Two-dimensional floating-gate transistors with sub-10 nm MoS_2 channel for artificial synapses

Jiequn Sun(孙杰群), Peng Song(宋鹏), Chengze Du(杜承泽), Jijia Guo(郭济嘉), Haoyu Wang(王浩宇), Hui Guo(郭辉), Haitao Yang(杨海涛), Lihong Bao(鲍丽宏), Hong-Jun Gao(高鸿钧)
Citation: Chin. Phys. B. 35, 108505 (2026). DOI: 10.1088/1674-1056/ae85fd | CSTR: 32038.14.CPB.ae85fd
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Antiferromagnetic ordering and strain effects in monolayer quadrilateral FeS_2

Zhihao Li(李志豪), Junlin Wang(王君林), Guanqi Li(李冠祺), Yao Li(黎遥), Liang He(何亮), Jing Wu, Yongbing Xu, Xianyang Lu(陆显扬)
Chin. Phys. B. 35, 098503 (2026). DOI: 10.1088/1674-1056/ae1c1f | CSTR: 32038.14.CPB.ae1c1f

Enhancing thermoelectric performance in Janus MoTeS through periodic structural alternations and van der Waals contact

Yi-Ming Chen(陈一鸣), Shi-Hua Tan(谭仕华), Xuan-Hao Cao(曹煊浩), Yan-Hong Zhou(周艳红)
Chin. Phys. B. 35, 078501 (2026). DOI: 10.1088/1674-1056/ae1203 | CSTR: 32038.14.CPB.ae1203

Work function engineering of MXene contacts for high-performance, self-powered AlGaN solar-blind photodetectors

Pan Dai(代盼), Dengshan Cai(蔡登山), Wenxian Yang(杨文献), Ying Gu(顾颖), Haowen Hua(华浩文), Mengyang Huang(黄梦洋), Peng Zhang(张鹏), Xueyan Feng(冯雪雁), Sijia Wei(魏思嘉), Zheng Zhong(钟政), Yi Gong(龚毅), Jianjun Zhu(朱建军), Shan Jin(金山), Shulong Lu(陆书龙), Min Jiang(蒋敏)
Chin. Phys. B. 35, 068503 (2026). DOI: 10.1088/1674-1056/ae617d | CSTR: 32038.14.CPB.ae617d

Single event effects evaluation on convolution neural network in Xilinx 28 nm system on chip

Xu Zhao(赵旭), Xuecheng Du(杜雪成), Xu Xiong(熊旭), Chao Ma(马超), Weitao Yang(杨卫涛), Bo Zheng(郑波), Chao Zhou(周超)
Chin. Phys. B. 33, 078501 (2024). DOI: 10.1088/1674-1056/ad3b82

Recent progress on two-dimensional ferroelectrics: Material systems and device applications

Zhiwei Fan(范芷薇), Jingyuan Qu(渠靖媛), Tao Wang(王涛), Yan Wen(温滟), Ziwen An(安子文), Qitao Jiang(姜琦涛), Wuhong Xue(薛武红), Peng Zhou(周鹏), Xiaohong Xu(许小红)
Chin. Phys. B. 32, 128508 (2023). DOI: 10.1088/1674-1056/ad08a4

RAPID COMMUNICATION

Two-dimensional floating-gate transistors with sub-10 nm MoS_2 channel for artificial synapses

Jiequn Sun(孙杰群)[1,2], Peng Song(宋鹏)[1,2], Chengze Du(杜承泽)[1,2], Jijia Guo(郭济嘉)[1,2], Haoyu Wang(王浩宇)[1,2], Hui Guo(郭辉)[1,2,3], Haitao Yang(杨海涛)[1,2,3], Lihong Bao(鲍丽宏)[1,2,3,†], and Hong-Jun Gao(高鸿钧)[1,2,3]

[1] *Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing* 100049, *China*
[3] *Hefei National Laboratory, Hefei* 230088, *China*

(Received 26 May 2026; revised manuscript received 23 June 2026; accepted manuscript online 3 July 2026)

High-density integration of synaptic transistors is crucial for the development of neuromorphic computing. However, aggressively scaling conventional silicon-based floating-gate devices to the sub-10 nm regime is hindered by severe short-channel effects. Here, we report two-dimensional floating-gate artificial synapses based on monolayer MoS_2 with channel lengths scaled down to the sub-10 nm limit. By integrating a high-κ hybrid Y_2O_3/HfO_2 dielectric layer, the sub-10 nm transistor achieves excellent electrostatic control, yielding an ultra-high on/off ratio exceeding 10^9. The sub-10 nm channel devices exhibit robust non-volatile memory performance, featuring a data retention exceeding 10^5 s and an endurance of over 5500 cycles. Furthermore, the sub-10 nm channel memory can work as an artificial synapse and demonstrates highly linear weight updates across 64 distinct conductance states with nonlinearity of ~ 0.50 for potentiation and ~ 0.13 for depression. Implementing these experimental characteristics into a VGG-11 convolutional neural network yields a high image recognition accuracy of $\sim 93.35\%$ on the CIFAR-10 dataset. Our work demonstrates that ultra-scaled 2D floating-gate transistors offer a viable and highly robust platform for future high-density, energy-efficient neuromorphic hardware.

Keywords: two-dimensional materials, short channel, sub-10 nm, floating-gate memory, synapse

PACS: 85.35.–p, 73.63.–b, 73.40.Qv, 85.30.De

DOI: 10.1088/1674-1056/ae85fd **CSTR:** 32038.14.CPB.ae85fd

1. Introduction

The rapid advancement of artificial intelligence (AI) has placed significant demands on current computational systems.[1] Traditional von Neumann systems, characterized by the spatial separation of processing and memory units, face severe latency and energy bottlenecks due to frequent data movement.[2,3] Neuromorphic computing offers a promising path to overcome these limitations, as it emulates the human brain's highly parallel and energy-efficient structure.[4] Synaptic transistors are the core building blocks of such hardware, and floating-gate (FG) devices have emerged as excellent candidates owing to their ability for logic operation and non-volatile synaptic weight storage.[5–7] Despite these advantages, the demand for high-density integration in neuromorphic systems places stringent requirements on device scaling. However, due to the unavoidable dangling bonds at the interface, scaling of conventional silicon-based floating-gate transistors suffers from severe short-channel effects (SCEs), including drain-induced barrier lowering and limited data retention and endurance.[8,9]

Two-dimensional (2D) materials and their van der Waals (vdW) heterostructures offer a promising solution, as their inherently dangling-bond-free surfaces naturally provide the atomically sharp interfaces required for high-performance electronic[10–15] and optoelectronic devices,[16–18] memories,[19–24] and integrated circuits.[25–28] Leveraging their atomically thin body and effective electrostatic control, 2D materials have been widely explored for scaling transistors down to the sub-10 nm regime. To achieve such aggressive dimensions, various nanofabrication strategies have been extensively investigated, including double exposure lithography,[29,30] carbon nanotube masks,[31] hydrogen plasma etching of grain boundaries,[32] shadow evaporation,[33,34] and vertical sidewall architectures.[35–37] Meanwhile, floating-gate synaptic devices based on 2D materials have been developed to emulate biological synapses with reliable non-volatile weight storage.[38–40] However, despite these significant advancements in both short-channel 2D logic devices and functional synaptic transistors, achieving a harmonious balance between aggressive channel scaling and robust synaptic performance remains a formidable challenge. The human brain contains approximately 10^{15} synapses, demanding extremely high integration density, which in turn ne-

†Corresponding author. E-mail: lhbao@iphy.ac.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

cessitates aggressive device scaling. In the past few years, it has been demonstrated that the limited charges in the memory stack can still effectively modulate the threshold voltage of the ultrashort-channel floating-gate memory,[34] and the utilization of vertical-channel structures has been demonstrated to reduce the channel length to 40 nm.[41] But the specific case of planar sub-10 nm synapse transistors remains unexplored. It is necessary to demonstrate that the ultrashort-channel floating-gate synapse can still emulate biological synaptic plasticity and perform near-linear weight updates to meet the demands of neuromorphic computing.

In this work, we overcome this gap by demonstrating 2D floating-gate transistors for artificial synapses with channel lengths scaled down to the sub-10 nm regime. By utilizing a high-κ hybrid Y_2O_3/HfO_2 dielectric, the sub-10 nm transistor achieves superior gate control, yielding an ultra-high on/off ratio exceeding 10^9. The sub-10 nm device functions effectively as a robust non-volatile memory, achieving a data retention exceeding 10^5 s and an endurance over 5500 cycles. Furthermore, the sub-10 nm channel memory can work as an artificial synapse and exhibits highly linear weight updates with a nonlinearity of ~ 0.50 for potentiation and ~ 0.13 for depression across 64 distinct conductance levels. Finally, incorporating these experimental device characteristics into a VGG-11 convolutional neural network simulation yields a recognition accuracy of $\sim 93.35\%$ on the CIFAR-10 dataset. These results show that scaled 2D floating-gate transistors can provide a viable and robust platform for future high-density, energy-efficient neuromorphic hardware.

2. Results and discussion

Figure 1(a) schematically illustrates the fabrication process of the floating-gate transistor with a sub-10 nm channel. First, the floating-gate electrode array was defined by laser direct writing (LDW) on a clean Si/SiO_2 (285 nm) substrate, followed by the thermal evaporation of Cr (1 nm)/Au (14 nm) and a standard lift-off process. Subsequently, yttrium was deposited by thermal evaporation and exposed to the atmosphere for more than 24 hours to ensure complete oxidation into a 4-nm Y_2O_3 seed layer. Then a 12-nm HfO_2 was deposited via atomic layer deposition (ALD), forming a high-κ hybrid Y_2O_3/HfO_2 dielectric layer.[42] Large-area monolayer MoS_2 films grown by chemical vapor deposition (CVD)[43] were then transferred onto the dielectric surface using a water-assisted transfer method[44] and patterned into 10×15 μm^2 rectangles via electron beam lithography (EBL) and reactive ion etching (RIE).[45] Finally, Cr (3 nm)/Au (27 nm) was deposited through double EBL/thermal evaporation/lift-off processes to form the source and drain electrodes.[30] By controlling the design widths of the floating gate (3 μm) and the source/drain electrodes (0.5 μm), a series of floating-gate transistors with various channel lengths can be fabricated on a single MoS_2 flake, as shown in Fig. 1(b). Figure 1(c) displays the scanning electron microscopy (SEM) image of the source and drain electrodes with a channel length of approximately 9 nm. A 280k× magnified view and SEM images of more sub-10 nm channel devices are provided in Fig. S1.

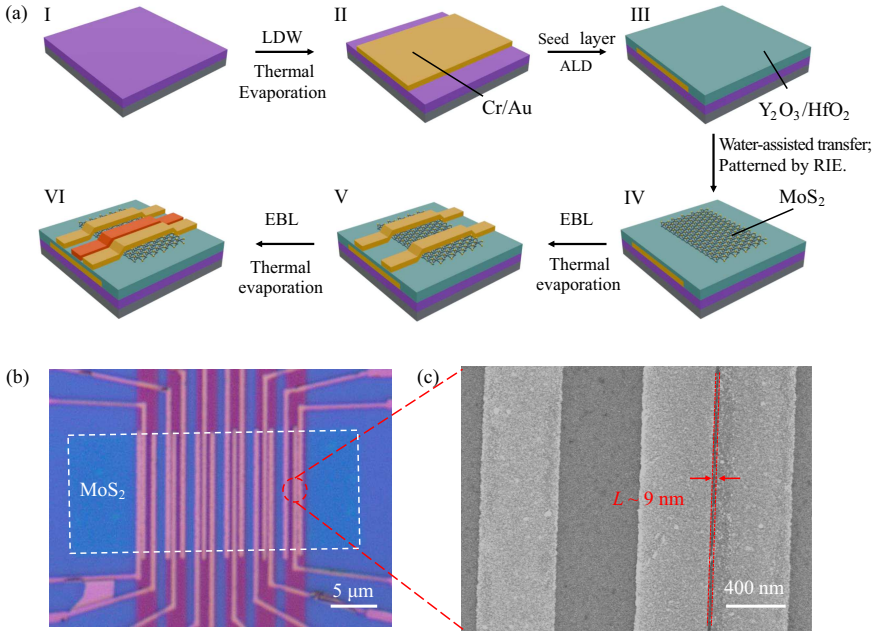

Fig. 1. The fabrication process of the floating-gate transistor with sub-10 nm channel length. (a) Schematic illustration of the fabrication steps for the sub-10 nm channel floating-gate memory. I: Clean Si substrate with a 285 nm thick SiO_2 layer. II: Patterning of the Cr/Au floating gate by laser direct writing and thermal evaporation. III: Y_2O_3 seed layer and HfO_2 facilitated by ALD. IV: Transfer of the monolayer MoS_2 film grown by CVD. V–VI: Definition and deposition of the Cr/Au source and drain electrodes via a double EBL/thermal evaporation/lift-off process. (b) Optical micrograph of a typical short-channel device. (c) SEM image of the short channel area, demonstrating a precisely defined channel length of approximately 9 nm.

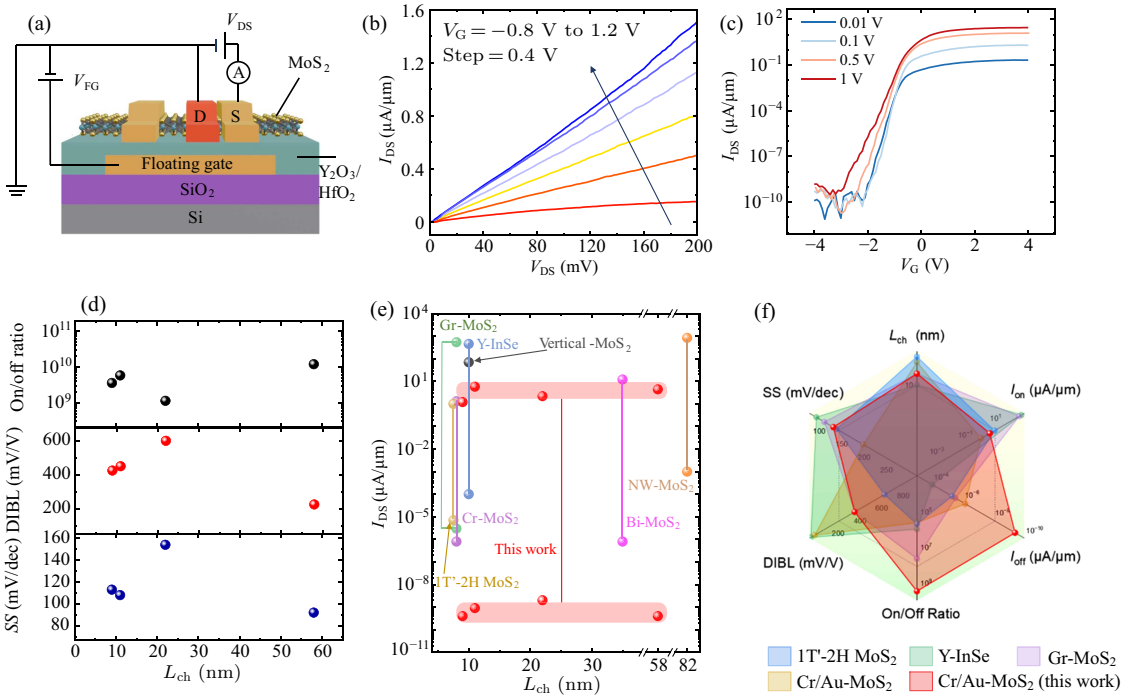

Fig. 2. Electrical characteristics of the sub-10 nm FET device. (a) Schematic of the sub-10 nm FET device. (b) Output characteristics of the sub-10 nm FET device with V_G ranging from −0.8 V to 1.2 V with steps of 0.4 V. (c) Transfer characteristics of the sub-10 nm FET device measured at $V_{DS} = 0.01$ V, 0.1 V, 0.5 V and 1 V, respectively. (d) Summary of key device parameters measured across devices with various channel lengths L_{ch}, including on/off ratio (top), drain-induced barrier lowering (DIBL) (middle), and subthreshold swing (SS) (bottom). (e) Benchmarking extinction ratio versus channel lengths in a few representative reports.[30–32,34,35,46,47] A series of devices with channel lengths ranging from sub-10 nm to 58 nm have demonstrated an on/off current ratio exceeding 10^9. (f) Radar plot comparing multiple key performance metrics — including L_{ch}, SS, DIBL, on/off Ratio, I_{on}, and I_{off} — among this work and other representative short-channel FET technologies.[30,32,34,47] The large shaded area of this work indicates superior overall performance across multiple parameters.

A schematic drawing of the short-channel transistor structure with electrical connections used in this work is shown in Fig. 2(a). Figure 2(b) displays the output characteristics of the MoS_2 transistor with a channel length of ∼ 9 nm. The drain–source voltage (V_{DS}) is swept from 0 V to 200 mV under varying gate voltages (V_G ranging from −0.8 V to 1.2 V with a step of 0.4 V). The linear curves at low bias indicate efficient carrier injection and relatively low contact resistance at the MoS_2 and Cr/Au interface.[13] Figure 2(c) shows the transfer characteristics of the sub-10 nm transistor at $V_{DS} = 0.01$ V, 0.1 V, 0.5 V and 1 V. The device exhibits excellent extinction behavior with a high current on/off ratio exceeding 10^{10} and an ultra-low off-state leakage current, demonstrating the robust gate control enabled by the high-κ hybrid Y_2O_3/HfO_2 dielectric layer and the excellent gate-tunability of monolayer MoS_2. The dual-sweep transfer curve of the 9-nm channel FET at a V_{DS} of 0.1 V is provided in the supplementary information (Fig. S2). The results exhibit a narrow hysteresis window of approximately 0.26 V.

To further investigate the channel length scaling behavior of the short-channel FET, key device parameters were extracted and summarized versus channel length ($L_{ch} =\sim 9$ nm, 11 nm, 22 nm, and 58 nm), as shown in Fig. 2(d). The on/off ratio remains consistently high ($> 10^9$) across the entire scaling range down to sub-10 nm, and exceeds 10^{10} in the 58-nm channel FET device. The drain-induced barrier lowering (DIBL) and subthreshold swing (SS) show a slight increase as L_{ch} decreases, while the device maintains an SS of ∼ 113 mV/dec and a DIBL of ∼ 425 mV/V with $L_{ch} =$ ∼ 9 nm. These excellent characteristics highlight the effectiveness of the FET devices in suppressing the short-channel effects (SCEs).[32] More data and analysis of the calculation of these key parameters are shown in Figs. S3 and S4.

Figure 2(e) benchmarks the extinction ratio of our short channel FETs and previous reports.[30–32,34,35,46,47] The on/off current used for comparison in this study was obtained at $V_{DS} = 0.1$ V. To ensure a fair and valid benchmarking, data from reported literature were also selected based on a drain–source voltage of $V_{DS} \geq 0.1$ V. Our devices demonstrate an ultralow off-state current (I_{off}) while maintaining a competitive on-state current (I_{on}), yielding an ultrahigh on/off ratio exceeding 10^9. Finally, a comprehensive performance comparison is presented in the radar plot in Fig. 2(f).[30,32,34,47] By integrating multiple metrics including L_{ch}, SS, DIBL, on/off ratio, I_{on}, and I_{off}, it is evident that our work (highlighted by the large red shaded area) achieves a superior balance between channel scaling and device performance. The specific numerical values are shown in Table S1. This underscores the potential of our sub-10 nm MoS_2 FETs for next-generation, low-power and high-performance logic applications. The transfer characteristics and memory characteristics of D2 and D3 from two more batches are shown in Fig. S5, assessing the reproducibil-

ity for our sub-10 nm channel floating-gate memory device.

A systematic evaluation of the non-volatile memory characteristics and reliability was further conducted to explore the potential of sub-10 nm MoS_2 FETs for data storage applications. Figure 3(a) illustrates the circuit configuration of the 9-nm channel floating-gate memory device. The drain–source current (I_{DS}) is regulated by a voltage pulse applied to the control gate, and the source electrode is grounded. During standard operation, the Au floating gate remained electrically isolated, being connected only temporarily to discharge stored charges when necessary.[19] Figure 3(b) displays the typical dual-sweeping transfer curves of the device under varying back-gate voltage (V_{BG}) sweep ranges. The device exhibits a pronounced hysteresis loop that expands monotonically with increasing sweeping voltage, revealing a large and highly tunable memory window. The extraction of the memory windows (ΔV) as a function of the maximum value of V_{BG} is shown in Fig. S6. A substantial memory window emerges when the back-gate sweep voltage surpasses 12 V, while no significant window appears below 6 V, confirming the precise control over the threshold electric field required for charge tunneling.[48]

The dynamic programming and erasing operations are demonstrated in Fig. 3(c). By applying a positive back-gate voltage pulse ($V_{BG,pulse}$ of +14 V, 1 s), electrons are injected into the floating gate and remain securely confined after the pulse due to the high energy barrier and minimal leakage current of the dielectric, shifting the transfer curve to state 0 (the positive V_{TH} state). The flat-band structure of the sub-10 nm channel floating gate memory is shown in Fig. S7, and the $Au/Y_2O_3/HfO_2$ barrier is 3.7 eV. Conversely, applying a negative back-gate voltage pulse ($V_{BG,pulse}$ of −14 V, 0.1 s) ejects the trapped electrons, shifting the curve negatively to state 1 (the negative V_{TH} state).[19–21]

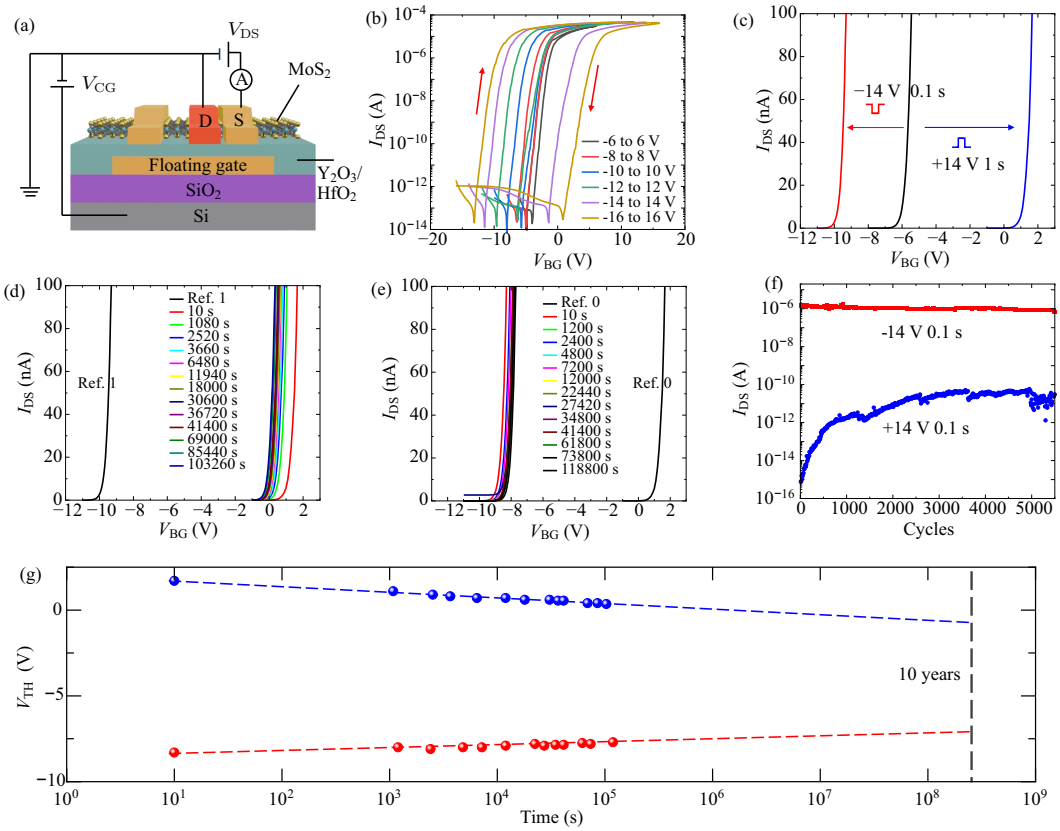

Fig. 3. Non-volatile memory characteristics and reliability of the short channel floating-gate memory device. (a) Schematic illustration of the circuit configuration for the short channel floating-gate memory. (b) Typical dual-sweeping transfer curves of the floating-gate memory device, exhibiting a large tunable memory window. (c) States 0 and 1 achieved with a +14 V (1 s) pulse and a −14 V (0.1 s) pulse, respectively. (d), (e) Demonstration of non-volatility of the sub-10 nm floating-gate memory device. The transfer curves (V_{DS} = 0.1 V) of state 0 (d) and state 1 (e) show a negligible shift even after 10^5 s. (f) Endurance of the sub-10 nm floating-gate memory. The device was repeatedly programmed to states 0 and 1 using cycles with 14 V (0.1 s) pulses and −14 V (0.1 s) pulses, respectively. The sub-10 nm floating-gate memory device worked well over 5500 cycles. (g) Data retention for states 0 and 1. Threshold voltage data are extracted from (d) and (e), respectively.

Figures 3(d) and 3(e) demonstrate the exceptional non-volatile data retention capability of the sub-10 nm floating-gate memory. The transfer curves measured at V_{DS} = 0.1 V for both states 0 and 1 exhibit negligible degradation even after 10^5 seconds. This robust data retention capability is further confirmed by Fig. 3(g), which plots the extracted threshold voltage (V_{TH}) as a function of time. The V_{TH} values for both the programming and erasing states remain stable, and the extrapolated trend clearly suggests that the significant V_{TH} window will persist for over 10 years.

The robust endurance of the sub-10 nm memory device was also evaluated with $V_{BG,pulse}$ of +14 V, 0.1 s and −14 V, 0.1 s. As presented in Fig. 3(f), the sub-10 nm floating-gate memory device worked well over 5500 cycles. The off-state current stabilizes at a level below 10^{-10} A after a rapid initial increase during the first few cycles. The degradation during the initial cycling stage can be attributed to the filling of pre-existing border traps[49] arising from inevitable lattice mismatch and slight stoichiometry deviations. During the initial cycling stage, these traps capture injected charge carriers. Once these traps are saturated, the charge injection/emission becomes dominated by the floating-gate, and the off-state current stabilizes. Meanwhile, the on-state current remains remarkably stable throughout the test. Consequently, a high on/off ratio exceeding 10^4 is maintained after the endurance test, demonstrating the excellent operational durability and defect tolerance of the sub-10 nm floating-gate memory.

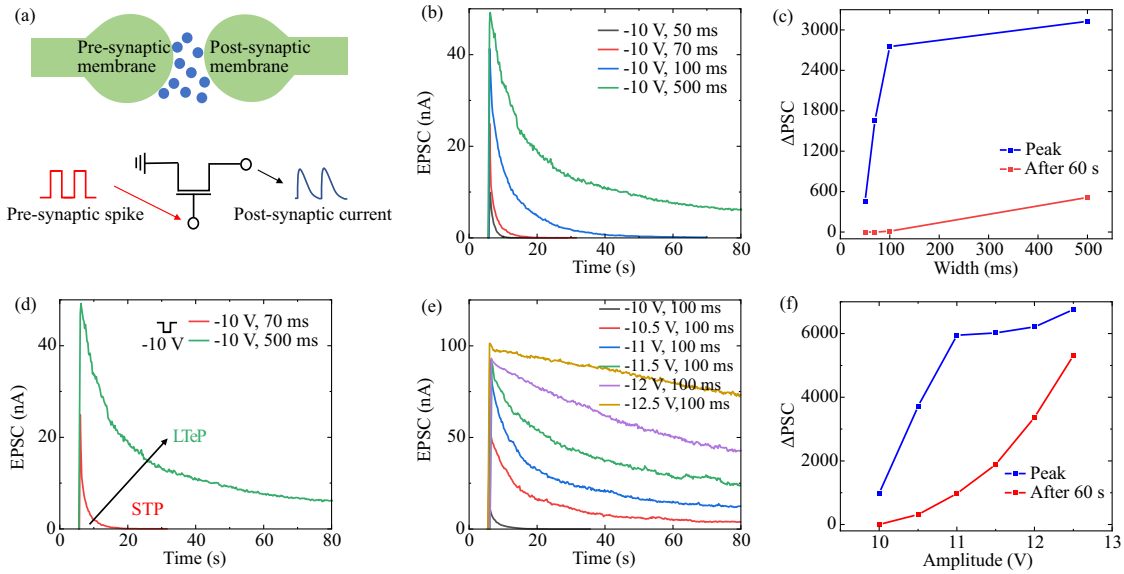

Fig. 4. Artificial synaptic behaviors of the short channel floating-gate transistor. (a) Schematic illustration of a biological synapse and the artificial synapse. (b) EPSC generated by single V_{BG} spikes with different pulse widths from 50 ms to 500 ms at a fixed amplitude of −10 V. The V_{DS} is set at 0.1 V during all reading processes. (c) The change in postsynaptic current as a function of pulse width. (d) Short-term/long-term plasticity of the short channel floating-gate transistor. (e) EPSC generated by single V_{BG} spikes with different amplitudes from −10 V to −12.5 V at a fixed width of 100 ms. (f) The change in postsynaptic current as a function of pulse amplitude.

The superior non-volatile memory characteristics provide a robust foundation for artificial synapses for neuromorphic computing applications. Figure 4(a) illustrates the analogy drawn between a biological synapse and our artificial electronic synapse.[1] In this configuration, the device gate serves as the pre-synaptic membrane receiving stimulus spikes, the active channel acts as the post-synaptic membrane, and the source–drain current represents the postsynaptic current (PSC).[50] The channel conductance is used to mimic the synaptic weight. Figure 4(d) displays the typical temporal response of the excitatory postsynaptic current (EPSC) generated by single gate spikes. The V_{DS} is set as 0.1 V during all reading processes. The EPSC exhibits a characteristic synapse-like behavior in that a rapid rise upon spike arrival is followed by a slow relaxation process back towards base current levels. A short duration spike (−10 V, 70 ms) induces a transient EPSC response that fully decays within 20 s, successfully emulating short-term plasticity (STP). Conversely, a longer spike (−10 V, 500 ms) consolidates this transient surge into a non-volatile weight update, characterized by a substantial persistent current remaining well after 60 s, which demonstrates effective long-term plasticity (LTeP).[4] The transformation from STP to LTeP occurs with increasing amplitudes and widths of the presynaptic pulse, as shown in Figs. 4(b) and 4(e). Figure 4(b) displays the typical temporal response of the EPSC generated by single gate spikes with widths varying from 50 ms to 500 ms at a fixed amplitude of −10 V. Wider pre-synaptic spikes induce larger initial current increases and lead to longer relaxation times, and this trend is quantitatively summarized in Fig. 4(c), plotting the total change in post-synaptic current ΔPSC measured at the peak and 60 s after the spike versus the pulse widths. The residual current after 60 s remains near zero for widths under 100 ms, while the residual current begins to increase notably with widths beyond 100 ms. Similarly, we investigated the influence of pre-synaptic pulse amplitude on synaptic weight modulation. Single spikes with varying negative amplitudes ranging from −10 V to −12.5 V were applied at a fixed width of 100 ms (Fig. 4(e)). Figure 4(f) displays the relationship between ΔPSC and pulse amplitudes. The peak PSC value approaches saturation as the pulse amplitude exceeds 11 V. In contrast, the PSC measured 60 s after the stimulus exhibits a notable increase with rising amplitude, illustrating a clear transition from STP to LTeP as the stimulus intensity increases from 10 V to 12.5 V. The energy efficiency

of the device is crucial for its application in low-power neuromorphic computing. The energy consumption calculation formula is

$$E = V_{\mathrm{DS}} \times I_{\mathrm{DS}} \times t_{\mathrm{spike}}. \tag{1}$$

Here, I_{DS} and t_{spike} represent the amplitude of the excitatory postsynaptic current and the pulse width, respectively, while V_{DS} denotes the operating voltage of the device. As shown in Fig. S8, we obtained an energy consumption of ~ 51 fJ from D4 with a channel length of ~ 26 nm, which is on the same order of magnitude as the energy range of biological synapses (1–10 fJ) and some state-of-the-art reports (0.03 fJ to 41 fJ).[41,51–53] This result demonstrates the promising potential of our short channel floating-gate transistor platform for low-power neuromorphic applications.

To evaluate the system-level viability of our sub-10 nm floating-gate synaptic device, we integrated its experimentally measured characteristics into a hardware-based artificial neural network (ANN) simulation. The potential for neuromorphic computing is heavily dependent on the device's ability to emulate synaptic weight updates, specifically through long-term potentiation (LTP) and depression (LTD) processes.[54,55] The conductance update behavior of our device is characterized in Fig. 5(a). Both the potentiation process and the depression process were achieved by applying programmed pulse schemes instead of identical pulses to improve the linearity and symmetry of LTP/D responses.[40,56] The potentiation process was achieved using a sequence of 64 pulses with amplitudes ramped from −7 V to −9.52 V (500 ms width), while the depression process was executed using 64 pulses with amplitudes from 2 V to 9.56 V (1 ms width). Figures S9 and S10 provide the potentiation weight states of the conductance measured under different pulse widths and amplitudes, and Fig. S11 shows the LTP/D characteristics measured with a different programmed 64-pulse scheme (−7.5 V to −8.76 V with a pulse width of 500 ms for the LTP process and 2 V to 9.56 V with a pulse width of 1 ms for the LTD process).

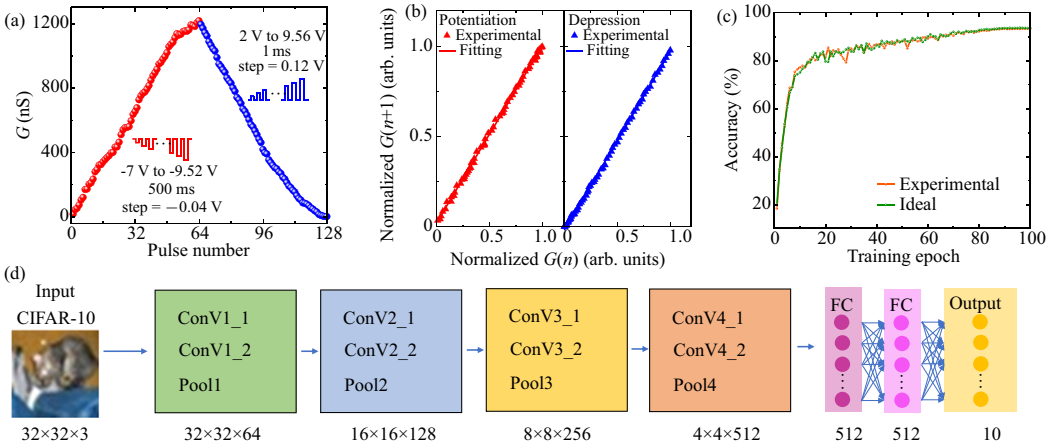

Fig. 5. Neuromorphic computing simulation based on the synaptic device. (a) LTP/D characteristics measured with a programmed pulse scheme. The LTP process is achieved by 64 pulses from −7 V to −9.52 V with a pulse width of 500 ms, and the LTD process is achieved by 64 pulses from 2 V to 9.56 V with a pulse width of 1 ms. (b) Experimental data and fitting curves of potentiation and depression. (c) The recognition accuracy evolution with training epochs based on the artificial synapse compared to the ideal synapse. (d) Schematic representation of a CNN based on a VGG-11 network for image recognition using the CIFAR-10 dataset.

The values of nonlinearity (N) of LTP and LTD were quantitatively analyzed by fitting experimental data with equations[54]

$$\begin{aligned} G(n+1) &= G(n) + \Delta G_{\mathrm{p}} \\ &= G(n) + M_{\mathrm{p}} \mathrm{e}^{-N_{\mathrm{p}}(G(n)-G_{\min}/G_{\max}-G_{\min})}, \\ &\quad \Delta G_{\mathrm{p}} > 0, \end{aligned} \tag{2}$$

$$\begin{aligned} G(n+1) &= G(n) + \Delta G_{\mathrm{d}} \\ &= G(n) + M_{\mathrm{d}} \mathrm{e}^{-N_{\mathrm{d}}(G(n)-G_{\min}/G_{\max}-G_{\min})}, \\ &\quad \Delta G_{\mathrm{d}} < 0, \end{aligned} \tag{3}$$

$$G_{\mathrm{normalized}} = \frac{G(n) - G_{\min}}{G_{\max} - G_{\min}}, \tag{4}$$

where N_{p} and N_{d} represent N for potentiation and depression processes, respectively, $G(n)$ is the channel conductance value after the n^{th} excitatory or inhibitory pulse, and $G_{\mathrm{normalized}}$ is the normalized conductance. Experimental data of LTP and LTD and the fitting curves are shown in Fig. 5(b). N_{p} and N_{d} extracted from the fitting curves are ~ 0.50 and 0.13, which indicates that our sub-10 nm device effectively mimics the behavior required for precise weight management in neuromorphic architectures. Based on the experimentally extracted weight update characteristics, we implemented a convolutional neural network (CNN) simulation using a VGG-11 architecture to perform image recognition tasks on the CIFAR-10 dataset. The architecture, as illustrated in Fig. 5(d), consists of eight convolutional layers for feature extraction, two fully connected hidden layers, and a softmax output layer for final classification. The training results depict the evolution of recognition accuracy over 100 training epochs as shown in Fig. 5(c).

At each epoch of learning, the simulator trains 50000 images randomly from the training data set, classifies 10000 images from the testing data set, and then tests the recognition accuracy of the classified images. The recognition accuracy of the neural network based on the sub-10 nm synaptic device improves from $\sim 18.31\%$ to $\sim 93.35\%$ with the increasing learning epoch, close to that of a neural network based on an ideal synaptic transistor with near-linear updates ($\sim 93.57\%$ after 90 epochs). The successful emulation suggests that the device inherent linearity and dynamic range of the sub-10 nm synapse are sufficient to support complex deep learning tasks without significant degradation in computational accuracy, confirming its suitability for high-performance, energy-efficient neuromorphic computing hardware.

3. Conclusion

In summary, extremely high-performance scaled monolayer MoS_2 FETs have been fabricated. The devices exhibit excellent electrostatic control, maintaining an ultra-high on/off ratio exceeding 10^9 even as channel lengths are scaled down to the sub-10 nm regime. The integration of a floating-gate architecture into the sub-10 nm device enables reliable non-volatile memory functionality, achieving a data retention exceeding 10^5 s and an endurance of over 5500 cycles. Furthermore, the sub-10 nm device can work as an artificial synapse in emulating the transition from STP to LTeP, and system-level simulations using a VGG-11 CNN on the CIFAR-10 dataset further validate the computational potential of the artificial synapses. These results underscore the potential of 2D semiconductors for nonvolatile memory and neuromorphic computing applications within a sub-10 nm footprint.

Appendix A: Experimental details

Device fabrication Floating-gate electrodes were patterned on cleaned Si/SiO_2 (285 nm) wafers using a LOR 5A/SPR220 3.0 bilayer resist system via laser direct writing, followed by the thermal evaporation of Cr (1 nm)/Au (14 nm) and a standard lift-off process. Subsequently, Y was deposited by thermal evaporation and exposed to the atmosphere for more than 24 hours to ensure complete oxidation into a 4-nm Y_2O_3 seed layer. A 12-nm HfO_2 layer was deposited using a low-temperature (110 °C) ALD process for 100 cycles. Large-area monolayer MoS_2 films grown by CVD were then transferred onto the dielectric surface using a water-assisted transfer method and patterned into 10×15 μm^2 rectangles via EBL and RIE. Finally, Cr (3 nm)/Au (27 nm) source/drain electrodes were fabricated through double EBL/thermal evaporation/lift-off processes.

Electrical measurement The electrical characteristics of the floating-gate transistors were determined on a Keithley 4200 semiconductor characterization system (4200-SCS, Keithley) equipped with a micromanipulator vacuum probe system (TTPX, Lake Shore Cryotronics, Inc.) under a vacuum of 10^{-4} Pa.

Neural network simulation The simulation of a CNN for image recognition tasks was carried out on a VGG-11 network using the CIFAR-10 dataset. The VGG-11 network is composed of 8 convolutional layers and 3 fully connected layers. Spatial downsampling was achieved through four stages of pooling operations, progressively reducing feature map dimensions from 32×32 to 4×4. The classifier consists of two fully connected hidden layers with 512 units each, followed by a 10-way softmax output layer. The normalized conductance of sub-10 nm floating-gate transistors was used to represent the synapse weight.

Acknowledgements

This work was supported by the National Key Research & Development Project of China (Grant No. 2022YFA1204100), the National Natural Science Foundation of China (Grant No. 62488201), CAS Project for Young Scientists in Basic Research (Grant No. YSBR-003), and Quantum Science and Technology – National Science and Technology Major Project of China (Grant No. 2021ZD0302700).

References

[1] Kim S G, Han J S, Kim H, Kim S Y and Jang H W 2018 *Adv. Mater. Technol.* **3** 1800457
[2] Xia Q and Yang J J 2019 *Nat. Mater.* **18** 309
[3] Sangwan V K and Hersam M C 2020 *Nat. Nanotechnol.* **15** 517
[4] Pan X, Jin T, Gao J, Han C, Shi Y and Chen W 2020 *Small* **16** 2001504
[5] Guo L, Leobandung E and Chou S Y 1997 *Science* **275** 649
[6] Gehring A and Selberherr S 2004 *IEEE Trans. Device Mater. Relib.* **4** 306
[7] Sebastian A, Le Gallo M, Khaddam-Aljameh R and Eleftheriou E 2020 *Nat. Nanotechnol.* **15** 529
[8] Ma R, Wu F, Zhang M, Lu Z, Wan J and Xie C 2019 *IEEE Access* **7** 44696
[9] Kim K S, Kwon J, Ryu H, Kim C, Kim H, Lee E K, Lee D, Seo S, Han N M, Suh J M, Kim J, Song M K, Lee S, Seol M and Kim J 2024 *Nat. Nanotechnol.* **19** 895
[10] Britnell L, Gorbachev R V, Jalil R, Belle B D, Schedin F, Mishchenko A, Georgiou T, Katsnelson M I, Eaves L, Morozov S V, Peres N M R, Leist J, Geim A K, Novoselov K S and Ponomarenko L A 2012 *Science* **335** 947
[11] Yu W J, Li Z, Zhou H, Chen Y, Wang Y, Huang Y and Duan X 2013 *Nat. Mater.* **12** 246
[12] Novoselov K S, Mishchenko A, Carvalho A and Castro Neto A H 2016 *Science* **353** aac9439
[13] Wu L, Shi J, Zhou Z, Yan J, Wang A, Bian C, Ma J, Ma R, Liu H, Chen J, Huang Y, Zhou W, Bao L, Ouyang M, Pantelides S T and Gao H J 2020 *Nano Res.* **13** 1127
[14] Wei C, Jiazila N, Liu X, Song P, Gao H, Sun J, Bao L, Lin X and Gao H J 2024 *Chin. Phys. Lett.* **41** 128501
[15] Liu X, Li L, Wei C, Song P, Gao H, Wu K, Jiazila N, Sun J, Guo H, Yang H, Zhou W, Bao L and Gao H J 2024 *Chin. Phys. B* **33** 127302
[16] Britnell L, Ribeiro R M, Eckmann A, Jalil R, Belle B D, Mishchenko A, Kim Y J, Gorbachev R V, Georgiou T, Morozov S V, Grigorenko A N, Geim A K, Casiraghi C, Neto A H C and Novoselov K S 2013 *Science* **340** 1311
[17] Yu W J, Liu Y, Zhou H, Yin A, Li Z, Huang Y and Duan X 2013 *Nat. Nanotechnol.* **8** 952

[18] Qu Z, Zhang Y, Zhao X, Wang Y, Yang F, Zhao W and Liu H 2025 *Chin. Phys. Lett.* **42** 058101

[19] Wu L, Wang A, Shi J, Yan J, Zhou Z, Bian C, Ma J, Ma R, Liu H, Chen J, Huang Y, Zhou W, Bao L, Ouyang M, Pennycook S J, Pantelides S T and Gao H J 2021 *Nat. Nanotechnol.* **16** 882

[20] Liu L, Liu C, Jiang L, Li J, Ding Y, Wang S, Jiang Y G, Sun Y B, Wang J, Chen S, Zhang D W and Zhou P 2021 *Nat. Nanotechnol.* **16** 874

[21] Wang H, Guo H, Guzman R, JiaziLa N, Wu K, Wang A, Liu X, Liu L, Wu L, Chen J, Huan Q, Zhou W, Yang H, Pantelides S T, Bao L and Gao H J 2024 *Adv. Mater.* **36** 2311652

[22] Xiang Y, Wang C, Liu C, Wang T, Jiang Y, Wang Y, Wang S and Zhou P 2025 *Nature* **641** 90

[23] Gao H, Liu X, Song P, Wei C, Jiazila N, Sun J, Wu K, Guo H, Yang H, Bao L and Gao H J 2024 *Chin. Phys. B* **33** 127201

[24] Song P, Liu X, Sun J, Jiazila N, Wei C, Gao H, Du C, Guo H, Yang H, Bao L and Gao H J 2025 *Chin. Phys. Lett.* **42** 068101

[25] Migliato Marega G, Zhao Y, Avsar A, Wang Z, Tripathi M, Radenovic A and Kis A 2020 *Nature* **587** 72

[26] Li N, Wang Q, Shen C, Wei Z, Yu H, Zhao J, Lu X, Wang G, He C, Xie L, Zhu J, Du L, Yang R, Shi D and Zhang G 2020 *Nat. Electron.* **3** 711

[27] Migliato Marega G, Ji H G, Wang Z, Pasquale G, Tripathi M, Radenovic A and Kis A 2023 *Nat. Electron.* **6** 991

[28] Shen Z, Wu H, Liu C, Liu Z, Jiang Y, Wang T and Zhou P 2026 *Nat. Mater.* **25** 199

[29] Zhu Y, Li Y, Arefe G, Burke R A, Tan C, Hao Y, Liu X, Liu X, Yoo W J, Dubey M, Lin Q and Hone J C 2018 *Nano Lett.* **18** 3807

[30] Jiang J, Xu L, Qiu C and Peng L M 2023 *Nature* **616** 470

[31] Liu Y, Guo J, Wu Y, Zhu E, Weiss N O, He Q, Wu H, Cheng H C, Xu Y, Shakir I, Huang Y and Duan X 2016 *Nano Lett.* **16** 6337

[32] Xie L, Liao M, Wang S, Yu H, Du L, Tang J, Zhao J, Zhang J, Chen P, Lu X, Wang G, Xie G, Yang R, Shi D and Zhang G 2017 *Adv. Mater.* **29** 1702522

[33] Guo W, Dong Z, Xu Y, Liu C, Wei D, Zhang L, Shi X, Guo C, Xu H, Chen G, Wang L, Zhang K, Chen X and Lu W 2020 *Adv. Sci.* **7** 1902699

[34] Jiang Y, Liu C, Cao Z, Li C, Liu Z, Wang C, Xiang Y and Zhou P 2024 *Nat. Electron.* **7** 868

[35] Jiang J, Doan M H, Sun L, Kim H, Yu H, Joo M K, Park S H, Yang H, Duong D L and Lee Y H 2020 *Adv. Sci.* **7** 1902964

[36] Zhang J, Gao F and Hu P 2021 *Nat. Electron.* **4** 325

[37] Yang X, He R, Lu Z, Chen Y, Liu L, Lu D, Ma L, Tao Q, Kong L, Xiao Z, Liu S, Li Z, Ding S, Liu X, Li Y, Wang Y, Liao L and Liu Y 2024 *Nat. Commun.* **15** 7676

[38] Bach T P A, Cho S, Kim H, Nguyen D A and Im H 2024 *ACS Nano* **18** 4131

[39] Cho H, Lee D, Ko K, Lin D Y, Lee H, Park S, Park B, Jang B C, Lim D H and Suh J 2023 *ACS Nano* **17** 7384

[40] Jin T, Zheng Y, Gao J, Wang Y, Li E, Chen H, Pan X, Lin M and Chen W 2021 *ACS Appl. Mater. Interfaces* **13** 10639

[41] Jang N, Kwon Y H, Seong N J, Choi K and Yoon S M 2025 *ACS Appl. Mater. Interfaces* **17** 40788

[42] Zou X, Wang J, Chiu C H, Wu Y, Xiao X, Jiang C, Wu W W, Mai L, Chen T, Li J, Ho J C and Liao L 2014 *Adv. Mater.* **26** 6255

[43] Ma J J, Wu K, Wang Z Y, Ma R S, Bao L H, Dai Q, Ren J D and Gao H J 2022 *Chin. Phys. B* **31** 088105

[44] Schneider G F, Calado V E, Zandbergen H, Vandersypen L M K and Dekker C 2010 *Nano Lett.* **10** 1912

[45] Xiao S, Xiao P, Zhang X, Yan D, Gu X, Qin F, Ni Z, Han Z J and Ostrikov K 2016 *Sci. Rep.* **6** 19945

[46] Shen P C, Su C, Lin Y, Chou A S, Cheng C C, Park J H, Chiu M H, Lu A Y, Tang H L, Tavakoli M M, Pitner G, Ji X, Cai Z, Mao N, Wang J, Tung V, Li J, Bokor J, Zettl A, Wu C I, Palacios T, Li L J and Kong J 2021 *Nature* **593** 211

[47] Nourbakhsh A, Zubair A, Sajjad R N, Tavakkoli K G A, Chen W, Fang S, Ling X, Kong J, Dresselhaus M S, Kaxiras E, Berggren K K, Antoniadis D and Palacios T 2016 *Nano Lett.* **16** 7798

[48] Hong A J, Song E B, Yu H S, Allen M J, Kim J, Fowler J D, Wassei J K, Park Y, Wang Y, Zou J, Kaner R B, Weiller B H and Wang K L 2011 *ACS Nano* **5** 7812

[49] Maji D, Duttagupta S P, Rao V R, Yeo C C and Cho B J 2007 *IEEE Electron Device Lett.* **28** 731

[50] Zucker R S and Regehr W G 2002 *Annu. Rev. Physiol.* **64** 355

[51] Zhi J, Wen Y, Chen J, Cai C, Yang S, Zhu H, Cheng R, Yin L, Liang S and He J 2026 *Nat. Commun.* **17** 3974

[52] Kim T J, Jeong H L, Song S W, Kwon D, Lee S H and Joo J 2026 *Adv. Sci.* **13** e17149

[53] Dang Z, Guo F, Duan H, Zhao Q, Fu Y, Jie W, Jin K and Hao J 2023 *Nano Lett.* **23** 6752

[54] Seo S, Kang B S, Lee J J, Ryu H J, Kim S, Kim H, Oh S, Shim J, Heo K, Oh S and Park J H 2020 *Nat. Commun.* **11** 3936

[55] Ambrogio S, Narayanan P, Tsai H, Shelby R M, Boybat I, Di Nolfo C, Sidler S, Giordano M, Bodini M, Farinha N C P, Killeen B, Cheng C, Jaoudi Y and Burr G W 2018 *Nature* **558** 60

[56] Zhou W, Li Y, Sun F, Gou X, Guan K, Liu L, Wang Y, Liu F, Tian Z, He X, Xia Y, Zhang C, Wang X and Zhang T 2026 *ACS Appl. Mater. Interfaces* **18** 30063