

Supporting Information

Dual-Gate Manipulation of HfZrOx-Based MoS₂ Field-Effect Transistor Towards Enhanced Neural Network

Yilun Liu, Qingxuan Li, Hao Zhu*, Li Ji, Qingqing Sun, David Wei Zhang, and Lin Chen*

State Key Laboratory of ASIC and System, School of Microelectronics,

Fudan University, Shanghai 200433, P. R. China.

| | LTP Linearity | Minimum EPSC Time | Image Classification Feasibility | CMOS Process Compatibility |
|---|---------------|-------------------|----------------------------------|----------------------------|
| MoS ₂ Synaptic Transistors ²² | ✓ | 50 ms | ✗ | ✗ |
| Dual-Gated MoS ₂ Neuristor ²³ | ✓ | 100 ms | ✗ | ✗ |
| Laterally Coupled MoS ₂ Transistor ²⁴ | ✗ | 100 ms | ✗ | ✗ |
| Polycrystalline MoS ₂ Artificial Synapses ²⁵ | ✗ | 1 ms | ✓ | ✓ |
| MoS ₂ Charge-trapping synaptic device ⁸ | ✗ | 10 ms | ✓ | ✓ |
| HZO Ferroelectric Tunneling Junctions ²⁶ | ✗ | 10 μs | ✗ | ✓ |
| α-In ₂ Se ₃ Ferroelectric Channel Transistors ²⁷ | ✗ | 30 ms | ✗ | ✓ |
| Waterproof Flexible PEDOT Synapses ²⁸ | ✗ | 50 ms | ✗ | ✗ |
| This work | ✓ | 10 ms | ✓ | ✓ |

Figure S1 The benchmark of emerging synaptic devices

Figure S1 shows the benchmark of emerging synaptic devices. We summarize some emerging synaptic devices with different structures. The LTP linearity, the shortest EPSC Time, the feasibility of image recognition and the compatibility with CMOS technology are compared in the following table.

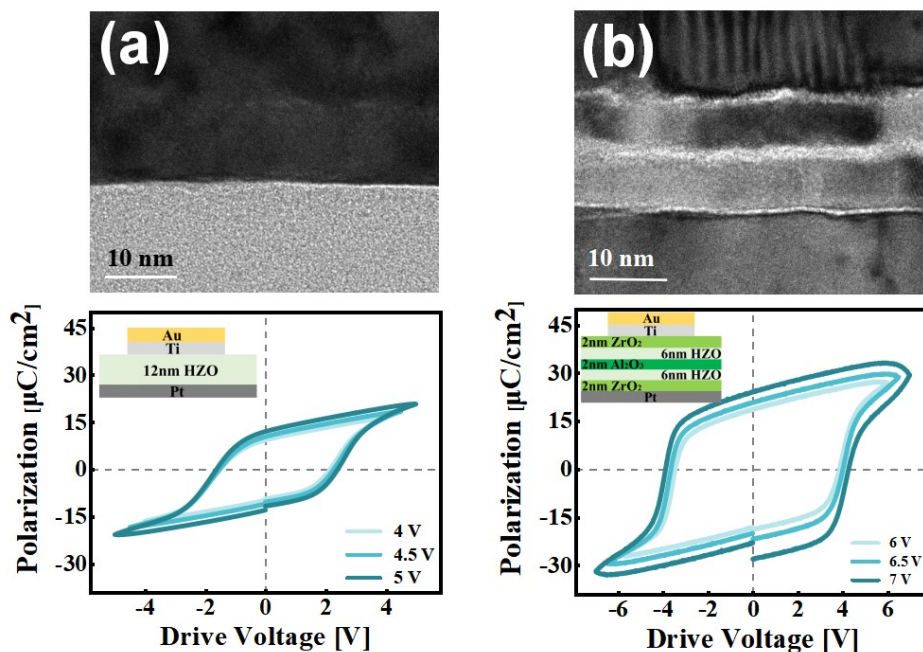


Figure S2 The P-V characteristics of (a) 12 nm HZO reference film and (b) crested symmetric structure film.

Figure S2 shows the P-V characteristics of the Zr-doped HfO₂ (HZO)-involved crested symmetrical stack and control sample with pure 12 nm HZO in capacitor structures. The inset shows the schematic test structure. It is observed from the curves that the control device with pure HZO layer exhibits a remanent polarization (Pr) of only 12 $\mu\text{C}/\text{cm}^2$, while for the stack with crested symmetry structure, the Pr can reach 24 $\mu\text{C}/\text{cm}^2$.

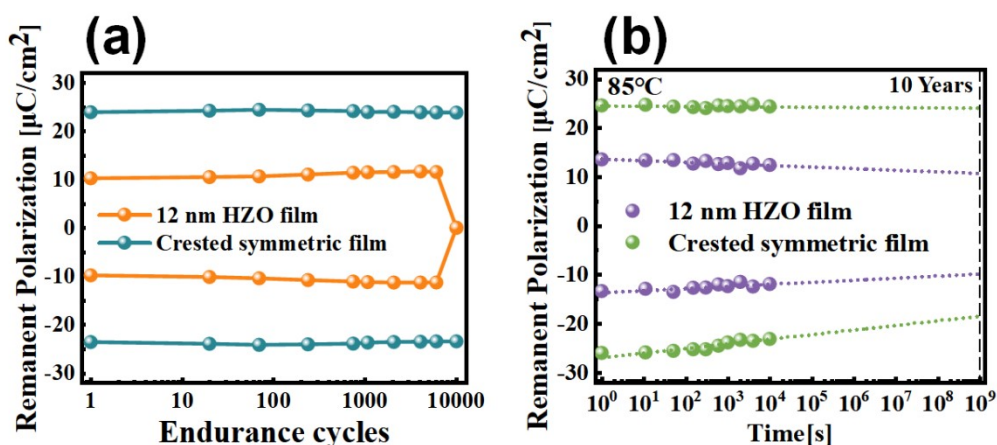


Figure S3 (a) Fatigue tests of the 2 film. (b) Retention properties of the 2 film at 85 °C.

Figure S3 shows the crested symmetric film can still maintain 24 $\mu\text{C}/\text{cm}^2$ with negligible degradation after 10⁴ cycles. The 12 nm HZO fails to maintain the Pr after 10⁴ pulse cycles, which can be attributed to the accumulation of defects during the stress process forming leakage path and a final breakdown. The loss of 2Pr of the devices with crested symmetric after 10-year extrapolation are 16% with respect to the initial value, while the loss of remanent polarization is over 25% in 12 nm HZO film.

We provide hysteresis loops, fatigue and retention characteristics of polarization, mainly to show that the stack structure we use can enhance the polarization characteristics of the film in many ways. In this way, the device can obtain a sub threshold swing lower than the physical limit 60mV/dec, effectively reducing the gate leakage and the static power consumption of the device. Neurons are the computing engine of the brain, they receive input signals from thousands of synapses and can complete complex tasks such as learning, memory, computing and parallel processing with low power consumption. Our double gate device and the artificial neural network based on it are mainly used to simulate the human brain. Therefore, the lower the static power consumption of the device, the more suitable for the simulation of biological synapses and the construction of brain like neural networks.

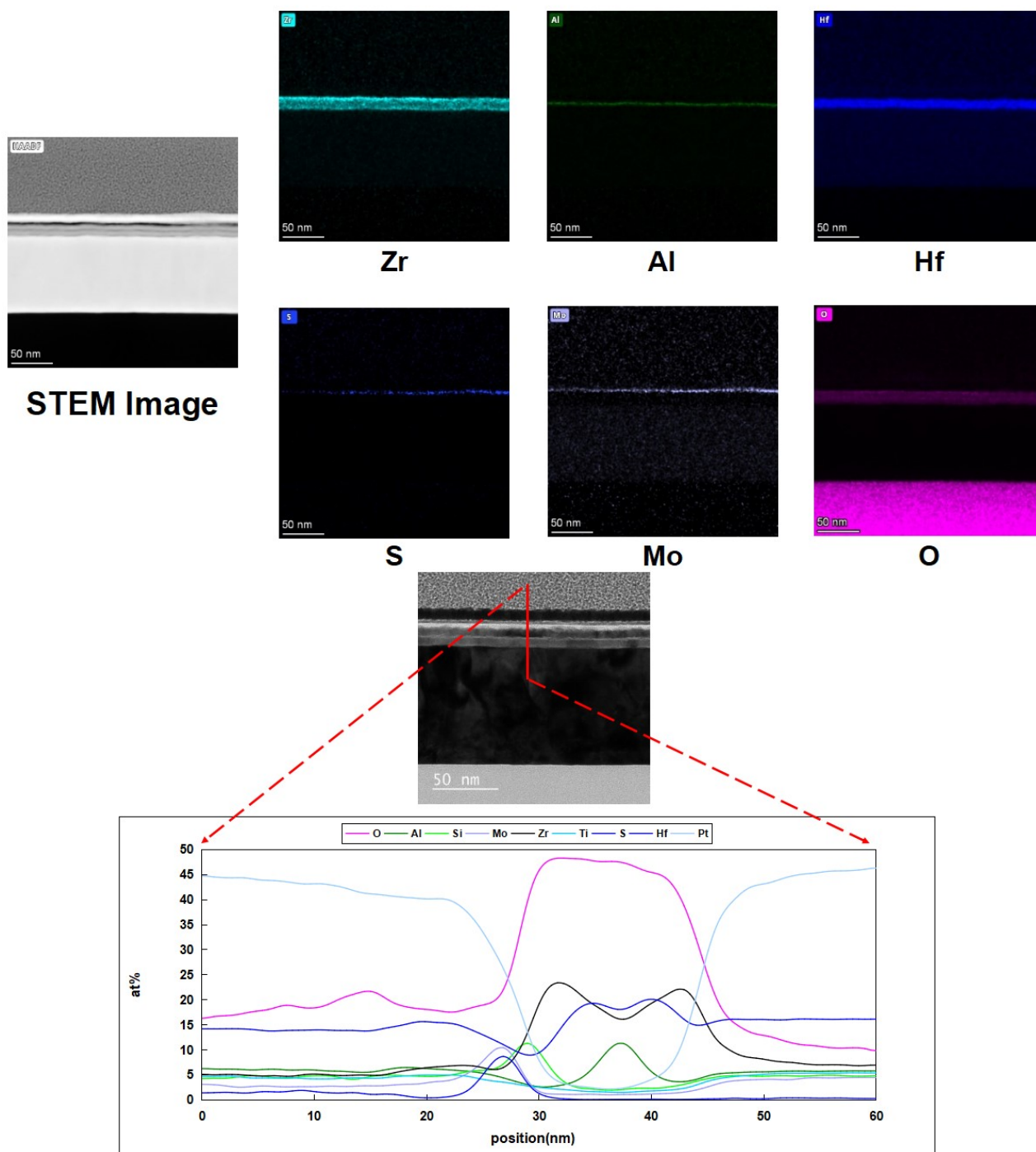


Figure S4. Energy Dispersive X-ray Spectroscopy (EDX) at gate.

It can be seen from Figure S4 that the layered distribution of the elements in the stack is basically consistent with that in the structure diagram. It can be seen that the growth of the stack is basically flat, without large fluctuations. And it can be seen that the layered part at the dotted line has no introduction of other elements

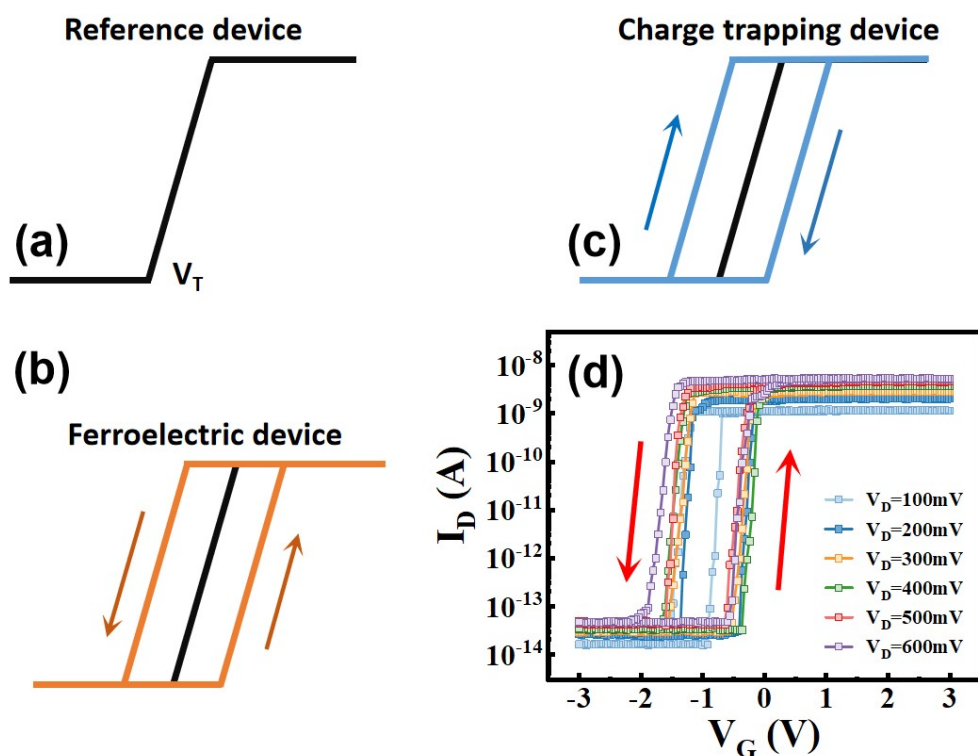


Figure S5. Schematic diagram of the transfer characteristics of several devices, (a) reference transistor, (b)ferroelectric device and (c)charge trapping device.(d)The hysteresis curve of the transfer characteristic.

The counterclockwise switching is an intrinsic characteristic of the FeFET, whereas a clockwise one is to be expected for charge trapping device. The V_T shift in charge trapping devices is determined by the amount of charges trapped. In FeFET, the V_T shift is controlled by FE polarization switching between two states where the energy of the FE system becomes minimal.

For instance, when the gate voltage is negative, for ferroelectric transistors, the direction of the electric domain in the ferroelectric layer with the applied electric field is downward, and positive charges are induced in the channel layer. To obtain the carrier concentration of the device to turn on, a larger inversion voltage is required, so V_T increases, as shown in Figure S4 (b), while for charge trapping devices, the negative gate voltage will cause the trapped electrons in the dielectric layer to be repelled and tunnel into the channel area, increasing the carrier concentration in the channel and reducing the required inversion electron concentration, so V_T will decrease, as shown in Figure S5 (c). Likewise, when the gate voltage is positive, the V_T of the ferroelectric device decreases and the V_T of the charge trapping device increases. We tested the hysteresis curve of the transfer characteristic of the device in Figure S5 (d) and found it to be counterclockwise, which is consistent with the above conclusion, indicating that the ferroelectric effect rather than charge trapping occurs in our device. At the same time, the device exhibits a subthreshold swing below the Boltzmann limit 60mV/dec.

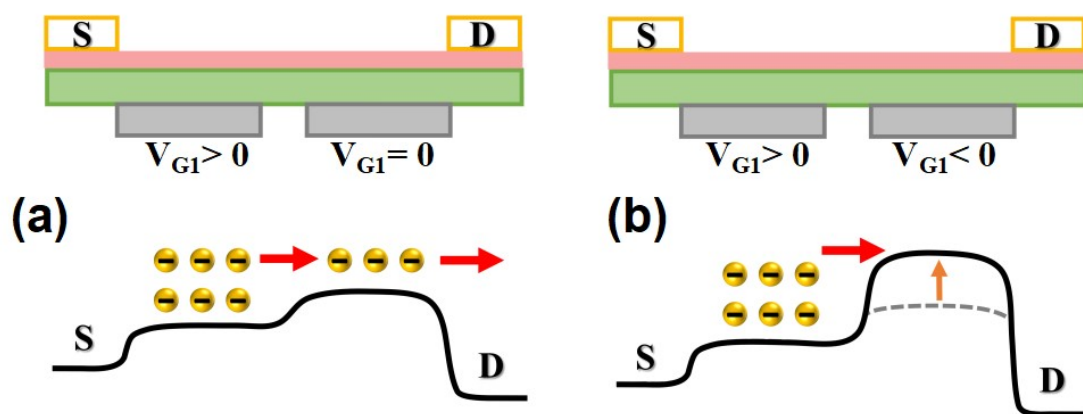


Figure S6. Band diagram of the dual-gate MoS₂ device under 2 gate control.

Figure S6 illustrating the energy band diagrams of the device, positive V_{G1} ($V_{G2}=0$) reduces the barrier height and makes it easier for electrons to pass through the barrier. When an increasing negative voltage is applied on G2, the barrier height controlled by G2 increases, making it more difficult for electrons to pass through. Therefore, as the negative voltage on G2 increases, the I_{DS} is gradually reduced until the device is turned off. With more negative V_{G2} , the barrier height at G2 increases, so that the I_{DS} change caused by the pulse signal on G1 is partially shielded by the barrier at G2, and the reduction of I_{DS} peak reduces the conductance state range of the device. This means the G2 electrode can be taken as a regulatory terminal which manipulate the channel current of the device together with G1.

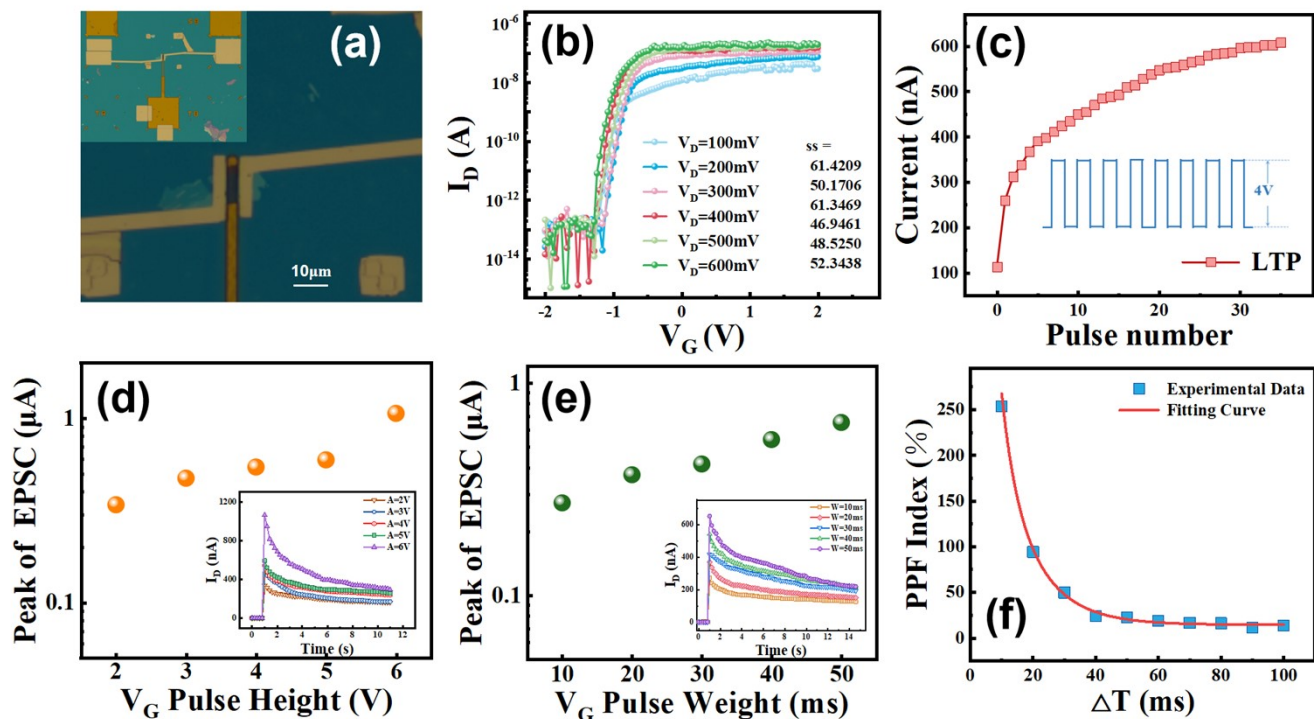


Figure S7. (a) Optical microscope image of a single-gate MoS₂ FET on the 300 nm SiO₂/Si substrate. (b) Transfer characteristic (I_D - V_G) of the single-gate MoS₂ FET. (c) LTP behaviors of the device under successive pulses. The inset shows the schemes of applied voltage for the LTP. Peak of EPSC curves caused by gate pulses with (d) different pulse amplitude and (e) different pulse width. (f) The paired-pulse facilitation (PPF) of the stacked MoS₂ transistor.

The optical microscope image, transfer characteristic curves and synaptic characteristics of a typical single-gate transistor are shown in Figure S6. It can be seen that the device can achieve a switching ratio of 10^6 and a subthreshold swing below 60 mV/dec. The I_{DS} of the device increases with the pulse width and amplitude of V_G . Like the double gate device, the LTP curve of typical single gate device also has the nonlinear problem.

Whether it is a single gate device or a double gate device, we try our best to make the drain source electrode cling to the gate in the vertical direction, to avoid overlap as much as possible. For the double gate device, the reactive ion etching is required after the photoresist pattern is prepared to dig out the gate groove, so the distance between the two gates cannot be too close, otherwise the two gates cannot be separated during the life-off process after the subsequent metal growth. For our double-gate device, the distance between the two gate electrodes is 4 μm . If the distance is too large, the on current of the device will be too small; If the distance is too small, the two gates will be connected and the device cannot be prepared further.

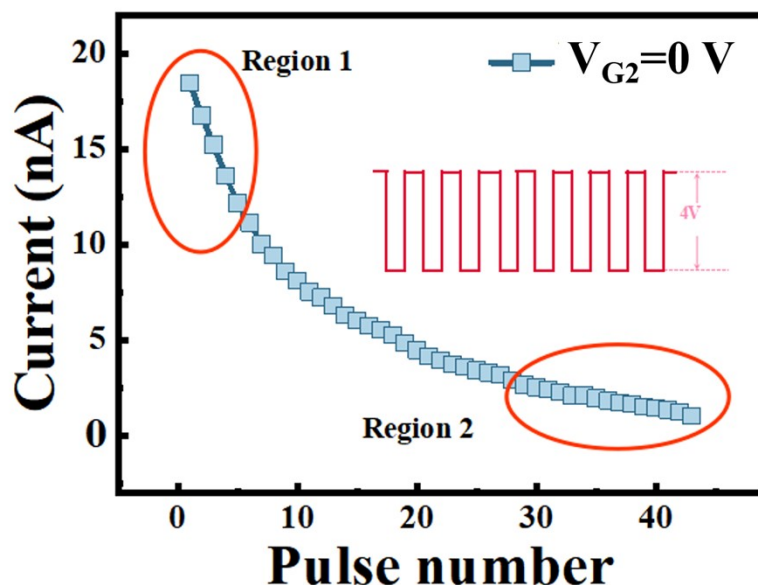


Figure S8. LTD behaviors of the device ($V_{G2}=0$ V) under successive pulses

The LTD characteristics of a single-gate-controlled device under the continuous pulse. A total of 43 identical negative pulse sequences are applied on G1 ($V_{G2}=0$). In region-1 under the initial stimulation of pulse set, I_D exhibits a large change, and there are few states in a certain current interval. In the last few pulse stimulation (like in region 2), the I_D of the device can increase linearly and intensively with the increase in the number of pulses, and more importantly, the amount of conduction states in this stage is larger within the same current interval as compared to that in the initial stage. This is consistent with the characteristics of LTP.

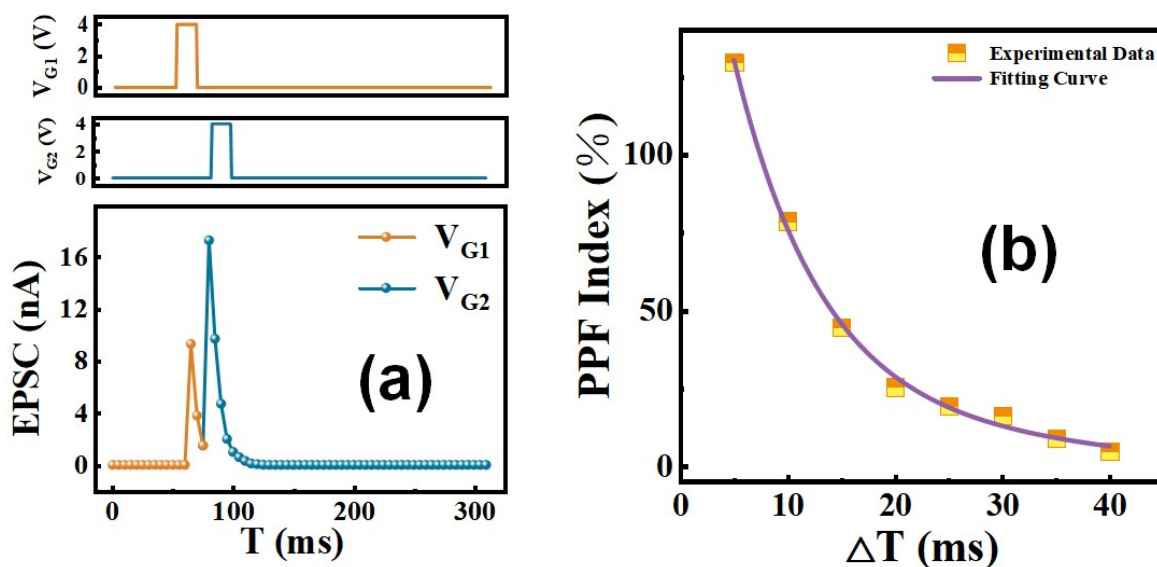


Figure S9. Paired pulse facilitation(PPF) characteristics of double-gate devices.

Figure S9 shows the paired-pulse facilitation (PPF) of the stacked MoS_2 transistor under dual gate control. This phenomenon is the application of two consecutive pulses to the two gates in sequence. It can be seen from Figure S5 (a) that the EPSC peak value induced by the second pulse is significantly larger than the EPSC amplitude induced by the first pulse, which indicates that this phenomenon of this device is basically consistent with the PPF phenomenon of neural synapses in biological systems. This phenomenon occurs because the time interval between the two gate pulse voltages is short, so that the dynamic changes of EPSC are coupled in time. This could also illustrate the effect of G2 on the synaptic characteristics of the whole device.

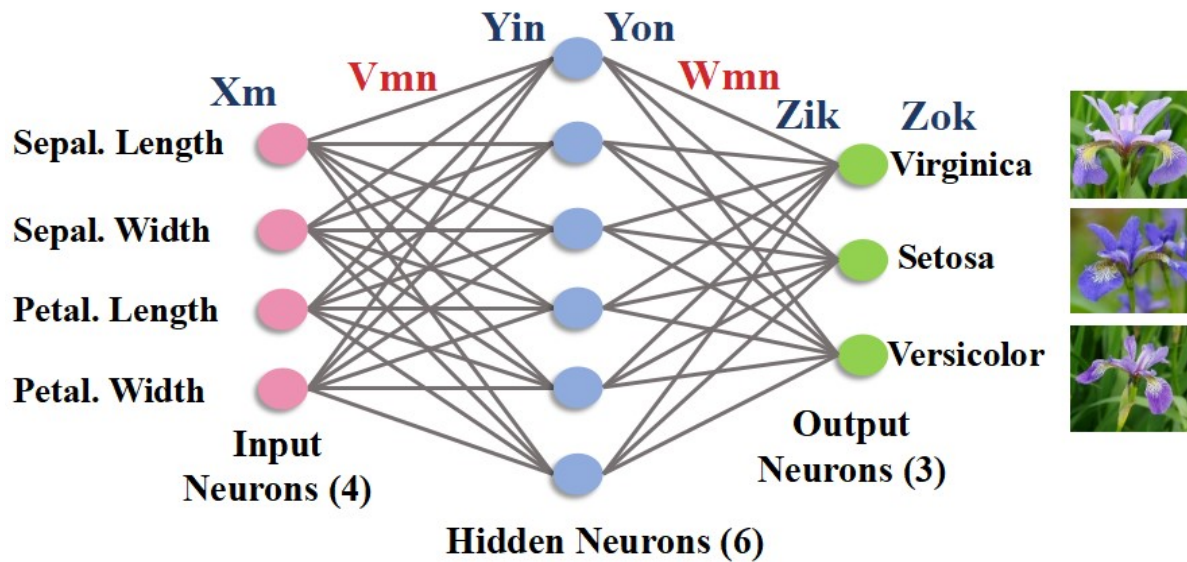


Figure S10. ANN with three layers, containing 4 input neurons, 6 hidden neurons, and 3 output neurons for iris classification.

In the ANN network training, 112 pieces of data in the database are used for training, the back propagation algorithm is used to update the weights after each training epochs, and the remaining 38 pieces of data are used to test the recognition rate of the network. Figure 5 shows flowchart of the training process. In our ANN testing process, X_m represents an input neuron, so the information passed from the input layer to the hidden layer is expressed as:

$$Y_{in} = \sum_{m=1}^4 X_m V_{mn}$$

where V_{mn} represents the weight value between an input neuron X_m and a hidden neuron Y_{in} , and all V_{mn} form the matrix V with a total of 4×6 weight values. The hidden layer uses the sigmoid activation function, so hidden layer neurons Y_{on} can be expressed as:

$$Y_{on} = \frac{1}{1 + e^{-Y_{in}}}$$

Thus, the input value of an output neuron can be expressed as:

$$Z_{ik} = \sum_{n=1}^6 Y_{on} W_{nk}$$

where W_{nk} represents the weight value between a hidden neuron Y_{on} and an output neuron Z_{ik} , and all W_{nk} form the matrix W with a total of 6×3 weight values. The output layer uses the sigmoid activation function, so output layer neurons Z_{ok} can be expressed as:

$$Z_{ok} = \frac{1}{1 + e^{-Z_{ik}}}$$

By comparing it with the correct result, the current output error of the neural network can be obtained:

$$E = \frac{1}{2} \sum_{k=1}^3 (O_k - Z_k)^2$$

where O_k is the correct output value. The value of it is 0 or 1. When the input picture is judged as this kind of flower, O_k is 1, otherwise 0.

Next, the weight value is updated based on the derivative of the error during the back-propagation process. ΔV_{mn} and ΔW_{nk} is used to represent the modification amount of the weight, and the modified weight is represented as V_{mn}' and W_{nk}' .

$$\Delta V_{mn} = \mu \frac{\partial E}{\partial V_{mn}}$$

$$\Delta W_{nk} = \mu \frac{\partial E}{\partial W_{nk}}$$

$$V_{mn}' = V_{mn} + \Delta V_{mn}$$

$$W_{nk}' = W_{nk} + \Delta W_{nk}$$

The modified weight value is the weight parameter of the neural network in the next forward propagation. After the set training epoch is completed, the network training is completed.

Then we used our double-gate device to simulate the image recognition function. With V_{G2} adjustment, the nonlinearity of LTP curve at the beginning of continuous pulse is improved without changing dynamic range of conductance (without V_{G2} : 0.5nA-18.6nA, with different V_{G2} : 0.2nA-18.6nA). In general, the recognition rate obtained by simulating the recognition function will be lower than the ideal recognition rate obtained from network training, because not all weights can be accurately replaced by the

conductance state. For our ANN based on double gate devices, the actual recognition rate can reach 100% ,which is the same as training process.

| | Current(nA) | |
|---------------------------|-------------|------|
| $V_{G2} = -0.5 \text{ V}$ | 1.9 | 4.93 |
| $V_{G2} = -0.4 \text{ V}$ | 4.93 | 6.87 |
| $V_{G2} = -0.3 \text{ V}$ | 6.87 | 9.18 |
| $V_{G2} = -0.2 \text{ V}$ | 9.18 | 11.3 |
| $V_{G2} = -0.1 \text{ V}$ | 11.3 | 14.8 |
| $V_{G2} = 0 \text{ V}$ | 14.9 | 18.6 |

Figure S11. The V_{G2} switching table with increasing I_D

The Figure S11 shows the increasing process of V_{G2} as I_D increases to obtain a more linear conductance state.

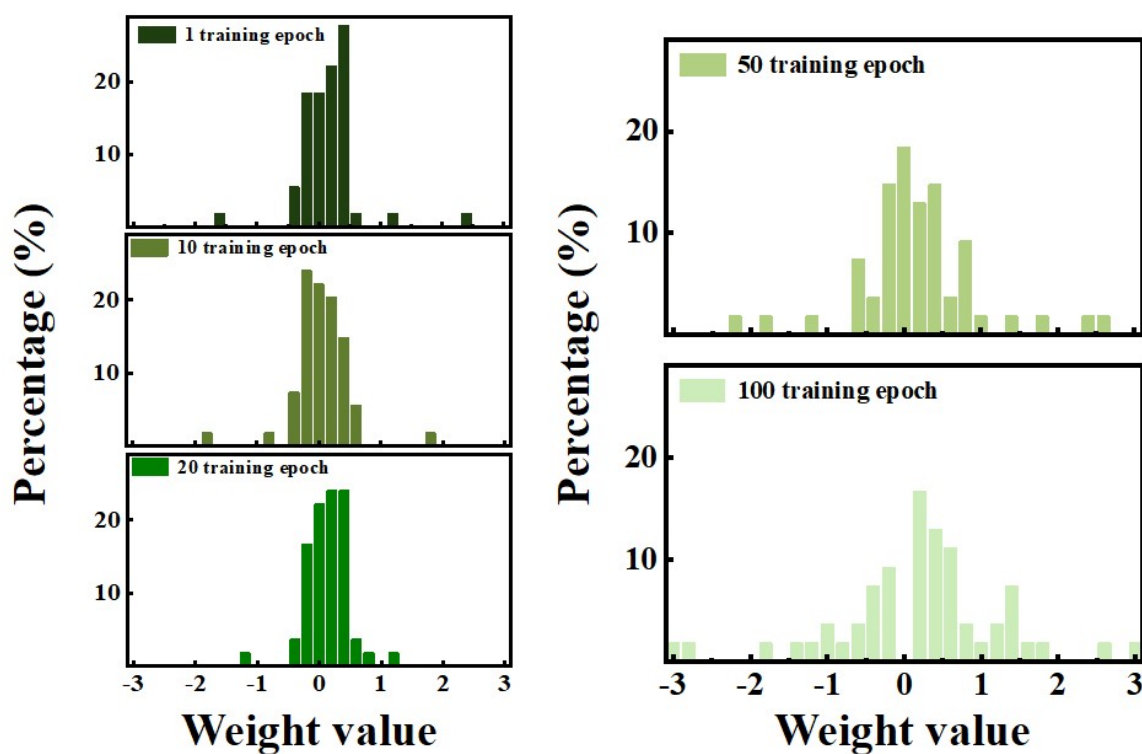


Figure S12. The weight value distribution changes with the number of training learning epochs

Figure S12 show the statistical distribution of the synaptic weights of the neural network at 1, 10, 20, 50, and 100 epochs of training. It can be seen that the distribution of weights has gradually shifted from being more concentrated to a quasi-normal distribution

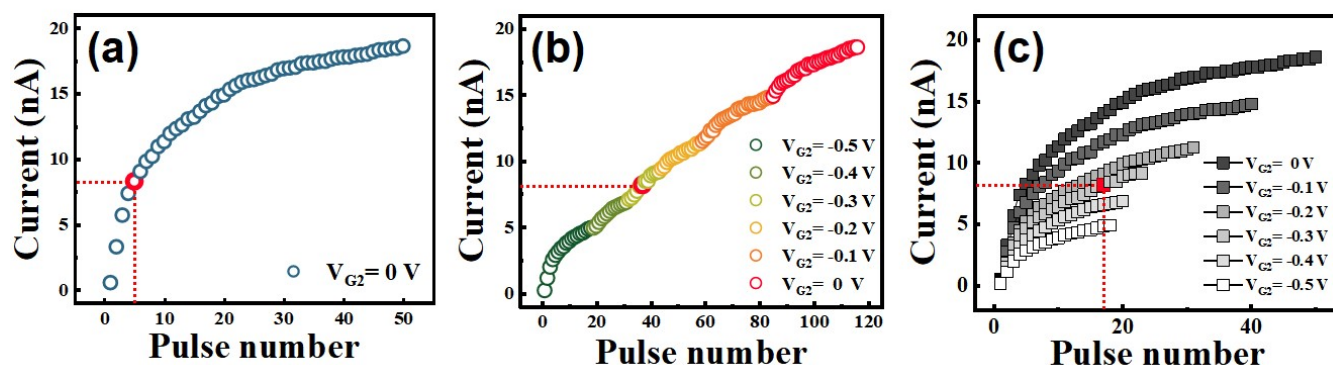


Figure S13. (a) LTP curve without G2 voltage ($V_{G2}=0$ V).
 (b) V_{G2} voltage corresponding to I_{DS} state (the conductance state).
 (c) LTP curves with different V_{G2} .

The key of building artificial neural network with neural synaptic devices is to effectively control the I_{DS} of devices (the conductance state), by adjusting the number of input pulses. For devices without G2 regulation, as shown in Figure S7(a), if 8.2 nA current is required in the use of the network, it can be obtained by applying 5 consecutive and identical pulses to the device. For a double gate device with G2 regulation, if it is necessary to obtain a current state of 8.2 nA, first judge the corresponding V_{G2} under this current, which can be obtained by comparing Figure S13(b) or Figure S11. It can be obtained that -0.3 V should be applied to G2 at this time, corresponding to Figure S13(c), and it can be seen that 17 continuous pulses should be applied to G1 at this time with $V_{G2} = -0.3$ V.

The conductance state of the whole system is adjusted by LTP curves under different G2, and there are different LTP curves corresponding to specific current states, so our devices can achieve continuous weight modulation. The weight updating method is also directly controlled by the number of pulses applied on the G1, which is consistent with the situation without G2 control.

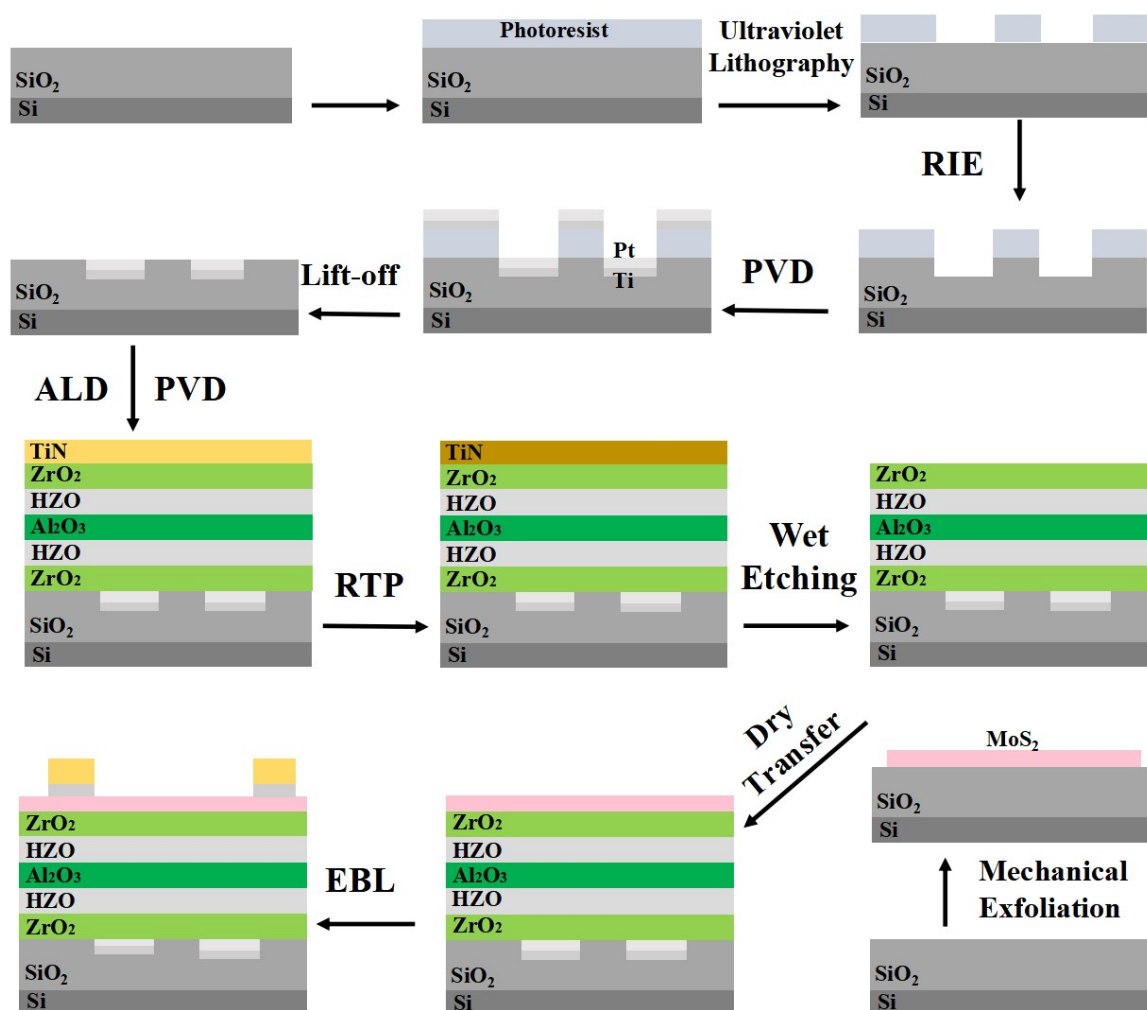


Figure S14. Device fabrication process flow.

Figure S14 schematically shows the process flow of device fabrication. The HZO ferroelectric layer and the dielectric stack are deposited by atomic layer deposition (ALD) at 300 °C with trimethylaluminium (TMA), tetrakisethylmethylamino-hafnium (TEMAH) and tetrakis-ethylmethylamino zirconium (TEMAZ) as the precursors for Al, Hf and Zr, respectively. H₂O is used as O precursor.