best-in-class performance with transconductance of 185 $\mu$S/\mu m and a minimum subthreshold swing (SS) of 86 mV/dec. We find that scaling the top-contact length has no impact on the contact resistance and electrostatics of three monolayers MoS$_2$ transistors, because edge injection is dominant. Further, we identify that SS degradation occurs at short channel length and can be mitigated by reducing the CET and lowering the Schottky barrier height. Finally, using a power performance area (PPA) analysis, we present a roadmap of material improvements to make 2D devices competitive with silicon gate-all-around devices.

CMOS technology has advertently followed Moore's law of device scaling for the past 50 years to achieve higher transistor density, higher speed and power improvements. A significant part of this device scaling, especially for the planar Metal–Oxide–Semiconductor–Field–Effect–Transistor (MOSFET) was achieved by scaling the gate length$^1$. This scaling is reaching its limits as short channel effects (SCE) significantly degrade the device performance. To partially overcome SCE, the tri-gate (FinFET) structure has been introduced$^2$. For future technology nodes, the gate-all-around nanosheet FET, which sandwiches thin layers of silicon channel between multiple gates, is expected to provide additional improvements. Both configurations enhance the electrostatic control over the channel and allow for further gate length scaling. However, it has been reported$^3$ that the required silicon channel thickness scaling below 10 nm severely degrades the carrier mobility due to increased surface-roughness scattering. In this context, two-dimensional (2D) semiconducting materials such as transition metal-dichalcogenides (TMDs) are considered to be ideal candidates due to their naturally passivated surface and ultra-thin body (1 monolayer MoS$_2$ ~ 0.65 nm), providing excellent gate-control and enhanced transport$^4$–$^7$. However, since many studies are performed with manually exfoliated flakes and collecting large datasets is very labor-intensive, there has been a strong focus on only selecting top performing devices, at the cost of less device understanding. Until recently, only a few TMD studies$^8$–$^{10}$ have focused on devices fabricated using large area grown films. Especially for device scaling$^{11}$, a statistically significant set of data is still lacking.

Therefore, we carry out a study of the impact of geometrical scaling on an extensive data set of large-area grown tri-layer MoS$_2$ MOSFETs (1300 devices). We investigate the impact of scaling the channel length ($L_{ch}$) and width ($W_{ch}$), contact length ($L_{cont}$) and effective oxide thickness (EOT) on various device performance metrics such as the on- and off-current ($I_{on}$, $I_{off}$), contact resistance ($R_c$), subthreshold swing (SS), interface trap density ($D_{it}$) and threshold voltage ($V_{T}$). We demonstrate that scaling the contact length down to 13 nm has no impact on the device performance. This confirms that carrier injection occurs exclusively from the edge of the metal directly into the thin TMD channel, which is in line with our TCAD simulations. Further, using our large data set, we make a detailed assessment on the scaling trends of SS and $V_T$ with device dimensions. We identify the variation in the number of MoS$_2$ layers in the channel and contact regions as a possible source for SS degradation and $V_T$ variability for ultra-scaled TMD MOSFETs. Such insights are crucial for device understanding and enables device architectures such as double-gate$^{12}$ or stacked TMD FETs to outperform Si FETs$^{13}$. This article is an extension of our previous work presented at IEDM 2019$^{10}$.

$^1$IMEC, Leuven, Belgium. $^2$KU Leuven, Leuven, Belgium. $^\ast$email: goutham.arutchelvan@imec.be
Results and discussion

We employ large area MoS$_2$ grown on a 2″ c-plane sapphire template by metal–organic chemical vapor deposition (MOCVD) process using molybdenum hexacarbonyl and dihydrogen sulfide as the precursors. Atomic force microscopy (AFM) shows the MoS$_2$ is composed of 3 monolayers (ML) fully closed and continuous film, with nucleation of 4 ML and 5 ML island regions (Fig. 1a). The average thickness is 3.6 ML, measured using Rutherford backscattering spectrometry (RBS). The device schematic is illustrated in Fig. 1b and details of the fabrication process (Fig. 1c) are discussed in the Methods section. Three different gate-oxides; (1) 50 nm SiO$_2$, (2) 12 nm HfO$_2$, and (3) 4 nm HfO$_2$ are used. An optical image after contact deposition is shown in Fig. 1d and cross-section TEM images of the final fabricated device are shown in Fig. 1e, f.

Direct current measurements are performed in N$_2$ ambient to avoid any impact of ambient humidity. A total of 1300 devices with varying $L_{ch}$ (30 nm to 5 μm), $L_{cont}$ (13 nm to 500 nm) and $W_{ch}$ (200 nm to 10 μm) are measured at two different drain-source bias ($V_{DS}$ = 0.05 V, 1 V). Back-gate leakage is low and below the tool noise range (< 1 pA) for the 50 nm SiO$_2$, 12 nm HfO$_2$, and 4 nm HfO$_2$ are used. An optical image after contact deposition is shown in Fig. 1d and cross-section TEM images of the final fabricated device are shown in Fig. 1e, f.

Channel edge effects are negligible, as confirmed by the constant on-state current density for several $W_{ch}$ (Fig S1). Devices with short $L_{ch}$, wide $W_{ch}$, and therefore high absolute current, show a large parasitic voltage drop over the source-drain metal probes, and are therefore omitted from the analysis. The threshold voltage of the FETs for $V_{DS}$ = 0.05 V, 1 V is obtained by both the linear extrapolation from peak-transconductance ($V_{TH}$)
and constant-current method (V_{T,CC} extracted at I_D = 10 nA * W/L) for L_{ch} > 500 nm and saturates for L_{ch} < 50 nm. I_{off} is higher for HfO2 compared to SiO2. (b) I_{D}, V_{DS} for L_{ch} = 500 nm shows linear triode regime and saturation at high V_{DS}. The dashed line follows the current at V_{DS} = V_{OV}. While the onset of saturation follows the V_{OV} at low V_G, it saturates at V_{DS} = 2.4 V for high V_G. The saturation current roughly scales V_{OV} and V_{OV}'. at low and high V_G, respectively. (c) I_{D}, V_{DS} for L_{ch} = 30 nm shows non-linear triode regime due to Schottky contacts and saturation at high V_{DS}. Saturation current follows a similar trend as I_{off} = 500 nm but V_{DS} at onset of velocity saturation is reduced to 1.4 V. (d) Conduction band profile for L_{ch} = 30 nm device with Schottky contacts shown for low and high V_{DS}. The Fermi-level at the source and drain are indicated by E_{S} and E_{D}, respectively. The Schottky barrier is shown as the abrupt potential change at the contact-channel interface. At low V_{DS}, I_{DS} is determined by Schottky contacts. At high V_{DS}, though the potential drops significantly at the source contact, velocity saturation or pinch-off near the drain determines the I_{D} characteristics.

Figure 2. (a) Scatter plot (with median line) showing I_{on} extracted at n_s = 1e13 cm^{-2} and I_{off} at a fixed displacement field of 0.4 V/nm below V_{T,CC} for V_{DS} = 1 V. I_{on} for the 50 nm SiO2 and 12 nm HfO2 devices overlap indicating no impact on low-field mobility and contact barrier. I_{on} roughly scales as 1/L_{ch} for L_{ch} > 500 nm and saturates for L_{ch} < 50 nm. I_{off} is higher for HfO2 compared to SiO2. (b) I_{D}, V_{DS} for L_{ch} = 500 nm shows linear triode regime and saturation at high V_{DS}. The dashed line follows the current at V_{DS} = V_{OV}. While the onset of saturation follows the V_{OV} at low V_G, it saturates at V_{DS} = 2.4 V for high V_G. The saturation current roughly scales V_{OV} and V_{OV}'. at low and high V_G, respectively. (c) I_{D}, V_{DS} for L_{ch} = 30 nm shows non-linear triode regime due to Schottky contacts and saturation at high V_{DS}. Saturation current follows a similar trend as I_{off} = 500 nm but V_{DS} at onset of velocity saturation is reduced to 1.4 V. (d) Conduction band profile for L_{ch} = 30 nm device with Schottky contacts shown for low and high V_{DS}. The Fermi-level at the source and drain are indicated by E_{S} and E_{D}, respectively. The Schottky barrier is shown as the abrupt potential change at the contact-channel interface. At low V_{DS}, I_{DS} is determined by Schottky contacts. At high V_{DS}, though the potential drops significantly at the source contact, velocity saturation or pinch-off near the drain determines the I_{D} characteristics.

**Scaling of on- and off-state currents.** From the representative transfer characteristics in Fig. 1g, we observe that the off-state current significantly increases as I_{on} is scaled, as a result of a loss of gate control. Accordingly, we extract the minimum current in the entire back gate sweep (I_{on}), and we observe that it is the same for both oxides and lower than the noise floor of the tool (< 1 pA). However, when comparing the I_{on} in the scatterplot Fig. 2a, which is extracted at a fixed displacement field of 0.4 V/nm below V_{T,CC} (i.e.,V_{GS} = V_{T,CC}/CET = 0.4 V/nm), we note that the HfO2 sample exhibits higher I_{off} compared to the SiO2 sample. This suggests that the subthreshold swing is limited by the high interface trap density (see Section D). We also note that for both oxides, I_{off} degrades with smaller L_{ch}. This is mainly due to SS degradation observed for short L_{ch} devices, and will be further discussed in Section E.

Next, we evaluate the I_{on} at a fixed charge density (n_s) of 10^{13} cm^{-2} and do not observe any difference between the 50 nm SiO2 and 12 nm HfO2 samples (Fig. 2a). This indicates that the carrier transport in the MoS2 channel is predominantly limited by charged impurities in the MoS2 or at the interfaces, and not by remote phonons in the gate oxide.
For the Ion, two distinct channel length scaling regimes can be identified in Fig. 2a. In the long-channel limit (\( \sim L_{ch} > 500 \text{ nm} \)), the Ion increases roughly proportional to \( 1/L_{ch} \) and the device operates in the triode region (illustrated in Fig. 2b for the 12 nm HfO2 sample and \( L_{ch} = 500 \text{ nm} \)) i.e. gate-overdrive \( V_{OV} (= V_{GS} − V_T) > V_{DS} \) for both oxides. The drain current also exhibits strongly linear dependence with \( V_{DS} \) in the triode region (Fig. 2b), suggesting that the channel resistance is dominant for this \( L_{ch} \) and beyond. We also extract a low-field-effect mobility of \( \sim 15 \text{ cm}^2/\text{V.s} \) (inset of Fig. 3c) using the transfer length method (TLM) for both the samples with 12 nm HfO2 and 50 nm SiO2. At higher lateral electric field (higher \( V_{DS} \)), Ion saturates (Fig. 2b), and the saturation current scales quadratically with \( V_{OV} \) (here \( V_{T,CC} = −0.4 \text{ V} \)) due to channel pinch-off near the drain. However, for the highest \( V_{OV} \) (\( \sim 2 \) to 2.4 V), the saturation current scales roughly linear with \( V_{OV} \), indicating that it is limited by saturation of drift velocity at high lateral-field16 (\( F_{LATERAL} > 5 \text{ V/μm} \)).

In the short-channel limit (\( \sim L_{ch} < 50 \text{ nm} \)), the dependence of Ion on \( L_{ch} \) saturates (Fig. 2a). Accordingly, in the output characteristics for \( L_{ch} = 500 \text{ nm} \), the Ion increases roughly proportional to \( 1/L_{ch} \) and the device operates in the triode region (illustrated in Fig. 2b for the 12 nm HfO2 sample and \( L_{ch} = 500 \text{ nm} \)) i.e. gate-overdrive \( V_{OV} (= V_{GS} − V_T) > V_{DS} \) for both oxides. The drain current also exhibits strongly linear dependence with \( V_{DS} \) in the triode region (Fig. 2b), suggesting that the channel resistance is dominant for this \( L_{ch} \) and beyond. We also extract a low-field-effect mobility of \( \sim 15 \text{ cm}^2/\text{V.s} \) (inset of Fig. 3c) using the transfer length method (TLM) for both the samples with 12 nm HfO2 and 50 nm SiO2. At higher lateral electric field (higher \( V_{DS} \)), Ion saturates (Fig. 2b), and the saturation current scales quadratically with \( V_{OV} \) (here \( V_{T,CC} = −0.4 \text{ V} \)) due to channel pinch-off near the drain. However, for the highest \( V_{OV} \) (\( \sim 2 \) to 2.4 V), the saturation current scales roughly linear with \( V_{OV} \), indicating that it is limited by saturation of drift velocity at high lateral-field16 (\( F_{LATERAL} > 5 \text{ V/μm} \)).

We can further understand both these observations from the simulated conduction band profile of \( L_{ch} = 30 \text{ nm} \) device (Fig. 2d) for low and high \( V_{DS} \). In the linear regime (\( V_{DS} = 0.2 \text{ V} \) and \( V_{OV} > V_{DS} \)), the drain-source potential
is predominantly dropped across the reverse-biased source and forward-biased drain Schottky contacts. With increasing $V_{DS}$ (higher lateral field), the transmission probability across the Schottky contacts increases rapidly, especially across the reverse-biased source, giving rise to the super-linear dependence of $I_{on}$ with $V_{DS}$. At even higher $V_{DS}$ ($V_{DS} = 1.2$ V), the electric field in the channel near the drain is large enough to cause either pinch-off at low $V_{OV}$ or saturation of the carrier drift velocity at high $V_{OV}$. Then, this results in saturation of the current.

**Contact length scaling.** Figure 3a shows that $I_{on}$ (@ $n_s = 10^{13}$ cm$^{-2}$) does not degrade as $L_{cont}$ is scaled down to 13 nm. This agrees with TCAD simulations$^{10,17,18}$ that predict contact edge injection of carriers for 1-3 layers of MoS$_2$ channel. This observation holds true for three different $L_{ch}$ (30 nm, 100 nm, 500 nm) over a wide range of $L_{cont}$ (500 nm to 13 nm) and for varying lateral field ($V_{DS} = 0.05$ V, 1 V). In all three cases, as predicted, we do not observe any systematic degradation of $I_{on}$ by scaling down $L_{cont}$ from 500 to 13 nm. Even for the shortest $L_{ch} = 30$ nm, where the channel resistance is negligible and the device is Schottky contact limited ($I_{on}$ is super-linear at low $V_{DS}$ in Fig. 2c), the contact resistance is independent of $L_{cont}$. Moreover, the electrostatic properties of the device are also unaffected by scaling down $L_{cont}$ as can be seen in Fig. 3b from the trend of $SS_{SCC}$ and $V_{TCC}$ (@ $V_{DS} = 1$ V) with $L_{ch}$ for two extreme contact lengths. The SS degradation and $V_T$ roll-off with shorter $L_{ch}$ are independent of the contact length. The insensitivity to $L_{cont}$ scaling also holds for other gate-oxides and charge densities (plots not shown). In summary, for 3 ML MoS$_2$, the active region of MoS$_2$ under the metal contact where most of the electrons get injected (called the transfer length $L_T$) is at least below 13 nm.

These results agree very well with our previous TCAD simulations with overlapping back-gate. For thin MoS$_2$ (1–3 ML), these predict $L_T$ smaller than the minimum simulated $L_{cont}$ of 2 nm (Fig S2). This is caused by the Schottky barrier (SB) at metal-MoS$_2$ interface, which depletes the MoS$_2$ underneath even at a high field and prevents vertical electron injection. Therefore, injection is only allowed from the edge of the metal contact directly into the carrier-rich channel, which is also predicted in other work$^{10,18}$. For thicker MoS$_2$ (more than 5 ML), the MoS$_2$ region underneath the contact is no longer depleted close to the oxide interface, and a longer section of the contact contributes to carrier injection$^{10,18–22}$. In a top-gate-only configuration, the absence of gate field under the contact would cause the vertical injection to become even more ineffective for both thick and thin MoS$_2$ channels. As a result, the contact length can also be downscaled for top-gated devices without any performance penalty (Fig S2). Moreover, reduction of contact barrier or MoS$_2$ sheet resistance under the contact does not increase the $L_T$ for 1–3 ML MoS$_2$ as the oblique trajectory still provides the least resistive path for carrier injection (Fig S2). However, such improvements could increase $L_T$ for thicker MoS$_2$ where a substantial carrier injection happens under the contact$^{19}$.

In other work$^{21,23–25}$, transfer lengths of 80 nm to 630 nm have been calculated using the transfer length method ($\lambda$-axis intercept), but those values are in contradiction with our results. As argued elsewhere$^{26}$, this method should not be used for thin TMD layers and Schottky contacts. The Schottky barrier fully depletes the TMD below, therefore the sheet resistance below the contact and in the channel are not the same, which is a requirement of the transfer length method. However, the transfer length method can still be reliably used for mobility calculation, because it does not have this requirement of identical TMD sheet resistance in the channel below the metal.

**Contact resistance extraction.** As we found in Section A that devices become more contact dominated as $L_{ch}$ is scaled, we now take a closer look at the value of the contact resistance. We extract the contact resistance directly as half of the total device resistance ($R_{tot}/2$) for devices with the shortest $L_{ch} = 30$ nm, without any need for extrapolation like in the TLM method. By considering $R_c = R_{tot}/2$, an upper limit is obtained for $R_c$, as it assumes negligible channel resistance. Figure 3c shows a plot of $R_{tot}/2$ at a charge density of $10^{13}$ cm$^{-2}$ vs $L_{ch}$. For $L_{ch} < 50$ nm, the $R_{tot}/2$ saturates, and we obtain a median Nickel-MoS$_2$, $R_c \sim 3$ kΩµm (at $n_s = 10^{13}$ cm$^{-2}$), which is in good agreement with $R_c$ extracted using TLM (inset of Fig. 3c). Our $R_c$ values are comparable to the state-of-the-art devices which have been demonstrated with Au$^{20}$ or Indium$^{27}$ contact metals. For increased $V_{OV}$, the contact resistance further drops due to better carrier injection into the accumulated channel, and we obtain $R_c \sim 1.2–2$ kΩµm @ $n_s = 2 \times 10^{13}$ cm$^{-2}$ (Fig 3d). For even higher carrier densities (compare $n_s = 2 \times 10^{13}$ cm$^{-2}$ to $2.7 \times 10^{13}$ cm$^{-2}$), $R_c$ no longer improves significantly. Significant device-to-device variation in contact resistance is observed, possible due to polymer residues between the metal contact and the MoS$_2$, which were not completely removed after the transfer and contact lithography steps of the fabrication flow.

**Long channel electrostatics and $D_{th}$ extraction.** Figure 4a shows that the subthreshold swing $SS_{SCC}$ obtained at $V_{DS} = 0.05$ V for different $L_{ch}$ improves with thinner back-gate oxide due to better gate control of the charge in the channel. Consequently, we achieve the best subthreshold swing for the devices on 4 nm HfO$_2$ substrate (Fig S3) with median $SS_{min} = 90$ mV/dec and 110 mV/dec (at $V_{DS} = 0.05$ V) for $L_{ch} = 50$ nm and 30 nm, respectively.

In the long-channel limit i.e., $L_{ch} > 1$ µm, $SS_{SCC}$ saturates to a constant median value of 80 mV/dec, 150 mV/dec, 1800 mV/dec for 4 nm HfO$_2$, 12 nm HfO$_2$, and 50 nm SiO$_2$ respectively. This is determined by the charging of MoS$_2$/oxide interface and channel defects (60° grain boundaries$^{28}$, and point defects$^{29}$), for which we calculate a trap density ($D_{th}$) of $4.5–7 \times 10^{12}$ cm$^{-2}$ eV$^{-1}$ from $SS_{min}$. This range of $D_{th}$ value is roughly similar across the different dielectrics. We also obtain this $D_{th}$ value using multi-frequency C-V measurements of TiN/HfO$_2$/MoS$_2$ MOSCap$^{2}$, where we obtain an acceptor-type trap density of $3.2–6 \times 10^{12}$ cm$^{-2}$ eV$^{-1}$ with energy levels near the midgap.

From C-V measurements, we find that the MOS capacitance is systematically lower than the target oxide capacitance due to exposure to water and/or atmospheric carbon during the wet transfer process from the sapphire template to the target substrates. Figure 4b shows how the maximum accumulation capacitance ($C_{max}$)
measured from TiN/HfO2/MoS2 MOScap (shown as the red diamonds) is lower than the value of Cox measured from TiN/HfO2/TiN MIMcap (without MoS2, shown as the black line). Equivalently, the capacitance equivalent oxide thickness (CET) values for MOScap (1.9 nm, 2.7 nm, and 3.8 nm) are systematically 1 nm higher than the EOT values of the MIMcaps (0.9 nm, 1.7 nm, and 2.6 nm). We calculate that the effect of quantum capacitance due to the limited density of states in MoS2, and the effect of charge centroid being a few angstrom away from the interface, are insufficient to account for this 1 nm difference. As the MIMcaps are not exposed to water or polymer during the fabrication, Fig. 4b shows the difference between the CET and EOT values can be explained by a 0.4 nm thick layer of water or hydrocarbons adsorbed from the ambient, or a combination thereof. In the future, we expect dry transfer in a controlled ambient will lower the CET, closer to the nominal EOT.

Short channel electrostatic degradation and variability. In the short-channel limit, i.e., Lch < 100 nm, Fig. 4a shows a degradation of median SS, but also increased scatter (SS at VDS = 1 V in Fig S4). A similar trend is also seen for SSmin (Fig S3). We hypothesize that the increased median and scatter could both be caused by the Schottky contacts, where the median SS degradation with shorter Lch is related to the relative increase of depletion regions from the Schottky contacts, while the scatter could be due to the variation in Schottky barrier height induced by the non-uniform thickness of the MoS2 seen in the AFM image in Fig. 1a.

We first verify the hypothesis of degraded median SS for shorter Lch by comparing representative experimental SS versus ID curves to simulations in Fig. 4c. We consider full SS–ID curves instead of extracting SS at a single
current level to understand the injection mechanism in a wider operation range. The simulations are performed for a SBH = 0.45 eV and uniform 3 ML MoS₂ channel. We observe two different regimes for SS for both the simulated and experimental data. In the first low-current regime (\(I_D < 1 \times 10^{-9} \text{ A/μm}\)), the current is limited by the thermionic emission of carriers from the metal into the channel. Here, the barrier for electrons consists of the highest position of conduction band edge inside the channel determined by the gate-bias. In this low-current regime, SS is determined by the change in the conduction band edge with gate-bias. As discussed in section D, the lower limit for SS (which corresponds to \(SS_{\text{min}}\) in Fig S3), is defined by the interface trap density. The degradation of \(SS_{\text{min}}\) for short-\(L_{\text{ch}}\) devices is due to the electrostatic potential of the source and drain metallurgical junctions influencing the channel potential and degrading the gate control. This is illustrated in Fig S5 where the conduction band energy is flat over most of the device for \(L_{\text{ch}} = 100\) nm, while it is lowered for \(L_{\text{ch}} = 30\) nm with the region of maximum barrier reducing to a small portion near the center of the device. Note that this effect is similar to conventional MOSFETs.

The second regime (\(I_D > 1 \times 10^{-9} \text{ A/μm}\)) is reached when the conduction band in the channel is lowered further, and carriers can efficiently tunnel through the SB (Fig S6). Here, the thermionic component over the barrier saturates and the tunneling path length determines the current. Because it continuously changes with higher \(V_{GS}\), the SS is worse than the first regime. Correspondingly, in the experimental devices, the \(SS_{\text{CC}}\) extracted at \(I_D > 1 \times 10^{-8} \text{ A/μm}\) (for \(L_{\text{ch}} < 100\) nm) shows a higher value than \(SS_{\text{min}}\) and stronger degradation with \(L_{\text{ch}}\). The SS for a given \(I_D\) also becomes nearly independent of \(L_{\text{ch}}\), because the tunneling path length depends only on the gate voltage and the thicknesses and dielectric permittivities of the TMD and oxide, for the low lateral electric field (\(V_{DS} = 0.05\) V). This is illustrated in Fig S6 where the conduction band energy and tunneling rate are plotted along the edge carrier injection path for \(L_{\text{ch}} = 30\) nm and 100 nm, showing no significant difference. With further reduction in SBH, the SS value in the second regime improves, reaching closer to the thermionic limit of the first regime.

We study the increased SS scatter for short \(L_{\text{ch}}\) seen experimentally, using simulations of devices with different uniform MoS₂ channel thickness and SBH. Figure 5a shows the simulated SS value for two different SBH (0.45 eV, 0.75 eV) and three different uniform thicknesses (1, 3 and 5 layers) of MoS₂ for \(L_{\text{ch}} = 30\) nm. Similar to the above case, we note two different regimes for SS irrespective of the barrier height. For the first regime of low \(I_D ( < 1 \times 10^{-8} \text{ A/μm}\) for \(SBH = 0.45\) eV and \(< 1 \times 10^{-11} \text{ A/μm}\) for \(SBH = 0.75\) eV), the SS is determined only by thermionic emission over the channel barrier. Therefore, the SS is independent on the channel thickness. However, the SS degrades for \(SBH = 0.75\) eV compared to 0.45 eV, because the higher Schottky barrier field penetrates deeper into the channel. For the second regime of high \(I_D (> 1 \times 10^{-8} \text{ A/μm}\) for \(SBH = 0.45\) eV and \(> 1 \times 10^{-10} \text{ A/μm}\) for \(SBH = 0.75\) eV), the SS is dependent on the tunneling length which is sensitive to the thickness of the semiconductor among other parameters. Subsequently, the gate control over the Schottky barrier, and hence the tunneling length, reduces with thicker MoS₂, resulting in poor SS for the 5 ML MoS₂ (Fig S7). In agreement with this observation, we also note that the difference in SS between the layers is more pronounced for the higher SBH of 0.75 eV.

In our experiments, we have even more variability due to non-uniform thickness within a single device. Even for the smallest functional device footprint (\(L_{\text{ch}} ~ 30\) nm \(* W_{\text{ch}} ~ 200\) nm), we always have a high probability (~ 70%) of having a mixed device i.e., regions of 3, 4 and 5 layers of MoS₂ within the same device. This is illustrated in Fig. 5b where the representative AFM (Fig 1a) image of the material was used to compute the probability of fabricating devices with different dimensions on only 3 (or) 4 (or) 5 or a mix of those layers. These mixed-thickness devices, together with the associated SBH variations, would result in non-uniform gate control and large scatter in the SS values of experimental devices. Also, note that the grain size and defects in the closed layers (1–3 ML) could additionally impact the device variability.

**Threshold voltage control.** We analyse \(V_T\) control for decreasing channel length, and Fig. 5c shows that there is no significant median \(V_T\) roll-off at \(V_{DS} = 0.05\) V. With a higher \(V_{GS} = 1\) V, we notice a \(V_T\) roll-off of about 200 mV from \(L_{\text{ch}} = 500\) nm to 30 nm. We attribute this roll-off to the higher lateral electric field across the reverse-biased Schottky contact, because \(V_{DS}\) is fixed at 1 V for all \(L_{\text{ch}}\). This higher electric field allows for increased carrier injection in short channel devices, which lowers \(V_T\). This roll-off could be mitigated by improving the gate control through gate-oxide scaling, or by reducing the amount of defects at the MoS₂/oxide interface. \(V_T\) control for decreasing channel width is also shown in Fig. 5c, and no systematic impact is seen as \(W_{\text{ch}}\) is scaled from 1 μm down to 200 nm. However, we note that the narrow devices (\(W_{\text{ch}} = 200\) nm) show higher \(V_T\) variability than wider devices (\(W_{\text{ch}} = 1\) μm), especially at \(V_{DS} = 0.05\) V. This increased \(V_T\) variability could be attributed to the higher probability of finding devices on discrete layers (Fig 5b) for narrower channel compared to a wider channel where the devices are always mixed. Other sources of variability such as bias-temperature instability, non-uniformity of the MoS₂ grains etc. could also impact the \(V_T\) variability and more dedicated experiments are required.

**Benchmark, projection and conclusion**

We present a benchmark chart (Fig. 5d) to compare the performance of our devices against flake and CVD 2D material FETs in literature. We choose the peak of transconductance (\(g_{m,max}\)) measured at \(V_{DS} = 1\) V and \(SS_{\text{min}}\) as the two metrics for comparison, similar to conventional Si transistors. The best corner is on the top-left since low \(SS_{\text{min}}\) and high \(g_{m,max}\) are desired. Our SiO₂ devices, owing to the thick EOT, provide low transconductance even for the shortest \(L_{\text{ch}}\) devices. Scaling the EOT (12 nm HfO₂ and 4 nm HfO₂) and using an optimized process flow (see Methods), we gain both in transconductance and SS, achieving a \(R_{\text{on}} < 2\) kΩ·μm for Ni contact metal and \(D_{\text{ox}} < 5 \times 10^{12} \text{ cm}^{-2} \text{eV}^{-1}\) for a CET of 1.9 nm. We demonstrate the highest \(g_{m,max} = 185 \mu\text{S/μm}\) at \(V_{DS} = 1\) V and a minimum SS of 86 mV/dec for 4 nm HfO₂. We also achieve \(I_{\text{max}} = 400 \mu\text{A/μm}\) at \(V_{DS} = 1\) V and \(V_{GS} = 4\) V for our 12 nm HfO₂ samples (Fig S8).
Despite the fact our 2D performance is among the best in literature, significant improvements are still needed to make 2D materials competitive with silicon channel devices for high-performance logic applications. Therefore we propose a roadmap using the Power Performance Area (PPA) metric for technology comparison in Fig. 6. 2D-FET and silicon nanosheet technology are compared using an inverter-based ring oscillator circuit, where each device consists of 4 vertically stacked sheets with scaled \( L_g = 14 \) nm and gate-all-around structure, corresponding to the imec 2 nm node\(^ {44} \). All devices are retargeted to an \( I_{\text{off}} = 2 \) nA at \( V_{\text{dd}} = 0.7 \) V and the inverter-circuit area is kept the same for fairer comparison between technologies. Starting from the baseline case (A) where experimental channel and contact parameters are assumed, the performance strongly improves in (B) when the Schottky barrier height is reduced. In (C), improvements to the 2D channel mobility results in higher ring-oscillator operating frequency compared to silicon, owing to superior electrostatic control of the 2D devices at shorter gate lengths. In (D), the ideal performance is simulated with more aggressively optimized material parameters.

In conclusion, we have scaled down the different device dimensions of CVD-grown MoS\(_2\) FETs and demonstrated \( g_{\text{m, max}} = 185 \) \( \mu \text{S}/\mu\text{m} \) and \( S_{\text{SS,min}} = 86 \) mV/dec which are among the best in literature. Using our large dataset, we systematically identified the key obstacles to be tackled to outperform silicon. First, we showed that scaling \( L_{\text{cont}} \) for thin MoS\(_2\) does not impact the short channel performance, which allows for an overall reduction in the device footprint and enables device and circuit level gate optimization\(^ {45} \). Second, we identified that for \( L_{\text{ch}} < 100 \) nm, the on-current is currently limited by high Schottky contact resistance (\( R_C = 1–2 \) kΩ·μm) at low \( V_{\text{DS}} \), and by a combination of velocity saturation and the Schottky barriers at high \( V_{\text{DS}} \). Third, we identified that

---

**Figure 5.** (a) Simulated SS versus log \( (I_D) \) for a uniform layer of 1, 3 and 5 monolayers of MoS\(_2\) for \( S_{\text{BH}} = 0.45 \) eV and 0.75 eV. For \( I_D > 1 \times 10^{-9} \) A/μm, tunneling through Schottky barrier determines the SS. Subsequently, a thinner channel results in better gate control, shorter tunneling length and therefore better SS. (b) Probability distribution versus device dimensions \( (L_{\text{ch}} \times W_{\text{ch}}) \). AFM from Fig. 1a was used to compute the probability distribution for fabricating devices on only 3, 4, 5 or a combination of those (mixed). Our experimental devices have a 60–70% probability of being mixed, leading to non-uniform gate control across the channel and contact regions. (c) \( V_{\text{TCC}} \) versus \( L_{\text{ch}} \) for two different \( W_{\text{ch}} \) (200 nm, 1000 nm) and \( V_{\text{DS}} \) (0.05 V, 1 V). No \( V_T \) roll-off at \( V_{\text{DS}} = 0.05 \) V due to excellent gate control over the channel for 12 nm HfO\(_2\). \( V_T \) roll-off of about 200 mV for \( V_{\text{DS}} = 1 \) V due to higher lateral-field at the source contact allowing for more carrier injection. No systematic \( V_T \) deviation between \( W_{\text{ch}} = 1 \) μm and 200 nm. (d) Benchmark plot showing \( g_{\text{m, max}} \) versus \( S_{\text{SS,min}} \). All values are at \( V_{\text{DS}} = 1 \) V except\(^ {39} \)–\( V_{\text{DS}} = 0.1 \) V, \( V_{\text{DS}} = 0.5 \) V\(^ {40} \), \( V_{\text{DS}} = 1.2 \) V\(^ {37} \), –\( V_{\text{DS}} = 1.5 \) V. In this work, 4 nm HfO\(_2\) provides best SS = 86 mV/dec and \( g_{\text{m, max}} = 185 \) \( \mu \text{S/\mu} \text{m} \).
our devices suffer from short channel effects (SS degradation), caused by the Schottky barrier at intermediate current level and the thick CET at low current level. Reducing the CET is therefore crucial to keep optimal electrostatic control of the thin channel. We established that a 0.4 nm layer of water or adsorbed hydrocarbons (or combination thereof) at the HfO2/MoS2 interface is the root cause of a lower-than-expected CET. This value is consistent across different thicknesses of HfO2. Therefore, an optimized transfer process free of water and carbon is needed to enable gate stack scaling below 1 nm, and additionally allow upscaling to 300 mm-wafer processing. Finally, we have demonstrated using a PPA analysis that if the obstacles of Schottky contacts, gate stack scaling and mobility improvement can be tackled, MoS2 FETs will significantly outperform silicon GAA FETs at the imec 2 nm node and beyond. Therefore, they are excellent candidates to continue logic scaling.

Methods

Device fabrication. For the device design, we use the back-gate configuration with top-contacts (Fig. 1b). The fabrication flow is summarized in Fig. 1c. The MoS2 is delaminated from the sapphire growth substrate using water intercalation and transferred to three different target substrates; (1) Si/50 nm SiO2 (2) Si/50 nm SiO2/5 nm TiN/12 nm HfO2, or (3) Si/50 nm SiO2/5 nm TiN/4 nm HfO2. Before transfer, the target substrates are pre-cleaned using a solvent rinse, followed by an optimized forming gas anneal (FGA) or soft O2 plasma, for SiO2 and HfO2 back-gate oxides, respectively. The active channels are patterned using PMMA mask and e-beam lithography, followed by reactive ion etching (Cl2 + O2) of MoS2. Source and drain contacts of different lengths (Lcont) with different channel lengths (Lch) are subsequently defined on the active channel by another e-beam lithography exposure of ZEP520A-2 resist (ZEON Corp.), e-beam evaporation of 10 nm Ni, and metal lift-off in anisole. We ensure a low vacuum pressure < 10–6 Torr while depositing the Ni contact metal. Finally, in a third e-beam lithography step, thicker Ni/Pd contact pads are lifted off.

TCAD calibration. All simulations are performed in Sentaurus Synopsys Device. The low-field mobility (μeff) is calibrated from an experimental TLM fit shown in Fig. 3c and implemented under a constant mobility model. An estimate for Dit is obtained from multi-frequency CV measurements as discussed in section D. An acceptor trap distribution uniform over the entire bandgap is assumed with Dit = 3 × 1012 cm−2 eV−1. With μeff and Dit fixed by experiments, the Schottky barrier height is fitted to median transfer characteristics of Lch = 30 nm devices which are predominantly contact-limited. For the Schottky injection, the non-local tunneling model based on the Wentzel-Kramers-Brillouin approach is used. All the parameters used in the simulation correspond to their median values.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 6 January 2021; Accepted: 4 March 2021
Published online: 23 March 2021

References
1. Dennard, R. H. et al. Design of ion-implanted MOSFET’s with very small physical dimensions. IEEE J. Solid-State Circuits 9, 256–268 (1974).
2. Skotnicki, T., Hutchby, J. A., King, T. J., Wong, H. P. & Boeuf, F. The end of CMOS scaling: toward the introduction of new materials and structural changes to improve MOSFET performance. IEEE Circuits Devices Mag. 21, 16–26 (2005).
40. Lingming Yang
32. Laturia, A., Van de Put, M. L. & Vandenberghe, W. G. Dielectric properties of hexagonal boron nitride and transition metal dichalcogenide MOSFET in the Ultimate Physical Scaling Limit. IEEE Electron Device Lett. 35, 402–404 (2014).

38. Kai-Shin Li
31. Kwon, J.
34. Nourbakhsh, A.
33. Penumatcha, A. V., Salazar, R. B. & Appenzeller, J. Analysing black phosphorus transistors using an analytic Schottky barrier MOSFET model. Nat. Commun. 6, 6948 (2015).

36. Huyghebaert, C.
29. Zhou, W.
26. Cheng, Z.
28. Wang, Y.
22. de la Rosa, C. J. L.
20. English, C. D., Shine, G., Dorgan, V. E., Saraswat, K. C. & Pop, E. Improved contacts to MoS2 transistors by ultra-high vacuum thermal annealing. Nano Lett. 13, 2640–2646 (2013).

19. Arutchelvan, G.
18. Szabó, Á., Jain, A., Parzefall, M., Novotny, L. & Luisier, M. Electron transport through metal/MoS2 interfaces: edge- or area-dependent process?. Nano Lett. 19, 3641–3647 (2019).

17. Arutchelvan, G.
16. Smithe, K. K. H., English, C. D., Suryavanshi, S. V. & Pop, E. High-Field Transport and Velocity Saturation in Synthetic Monolayer MoS2. Nano Lett. 18, 4516–4522 (2018).

15. Novoselov, K. S.
6. Szabo, A., Rhyner, R. & Luisier, M. Ab-initio simulations of MoS2 transistors: From mobility calculation to device performance evaluation. in 2014 IEEE International Electron Devices Meeting 30.4.1–30.4.4 (2014). doi:https://doi.org/10.1109/IEDM.2014.7047142

14. Lee, Y., Fiore, S. & Luisier, M. Ab initio mobility of single-layer MoS2 and WS2: comparison to experiments and impact on the device characteristics. in 2019 IEEE International Electron Devices Meeting (IEDM) 5.7.1–5.7.4 (2019). doi:https://doi.org/10.1109/IEDM.2019.8870506.

13. Wang, Y.
12. Verreck, D., Arutchelvan, G., Heyns, M. M. & Radu, I. P. Device and Circuit Level Gate Configuration Optimization for 2D Materials: From the Metal to the Channel: a study of carrier injection through the metal/2D MoS2 interface. Nanoscale 9, 10869–10879 (2017).

11. Liu, H., Neal, A. T. & Ye, P. D. Channel length scaling of MoS2 MOSFETs. in 2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD) 1–4 (2019).

10. Jin, S., Fischetti, M. V. & Tang, T. Modeling of Surface-Roughness Scattering in Ultrathin-Body SOI MOSFETs. IEEE Trans. Electron Devices 54, 2191–2203 (2007).

9. Kang, K. et al. High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity. Nature 520, 656–660 (2015).

8. Smets, Q. et al. Ultra-scaled MoCVD MoS2 MOSFETs with 42nm contact and 230µA/µm drain current. in 2019 IEEE International Electron Devices Meeting (IEDM) 23.2.1–23.2.4 (2019). doi:https://doi.org/10.1109/IEDM.2019.8993650.

7. Majumdar, K., Hobbs, C. & Kirsch, P. D. Benchmarking Transition Metal Dichalcogenide MOSFET in the Ultimate Physical Scaling Limit. IEEE Electron Device Lett. 35, 402–404 (2014).

6. Novoselov, K. S.
5. Szabo, A., Rhyner, R. & Luisier, M. Ab-initio simulations of MoS2 transistors: From mobility calculation to device performance evaluation. in 2014 IEEE International Electron Devices Meeting 30.4.1–30.4.4 (2014). doi:https://doi.org/10.1109/IEDM.2014.7047142

4. Chhowalla, M., Jena, D. & Zhang, H. Two-dimensional semiconductors for transistors. Nat. Rev. Mater. 1, 16052 (2016).

3. Novoselov, K. S.
2. Laturia, A., Van de Put, M. L. & Vandenberghe, W. G. Dielectric properties of hexagonal boron nitride and transition metal dichalcogenide MOSFET in the Ultimate Physical Scaling Limit. IEEE Electron Device Lett. 35, 402–404 (2014).

2. Laturia, A., Van de Put, M. L. & Vandenberghe, W. G. Dielectric properties of hexagonal boron nitride and transition metal dichalcogenide from monolayer to bulk. NPJ 2D Mater. Appl. 2, 6 (2018).

1. Park, S. & Akinwande, D. First demonstration of high performance 2D monolayer transistors on paper substrates. in 2019 Symposium on VLSI Technology T244–T245 (2019). doi:https://doi.org/10.23919/VLSIT.2019.8776498.

0. Majumdar, K., Hobbs, C. & Kirsch, P. D. Benchmarking Transition Metal Dichalcogenide MOSFET in the Ultimate Physical Scaling Limit. IEEE Electron Device Lett. 35, 402–404 (2014).
46. Verreck, D. et al. The role of nonidealities in the scaling of MoS2 FETs. IEEE Trans. Electron Devices 65, 4635–4640 (2018).

**Author contributions**
G.A. and Q.S. conceived the experiments. G.A., with support from Q.S. and S.S., performed the device fabrication, characterization and analysis. G.A. and D.V. performed the TCAD simulations. Z.A. performed the PPA analysis. C-V measurements and analysis were performed by A.G., with support from D.L. The CVD MoS2 was grown by B.G. and e-beam litho support was provided by J.J. Prof. M.H., I.A. and I.R. managed the overall activities and allocation of resources under the beyond CMOS program at imec, Belgium. The manuscript was written by G.A. with contributions from Q.S. and D.V. The manuscript was revised by all co-authors.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-85968-y.

Correspondence and requests for materials should be addressed to G.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021