A Reliable All-2D Materials Artificial Synapse for High Energy-Efficient Neuromorphic Computing

J. Tang, Congli He,* Jianshi Tang, Kun Yue, Qingtian Zhang, Yizhou Liu, Qin Qin Wang, Shuoepi Wang, Na Li, Cheng Shen, Yanchong Zhao, Jieying Liu, Jiahao Yuan, Zheng Wei, Jiawei Li, Kenji Watanabe, Takashi Taniguchi, Dashan Shang, Shouguo Wang, Wei Yang, Rong Yang, Dongxia Shi, and Guangyu Zhang*

High-performance artificial synaptic devices are indispensable for developing neuromorphic computing systems with high energy efficiency. However, the reliability and variability issues of existing devices such as nonlinear and asymmetric weight update are the major hurdles in their practical applications for energy-efficient neuromorphic computing. Here, a two-terminal floating-gate memory (2TFGM) based artificial synapse built from all-2D van der Waals materials is reported. The 2TFGM synaptic device exhibits excellent linear and symmetric weight update characteristics with high reliability and tunability. In particular, the high linearity and symmetric synaptic weight realized by simple programming with identical pulses can eliminate the additional latency and power consumption caused by the peripheral circuit design and achieve an ultralow energy consumption for the synapses in the neural network implementation. A large number of states up to \( \approx 3000 \), high switching speed of 40 ns and low energy consumption of 18 fJ for a single pulse have been demonstrated experimentally. A high classification accuracy up to 97.7% (close to the software baseline of 98%) has been achieved in the Modified National Institute of Standards and Technology (MNIST) simulations based on the experimental data. These results demonstrate the potential of all-2D 2TFGM for high-speed and low-power neuromorphic computing.

1. Introduction

With the rapid development of computing technology, computing tasks such as image and speech recognition, game playing, and unmanned driving have become more and more complicated.[1,2] Amid the slowdown of Moore’s law scaling, conventional hardware based on the von Neumann architecture faces performance constraints in power dissipation and energy efficiency due to the physical separation of data storage and processing.[3–5] Inspired by the architecture and working principle of human brain, which is adaptive, massively parallel, and fault-tolerant, innovative neuromorphic computing systems have been proposed to address the rapidly growing computation power and efficiency requirements.[6,7] Extensive efforts have been devoted to develop electronic devices that are capable of mimicking the behaviors of biological...
neurons and synapses.[8–10] For example, a variety of devices, including memristors,[10,11] phase change memory (PCM),[12] spintronic devices,[13] and synaptic transistors (including floating-gate, ferroelectric-gate, electrolyte-gate, and opto-electronic synaptic transistors, etc.),[14–19] have been reported to emulate synaptic functions.

However, reliability and variability issues, for example, linear/symmetric weight update and number of states, are still the major hurdles in practical applications for energy-efficient neuromorphic computing,[20,21] In general, large nonlinearity leads to complex weight modulation as well as high energy and time costs in the training process. By contrast, a linear and symmetric weight update behavior with sufficient number of states can effectively improve the inference accuracy and reliability of neuromorphic computing.[22] Therefore, to build a low power and high-accuracy artificial neuromorphic network, it is necessary to improve the linearity and symmetry of an artificial synapse device.[22,23] To this end, various strategies have been developed.[24–26] For example, nearly linear and symmetric weight update were achieved by continuously ramping up the pulse amplitude or other complex programming pulse designs.[16,22] But the generation of such non-identical pulses puts an additional burden on the peripheral circuit design, such as the digital-to-analog converter (DAC) circuits, causing additional latency and power consumption. Thus, the search of an ideal artificial synapse device with good linear/symmetric weight update and low power consumption is still on the way.

Recently, 2D materials have been considered for application to neuromorphic computing systems due to the tremendous potentials related to their atomically ultra-thin body with facile electronic tunability.[27–26] Here, we fabricated all-2D 2TFGM synapse devices and realized highly symmetric and linear weight update with low power consumption and good reliability. Especially, the linearity, on/off ratio, and the number of analog states of the 2TFGM artificial synapse can be tuned by the amplitude and width of the operation pulses. Image classifications based on the experimental potentiation/depression (P/D) results with different linearity/symmetric weight update were achieved by continuously ramping up the pulse amplitude or other complex programming pulse designs.[16,22] But the generation of such non-identical pulses puts an additional burden on the peripheral circuit design, such as the digital-to-analog converter (DAC) circuits, causing additional latency and power consumption. Thus, the search of an ideal artificial synapse device with good linear/symmetric weight update and low power consumption is still on the way.

Recently, 2D materials have been considered for application to neuromorphic computing systems due to the tremendous potentials related to their atomically ultra-thin body with facile electronic tunability.[27–26] Here, we fabricated all-2D 2TFGM synapse devices and realized highly symmetric and linear weight update with low power consumption and good reliability. Especially, the linearity, on/off ratio, and the number of analog states of the 2TFGM artificial synapse can be tuned by the amplitude and width of the operation pulses. Image classifications based on the experimental potentiation/depression (P/D) results with different linearity/symmetric weight update were achieved by continuously ramping up the pulse amplitude or other complex programming pulse designs.[16,22] But the generation of such non-identical pulses puts an additional burden on the peripheral circuit design, such as the digital-to-analog converter (DAC) circuits, causing additional latency and power consumption. Thus, the search of an ideal artificial synapse device with good linear/symmetric weight update and low power consumption is still on the way.

Device structures and memory characteristics of 2TFGM are shown in Figure 1. Figure 1a shows a schematic view of the MoS 2-based all-2D 2TFGM fabricated through layer-by-layer stacking process (see Experimental Section and Figure S2, Supporting Information). Monolayer MoS 2 is used as the channel (see Figures S1 and S3, Supporting Information, for MoS 2 films and device characterizations), exfoliated few-layer graphene (FLG) are used both as the floating gate (FG) electrode and as contact electrode, and h-BN as the tunneling layer. Note that graphene contact can reduce the contact resistance, and the atomically flat surface of h-BN could avoid trapped charges and electron scattering at the interface.[37–39]

The device is measured by applying operation voltages on drain electrodes with source electrode grounded. Figure 1b shows the current–voltage (I–V) switching hysteresis loops of typical 2TFGMs with different h-BN tunneling layer thicknesses of 7, 10, and 15 nm (channel length/width of 3 µm/10 µm was kept unchanged). The sweeping directions are indicated by the dashed arrows. All 2TFGMs have nearly symmetric current hysteresis loops with ultra-high on/off ratios of >10 8 during the programming and erasing process. The operation voltages were reduced from ±20 to ±8 V, respectively, which could be ascribed to the decreasing width of the tunneling barrier due to adopting thinner tunneling layer. Figure 1c illustrates band diagrams of the drain/h-BN/FG in the programming (i) and erasing (ii) processes. For the programming process (Figure 1c(i)), a large potential drop between FG and drain is introduced by applying positive voltages (Vds), enabling electron tunneling from FG to drain, and making the FG positively charged (i.e., hole accumulation). Owing to the strong electrostatic gating of the MoS 2 channel through the thin h-BN dielectric layer, the device resistance state (RS) is changed from high-resistance state (HRS) to low-resistance state (LRS). For the erasing process (Figure 1c(ii)), electrons tunnel from drain to FG and are stored in FG when a negative voltage is applied, resulting a change from LRS to HRS. As the electron tunneling probability depends strongly on the barrier thickness, higher drain voltages are required for operating a device with thicker h-BN tunneling layers (Figure 1b). The charge tunneling and storage in FG are further confirmed in the 2TFGM device with extended FG. (See Figure S4, Supporting Information).

The retention characterizations after ±12 V programming/erasing process of a 2TFGM with 10 nm-thick h-BN were measured at different temperatures of 300, 400, and 500 K with Vread = 1 V (Figure 1d, also see Figure S5, Supporting Information, for more details). Both LRS and HRS feature a good retention property of >10 4 s at 300 K. This excellent retention is attributed to high barrier height of h-BN and the stored electrons/holes cannot dissipate from the embedded FG without external stimulations. A large memory window is still preserved, indicating the good robustness to thermal environment of the 2TFGM devices. Note that HRS increases obviously in
the initial 1000 s at 400 K, which could be ascribed to electrons tunneling back from FG to drain with higher kinetic energy through the thermal-emission process. The decrease of the on/off ratio and the deterioration of the retention property indicate that hot electron emission plays an important role in addition to the tunneling effect with the increase of temperature. The memory window almost disappears at 500 K. The retention characteristics of HRS/LRS become poorer for the device as decreasing the thickness of hBN (corresponding to narrower tunneling barrier, Figure S6, Supporting Information). Thus, the retention properties are mainly affected by the kinetic energy of the carriers and the tunneling barrier height/width. Our 2TFGM devices also show good endurance characteristic of at least $10^5$ cycles under programming voltage pulse amplitude of ±18 V and width of 100 ms with $V_{\text{read}} = 1$ V (Figure 1e). The on/off ratio remains over $10^4$ without degradation during the whole process. 2TFGM devices still show stable resistance switching under ±23 V with pulse width of 100/40 ns, indicating high-speed pulse programmability of our devices (Figure 1f). Good endurance ($>10^5$ cycles), long retention time ($>10^4$ s), high switching speed (40 ns) as well as good thermal stability are simultaneously realized, revealing such all-2D 2TFGMs as very promising candidates for digital data storage applications. We also demonstrate a $4 \times 4$ 2TFGM array fabricated from a standard planar fabrication technology (Figure 1g). Figure 1h shows the stored map of “N07” patterns (i–iii) by operating the 2TFGM devices array.
operating the 2TFGM devices array, suggesting good device uniformity and non-volatile properties (note that the pristine conductance states of the $4 \times 4$ 2TFGM array are shown in Figure S7, Supporting Information).

2.2. Multi-Bit Memory States Operation

Tunable conductance states can be achieved when different $V_	ext{ds}$ are applied for programming, enabling the multi-level states as illustrated in Figure 2a. Different conductance states for a device with 10 nm-thick h-BN are achieved by applying voltages from $+8$ to $+12$ V in the programming process, owing to different amount of holes stored in the FG under different applied voltages. The device conductance is unchanged for $V_	ext{ds}$ varying from 0 to $+7$ V, since electrons cannot tunnel from FG into drain through the barrier under small $V_	ext{ds}$. Multi-level resistance switching behaviors are also observed in the erasing process, as shown in Figure S8, Supporting Information. Good retention longer than $10^4$ s for each conductance state tested with 1 V read pulse is shown in Figure 2b. Furthermore, each conductance state shows a low drift coefficient ranging from 0.003 to 0.175, extracted using a power-law equation:

$$G(t) = G_0 \left( \frac{t}{t_0} \right)$$  \hspace{1cm} (1)$$

where $G_0$ is the initial conductance at $t_0$, and $\alpha$ is the conductance drift coefficient. Note that the highly reproducible multi-level resistive switching behavior can also be achieved by applying a series of voltage pulses with different amplitudes as shown in Figure S9, Supporting Information.

Quasi-continuous distinguishable conductance states can be realized in these 2TFGMs, as shown in Figure 2c. 131 quasi-continuous distinguishable conductance states with ultra-low drift between 0.1 to 13 nS with 0.1 nS per step are achieved successfully by applying a different number of pulses. The cumulative distribution of 16-level conductance states, which are distributed between 0.5 to 9 nS uniformly. The 2TFGM device could be programmed to a target conductance value simply by controlling the number of applied identical pulses of $\pm 12$ V.

Figure 2. Highly reliable multi-bit memory states of the all-2D FGM. a) $I-V$ hysteresis plots of a 2TFGM device with 10 nm-thick h-BN under different source-drain voltages $V_	ext{ds}$. $V_	ext{ds}$ gradually changed from $+4$ to $+12$ V for programming and $V_	ext{ds} = -12$ V for erasing. The device conductance is gradually increased with programming voltages, showing multi-level resistance switching. b) Retention time of more than $10^4$ s for multiple resistance states tested by 1 V pulses. Small conductance drift of each conductance state is calculated from the slope of logarithmic plot of conductance evolution over time, indicating the good stability of our 2TFGM device. c) 131 quasi-continuous distinguishable conductance states with ultra-low drift between 0.1 to 13 nS with 0.1 nS per step are achieved successfully by applying a different number of pulses. d) The cumulative distribution of 16-level conductance states, which are distributed between 0.5 to 9 nS uniformly. The 2TFGM device could be programmed to a target conductance value simply by controlling the number of applied identical pulses of $\pm 12$ V.
conductance states are shown in Figure 2d. The 2TFGM device can be programmed to a target conductance value by controlling the number of identical pulses applied. All these device characteristics indicate excellent reliability and stability of the 2TFGM devices.

2.3. Artificial Synaptic Behaviors Investigation

Such reliable multi-bit data storage is highly desirable for high-performance artificial synaptic devices. Figure 3a shows a schematic illustration of biological neurons, consisting of a soma, an axon and dendrites, and synapses, which are the conjunctions between two neighboring neurons. The enlarged area illustrates the synaptic chemical messengers transmission between the axon terminal of the pre-synaptic neuron and the dendrite of the post-synaptic neuron. If an action potential arrives at the pre-synaptic neuron, the synaptic membrane is depolarized and triggered neurotransmitters are released, diffusing across the synaptic cleft, docking with receptors on the post-synaptic neuron, and generating an excitatory postsynaptic current (EPSC) or inhibitory postsynaptic current (IPSC), whose amplitudes or intensities are determined by the synaptic weight. Long-term potentiation (LTP) and depression (LTD), which are two essential synaptic functions in learning, were also emulated by tuning the conductance state of the 2TFGM device gradually, as demonstrated in Figure 3b. 3000 quasi-continuous states were programmed for the LTP/LTD process through a series program/erase pulses of ±15 V (40 ns pulse width), suggesting that an extremely large number of states are available for neuromorphic computations in this 2TFGM-based artificial synapse. Figure 3c,d shows the transient electrical responses during the LTP/LTD process captured by oscilloscope. Figure 3e,f are the enlarged views of several cycles in Figure 3c,d, respectively, in which EPSC and IPSC biological behaviors are triggered by the pre-synaptic pulse of \( V_{ds} \). The energy consumption for a single pulse event can be calculated as \( I_{\text{peak}} \times t \times V_{ds} \), where \( I_{\text{peak}} \), \( t \), and \( V_{ds} \) are the peak value of ESPC (IPSC), the pulse width, and the operation voltage, respectively. The energy consumption for a single pulse is estimated to be 18 fJ (30 nA, ±15 V, 40 ns), which is much lower than that of conventional CMOS circuit (≈900 pJ). Such all-2D 2TFGM artificial synaptic devices hence exhibit great potential for high-speed and low-power neuromorphic computing.

2.4. Linear and Symmetric Weight Update Tuning

A linear and symmetric weight update behavior with sufficient number of states is critical for improving the inference accuracy and reliability of neural networks. While nearly linear and symmetric weight update can be realized by encoding the input pulse, it leads to restrictive circuit complexity, time delay, and additional energy consumption. Therefore, an
An ideal artificial synapse device that can be simply programmed with identical pulse is required for a low-power and high-accuracy neuromorphic circuit. As shown in Figures S11 and S12, Supporting Information, the conductance state of a 2TFGM device changes as a function of the pulse amplitude and width. Note that the 2TFGM device is always set to the same starting conductance state before applying the pulses. The dependences of the device conductance change on the programming pulse width and amplitude were systemically studied, as shown in Figure 4a,b. The conductance change ($\Delta G$) in both LTP/LTD processes increases with pulse width (amplitude) and ultimately saturates for larger pulse width (amplitude). The short dash lines are an exponential fitting to the experimental data. The saturation behavior is a consequence of the maximum number of trap charges in FG induced by a particular amplitude (width) of voltage. The plasticity of our 2TFGM artificial synapse is tuned by the amplitude of the operation pulse from $\pm 13$ to $\pm 10$ V (100 ms), as seen in Figure 4c. When the $\pm 13$ V pulses are applied, the resulted on/off ratio is 20, while the conductance saturates after only 20 pulses. More conductance states (up to 400) with small variation of conductance per pulse ($\Delta G < 0.1$ nS, Figure S13, Supporting Information) can also be realized by reducing the pulse amplitudes to $\pm 10$ V. The plasticity (including linearity, on/off ratio, and the number of states) in the 2TFGM artificial synapse was tuned by adjusting the amplitude of voltage pulses from $\pm 10$ to $\pm 13$ V (100 ms), demonstrating good reproducibility, linearity, and symmetry in the synaptic weight update. The pursuit for linear and symmetric weight update for neuromorphic computing. a,b) The changes in the device conductance as a function of pulse width (a) and amplitude (b) during the P/D process. c) The linearity, on/off ratio, and number of states in the P/D process could be tuned by adjusting the amplitude of voltage pulses from $\pm 10$ to $\pm 13$ V (100 ms). d) Nonlinearity analysis on the weight update of the different P/D curves in (c). The orange dashed lines represent the ideal linearity and symmetry of weight update. e) Cycled P/D operations of the 2TFGM artificial synapse. At least 50 states are programmed using a series of pulses with amplitude of $\pm 12$ V and width of 100 ms, demonstrating good reproducibility, linearity, and symmetry in the synaptic weight update. f) The image classification accuracy for hand-written digits from the MNIST database under different P/D processes as a function of the training epoch. A high recognition accuracy of 97.7% is achieved using the P/D process of 400 states and $V_{ds} = \pm 10$ V. g) The error rate after 20 training epochs corresponding to different P/D processes in (f). Inset illustrates the simulated neural network structure. Here a three-layer perceptron (including one hidden layer) is simulated with the standard backpropagation algorithm.
analog states) of the 2TFGM artificial synapse are tuned by the amplitude and width of the operation pulses.

Furthermore, quantitative analysis on the linearity of weight update behavior in the LTP/LTD processes is shown in Figure 4d. The nonlinearity factor ($\varphi$) \cite{18,23} which illustrates the nonlinear behavior of the weight update, is calculated based on the normalized conductance ($G_p$ or $G_d$) as a function of pulse number ($p$):

$$G_p = G_{\text{max}} + B \left(1 - e^{-\varphi p} \right)$$  \hspace{1cm} (2)$$

$$G_d = G_{\text{max}} - B \left(1 - e^{\varphi(p-p_{\text{max}})} \right)$$  \hspace{1cm} (3)$$

$$B = \frac{G_{\text{max}} - G_{\text{min}}}{1 - e^{-\varphi p_{\text{max}}}}$$  \hspace{1cm} (4)$$

where $G_{\text{max}}$ and $G_{\text{min}}$ represent the maximum and minimum conductance in the LTP/LTD processes, and $p_{\text{max}}$ is maximum pulse number. We found trade-offs among linearity, on/off ratio, and the number of states. The number of states could be increased by shortening the pulse length without sacrificing the nonlinearity and on/off ratio (Figure S14, Supporting Information). Comparing the nonlinearity of operation pulses (±11 V, 100 ms) with 150 states, 100 states, and 60 states, it is found that good linearity of weight update behavior could be obtained by reducing the number of states as well as the on/off ratios (Figure S15, Supporting Information). The best nonlinearity factor of $\varphi_p = 0.18$ for the potentiation process and $\varphi_d = -0.29$ for the depression process are achieved in our all-2D 2TFGM devices (Table S1, Supporting Information). Figure 4e shows the cycled LTP/LTD operations of the 2TFGM artificial synapse, demonstrating good reproducibility of the synaptic weight update behavior.

To further demonstrate the potential of 2TFGMs for neuromorphic computing, an artificial neural network has been simulated using the MNIST dataset based on the measured long-term plasticity characteristics. The schematic view of the simulated neural network is shown in the inset of Figure 4g, which is a three-layer perceptron (including one hidden layer), and the standard backpropagation algorithm is used for training. The learning curves are displayed in Figure 4f for different programming conditions. A high recognition accuracy of 97.7% (equivalently ~2.3% error rate as shown in Figure 4g), which approaches the software baseline accuracy of numerical simulations (~98%), is achieved after 20 training epochs using 400 states for the LTP/LTD processes under $V_{ds} = \pm 10$ V. This excellent performance is mainly attributed to the large number of conductance states available as well as the good linearity and symmetry in the weight update. In comparison with existing neuromorphic devices, such as typical redox, electrochemical, phase change, ferroelectric, magnetoresistive random access memory devices, and other memory devices based on 2D materials, the all-2D materials-based 2TFGM features plenty number of analog states (~3000) and low energy consumption per synaptic event (~18 fJ). Most importantly, high linearity and symmetric weight update behavior (0.18/−0.29 for P/D process) driven by the identical pulse is realized in the device, which has been pursued for the development of high energy-efficient artificial neural networks. See Figure S16, Supporting Information, for more details.

3. Conclusion

In summary, we demonstrated a high-performance artificial synapse of 2TFGM built from all-2D material van der Waals heterostructures. Ultralow drift coefficients, highly symmetric and linear synaptic weight update plasticity, large number of analog states, and fast updating speed have been exhibited in 2TFGMs, which are critical for accelerating the training efficiency, accuracy, and reducing the energy consumption in artificial neural networks. A high classification accuracy up to 97.7% has been achieved in the MNIST simulations. Our results demonstrate that such highly reliable all-2D 2TFGMs are very promising candidates for high-accuracy, low-energy consumption, and high-speed neuromorphic computing applications.

4. Experimental Section

Materials: High quality CVD-grown monolayer MoS2 films \cite{48} were used as channel materials in all-2D 2TFGMs. The growth process of MoS2 films was carried out in a CVD system with three temperature zones, during which S (Alfa, 99.5%, 8 g) powder and MoO3 (Alfa, 99.9995%, 30 mg) powder used as reaction sources and S-source, MoO3 individually and the pressure in the chamber was ~1 torr. The temperature was held at 130, 530, and 900 °C for 30 min. Thin h-BN flakes and FLG were exfoliated from h-BN and natural flaky graphite crystals (purchased from NGS Trading & Consulting GmbH, Germany). Large-scale thin FLG and h-BN flake could be obtained by poly-propylene-carbonate (PPC) assisted thermal exfoliation methods.

Device Fabrications: The MoS2 films on a sapphire substrate was first spin-coated by a polymethyl-methacrylate (PMMA) then etched in KOH solution (1 mol L$^{-1}$, 110 °C) for 30 min. The as-received MoS2 triangles supported on PMMA films, FLG, or h-BN flakes held by PPC films were stacked precisely through layer-by-layer stacking methods in our home-made transfer station. The sacrificing layer of PMMA or PPC could be removed by rinsing in acetone for >1 h at a room temperature. The floating gate layer and contact of FLG and the channel MoS2 layer were defined by electron beam lithography (EBL) and oxygen reactive ion etching (RIE) process. Devices were finally wired out by Ti (3 nm)/Au (30 nm) electrode for electrical measurements.

Sample Characterizations: Raman and photoluminescence (PL) spectra were acquired from a Horiba Jbin Yyon Lab RAM HR-Evolution Raman system with a 532-nm He-Ne laser (spot size=1 μm, power 10 mW) in ambient conditions. Surface morphology was characterized by atomic force microscope (Asylum Research Cypher S instruments) with AC160 TS tip under the tapping mode. The electrical measurements were carried out in a close-cycle cryogenic (70–500 K) probe station equipped with Agilent B1500 semiconductor parameter analyzer. All the measurements were operated with a base pressure of 10$^{-6}$ Torr under dark condition.

MNIST Simulations: The image classification simulations of handwritten digits were performed using the Modified National Institute of Standards and Technology (MNIST) dataset. The simulator used a 3-layer (784 × 200 × 10) perceptron neural network. The input layer with 784 neurons, representing the MNIST handwritten digit image (28 × 28 pixels), was fully connected to a hidden layer with 200 neurons, which was then fully connected to the output layer of 10 neurons. The total number of synapses was 784 × 200 + 200 × 10 = 158 800. Rectified
linear unit was used as the activation function of the hidden layer and softmax was used as the classification function of the output layer. For each synapse, the normalized conductance difference of a 2TFGM device and a reference resistor was used to represent its weight. All the weights were randomly initialized between $-1$ and $1$ and the stochastic gradient descent (SGD) with the cross-entropy loss function and a mini-batch size of 200 were used to train the network. The learning rate was 0.1. In each iteration, the update value of each synapse was first calculated according to the gradient and learning rate. Then a series of pulses were applied to each 2TFGM device, whose corresponding update value was larger than 0.001, until the exact weight change beyond the update value. The simulator could achieve a numerical accuracy limit of about 98% (baseline accuracy).

Supporting Information
Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements
The authors thank Dr. Aiwei Wang and Dr. Liangmei Wu in IOP-CAS (China) for fruitful discussions. This work is supported by the National Science Foundation of China (NSFC) under the grant No. 61888102, 11834017, 61734001, 51901023, 61974081, and 91964104, the Strategic Priority Research Program of Chinese Academy of Sciences (CAS) under the grant No. XDB30000000, the Key Research Program of Frontier Sciences of CAS under the grant No. QYZDB-SSW-SLH04, the National Key R&D program under grant No. 2016YFA0300904, and the Youth Innovation Promotion Association of CAS under the grant No. 2018013. Y.L. acknowledges the support from the Special Postdoctoral Researchers Program of RIKEN. K.W. and T.T. acknowledge the supports from the Frontier Sciences of CAS under the grant No. QYZDB-SSW-SLH004, the Priority Research Program of Chinese Academy of Sciences (CAS) under the grant No. QYZDB-SSW-SLH004, the NSFC under the grant No. 61888102, (China) for fruitful discussions. This work is supported by the National Science Foundation of China (NSFC) under the grant No. 61888102, 11834017, 61734001, and the CREST (JPMJCR15F3), JST.

Conflict of Interest
The authors declare no conflict of interest.

Author Contributions
J.T. and C.L.H. contributed equally to this work. G.Y.Z. supervised the experiments. Q.Q.W. supplied high-quality CVD grown monolayer MoS$_2$ samples. K.W. and T.T. supplied high quality h-BN crystals. J.T. fabricated the devices and carried out the electrical measurements with the assistance from C.L.H. J.S.T and Q.Z. performed the MNIST simulation. J.T., C.L.H., J.S.T, K.Y., Y.Z.L., and G.Y.Z. analyzed data, wrote and revised the manuscript. All authors discussed and commented on it.

Data Availability Statement
Research data are not shared.

Keywords
2D materials, artificial synapse, linear weight update, MoS$_2$

Received: December 24, 2020
Revised: February 18, 2021
Published online: March 24, 2021

[1] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessssche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Nature 2016, 529, 484.
[2] J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, L. Shi, Nature 2019, 572, 106.
[3] T. Skotnicki, J. A. Hutchby, T. J. King, H. S. P. Wong, F. Boeuf, IEEE Circuits Device 2005, 21, 16.
[4] M. A. Zidan, J. P. Strachan, W. D. Lu, Nat. Electron. 2018, 1, 22.
[5] W. Zhang, B. Gao, J. Tang, P. Yao, S. Yu, M. F. Chang, H. J. Yoo, H. Qian, H. Wu, Nat. Electron. 2020, 3, 371.
[6] C. Mead, Proc. IEEE 1990, 78, 1629.
[7] L. F. Abbott, W. G. Regehr, Nature 2004, 431, 796.
[8] J. Tang, F. Yuan, X. Shen, Z. Wang, M. Rao, Y. He, Y. Sun, X. Li, W. Zhang, Y. Li, B. Gao, H. Qian, B. I. Song, J. J. Yang, H. Wu, Adv. Mater. 2019, 31, 1902761.
[9] T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, M. Aono, Nat. Mater. 2011, 10, 591.
[10] S. H. Jo, T. Chang, I. Ebbing, B. B. Bhadviya, P. Mazumder, W. Lu, Nano Lett. 2010, 10, 1297.
[11] D. B. Strukov, G. S. Snider, D. R. Stewart, R. S. Williams, Nature 2008, 453, 80.
[12] T. Tura, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, Nat. Nanotechnol. 2016, 11, 693.
[13] N. Locatelli, V. Cross, J. Grollier, Nat. Mater. 2013, 13, 11.
[14] S. Kim, B. Choi, M. Lim, J. Yoon, J. Lee, H. D. Kim, S. J. Choi, ACS Nano 2017, 11, 2814.
[15] Q. A. Yu, Y. S. Shin, Y. R. Kim, V. L. Nguyen, W. T. Kang, H. Kim, D. H. Luong, I. M. Lee, K. Lee, D. S. Ko, J. Heo, S. Park, Y. H. Lee, W. J. Yu, Nat. Commun. 2016, 7, 12725.
[16] M. Jerry, P. Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, S. Datta, IEEE Int. Electron Devices Meet. 2017, 6.2.1.
[17] C. S. Yang, D. S. Shang, N. Liu, G. Shi, X. Shen, R. C. Yu, Y. Q. Li, Y. Sun, Adv. Mater. 2017, 29, 1700906.
[18] J. Tang, D. Bishop, S. Kim, M. Capel, T. Gokmen, T. Todorov, S. Shin, K. T. Lee, P. Solomon, K. Chan, W. Haensch, J. Rozen, IEEE Int. Electron Devices Meet. 2018, 13.1.1.
[19] S. Dai, Y. Zhao, Y. Wang, J. Zhang, L. Fang, S. Jin, Y. Shao, J. Huang, Adv. Funct. Mater. 2019, 29, 1903700.
[20] H. S. Wong, S. Salakhuddin, Nat. Nanotechnol. 2015, 10, 191.
[21] J. J. Yang, D. B. Strukov, D. R. Stewart, Nat. Nanotechnol. 2013, 8, 13.
[22] J. Woo, S. Yu, IEEE Nanotechnol. Mag. 2018, 12, 36.
[23] P. Y. Chen, B. Lin, I. T. Wang, T. H. Hou, J. Y. Ye, S. Vrudhula, J. S. Seo, Y. Cao, S. Yu, IEEE/ACM Int. Conf. Comput. Aided Des. 2015, 194.
[24] S. Choi, S. H. Tan, Z. Li, Y. Kim, C. Choi, P. Y. Chen, H. Yeon, S. Yu, J. Kim, Nat. Mater. 2018, 17, 335.
[25] E. J. Fuller, S. T. Keene, A. Melianas, Z. Wang, S. Agnarwal, Y. Li, Y. Tuchman, C. D. James, M. J. Marinella, J. J. Yang, A. Salleo, A. A. Talin, Science 2019, 364, 570.
[26] Y. Jeong, S. Kim, W. D. Lu, Appl. Phys. Lett. 2015, 107, 173105.
[27] C. Liu, H. Chen, S. Wang, Q. Liu, Y. G. Jiang, D. W. Zhang, M. Liu, P. Zhou, Nat. Nanotechnol. 2020, 15, 545.
[28] F. Zhou, J. Chen, X. Yao, X. Wang, Y. Chai, Research 2019, 2019, 9490413.
[29] D. Akinwande, N. Petrone, J. Hone, Nat. Commun. 2014, 5, 5678.
[30] V. K. Sangwan, H. S. Lee, H. Bergeron, I. Balla, M. E. Beck, K. S. Chen, M. C. Hersam, Nature 2018, 554, 500.
[31] Y. Shi, X. Liang, B. Yuan, V. Chen, H. Li, F. Hui, Z. Yu, F. Yuan, E. Pop, H. S. P. Wong, M. Lanza, Nat. Electron. 2018, 1, 458.
[32] M. Wang, S. Cai, C. Pan, C. Wang, X. Lian, Y. Zhao, K. Xu, T. Cao, X. Pan, B. Wang, S. J. Liang, J. J. Yang, P. Wang, F. Miao, Nat. Electron. 2018, 1, 130.
[33] X. Zhu, D. Li, X. Liang, W. D. Lu, Nat. Mater. 2019, 18, 141.

[34] T. Paul, T. Ahmed, K. Kanhaiya Tiwari, C. Singh Thakur, A. Ghosh, 2D Mater. 2019, 6, 045008.

[35] S. G. Yi, M. U. Park, S. H. Kim, C. J. Lee, J. Kwon, G. H. Lee, K. H. Yoo, ACS Appl. Mater. Interfaces 2018, 10, 31480.

[36] C. He, J. Tang, D. S. Shang, J. Tang, Y. Xi, S. Wang, N. Li, Q. Zhang, J. K. Lu, Z. Wei, Q. Wang, C. Shen, J. Li, S. Shen, J. Shen, R. Yang, D. Shi, H. Wu, S. Wang, G. Zhang, ACS Appl. Mater. Interfaces 2020, 12, 11945.

[37] X. Cui, G. H. Lee, Y. D. Kim, G. Arefe, P. Y. Huang, C. H. Lee, D. A. Chenet, X. Zhang, L. Wang, F. Ye, F. Pizzocchero, B. S. Jessen, K. Watanabe, T. Taniguchi, D. A. Muller, T. Low, P. Kim, J. Hone, Nat. Nanotechnol. 2015, 10, 534.

[38] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, J. Hone, Nat. Nanotechnol. 2010, 5, 722.

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

[40] M. Boniardi, D. Ielmini, S. Lavizzari, A. L. Lacaita, A. Redaelli, A. Pirovano, IEEE Trans. Electron Devices 2010, 57, 2690.

[41] K. Ding, J. Wang, Y. Zhou, H. Tian, L. Lu, R. Mazzarello, C. Jia, W. Zhang, F. Rao, E. Ma, Science 2019, 366, 210.

[42] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A. S. LaMantia, L. E. White, Neuroscience, Oxford University Press, Oxford 2017.

[43] G. Indiveri, E. Chicca, R. Douglas, IEEE Trans. Neural Networks 2006, 17, 211.

[44] M. Zhao, B. Gao, J. Tang, H. Qian, H. Wu, Appl. Phys. Rev. 2020, 7, 011301.

[45] Z. Wang, H. Wu, G. W. Burr, C. S. Hwang, K. L. Wang, Q. Xia, J. J. Yang, Nat. Rev. Mater. 2020, 5, 173.

[46] S. Park, A. Sheri, J. Kim, J. Noh, J. Jang, M. Jeon, B. Lee, B. R. Lee, B. H. Lee, H. Hwang, IEEE Int. Electron Devices Meet. 2013, 25.6.1.

[47] D. Kuzum, R. G. Jeyasingh, B. Lee, H. S. Wong, Nano Lett. 2012, 12, 2179.

[48] Q. Wang, N. Li, J. Tang, J. Zhu, Q. Zhang, Q. Jia, Y. Liu, Z. Wei, H. Yu, Y. Zhao, Y. Guo, L. Gu, G. Sun, W. Yang, R. Yang, D. Shi, G. Zhang, Nano Lett. 2020, 20, 7193.