ReS$_2$ Charge Trapping Synaptic Device for Face Recognition Application

Ze-Hui Fan, Min Zhang, Lu-Rong Gan, Lin Chen*, Hao Zhu, Qing-Qing Sun and David Wei Zhang

Abstract
Synaptic devices are necessary to meet the growing demand for the smarter and more efficient system. In this work, the anisotropic rhenium disulfide (ReS$_2$) is used as a channel material to construct a synaptic device and successfully emulate the long-term potentiation/depression behavior. To demonstrate that our device can be used in a large-scale neural network system, 165 pictures from Yale Face database are selected for evaluation, of which 120 pictures are used for artificial neural network (ANN) training, and the remaining 45 pictures are used for ANN testing. A three-layer ANN containing more than $10^5$ weights is proposed for the face recognition task. Also 120 continuous modulated conductance states are selected to replace weights in our well-trained ANN. The results show that an excellent recognition rate of 100% is achieved with only 120 conductance states, which proves a high potential of our device in the artificial neural network field.

Keywords: Charge trapping memory, Synaptic device, Two-dimension material, Artificial neural network

Background
Since the advent of modern computers, the von Neumann structure, wherein the arithmetic unit is separated from the memory, has been widely used. This kind of structure makes data transmission between the arithmetic unit and memory becomes a bottleneck, significantly limiting the improvement of computer performances [1, 2]. Meanwhile, the arithmetic unit and main memory are both volatile devices with high energy consumption, and information will disappear immediately if the power is cut off [3]. In contrast, the human brain is an efficient information storage and computing system with high fault tolerance and low power consumption, and information will be stored even if the power is off for years [4]. This is because the human brain consists of a highly interconnected, massively parallel, and structurally variable complex network consisted of about $10^{11}$ neurons and $10^{15}$ synapses [4, 5]. These neurons are considered to be the brain’s computational engines, receiving input signals from thousands of synapses in parallel. Synaptic plasticity is a biological process that changes synaptic weight through synaptic activities, and it is considered as a source of learning and memory [6].

The two-dimension (2D) materials with a small size and excellent electronic properties, such as graphene, transition metal dichalcogenides (TMDCs), and black phosphorus, have attracted significant attention and have been successfully implemented into synaptic devices [7, 8]. The TMDCs with the symmetric lattice, such as MoS$_2$ and WSe$_2$, have been widely studied [9, 10]. On the other hand, rhenium disulfide (ReS$_2$) with a distorted octahedral (1T) crystal structure has been rarely explored in the neuromorphic field. Most TMDs have a direct bandgap in the monolayer and an indirect bandgap in the multilayer, so a monolayer material that is difficult to obtain is needed for good device performance. However, ReS$_2$ within ten layers are all considered to have a direct bandgap [11], which means ReS$_2$ within ten layers can all perform well. Besides, the asymmetric lattice structure leads to weaker interlayer coupling energy, which benefits the exfoliation work, and thus makes the synaptic device much easier to fabricate [12–15]. In this study, ReS$_2$ film is used as a channel material. The crystal structure of monolayer ReS$_2$ is shown in Fig. 1a, where directions $a$ and $b$ denote the second shortest axis and the shortest axis in the basal plane, respectively. Based on the previous scientific researches and plenty of optical images of our exfoliated ReS$_2$ film [13], direction $b$ denotes the crystallographic orientation with the highest electron mobility. To illustrate the electrical characteristics of our ReS$_2$ synaptic device better, direction $b$ is considered as a direction of channel current, as shown in Fig. 1b.

* Correspondence: linchen@fudan.edu.cn
State Key Laboratory of ASIC and System, School of Microelectronics, Fudan University, Shanghai 200433, China
There have been many devices with different structures that successfully simulated synaptic dynamics, such as short-term plasticity (STP), long-term potentiation (LTP), and long-term depression (LTD) [16–18]. A MoS$_2$/PTCDA hybrid heterojunction synapse has been demonstrated with efficient photoelectric dual modulation [10]. A carbon nanotube synapse [19] and silicon-based MoS$_2$ synapse [20] showed dynamic logic. However, the mentioned studies focused only on the synaptic level. In some studies, different conductance states were realized to prove that their devices could be used to build artificial neural networks (ANNs), but they did not put the conductive states into the ANNs for calculation [21, 22]. In this work, 120 continuous conductance states are modulated, and the corresponding conductance values are used in the trained face recognition network for calculation; an excellent recognition rate of 100% is achieved.

**Methods**

The schematic structure of our synaptic device is shown in Fig. 1d, where it can be seen that a 70-nm ITO (indium tin oxide) film was deposited on the SiO$_2$/Si substrate as a back gate electrode. The substrate was a Si wafer with 200-nm SiO$_2$ on top. It was first cleaned with the acetone, isopropyl alcohol, and deionized water, and then dried with N$_2$ gas before the ITO deposition. The ITO layer was first deposited by sputtering and then annealed at 400 °C in the N$_2$ atmosphere for 10 min by rapid thermal processing (RTP). Transparent ITO electrodes are used in order to accurately fabricate source and drain electrodes using electron beam lithography. The Al$_2$O$_3$/ZrO$_2$/Al$_2$O$_3$ sandwiched structures with a thickness of 12 nm, 4 nm, and 4 nm were grown on the ITO by atomic layer deposition (ALD) as a barrier layer, an electron capture layer, and a tunneling layer respectively. Next, the mechanically exfoliated ReS$_2$ flakes with a thickness of about 3.6 nm were deposited as a channel under the patterned Ti/Au electrodes. The Ti/Au electrodes with 10-nm and 70-nm thickness were patterned using the electron beam lithography followed by the electron beam evaporation as a source and a drain, respectively. Figure 1c shows the atomic force microscope image of our 3.6-nm thickness ReS$_2$ film (about five layers); the channel length was designed to be 1.5 μm (see the inset in Fig. 1b). In this work, the ITO back gate acted as a presynapse neuron, and the Ti/Au electrodes acted as a postsynapse neuron. A small and constant voltage was applied between the source and drain electrodes, while the ITO back gate electrode was applied with pulses to modulate synaptic device performance.

**Results and Discussion**

Figure 2a shows the transfer characteristics of our synaptic device at a 2-V back gate voltage ($V_{bg} = 2$ V) under a fixed drain-to-source voltage ($V_{ds}$) changing from 100 to 700 mV with the step of 100 mV. An On/Off current ratio over $10^6$ could be observed. The curve displayed the drain-to-source current ($I_{ds}$), which first increased rapidly and then became saturated; the excellent saturation characteristics corresponded to the strong channel regulation by the ITO back gate electrode. Unlike the traditional transistors, which use silicon as a bottom gate electrode and SiO$_2$ as a dielectric at
the operation voltage of usually more than 20 V [23], the operation voltage of our synaptic device with only a 20-nm distance between the ReS₂ channel and ITO back gate electrode was below 5 V, significantly improving the efficiency of synaptic device. The inset in Fig. 2a shows the superlinear relationship under the low-\(V_{\text{ds}}\) regimes, which demonstrates a good Schottky contact between the ReS₂ channel and source and drain electrodes. As shown in Fig. 2b, \(I_{\text{ds}}-V_{\text{bg}}\) hysteresis curve could be observed when \(V_{\text{bg}}\) changed from −5 to 5 V and then reversed back at a constant bias of 0.1 V (\(V_{\text{ds}} = 0.1\) V). In the measurements, a small constant voltage of 0.1 V was applied between the source and drain electrodes to “read” the postsynaptic current. The memory window, which provided the basis for synaptic performance, was about 3.5 V; such a big memory window made our ReS₂ device very promising for synaptic applications [24]. Since the top of the valence band of ZrO₂ was higher than that of Al₂O₃, and the bottom of the conduction band was lower than that of Al₂O₃ (see the inset in Fig. 2c), ZrO₂ used as an intermediate layer sandwiched between alumina could capture charge effectively. The energy band diagrams under positive and negative back gate voltage are shown in Fig. 2c and d, respectively. When a positive voltage was applied, electrons in the ReS₂ channel would first tunnel through the Al₂O₃ tunneling layer, then be captured by the ZrO₂ trapping layer. On the contrary, when ITO was applied with a negative voltage, electrons gathered in the ZrO₂ layer would be sent to the ReS₂ channel; the energy bands bent in the direction of the channel.

In Fig. 3a, a typical excitatory postsynaptic current (EPSC) was detected after applying a negative input pulse (with the amplitude of −1 V and duration of 10 ms) at the ITO back gate. Also, an inhibitory postsynaptic responded to a positive voltage pulse (with the amplitude of 1 V and duration of 10 ms) was observed in Fig. 3b, which is similar to a biological synapse [25]. The pulse signal from the presynapse neuron was transmitted to the postsynapse neuron through the synapse and converted into the postsynaptic current (PSC) [26]. The PSC value was determined by pulse amplitude and duration. When the pulse was negative, the electrons from the defects of ZrO₂ gained enough energy to tunnel through the upper Al₂O₃ dielectric layer into the ReS₂ channel. The constant value of the current was slightly higher than the previous value (\(\Delta\text{PSC} = 0.04\) nA) and could maintain for a long time. This phenomenon corresponded to the long-term potentiation (LTP) in the biological synapse. However, when the pulse was positive, electrons in the ReS₂ channel tunnelled through the Al₂O₃ layer under the attraction of the electric field and were captured by the defects of ZrO₂. Thus,
the constant value of the current was slightly lower than the original value and could maintain the same for a long time (ΔPSC = 0.06 nA). This process corresponded to the long-term depression (LTD) in the biological synapse. The LTP and LTD provided a physiological substrate for learning and memory in synaptic devices. When the negative pulses with the amplitude of −2 V and duration of 10 ms were applied continuously, with a 1-s interval between pulses, the rising current in the two steps was observed, as shown in Fig. 3c. The rising current values were 1.6 nA and 1.4 nA, respectively. Therefore, a continuous and uniformly rising current could be obtained under the periodic gate voltage pulses, and the steady current after stimulation could last for a long time, as shown in Fig. 3d. This finding provided a basis for obtaining the multiple stable conductive states.

Figure 4a shows 120 current values after applying 120 negative pulses with an amplitude of −2 V and a duration of 10 ms and with a 1-s interval between pulses. Apparently, the current curve showed excellent linearity, 120 effective high-stability conductance states were obtained in each state. Different conductance states corresponded to different ANN weight values [27].

In this work, a three-layer artificial neural network for face recognition task is proposed, and its structure is presented in Fig. 4b, wherein it can be seen that the input layer consists of 1024 neurons that correspond to 1024 pixels of an image, the middle (hidden) layer consists of 256 neurons, and the output layer consists of 15 neurons that correspond to 15 classes of faces.

The development of the proposed ANN is as follows. A total of 165 pictures, including 15 types of pictures from Yale Face database [28] are used for ANN training and testing. Eight images of each type are used for ANN training, and the remaining three images of each kind are used for ANN testing. Given that the modules are smooth functions relative to their inputs and their internal weights, the multilayer architectures can be trained by simple stochastic gradient descent, and the gradients are generally computed by the backpropagation procedure [29]. Therefore, we use the classical backward propagation (BP) algorithm to build our network and show how the BP algorithm works for our ANN.

In this work, \( X_m \) represents an input neuron, so the input value of a hidden neuron can be expressed as:

\[
Y_{in} = \sum_{m=1}^{1024} X_m V_{mn}
\]

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 \( V \) having a total of \( 1024 \times 256 \) weight values; the initial value of this matrix is randomly assigned. The
The activation function of the hidden layer is the sigmoid function, so the output value of a hidden neuron is given by:

\[ Y_{on} = \frac{1}{1 + e^{Y_{in}}} \]

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

\[ Z_{ik} = \sum_{n=1}^{256} Y_{on} W_{nk} \]

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 \( 256 \times 15 \) weight values; the initial value of \( W_{nk} \) is also randomly assigned. Besides, we use the sigmoid function as an activation function of the output layer, so that the output value of an output neuron is given by:

\[ Z_{ok} = \frac{1}{1 + e^{Z_{ik}}} \]

Comparing the above-calculated output with the correct output, the total output error can be obtained, and it is expressed as:

\[ E = \frac{1}{2} \sum_{k=1}^{15} (O_k - Z_k)^2 \]

where \( O_k \) is the correct output value. So far, the forward propagation process of the network has been completely described. To improve the recognition rate, the backpropagation process is needed to calculate the errors of the weights, and they are used to update the network weights in the next iteration.

\[ \Delta V_{mn} = \mu \frac{\partial E}{\partial V_{mn}} \]

\[ \Delta W_{nk} = \mu \frac{\partial E}{\partial W_{nk}} \]

\[ V_{mn}' = V_{mn} + \Delta V_{mn} \]

\[ W_{nk}' = W_{nk} + \Delta W_{nk} \]

In the above mathematical expressions, \( \Delta V_{mn} \) and \( \Delta W_{nk} \) respectively represent the errors of \( V_{mn} \) and \( W_{nk} \); after adding the errors to the original weight, we get the updated weight \( V_{mn}' \) and \( W_{nk}' \); \( \mu \) is the learning rate.
and $\mu = 0.06$. After updating the weights, a new image is fed to the ANN, and the weight update process is repeated until all 120 images have been used for training. Next, we use the trained network to identify the remaining 45 images and calculate the recognition rate. The ANN testing process requires only the forward propagation process. Each image used for testing gets 15 output values after a forward propagation. The output value reflects the probability that the input image is of a certain type. The output with the maximum probability value is selected, and the corresponding type is the type of the input picture identified by the network. The recognition results are compared with the standard output; all correctly identified pictures are counted, and their total number is $n$. In each training-recognition cycle, the recognition rate $r$ is given by:

$$r = \frac{n}{45} \times 100\%$$

Generally, the recognition rate of the first recognition is very low, and in our ANN with 256 hidden neurons, the first recognition rate is only 17.78%. The above training-recognition process is repeated until the maximum recognition rate is obtained. The whole training-recognition cycle is shown in Fig. 4c.

As shown in Fig. 5a, during the ANN development process, the maximum recognition rate and rising speed of recognition rate (training speed) were different at a different number of hidden neurons. A larger number of hidden neurons led to a higher maximum recognition rate and a faster rising speed, but also increased energy consumption, so certain tradeoff should have to be made. In the case of 256 hidden neurons, the recognition rate reached 100% after 600 iterations of training, as shown in Fig. 5b. Since this was definitely the maximum recognition rate that could be achieved, in our ANN, we set the number of hidden neurons to 256. The distribution of weight values after different training-testing cycles is presented in Fig. 5c, and it indicates that the weights became more scattered after more cycles, that is to say, to reach a higher recognition rate, the weights in the ANN had to be adjusted. Once we achieved the maximum recognition rate, the matrices $V$ and $W$ having the optimal weight value were obtained. To demonstrate better that our ReS$_2$ device is suitable to be applied to ANNs, all weight values in the weight matrices $V$ and $W$ were replaced by device’s conductance values. We used $I_j(j = 1, 2, \cdots, 120)$ to represent 120 conductance values that were obtained after 120 cycles, and we made a linear transformation of the
original conductance values so that conductance range was consistent with the weight range, which was given by:

\[ C_j = AI_j + B \]

where \( C_j \) represented the weight value after the linear transformation. In the case of 600 cycles, the linear transformation coefficients were \( A = 1.3769 \times 10^{10} \) and \( B = -65.784 \). Next, we subtracted each \( C_j \) from each weight value and replaced the weight value with \( C_j \) that had the smallest absolute value after subtraction; namely, we calculated \( \min|V_{mn} - C_j|, \min|W_{nk} - C_j| \) and replaced each weight value with the corresponding \( C_j \). In this way, we obtained new \( V \) and \( W \) weight matrices wherein all the weight values were replaced by \( C_n \). Then, we used our new weight matrices in ANN testing, and the ANN recognition rate of 100% was achieved, which proved that our 120 conductance states could be perfectly used as weight values in the ANN. For the purpose of further analysis, we replaced the weight values after 100–500 training cycles (in steps of 100), and the identification results obtained after the replacement are completely consistent with the original one, as shown in Fig. 5d. This proves that these 120 current values could perfectly replace over \( 10^5 \) weight values for calculation. By further increasing the number of gate pulses, more conductance states could be obtained, which proved that our ReS\(_2\) device could be used in a large-scale neural network system.

Conclusions

In this work, we introduce a high-k dielectric stack based 2D ReS\(_2\) synaptic device and demonstrate some fundamental synaptic behaviors such as long-term potentiation and long-term depression. The results show that our ReS\(_2\) device can simulate synaptic performance well. Also, an ANN is constructed to prove the application of the proposed device in artificial neural networks. Applying 120 periodic gate voltage pulses, 120 effective, clearly distinguished conductance states are obtained, and they are used to replace more than \( 10^5 \) weights in the ANN for face recognition. The recognition rate of 100% is achieved after replacement. This excellent result demonstrates that our ReS\(_2\) synapse can be used to build an artificial neural network.

Abbreviations

2D: Two-dimension; ALD: Atomic layer deposition; ANN: Artificial neural network; LTD: Long-term depression; LTP: Long-term potentiation

Acknowledgements

We thank Ms. Jing Xu at the Fudan University for the guidance on device fabrication.

Authors’ Contributions

Z-HF and LC prepared the ReS\(_2\) synaptic devices and designed the electrical measurement method of synaptic plasticity. MZ and L-RG helped exfoliate ReS\(_2\) flakes. LC, HZ, Q-QS, and D-WZ supervised the whole work. All authors critically read and approved the final manuscript.

Funding

This work was supported by the NSFC (61704030 and 61522404), Shanghai Rising-Star Program (15QA140050), the Program of Shanghai Subject Chief Scientist (18XD1402800), and the Support Plans for the Youth Top-Notch Talents of China.

Availability of Data and Materials

The authors declare that the materials, data, and associated protocols are available to the readers, and all the data used for the analysis are included in this article.

Competing Interests

The authors declare that they have no competing interests.

Received: 27 June 2019 Accepted: 24 December 2019

Published online: 03 January 2020

References

1. Popper KR (1968) Birkhoff and von Neumann’s interpretation of quantum mechanics. Nature 219:682
2. Cheng S, Jheng S (2016) Physical realization of von Neumann lattices in rotating Bose gases with dipole interatomic interactions. Sci Rep-Uk 6:31801
3. MacPherson RD, Stolovitz DJ (2007) The von Neumann relation generalized to coarsening of three-dimensional microstructures. Nature 446:1053
4. Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE (2014) Intrinsic and task-evoked network architectures of the human brain. Neuron 83:238–251
5. Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. Nature 487:51
6. Bumstock G (2007) Physiology and pathophysiology of purinergic neurotransmission. Physiol Rev 87:659–797
7. Tian H, Guo Q, Xie Y, Zhao H, Li C, Cha J, Xia F, Wang H (2016) Anisotropic black phosphorous synaptic device for neuromorphic applications. Adv Mater 28:4991–4997
8. Liu C, Yan X, Song X, Ding S, Zhang DW, Zhou P (2018) A semi-floating gate memory based on van der Waals heterostructures for quasi-non-volatile applications. Nat Nanotechnol 13:404–410
9. Wang Y, Cong C, Yang W, Shang J, Peimyoo N, Chen Y, Kang J, Wang J, Huang W, Yu T (2015) Strain-induced direct–indirect bandgap transition and phonon modulation in monolayer WS\(_2\). Nano Res 8:2562–2572
10. Wang S, Chen C, Yu Z, He Y, Chen X, Wan Q, Shi Y, Zhang DW, Zhou H, Wang X (2019) Others: A MoS\(_2\)/PTCDIA hybrid heterojunction synapse with efficient photoelectric dual modulation and versatility. Adv Mater 31:1806237
11. Mak KF, Lee C, Hone J, Shan J, Heinz TF (2010) Atomically thin MoS\(_2\): a new direct-gap semiconductor. Phys Rev Lett 105:136805
12. Jadczyk J, Kutrowska-Girzycka J, Smole N, Ski T, Kossacki P, Huang YS, Bryja L (2019) Exciton binding energy and hydrogenic Rydberg series in layered ReS\(_2\). Sci Rep-UK 9:15178
13. Liu E, Fu Y, Wang Y, Feng Y, Liu H, Wan X, Zhou W, Wang B, Shao L, Ho C (2015) Others: Integrated digital inverters based on two-dimensional anisotropic ReS\(_2\) field-effect transistors. Nat Commun 6:6991
14. Tongay S, Sahin H, Ko C, Luce A, Fan W, Liu K, Zhou J, Huang Y, Ho C, Yan J (2014) Others: Monolayer behaviour in bulk ReS\(_2\) due to electronic and vibrational decoupling. Nat Commun 5:3252
15. De Sanctis A, Amit I, Hepplestone SP, Giacondi MF, Russo S (2018) Strain-engineered inverse charge-funnelling in layered semiconductors. Nat Commun 9:410
16. Prezioso M, Merrikh-Bayat F, Hoskins BD, Adam GC, Likharev KK, Strukov DB (2015) Training and operation of an integrated neuromorphic network based on metal-oxide memristors. Nature 521:61
17. Mak KF, Lee C, Hone J, Shan J, Heinz TF (2010) Atomically thin MoS\(_2\): a new direct-gap semiconductor. Phys Rev Lett 105:136805
18. De Sanctis A, Amit I, Hepplestone SP, Giacondi MF, Russo S (2018) Strain-engineered inverse charge-funnelling in layered semiconductors. Nat Commun 9:410
19. Kim K, Chen C, Truong Q, Shen AM, Chen Y (2013) A carbon nanotube synaptic device with dynamic logic and learning. Adv Mater 25:1693–1698
20. Jiang J, Guo J, Wan X, Yang Y, Xie H, Niu D, Yang J, He J, Gao Y, Wan Q (2017) 2D MoS2 neuromorphic devices for brain-like computational systems. Small 13: 1700933

21. Wang ZQ, Xu HY, Li XH, Yu H, Liu YC, Zhu XJ (2012) Synaptic learning and memory functions achieved using oxygen ion migration/diffusion in an amorphous InGaZnO memristor. Adv Funct Mater 22:2759–2765

22. Wang T, He Z, Liu H, Chen L, Zhu H, Sun Q, Ding S, Zhou P, Zhang DW (2018) Flexible electronic synapses for face recognition application with multimodulated conductance states. Acs Appl Mater Inter 10:37345–37352

23. Tian H, M W, Wang X, Zhao H, Xie Q, Li C, Li Y, Yang Y, Ren T (2015) Graphene dynamic synapse with modulatable plasticity. NANO LETT 15: 8013–8019

24. Arnold AJ, Razavieh A, Nasr JR, Schulman DS, Eichfeld CM, Das S (2017) Mimicking neurotransmitter release in chemical synapses via hysteresis engineering in MoS2 transistors. Acs Nano 11:3110–3118

25. Bean BP (2007) The action potential in mammalian central neurons. Nat Rev Neurosci 8:451

26. Shi Y, Kirwan P, Smith J, Robinson HP, Livesey FJ (2012) Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat Neurosci 15:477

27. De Domenico M, Nicosia V, Arenas A, Latora V (2015) Structural reducibility of multilayer networks. NAT COMMUN 6:6864

28. Belhumeur PN, Hespanha JAO, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis & Machine Intelligence:711–720

29. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436

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