nature portfolio

Peer Review File

Wafer-scale high-κ HfO2 dielectric films with sub-5-Å equivalent oxide thickness for 2D MoS_2 transistors

Corresponding Author: Professor Guangyu Zhang

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have revised the manuscript and addressed some of the questions raised by the reviewers.
In general, I am good with the technical discovery of the paper. However, I am not very happy with the writing style of this paper which tries to over-claim too much. In particular, the authors mixed up the discussions of long and short channel devices. I could not find the SS for short channel FET in the paper but the authors keep emphasizing SS~ 60 that I believe that should only bet he signatures of their long channel devices.
IRDS roadmap for future devices should also refer to short-channel FET. The long channel SS ~62 is an irrelevant argument. The authors should try to clarify this in their manuscript, ensuring not misleading the readers.

Reviewer #4

(Remarks to the Author)
It is great that the authors have included top gated device performance in this version of the manuscript. Some questions need to be addressed before the manuscript can be recommended for publication:
1. Figure 3 - back gated MoS2 FETs, Vt is negative, which is acceptable according to other published work. For Figure 4, how were the FET characteristics obtained? What was the voltage on the back gate Si/SiO2? Why is the Vt with top gate more positive than with back gate?
2. Compare the current at the same Vg-Vt for back gate and top gate configurations. Seems like with top gate, the current is higher. Why? Is the MoS2 getting doped by Sb or the MOALD process?

Reviewer #5

(Remarks to the Author)
In the manuscript “Sub-3-A EOT High-k Dielectrics for Advanced Transistors”, the authors claim to demonstrate a gate dielectric with sub 3A EOT via an ALD process at 200C. The work is relevant, timely and highly interesting for the field of ultra-scaled nanoelectronics based on 2D materials as channels. However, there are several central points that are presented in a misleading way, which render the manuscript unsuitable for publication in its present form. A scholarly clear presentation that accurately states the key performance metrics (including statistics and error bars) and makes the work reproducible is essential. There are several major issues:
1) The central claim of 2.5A EOT is based on capacitance-voltage measurements on Au-HfO2-Au MIM structures. However, in particular for ultra-thin gate dielectrics that should be used as gate dielectrics with van der Waals bonded 2D materials, like MoS2, the effective CET and EOT that will be measured on a MIS structure (i.e. Au-HfO2-MoS2) will be substantially different from the EOT measured on a MIM structure (i.e. Au-HfO2-Au) (see also Fig. 6a in A. Gaur et al., 2D Materials, 2019, 10.1088/2053-1583/ab20fb). However, what really matters for an ultra-scaled 2D field-effect transistor is the CET and EOT achieved in the actual gate stack, i.e. for the Au-HfO2-MoS2 system, as was also pointed out by reviewer 3. Hence, please report the following:
a) Measure capacitance(voltage) (C(V)) curves on the MIS stacks used for the back gated FETs in Fig. 3. Ideally report on

the C(V) for different device areas and measure the capacitance for frequencies from kHz to MHz, for at least 5 different devices. Then please report on the average CET and EOT for these samples including error-bars.
b) Measure C(V) curves on the MIS stacks used for the top gated FETs presented in Fig. 4. Again, report the C(V) curves for different device areas, for varying frequencies from kHz to MHz, for at least 5 different devices. Please report on the average CET and EOT for these samples.
c) In Supplementary Fig. 13 you show a C(V) measurement. Is this for top or for back gated devices? What is the device area? Please also show the curve of capacitance (in microF) as a function of gate voltage before normalizing to the area. Are there overlap capacitances of the gate contact with source and drain contacts included? Please show an optical microscopy image where you measure the active area of the device that is used for normalization. The estimated CET is between 3.1A and 4.3A depending on the sweep frequency.
d) Please create an overview Fig. like now presented in Fig. 2d where the CET and EOT values as measured on the MIM, and the 2 different MIS stacks (top and back gated) are compared directly (a Figure comparable to Fig. 6a in Gaur et al., 2D Materials, 10.1088/2053-1583/ab20fb).
e) Please use in the title and consistently throughout the manuscript the EOT value that is consistently achieved for the MIS stacks in both the top gate and the back gate design. Hence, if e.g. an EOT of 0.5nm can be reached consistently, please use sub-5A EOT.
f) Typically, the CET that is measured in a MIS stack is larger than the EOT, as there is an additional van der Waals gap between the 2D MoS2 and the HfO2. How large is the van der Waals gap between the MoS2 and the HfO2? What is the effective capacitance of the van der Waals gap Cvdw? Typically the van der Waals gap is on the order of about 0.3nm, much larger than the 0.05nm claimed in the rebuttal, see also Zhang et al., Nature Electronics 2022, 10.1038/s41928-022-00824-9.
g) Finally, also the quantum capacitance of monolayer MoS2 limits the overall measured gate capacitance in highly scaled gate stacks, see also R. Bennett et al., Nano Letters 2023, 10.1021/acs.nanolett.2c03913. When using a MoS2 channel the gate capacitance is reduced due to the quantum capacitance. Also, I would heavily dispute the claim of the authors “CET considers more complicated conditions related to semiconductor interface and quantum capacitance.” In fact, EOT is an ideal number that refers to the dielectric constant and the physical thickness of the insulator, whereas CET reflects the actual physical gate control aimed for. Hence, CET is the number specified in the IRDS, that should be smaller than 0.9nm for advanced technology nodes (see https://irds.ieee.org/editions/2023/20-roadmap-2023-edition/130-irds%E2%84%A2-2023-more-moore).
2) What is the thickness of the gate insulator used for the top gate configuration presented in Fig. 4? About 1nm SbOx and about 2nm HfO2? Please indicate the layer thicknesses together with the probable dielectric constants of both components of the gate stack (SbOx and HfO2) as well as the CET. The CET and EOT for the combined gate stack is certainly larger than 0.5nm - which is fine, as long as it is clearly stated.
3) As pointed out by Reviewer 2, comment 3, the previously reported DFT calculations based on PBE functionals severely underestimated the HfO2 band gap. The authors use DFT calculations to study the impact of oxygen vacancies on the band structure. However, this methodology has one major flaw. In fact they are investigating the impact of oxygen vacancies on crystalline HfO2. In reality the HfO2 films investigated are (as most HfO2 dielectrics used) amorphous. In order to study amorphous films with DFT one needs to use packages that can handle large supercell calculations like CP2K and a melt and quench method, see for example J. Strand et al., J. Phys. Cond. Matter 2018, 10.1088/1361-648X/aac005. I do not think that the DFT calculations performed for crystalline HfO2 have any meaning for the amorphous HfO2 films investigated. Hence, please either repeat your calculations for amorphous HfO2 supercells or completely remove the paragraph about oxygen vacancies and the DFT data from the supplementary information.
4) As pointed out by Reviewer 2, comment 4 the authors talk about FN vs direct tunneling but do not show any modeled leakage currents. Please include a calculated tunnel current based on FN tunneling, direct tunneling or a combination of both of them according to a Tsu-Esaki model, see also Tsu et al., Applied Phys, Letters 1973, 10.1063/1.1654509 or T. Knobloch et al., Nature Electronics 2021, 10.1038/s41928-020-00529-x. Please calculate the tunneling currents using a model for direct tunneling. In this way the authors will also see that the statement “The direct tunneling current is suppressed below the detection limit due to the large bandgap and high-quality of the sample.” is just plainly wrong. The direct tunnel currents depend a lot on the dielectric thickness and the effective tunnel masses (see for example F. Sacconi et al., IEEE TED 2007, 10.1109/TED.2007.908880) and will be sizable for a film of only 1.3nm thickness - even in the theoretical case without any defects, or trap assisted tunneling.
5) In Figure 2c for evaluating the capacitance of MIM structures the C(V) curves should be shown at least up to +- 0.6V. Ideally the capacitance on at least 5 devices should be compared as well as the leakage currents on 5 MIM structures in a range from +- 0.6V.
6) As pointed out by Reviewer 2, comment 12, please show a cross section of your devices, including all the dimensions. The Fig 3a in the new manuscript includes an artistic representation but I am still missing many important device dimensions, i.e. contact length, overall gated area, BG Au thickness, HfO2 thickness used for devices, source/drain dimensions (i.e. width, length, thickness), overlap areas of contacts with back gate etc. Please include an optical microscopy image and/or a simple 2D schematic including all important dimensions.
7) The subthreshold swing that the authors report should be averaged over at least 2 decades of current, see also Z. Cheng et al. Nature Electronics 2022, doi:10.1038/s41928-022-00798-8. Taking the minimum value of SS at a small drain current of 1e-4 uA/um is meaningless. Please average over 2-3 decades in Fig. 3e and update Fig. 3g accordingly.
8) I do not see any evidence presented for the allegedly negligible hysteresis. Please show a double-sweep IdVg for a range of different sweep frequencies. Also the hysteresis is typically evaluated close to the threshold voltage. As the hysteresis depends strongly on the sweep frequency a small hysteresis at one frequency cannot give a meaningful conclusion about the hysteresis further away. Please evaluate the hysteresis width as a function of the sweep frequency spanning at least 2 orders of magnitude in sweep frequency, see for example Illarionov et al., npj 2D Mater. Appl. 2024, 10.1038/s41699-024-00445-0.

And a few minor issues:
9) When responding to reviewer 2, question 7, the authors included in the rebuttal FigR12 c and d. In particular Fig. R12d should be included either in the main manuscript or the supplementary including a scale bar. What is the width and length of the devices used for the circuits? What is the active area covered by the back gate and the MoS2?
10) The optical images shown in Fig. 5e, 5h and 5j are taken with a large magnification. It would be much more relevant to use a magnification where one can see one cell of an inverter, one cell of an AND logic gate etc., hence please zoom in more and include these images instead.

Version 1:

Reviewer comments:

Reviewer #5

(Remarks to the Author)
In the manuscript “Sub-5-Å EOT High-κ Dielectrics for Advanced Transistors” the authors present an extremely scaled gate dielectric. During the previous review round, the authors provided comprehensive additional C(V) measurement data to improve the reproducibility of this very relevant and timely study. In general I think that the study improved considerably by the revisions made. However, I agree with the other reviewers that still I am not happy with the language and the metrics used to present the paper. I think the authors are unnecessarily overselling the performance of their devices, thereby misleading the readers. In particular, please address the following issues:
1) As mentioned by reviewer 1 the long channel minimum subthreshold slope of ~62mV/dec is irrelevant. The relevant metric is the averaged subthreshold slope over at least 2 orders of magnitude (better 5 or more) for the short channel devices, see the IRDS Roadmap (https://irds.ieee.org/editions), which is about 110mV/dec based on the data shown in Fig. 16. Please remove the claim of 60mV/dec. from the abstract. (p1, line 37). The same holds true for Fig. 4k - please compare instead of the minimum subthreshold slope, the average subthreshold slope, as this is the relevant and well reproducible metric. Please also remove the ideal SS claim from line 272 and instead give the average SS.
2) Can you please briefly comment on the more positive Vt for the top gated FETs in comparison to back gated FETs and the comparable albeit slightly smaller on current level for top gated FETs in the main text? I think the answers to the points raised by Referee 4 are accurate and sufficient but should be briefly, ie. in one sentence each be reflected in the main text.
3) I think that all the C(V) data reported now in the Supplementary in response to the previous review round is a major improvement of the manuscript. However, these changes are mostly presented in the Supplementary, I would suggest the following changes in the main manuscript. Please include as additional figures in Fig. 2 the C(V) data, for example, please add Supplementary Fig. 13 a and f as well as Supp. Fig. 20. As stated before, the CET value is much more relevant than the small EOT. As the small CET/EOT is one of the main achievements of the work presented, I would argue that supplementary Fig. 20 is from my perspective one of the most important figures of the whole manuscript and should be presented alongside some of the raw C(V) curves that were evaluated to create this Figure in the main manuscript.

Referee #1 (Remarks to the Author):

The authors have revised the manuscript and addressed some of the questions raised by the reviewers.

In general, I am good with the technical discovery of the paper. However, I am not very happy with the writing style of this paper which tries to over-claim too much. In particular, the authors mixed up the discussions of long and short channel devices. I could not find the SS for short channel FET in the paper but the authors keep emphasizing SS~ 60 that I believe that should only be the signatures of their long channel devices.

IRDS roadmap for future devices should also refer to short-channel FET. The long channel SS ~62 is an irrelevant argument. The authors should try to clarify this in their manuscript, ensuring not misleading the readers.

Response:

We thank the referee for their critical feedback. We have corrected some exaggerated statements in the revised manuscript, such as SS is 60 mV/dec. According to the referee's suggestion, we have also included the SS data in short channel devices, as shown in the below Fig. R1. In this device, the channel length is 100 nm. From 10^{-4} μA/μm to 10^{-2} μA/μm, The SS value is ~70 mV/dec as shown in the Fig. R1 and *Supplementary Fig. 16* in line 29, page 7 in revised manuscript: "The sweeping measurements (Fig. 3h) also indicate an exceedingly small hysteresis (<10 mV) and low SS value (70 mV/dec, Supplementary Fig. 16)."

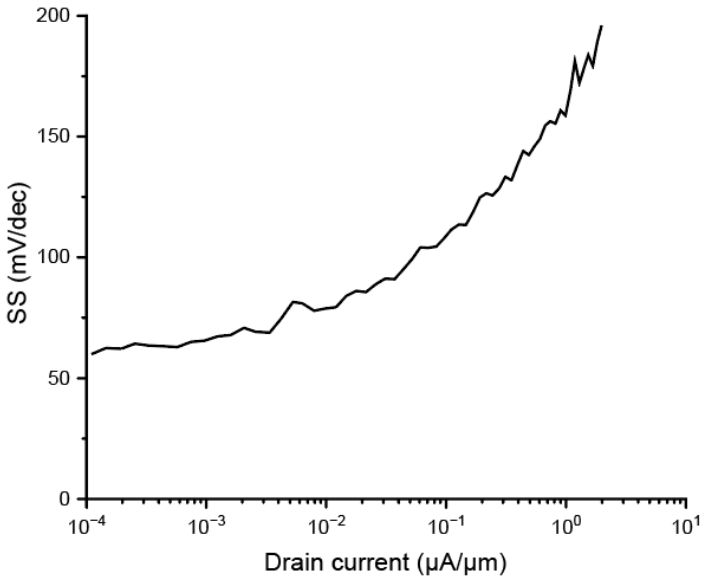

Fig. R1 | Corresponding subthreshold swing curve with 0.1 V source-drain bias in 100-nm channel length MoS_2 FETs.

Referee #4 (Remarks to the Author):

It is great that the authors have included top gated device performance in this version of the manuscript. Some questions need to be addressed before the manuscript can be recommended for publication:

Response:

We thank the referee for their positive comment on the inclusion of the top-gated device data. We have addressed all the specific questions raised below point-by-point in the revised manuscript and believe these changes have further improved the quality of our work.

1. Figure 3 - back gated MoS_2 FETs, V_t is negative, which is acceptable according to other published work. For Figure 4, how were the FET characteristics obtained? What was the voltage on the back gate Si/SiO_2? Why is the V_t with top gate more positive than with back gate?

Response:

We thank the referee for these insightful questions. Regarding the measurement configuration for the top-gated FETs in Figure 4, the device characteristics were obtained by applying a voltage solely to the top gate (Au), with the SbO_x/HfO_2 bilayer serving as the gate dielectric. The back gate Si/SiO_2 was electrically floated (i.e., no voltage was applied) during these measurements to ensure that the recorded performance reflects the electrostatic control of the top gate only.

The observed positive shift in the V_t for the top-gated device, compared to the back-gated one, is a key finding. We attribute this shift primarily to the incorporation of oxygen doping within the MoS_2 channel during the fabrication of the top-gate stack. This intentional doping introduces n-type carriers, which effectively shifts the turn-on voltage in the positive direction. This method of threshold voltage modulation via channel doping is well-established, as supported by prior literature (e.g., *Small* 16, 2004276 (2020)).

2. Compare the current at the same V_g-V_t for back gate and top gate configurations. Seems like with top gate, the current is higher. Why? Is the MoS_2 getting doped by Sb or the MOALD process?

Response:

We thank the referee for their comment. Output curves for back and top-gated transistors in Fig. 3d and Fig. 4h show that the current for back-gate transistors (25 μA/μm) is larger than that for top-gated transistors (15 μA/μm) under the same 0.5 V source voltage and 1 V V_g-V_t. The reason may come from the small contact resistance variation (*Nature* ***613****, 274–279 (2023)*) or gate controllability difference (0.34 nm CET for back-gated devices and 0.50 nm CET for top-gated devices).

Referee #5 (Remarks to the Author):

In the manuscript "Sub-3- Å EOT High-k Dielectrics for Advanced Transistors", the authors claim to demonstrate a gate dielectric with sub 3Å EOT via an ALD process at 200℃. The work is relevant, timely and highly interesting for the field of ultra-scaled nanoelectronics based on 2D materials as channels. However, there are several central points that are presented in a misleading way, which render the manuscript unsuitable for publication in its present form. A scholarly clear presentation that accurately states the key performance metrics (including statistics and error bars) and makes the work reproducible is essential. There are several major issues:

Response:

Thank the referee for highlighting the need for greater clarity and precision in our manuscript. We sincerely apologize for any misleading impressions in our original presentation. We have thoroughly revised the manuscript to address these concerns with utmost seriousness.

1) The central claim of 2.5Å EOT is based on capacitance-voltage measurements on Au-HfO_2-Au MIM structures. However, in particular for ultra-thin gate dielectrics that should be used as gate dielectrics with van der Waals bonded 2D materials, like MoS_2, the effective CET and EOT that will be measured on a MIS structure (i.e. Au-HfO_2-MoS_2) will be substantially different from the EOT measured on a MIM structure (i.e. Au-HfO_2-Au) (see also Fig. 6a in A. Gaur et al., 2D Materials, 2019, 10.1088/2053-1583/ab20fb). However, what really matters for an ultra-scaled 2D field-effect transistor is the CET and EOT achieved in the actual gate stack, i.e. for the Au-HfO_2-MoS_2 system, as was also pointed out by reviewer 3. Hence, please report the following:

Response:

We acknowledge careful review by the referee. We totally agree that the CET and EOT achieved in the actual gate is that really matters.

a) Measure capacitance(voltage) (C(V)) curves on the MIS stacks used for the back gated FETs in Fig. 3. Ideally report on the C(V) for different device areas and measure the capacitance for frequencies from kHz to MHz, for at least 5 different devices. Then please report on the average CET and EOT for these samples including error-bars.

Response:

We thank the referee for this insightful and constructive suggestion. We agree that a detailed characterization of the MIS (Metal-Insulator-Semiconductor) capacitor stacks is crucial for a complete understanding of the device performance presented in Fig. 3. According to the referee's suggestion, we have performed comprehensive Capacitance-Voltage (C-V) measurements on these structures.

The C-V measurements were performed on MIS capacitor structures. The voltage was swept from accumulation to inversion (from 0 V to 1 V for MoS_2 FETs with 1.3 nm HfO_2 dielectrics; from 0

V to 1.5 V for MoS_2 FETs with 2.6 nm HfO_2 dielectrics; from 0 V to 2 V for MoS_2 FETs with 3.9 nm HfO_2 dielectrics). Measurements were conducted on 5 devices with varying areas (from 40 μm^2 to 200 μm^2 depending on devices) at multiple frequencies ranging from 5 kHz to 1 MHz. The representative C-V characteristics are presented in Fig. R1~R3 and *Supplementary Fig. 13-15* in the revised Supplementary Materials.

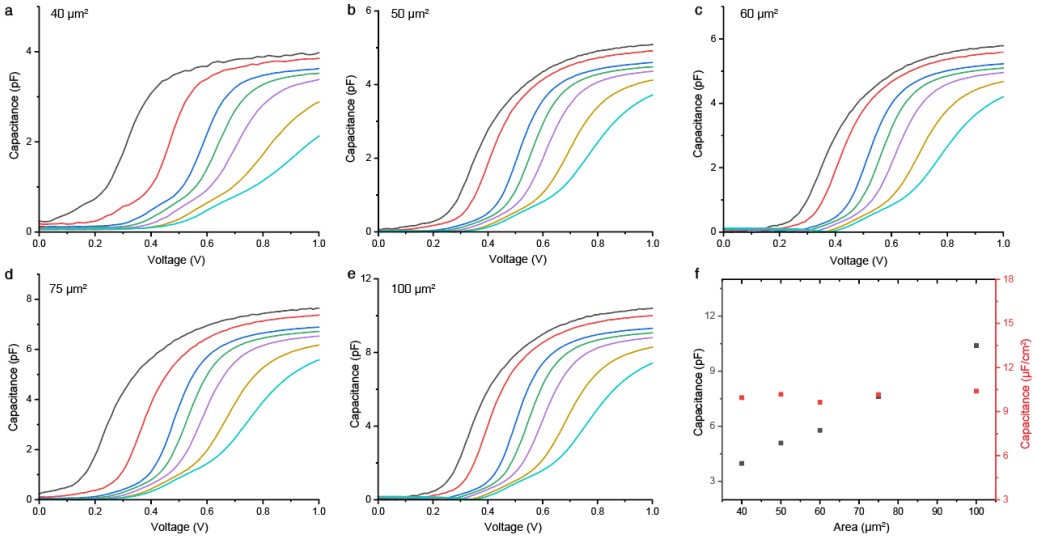

Fig. R1 | Effective capacitance tested in MoS_2 gate-first transistors with 1.3 nm HfO_2 dielectrics. Frequencies vary from 5 kHz to 1 MHz in devices with various areas from (a) 40 μm^2 to (e) 200 μm^2. (f) Statistic capacitance and unit capacitance for all five devices

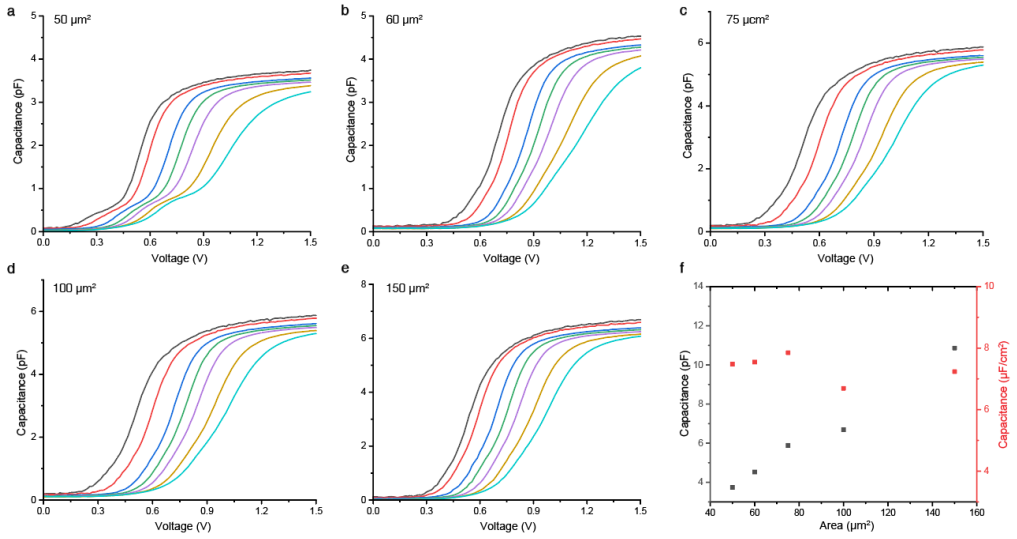

Fig. R2 | Effective capacitance tested in MoS_2 gate-first transistors with 2.6 nm HfO_2 dielectrics. Frequencies vary from 5 kHz to 1 MHz in devices with various areas from (a) 40 μm^2 to (e) 200 μm^2. (f) Statistic capacitance and normalized capacitance for all five devices

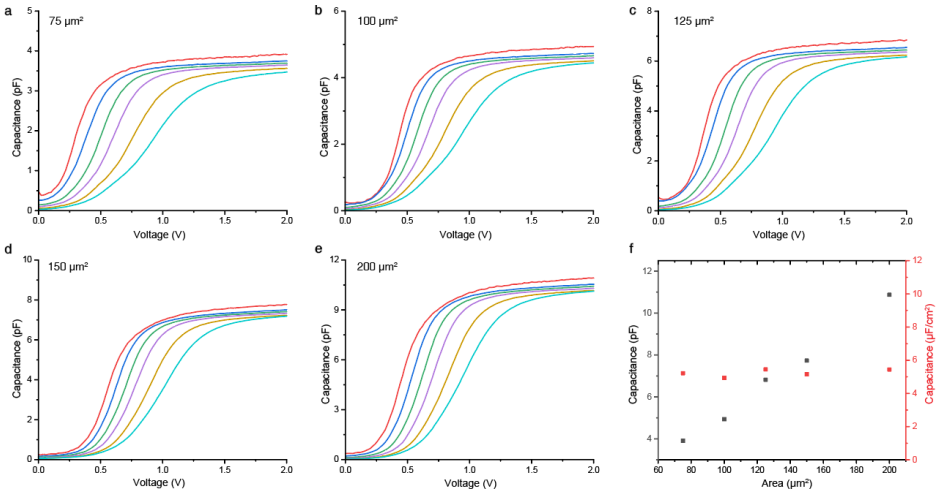

Fig. R3 | Effective capacitance tested in MoS_2 gate-first transistors with 3.9 nm HfO_2 dielectrics. Frequencies vary from 5 kHz to 1 MHz in devices with various areas from (a) 40 μm^2 to (e) 200 μm^2. (f) Statistic capacitance and normalized capacitance for all five devices

From the accumulation capacitance (C_{ox}) measured at 1 MHz, we calculated the CET using the formula for a parallel plate capacitor:

$$CET = C_{ox}\epsilon_0 / A$$

where ϵ_0 is the vacuum permittivity (8.854×10^{-12} F/m) and A is the capacitor area.

The extracted CET values for 5 representative devices are summarized in **Table R1**.

Oxide Thickness	Area (μm^2)	C_{ox} @ 5 kHz (pF)	CET (nm)	Average (nm)	Std. Dev. (nm)
1.3	40	3.98	0.347	0.3432	0.0098
1.3	50	5.09	0.339		
1.3	60	5.78	0.358		
1.3	75	7.62	0.34		
1.3	100	10.4	0.332		
2.6	50	3.74	0.462	0.4704	0.0287
2.6	60	4.53	0.457		
2.6	75	5.89	0.44		

2.6	100	6.69	0.516		
2.6	150	10.86	0.477		
3.9	75	3.91	0.662	0.6596	0.0272
3.9	100	4.94	0.699		
3.9	125	6.82	0.633		
3.9	150	7.74	0.669		
3.9	200	10.88	0.635		

Table R1 Summary of extracted parameters from C-V measurements on five different MIS capacitors.

These measurements show an average CET of 0.34 ± 0.01 nm for 1.3 nm HfO_2 dielectrics, 0.47± 0.03 nm for 2.6 nm HfO_2 dielectrics, 0.66± 0.03 nm for 3.9 nm HfO_2 dielectrics. The small standard deviation across multiple devices highlights the uniformity and reproducibility of our dielectric deposition process. We believe these additional data and analyses substantially strengthen the paper by providing a quantitative evaluation of the gate stack, which directly supports the high-performance FET characteristics presented. We once again thank the referee for their valuable advice.

b) Measure C(V) curves on the MIS stacks used for the top gated FETs presented in Fig. 4. Again, report the C(V) curves for different device areas, for varying frequencies from kHz to MHz, for at least 5 different devices. Please report on the average CET and EOT for these samples.

Response:

We thank the referee for their continued diligence and for the excellent suggestion to characterize the top-gated MIS stacks. This provides a direct assessment of the gate module quality for the devices presented in Fig. 4. In response, we have conducted the requested C-V measurements on the top-gated structures at five devices with different areas, sweeping the frequency from 5 kHz to 1 MHz as shown in the following Fig. R4 and *Supplementary Fig. 15* in the revised Supplementary Materials.

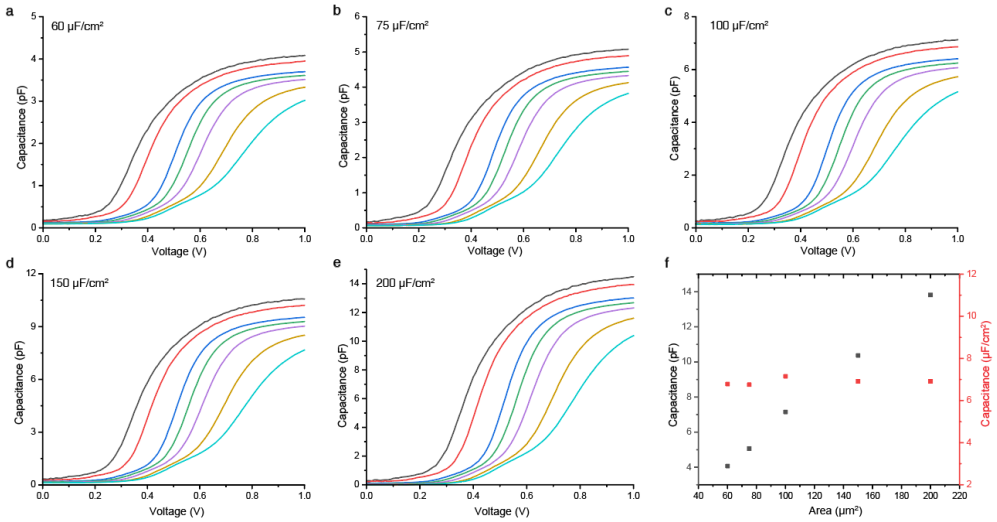

Fig. R4 | Effective capacitance tested in MoS_2 gate-last transistors with 1 nm SbO_x seeding layer and 1.3 nm HfO_2 dielectrics. Frequencies vary from 5 kHz to 1 MHz in devices with various areas from (a) 60 μm^2 to (e) 200 μm^2. (f) Statistic capacitance and normalized capacitance for all five devices

The extracted CET values for five representative devices are summarized in **Table R2**.

Oxide Thickness	Area (μm^2)	C_{ox} @ 5 kHz (pF)	CET (nm)	Average (nm)	Std. Dev. (nm)
2	60	4.07	0.509	0.5006	0.01106
2	75	5.07	0.511		
2	100	7.15	0.483		
2	150	10.37	0.500		
2	200	13.82	0.500		

Table R2 | Summary of extracted parameters from C-V measurements on five different MIS capacitors.

Based on these results, the average CET for the top-gated stack is 0.50 ± 0.01 nm. The consistent values obtained across the wafer underscore the robustness of our fabrication process. We are confident that this additional characterization provides a more complete and robust validation of our device quality. We thank the referee for pushing for this higher standard of evidence.

c) In Supplementary Fig. 13 you show a C(V) measurement. Is this for top or for back gated devices? What is the device area? Please also show the curve of capacitance (in microF) as a function of gate voltage before normalizing to the area. Are there overlap capacitances of the gate contact with source and drain contacts included? Please show an optical microscopy image where you measure the active area of the device that is used for normalization. The estimated CET is between 3.1Å and 4.3Å depending on the sweep frequency.

Response:

We thank the referee for these insightful and detailed questions regarding the C-V characterizations. The C-V measurement in the former *Supplementary Fig. 13* was indeed performed on a complete gate-first FET structure. As the referee noted, and as confirmed by the optical image in Fig. R5, this device configuration includes significant gate-to-source/drain overlap. We fully agree that the parasitic capacitance introduced by this overlap leads to an artificially inflated total capacitance measurement and, consequently, an underestimation of the true CET.

To address this critical point and accurately determine the intrinsic properties of our gate dielectric, we have fabricated and characterized new MIS capacitor test structures that are free from any overlap. The complete and rigorous characterization of these new devices is now presented in the revised manuscript and the above Figs. R1-R4.

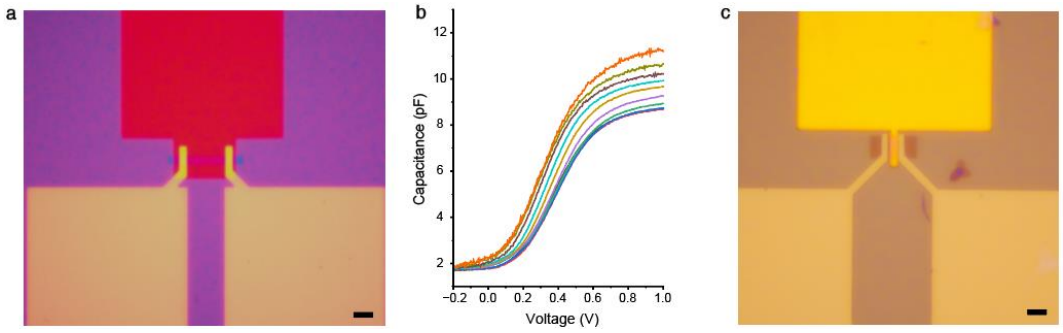

Fig. R5 | **a**, Optical image of back-gated MoS_2 transistors in Supplementary Fig. 13. **b**, C-V measurement of MIS structure before normalizing the area. **c**, Typical optical image of gate-source/drain overlap-free transistors for MIS C-V measurement in Figs. R1~R4. Scale bar: 10μm

d) Please create an overview Fig. like now presented in Fig. 2d where the CET and EOT values as measured on the MIM, and the 2 different MIS stacks (top and back gated) are compared directly (a Figure comparable to Fig. 6a in Gaur et al., 2D Materials, 10.1088/2053-1583/ab20fb).

Response:

Per the referee's recommendation, we have created a new overview figure (Fig. R6), which has been added to the revised Supplementary Materials as *Supplementary Fig.20.*

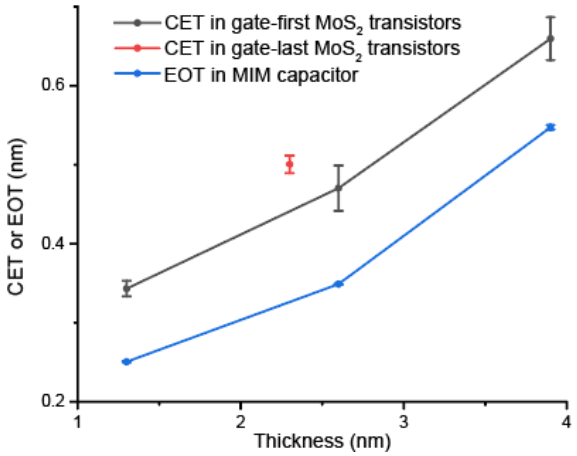

Fig. R6 | Overview of CET or EOT of ultrathin dielectrics in MIM capacitor and MIS transistors

e) Please use in the title and consistently throughout the manuscript the EOT value that is consistently achieved for the MIS stacks in both the top gate and the back gate design. Hence, if e.g. an EOT of 0.5nm can be reached consistently, please use sub-5Å EOT.

Response:

We thank the referee for this excellent suggestion to improve the consistency and impact of our manuscript. Based on the comprehensive C-V analysis presented for both our top-gated and back-gated MIS stacks, we can confirm that an EOT of approximately 0.5 nm is consistently achieved. Following the referee's valuable guidance, we have adopted the "sub-5Å EOT" metric to describe this key achievement.

f) Typically, the CET that is measured in a MIS stack is larger than the EOT, as there is an additional van der Waals gap between the 2D MoS_2 and the HfO_2. How large is the van der Waals gap between the MoS_2 and the HfO_2? What is the effective capacitance of the van der Waals gap C_{vdw}? Typically the van der Waals gap is on the order of about 0.3nm, much larger than the 0.05nm claimed in the rebuttal, see also Zhang et al., Nature Electronics 2022, 10.1038/s41928-022-00824-9.

Response:

We sincerely thank the referee for the expertise on this critical topic and for pointing out the important role of the van der Waals (vdW) gap. The referee is correct that the physical van der Waals (vdW) gap is typically on the order of 0.3 to 0.35 nm (3.0 to 3.5 Å).

Assuming the vdW gap is essentially a vacuum, its dielectric constant (k_{vdW}) is ≈ 1. We can then calculate its capacitance per unit area (C_{vdW}/A) and its contribution to the EOT (EOT_{vdW}).

$$C_{vdw}/A = k_{vdW}\,\epsilon_0/d_{vdW} = 2.95*10^{-2}\ \mu F/cm^2$$

$$EOT_{vdW} = d_{vdW}k_{SiO2}/k_{vdW} = 1.17\ nm$$

This analysis shows that the vdW gap makes a very significant contribution of ~1.17 nm to the total measured EOT. However, for accurate electrostatic and capacitance calculations, one must use the electrical-equivalent gap, which is significantly smaller. As demonstrated by Vaidya et al., the tails of the ionic potentials from the 2D material protrude into the interlayer gap, greatly reducing the effective electrical separation compared to the physical one (*Phys. Rev. Applied* ***10****, 034070*). The paper shows that a physical separation of 3.0 Å (0.3 nm) corresponds to an electrical-equivalent separation of just 0.6 Å (0.06 nm). This also explains why in some literature reports, despite the existence of vdW gap, CET is still less than 1nm. (*Nat. Electron.* **7**, 1126–1136 (2024), and *Nature* **605**, 262–267 (2022).)

g) Finally, also the quantum capacitance of monolayer MoS_2 limits the overall measured gate capacitance in highly scaled gate stacks, see also R. Bennett et al., Nano Letters 2023, 10.1021/acs.nanolett.2c03913. When using a MoS_2 channel the gate capacitance is reduced due to the quantum capacitance. Also, I would heavily dispute the claim of the authors "CET considers more complicated conditions related to semiconductor interface and quantum capacitance." In fact, EOT is an ideal number that refers to the dielectric constant and the physical thickness of the insulator, whereas CET reflects the actual physical gate control aimed for. Hence, CET is the number specified in the IRDS, that should be smaller than 0.9nm for advanced technology nodes (see https://irds.ieee.org/editions/2023/20-roadmap-2023-edition/130-irds%E2%84%A2-2023-more-moore).

Response:

We thank the referee for their comment. In line 10, page 2 in revised manuscript, we have modified the sentence as "EOT of a high-κ dielectric refers to the equivalent thickness of a SiO_2 film having the same specific capacitance with the high-κ dielectric film, while capacitance equivalent thickness (CET) reflects the actual physical gate controllability in devices."

2) What is the thickness of the gate insulator used for the top gate configuration presented in Fig. 4? About 1nm SbO_x and about 2nm HfO_2? Please indicate the layer thicknesses together with the probable dielectric constants of both components of the gate stack (SbO_x and HfO_2) as well as the CET. The CET and EOT for the combined gate stack is certainly larger than 0.5nm - which is fine, as long as it is clearly stated.

Response:

We thank the referee for this question. As shown in the line 11, page 8, we have demonstrated the vertical structure of the top-gated devices as "The total thickness of the device is only ~3 nm, including 0.7 nm MoS_2, 1 nm SbO_x seeding layer and 1.3 nm HfO_2, which is confirmed with CS-TEM and STEM-EDS mapping (Fig. 4e and 4f)." The CET is measured as ~ 0.50 nm and presented in the above Fig. R4. The dielectric constant is ~17.9.

3) As pointed out by Reviewer 2, comment 3, the previously reported DFT calculations based on PBE functionals severely underestimated the HfO_2 band gap. The authors use DFT calculations to study the impact of oxygen vacancies on the band structure. However, this methodology has one major flaw. In fact they are investigating the impact of oxygen vacancies on crystalline HfO_2. In reality the HfO_2 films investigated are (as most HfO_2 dielectrics used) amorphous. In order to study amorphous films with DFT one needs to use packages that can handle large supercell calculations like CP2K and a melt and quench method, see for example J. Strand et al., J. Phys. Cond. Matter 2018, 10.1088/1361-648X/aac005. I do not think that the DFT calculations performed for crystalline HfO_2 have any meaning for the amorphous HfO_2 films investigated. Hence, please either repeat your calculations for amorphous HfO_2 supercells or completely remove the paragraph about oxygen vacancies and the DFT data from the supplementary information.

Response:

Thanks for the reviewer's comments. HfO_2 serves as a high dielectric layer in CMOS and MIM systems. It was often studied in amorphous form. So, we recalculate the bandgaps and densities of states (DOS) of HfO_2 under amorphous structures. In previous results, we have compared the effects of HSE and PBE functionals to bandgaps in HfO_2. So here we directly use PBE functional to calculate the electronic properties. We find that the measured bandgaps keep the same variation law, which validates the reliability of the changes of bandgaps when increasing O vacancies concentrations. The bandgaps were measured with the help of inverse participation ratio (IPR) of densities of states, which can reflect the localization degree of densities of states (DOS).

$$IPR(\mathrm{eV}) = \frac{\sum_N DOS_i(eV)^2}{(\sum_N DOS_i(eV))^2}$$

Where $DOS_i(eV)$ represents the DOS of each atom. In amorphous structures, due to the disorder, the band edges often show localized electronic states (*Phys.Rev. B 102, 054205 (2020)., Phys. Rev. B. 111, 035203(2025), Phys. Rev. B. 89, 125201*). Compared with crystalline structure, the bandgap (E_g) becomes mobility bandgap (ME_g), and the valence band maximum and conduction band minimum are switched as valence band mobility edge (VBME) and conduction band mobility edge (CBME). Thus, here we use IPR to analyze the band edge of amorphous structure (Figs. R7~R10), Fig. 1e and *Supplementary Figs. 3~5* in revised manuscript.. In our analysis, the mid-energy points of the IPR within the valence and conduction bands were designated as the valence band mobility edge (VBME) and conduction band mobility edge (CBME), respectively. Consequently, the resulting mobility band edges and mobility bandgap, which were found to decrease systematically with increasing oxygen vacancy concentration (Tables R3-R8). The original conclusion in the part remains unchanged in revised manuscript.

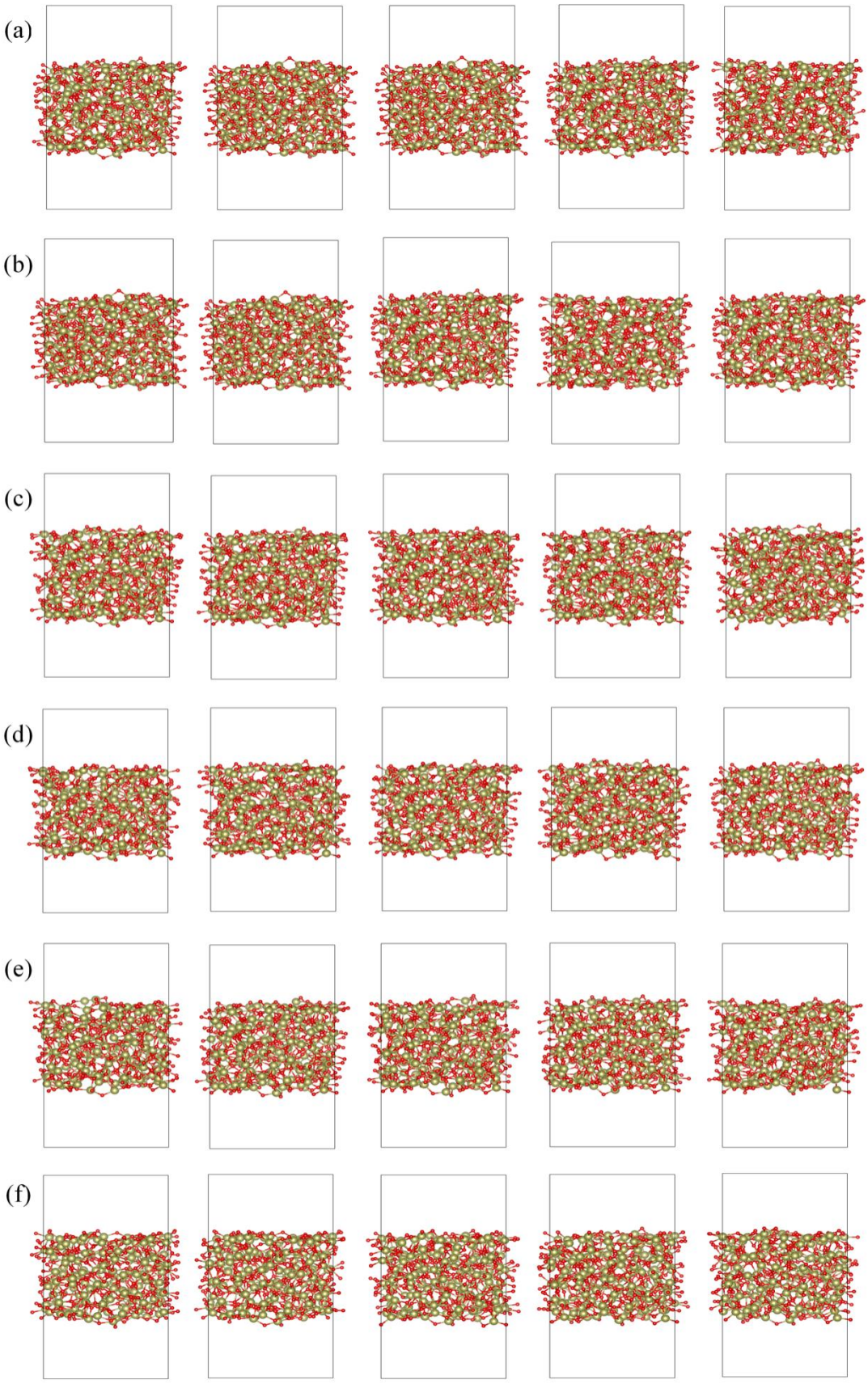
(a)
(b)
(c)
(d)
(e)
(f)

Fig. R7 | The amorphous structures of the HfO_2 slabs with different O vacancies concentrations. **a**, perfect, **b**, 2.2%, **c**, 4.4%, **d**, 7.2%, **e**, 10% and **f**, 15.6%.

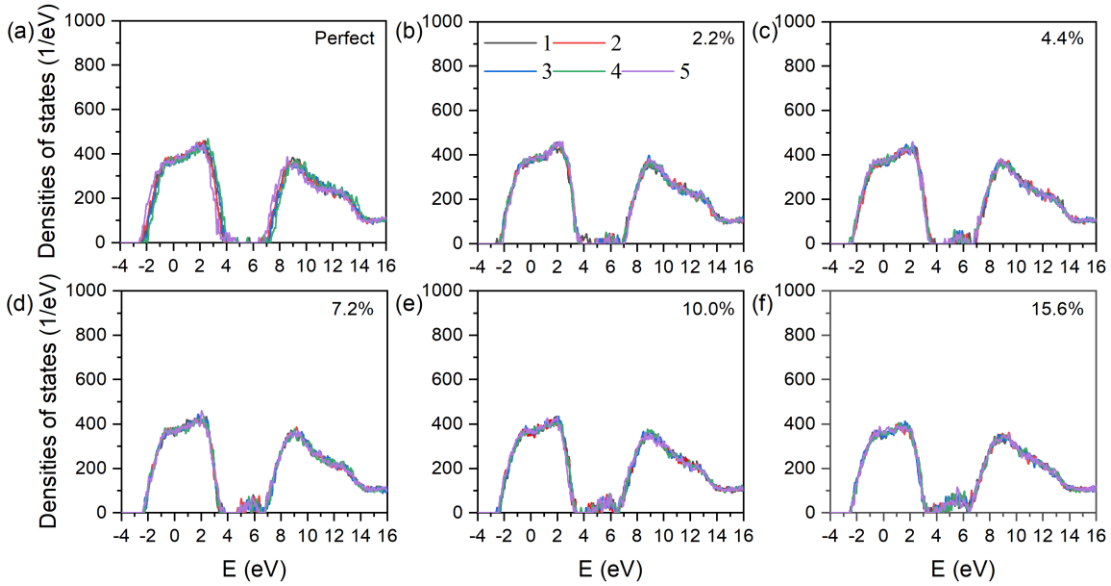

Fig. R8 | The density of states of the HfO_2 slabs at different O vacancies concentrations.

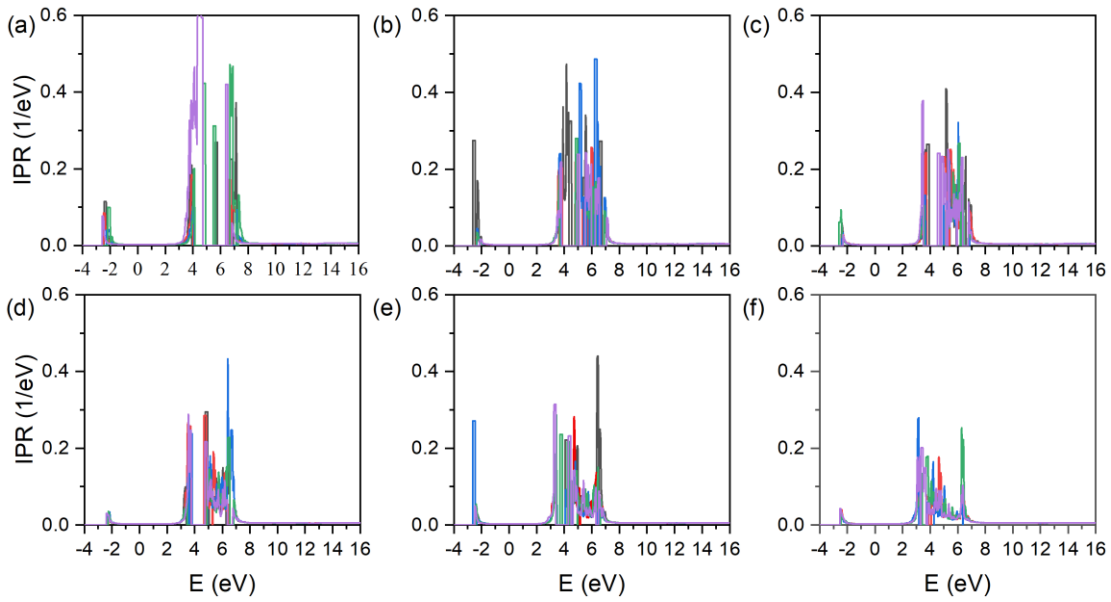

Fig. R9 | The Inverse participation ratio (IPR) value for the densities of states of the HfO_2 slabs at different O vacancies concentrations.

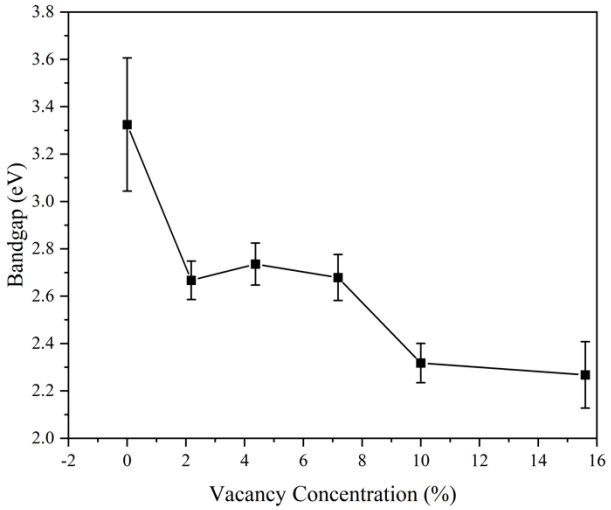

Fig. R10 | The bandgaps of amorphous HfO_2 slabs with different O_v defect concentrations calculated by considering PBE functional

Table R3 | the measured bandgaps from densities of states of the HfO_2 slabs without O vacancies concentrations.

Structure	VBME (PBE)	CBME (PBE)	ME_g (PBE)
1	3.535	7.095	3.560
2	3.665	7.060	3.395
3	3.800	7.405	3.605
4	4.285	7.260	2.975
5	3.945	7.035	3.090
Average	3.846	7.171	3.325

Table R4 | the measured bandgaps from densities of states of the HfO_2 slabs with 2.2% O vacancies concentrations.

Structure	VBME (PBE)	CBME (PBE)	ME_g (PBE)
1	3.545	6.195	2.650
2	3.455	6.095	2.640
3	3.415	6.215	2.800
4	3.535	6.115	2.580
5	3.530	6.195	2.665
Average	3.496	6.163	2.667

Table R5 | the measured bandgaps from densities of states of the HfO_2 slabs with 4.4% O vacancies concentrations.

Structure	VBME (PBE)	CBME (PBE)	ME_g (PBE)
1	3.565	6.305	2.740
2	3.490	6.195	2.705
3	3.375	6.135	2.760
4	3.330	6.190	2.860
5	3.285	5.900	2.615
Average	3.409	6.145	2.736

Table R6 | the measured bandgaps from densities of states of the HfO_2 slabs with 7.2% O vacancies concentrations.

Structure	VBME (PBE)	CBME (PBE)	ME_g (PBE)
1	3.355	6.030	2.675
2	3.355	5.940	2.585
3	3.410	6.140	2.730
4	3.255	6.070	2.815
5	3.405	5.995	2.590
Average	3.356	6.035	2.679

Table R7 | the measured bandgaps from densities of states of the HfO_2 slabs with 10% O vacancies concentrations.

Structure	VBME (PBE)	CBME (PBE)	ME_g (PBE)
1	3.300	5.645	2.345
2	3.335	5.695	2.360
3	3.360	5.555	2.195
4	3.425	5.705	2.280
5	3.240	5.650	2.410
Average	3.332	5.650	2.318

Table R8 | the measured bandgaps from densities of states of the HfO_2 slabs with 15.6% O vacancies concentrations.

Structure	VBME (PBE)	CBME (PBE)	ME_g (PBE)
1	3.230	5.355	2.125
2	3.050	5.380	2.330
3	2.905	5.385	2.480
4	3.130	5.350	2.220
5	3.175	5.360	2.185
Average	3.098	5.366	2.268

4) As pointed out by Reviewer 2, comment 4 the authors talk about FN vs direct tunneling but do not show any modeled leakage currents. Please include a calculated tunnel current based on FN tunneling, direct tunneling or a combination of both of them according to a Tsu-Esaki model, see also Tsu et al., Applied Phys, Letters 1973, 10.1063/1.1654509 or T. Knobloch et al., Nature Electronics 2021, 10.1038/s41928-020-00529-x. Please calculate the tunneling currents using a model for direct tunneling. In this way the authors will also see that the statement "The direct tunneling current is suppressed below the detection limit due to the large bandgap and high-quality of the sample." is just plainly wrong. The direct tunnel currents depend a lot on the dielectric thickness and the effective tunnel masses (see for example F. Sacconi et al., IEEE TED 2007, 10.1109/TED.2007.908880) and will be sizable for a film of only 1.3nm thickness - even in the theoretical case without any defects, or trap assisted tunneling.

Response:

We thank the reviewer for these insightful and constructive comments. We fully agree that the leakage current depends on multiple factors beyond the bandgap, including the effective tunneling mass, barrier height, and dielectric thickness. Based on our previous experience, we found that the WKB or Tsu–Esaki tunneling models are most reliable for relatively thick dielectric layers, where transport is dominated by bulk-like barrier transmission. In the present case with an ultrathin dielectric layer of ~1.3 nm, the tunneling process is strongly influenced by interfacial atomic configurations and the overlap between the electronic wavefunctions of the contact and dielectric layers. Such effects cannot be accurately captured by simple analytical models with homogeneous barrier parameters, which are typically calibrated for thicker systems.

A rigorous and quantitative description of tunneling in this ultrathin regime would require atomistic-level transport simulations (for example, first-principles nonequilibrium Green's function calculations) that explicitly include the contact–dielectric interfaces. These simulations are highly computationally demanding and beyond the current scope of this work, which primarily

focuses on the experimental demonstration and characterization of the dielectric properties. Nevertheless, we fully recognize the importance of such modeling and plan to pursue detailed atomistic simulations in our future studies to further understand the interface-dependent tunneling mechanisms.

In the present work, the experimentally observed suppression of leakage current is consistent with the large bandgap of the dielectric with minimal oxygen vacancies and the oxygen-rich surface, which effectively reduces wavefunction overlap at the contact–dielectric interface. We have clarified these points in the revised manuscript.

5) In Figure 2c for evaluating the capacitance of MIM structures the C(V) curves should be shown at least up to +- 0.6V. Ideally the capacitance on at least 5 devices should be compared as well as the leakage currents on 5 MIM structures in a range from +- 0.6V.

Response:

We thank the referee for this valuable feedback, which will certainly help improve the clarity and completeness of our results. We agree that providing a wider voltage range and a statistical analysis of device performance is essential. As shown in Fig. R11, Fig. 2c and *Supplementary Fig. 8a* in the revised manuscript, we have re-measured the MIM capacitance and leakages.

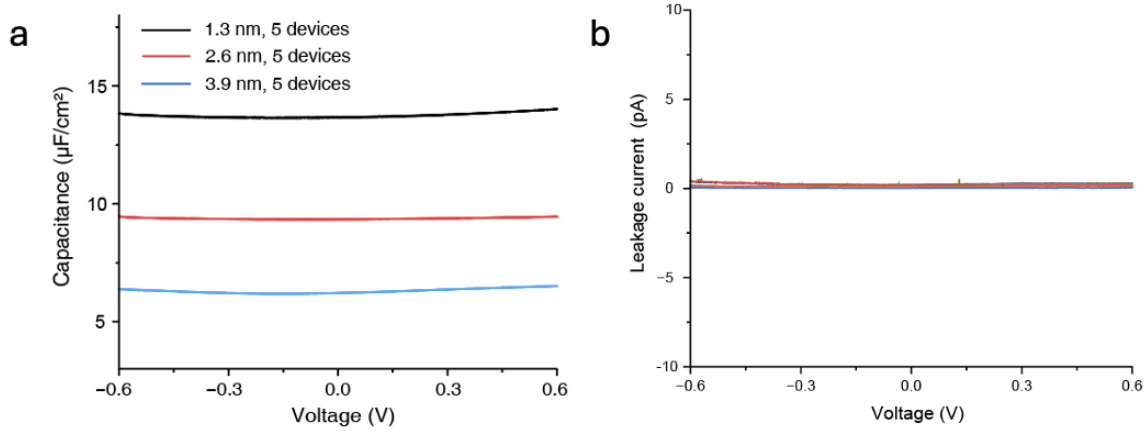

Fig. R11 | **a**, Electrical capacitance measurement and **b**, corresponding leakage current of 1.3 nm ~ 3.9 nm HfO_2 in physical thickness

6) As pointed out by Reviewer 2, comment 12, please show a cross section of your devices, including all the dimensions. The Fig 3a in the new manuscript includes an artistic representation but I am still missing many important device dimensions, i.e. contact length, overall gated area, BG Au thickness, HfO_2 thickness used for devices, source/drain dimensions (i.e. width, length, thickness), overlap areas of contacts with back gate etc. Please include an optical microscopy image and/or a simple 2D schematic including all important dimensions.

Response:

We thank the referee for this comment. We have added dimensions in revised device schematic illustration. Optical image is also shown in Fig. R12 and *Supplementary Fig. 11* revised Supplementary materials.

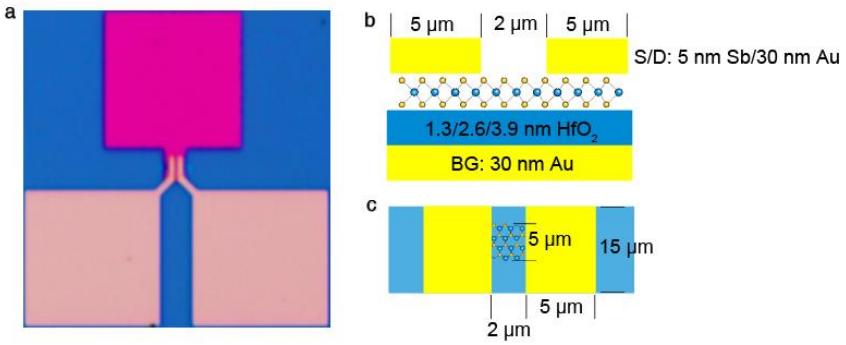

Fig. R12 | a, Optical image of back-gated MoS_2 transistor. Schematic illustration of b, side view and c, top view of the back-gated device in a. Contact length is 5 μm, overall gated area is 225 μm^2, BG Au thickness is 30 nm, HfO_2 thickness used for devices are 1.3 nm, 2.6 nm and 3.9 nm, respectively. Width, length, thickness for Source/drain contact are 5 μm, 2 μm and 35 nm. Overlap area of contacts with back gate is 150 μm^2.

7) The subthreshold swing that the authors report should be averaged over at least 2 decades of current, see also Z. Cheng et al. Nature Electronics 2022, doi:10.1038/s41928-022-00798-8. Taking the minimum value of SS at a small drain current of 1e-4 uA/um is meaningless. Please average over 2-3 decades in Fig. 3e and update Fig. 3g accordingly.

Response:

We have re-estimated and updated the SS value as shown in Fig. R13 and updated Fig. 3g in revised manuscript.

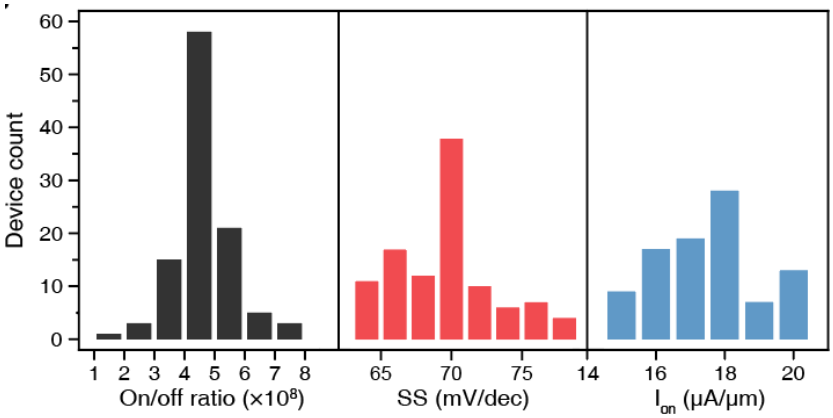

Fig. R13 | The corresponding statistical distribution of on/off ratio, SS, and on-state current

8) I do not see any evidence presented for the allegedly negligible hysteresis. Please show a double-sweep I_dV_g for a range of different sweep frequencies. Also the hysteresis is typically evaluated

close to the threshold voltage. As the hysteresis depends strongly on the sweep frequency a small hysteresis at one frequency cannot give a meaningful conclusion about the hysteresis further away. Please evaluate the hysteresis width as a function of the sweep frequency spanning at least 2 orders of magnitude in sweep frequency, see for example Illarionov et al., npj 2D Mater. Appl. 2024, 10.1038/s41699-024-00445-0.

Response:

Thank you for these valuable suggestions. We have included the hysteresis data under sweep frequency of 0.01 V/s, 0.01 V/s and 0.1 V/s, as shown in Fig. R14 and *Supplementary Fig. 12* in revised manuscript. A small hysteresis of 11 mV, 9 mV and 11 mV are obtained at current of 1nA, which is slightly dependent on the sweeping frequency.

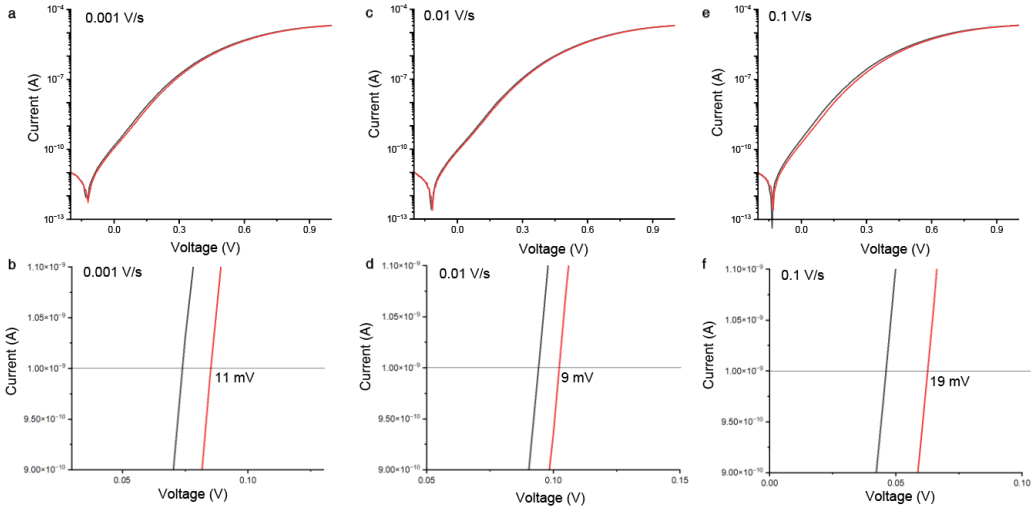

Fig. R14 | Bidirectional transfer curve of MoS_2 FETs with 2.5-Å-EOT MOALD HfO_2 with V_{ds} of 0.1 V at sweep frequencies of **a**, 0.001 V/s, **c**, 0.01 V/s and **e**, 0.1 V/s. **b**, **d**, **f**, Corresponding hysteresis at 1 nA for a, c and e, respectively.

And a few minor issues:

9) When responding to reviewer 2, question 7, the authors included in the rebuttal FigR12 c and d. In particular Fig. R12d should be included either in the main manuscript or the supplementary including a scale bar. What is the width and length of the devices used for the circuits? What is the active area covered by the back gate and the MoS_2?

Response:

We acknowledge the referee's feedback. We have added former Fig. R12c and R12d into *Supplementary Fig. 22* in revised manuscript (Fig. R15 below). The width and length of the devices for circuits are 10 μm and 30 μm, respectively. Back gate covered all active area and MoS_2, around 600 μm^2.

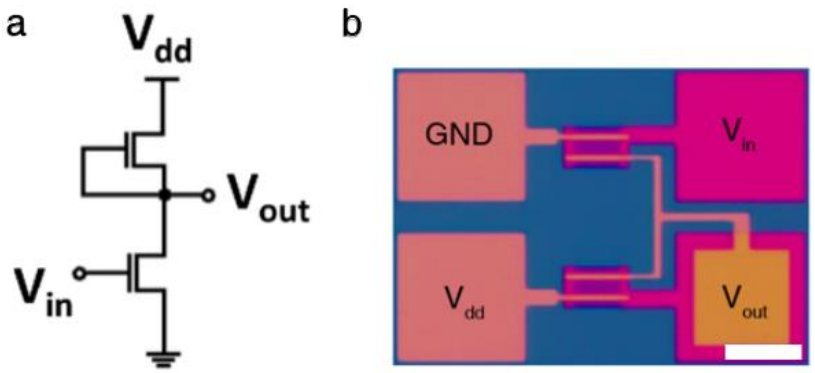

Fig. R15 | a, Equivalent circuit and **b**, true device for an all-NMOS inverter. scale bar: 50 μm

10) The optical images shown in Fig. 5e, 5h and 5j are taken with a large magnification. It would be much more relevant to use a magnification where one can see one cell of an inverter, one cell of an AND logic gate etc., hence please zoom in more and include these images instead.

Response:

We thank the referee's suggestions. We have added enlarging optical images in Fig. 5e, 5h and 5j in Fig. R16 and updated Fig. 5 the revised manuscript.

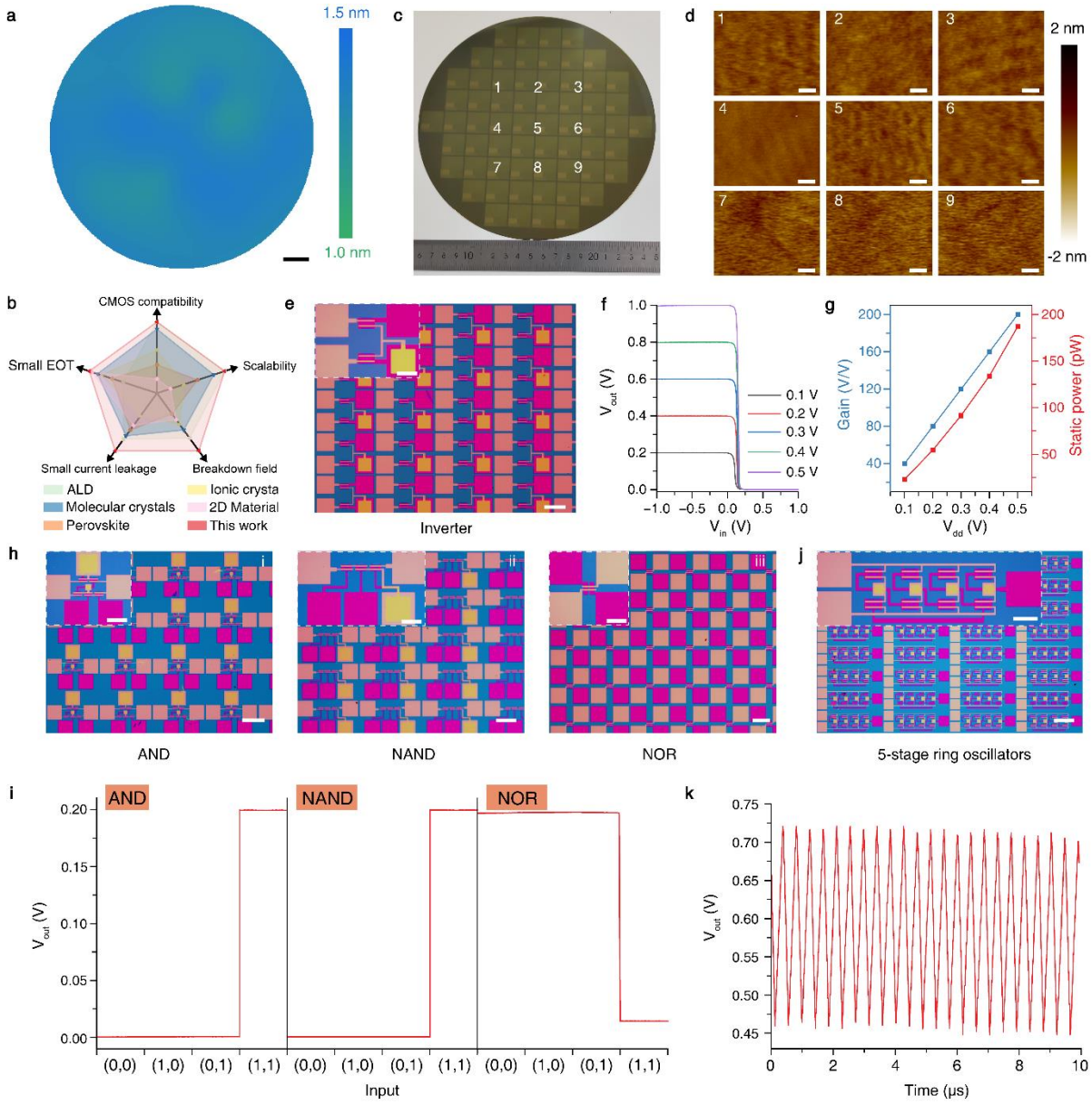

Fig. R16 | MoS_2 FETs and circuits at 2.5-Å-EOT and 8-inch wafer scale. a, Thickness mapping of 8-inch 1.3-nm HfO_2 wafer, scale bar 2 cm. **b**, A radar comparison between our and other dielectrics in terms of EOT, current leakage, breakdown field, scalability and CMOS compatibility. **c**, Photos of 8-inch MoS_2 FETs and circuits. **d**, AFM tomography of HfO_2 surface in 9 different positions in c, scale bar: 2 μm. **e**, Optical images of inverter at small (scale bar: 50 μm) and large magnification (scale bar: 100 μm). **f**, Output curve, **g**, voltage gain and static power of an inverter at supply voltage from 0.1 V to 0.5 V. **h**, Optical images at small (scale bar: 50 μm) and large magnification (scale bar: 100 μm) and **i**, Output characteristics of logic AND, NAND and NOR devices at supply voltage of 0.2 V. **j**, Optical images at small (scale bar: 50 μm) and large magnification (scale bar: 200 μm) and **k**, Output curve of a 5-stage ROs at supply voltage of 0.5 V.

Referee #5 (Remarks to the Author):

In the manuscript “Sub-5-Å EOT High-κ Dielectrics for Advanced Transistors” the authors present an extremely scaled gate dielectric. During the previous review round, the authors provided comprehensive additional C(V) measurement data to improve the reproducibility of this very relevant and timely study. In general, I think that the study improved considerably by the revisions made. However, I agree with the other reviewers that still I am not happy with the language and the metrics used to present the paper. I think the authors are unnecessarily overselling the performance of their devices, thereby misleading the readers. In particular, please address the following issues:

Response:

We sincerely thank the reviewer for acknowledging the improvements made in the previous round. We take your concerns regarding the language, metrics, and the presentation of our results very seriously. We realize that our previous enthusiasm for the results may have led to subjective phrasing that detracted from the scientific objectivity of the study.

In this revision, we have thoroughly rewritten the manuscript to adhere to a strictly objective tone. We have removed subjective qualifiers and "overselling" claims. Point-by-point responses are listed below.

1) As mentioned by reviewer 1 the long channel minimum subthreshold slope of ~62mV/dec is irrelevant. The relevant metric is the averaged subthreshold slope over at least 2 orders of magnitude (better 5 or more) for the short channel devices, see the IRDS Roadmap (https://irds.ieee.org/editions), which is about 110 mV/dec based on the data shown in Fig. 16. Please remove the claim of 60 mV/dec. from the abstract. (p1, line 37). The same holds true for Fig. 4k - please compare instead of the minimum subthreshold slope, the average subthreshold slope, as this is the relevant and well reproducible metric. Please also remove the ideal SS claim from line 272 and instead give the average SS.

Response:

We acknowledge that for short-channel devices, the minimum subthreshold slope (SS_{min}) does not adequately represent the device's switching performance. We agree that the average subthreshold slope (SS_{avg}), extracted over several decades of drain current, is the scientifically rigorous metric, consistent with the benchmarks set by the IRDS Roadmap.

To address this, we have revised the manuscript to strictly report SS_{avg} instead of SS_{min} for our short-channel devices. Specifically:

1. **Abstract (Page 1, Line 38):** We have removed the claim of 62 mV/dec (which corresponded to the minimum value). We now state the SS_{avg} value of 75 mV/dec, which represents the average slope over **3** decades of current.

2. **Figure 4i Revision:** We have redrawn Figure 4i. Instead of comparing the "best-case" SS_{min}, we now benchmark our device against the state-of-the-art using SS_{avg} (71 mV/dec) as the metric. This provides a fair and reproducible comparison.
3. **Line 221 Revision:** We have deleted the phrase "ideal SS". The text now reads: *" Fig. 4g shows the transfer curve of an as-fabricated gate-last MoS_2 transistor, revealing a high on/off ratio of 10^8 and average SS of 71 mV/dec"*

We believe these changes eliminate the ambiguity and "overselling" concerns, providing a transparent view of the device physics and performance.

2) Can you please briefly comment on the more positive Vt for the top gated FETs in comparison to back gated FETs and the comparable albeit slightly smaller on current level for top gated FETs in the main text? I think the answers to the points raised by Referee 4 are accurate and sufficient but should be briefly, ie. in one sentence each be reflected in the main text.

Response:

As suggested, we have incorporated the explanations provided in our previous response to Referee 4 directly into the main text. We have added two concise sentences to the revised manuscript in **Line 223** to clarify the physical origins of the positive V_{th} shift and the slightly reduced I_{on} in the gate-last configuration:

"Note that the V_{th} is more positive in gate-last transistors than that in gate-first transistors. We attribute this shift primarily to the incorporation of oxygen doping within the MoS_2 channel during the fabrication of the gate-last stacks. This doping introduces n-type carriers, which effectively shifts the turn-on voltage in the positive direction[7]. The output curves of this device is shown in Supplementary Fig. 19, illustrating a high current density of 15 μA/μm at 0.5 V source bias, which is slightly smaller than that in gate-first transistors. The reason may come from the small contact resistance variation[38] or gate controllability difference between 0.34 nm CET for gate-first devices and 0.50 nm CET for gate-last devices."

3) I think that all the C(V) data reported now in the Supplementary in response to the previous review round is a major improvement of the manuscript. However, these changes are mostly presented in the Supplementary, I would suggest the following changes in the main manuscript. Please include as additional figures in Fig. 2 the C(V) data, for example, please add Supplementary Fig. 13 a and f as well as Supp. Fig. 20. As stated before, the CET value is much more relevant than the small EOT. As the small CET/EOT is one of the main achievements of the work presented, I would argue that supplementary Fig. 20 is from my perspective one of the most important figures of the whole manuscript and should be presented alongside some of the raw C(V) curves that were evaluated to create this Figure in the main manuscript.

Response:

We appreciate this suggestion and fully agree. We recognize that the aggressive scaling of the dielectric (sub-5 Å EOT) is the central achievement of this work, and the supporting C-V measurements are indeed the most critical evidence.

Per your recommendation, we have significantly restructured **Figs. 3 and 4** in the main manuscript to include the key C-V characterization data:

1. **Raw C-V Data:** We have moved the representative raw C-V curves (previously Supplementary Figs. 13c, 13f, 14d and 15a) into the main text as **Figs. 3h~3k**. This allows readers to directly assess the quality and frequency dispersion of the dielectric capacitance.
2. **CET Scaling Trend:** We agree that the Capacitance Equivalent Thickness (CET) is the more direct and robust metric compared to the derived EOT. We have moved the CET vs. Thickness plot (previously Supplementary Fig. 20) to **Fig. 4j**. This figure now clearly demonstrates the scaling trend and validates the sub-5 Å EOT extraction alongside the measured CET values.

By moving these figures to the main text, we believe the manuscript now provides a much more compelling and transparent demonstration of the dielectric performance.