ReSe$_2$-Based RRAM and Circuit-Level Model for Neuromorphic Computing

Yifu Huang$^1$, Yuqian Gu$^1$, Xiaohan Wu$^1$, Ruijing Ge$^1$, Yao-Feng Chang$^2$, Xiyu Wang$^3$, Jiahao Zhang$^1$, Deji Akinwande$^1$ and Jack C. Lee$^1$

$^1$Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, United States, $^2$Intel Corporation, Hillsboro, OR, United States, $^3$Oracle Corporation, Austin, TX, United States

Resistive random-access memory (RRAM) devices have drawn increasing interest for the simplicity of its structure, low power consumption and applicability to neuromorphic computing. By combining analog computing and data storage at the device level, neuromorphic computing system has the potential to meet the demand of computing power in applications such as artificial intelligence (AI), machine learning (ML) and Internet of Things (IoT). Monolayer rhenium diselenide (ReSe$_2$), as a two-dimensional (2D) material, has been reported to exhibit non-volatile resistive switching (NVRS) behavior in RRAM devices with sub-nanometer active layer thickness. In this paper, we demonstrate stable multiple-step RESET in ReSe$_2$ RRAM devices by applying different levels of DC electrical bias. Pulse measurement has been conducted to study the neuromorphic characteristics. Under different height of stimuli, the ReSe$_2$ RRAM devices have been found to switch to different resistance states, which shows the potentiation of synaptic applications. Long-term potentiation (LTP) and depression (LTD) have been demonstrated with the gradual resistance switching behaviors observed in long-term plasticity programming. A Verilog-A model is proposed based on the multiple-step resistive switching behavior. By implementing the LTP/LTD parameters, an artificial neural network (ANN) is constructed for the demonstration of handwriting classification using Modified National Institute of Standards and Technology (MNIST) dataset.

Keywords: RRAM, 2D material, ReSe$_2$, neuromorphic computing, verilog-a, artificial neural network

INTRODUCTION

With the rapid development of artificial intelligence (AI), machine learning (ML) and Internet of Things (IoT), novel computing technology for information processing is becoming crucial (Moh and Raju, 2018; Mohanta et al., 2020; Zhu et al., 2020). Conventional circuits based on von Neumann architecture are facing challenges including physical separation of computing units and memory, and low density of on-chip memories, which in turn lead to high energy consumption and low operation efficiency (Du Nguyen et al., 2017; Zidan et al., 2018; Amirsoleimani et al., 2020; Sebastian et al., 2020). Therefore, neuromorphic computing, which is inspired by the operation of human brains, has been proposed as a promising computing paradigm to overcome the bottleneck of von Neumann architecture (Kuzum et al., 2013). To physically realize the design of neuromorphic computing circuits, development of novel devices that can directly mimic brain-like long-term potentiation and depression (LTP/LTD) behaviors therefore arise significant interest.

Resistive random-access memory (RRAM), also known as memristors, is one of the most competitive candidates for the application of neuromorphic computing (Hong X. et al., 2018).
The operation of RRAM devices is based on the non-volatile resistive switching (N VRS) phenomenon of the active layer, typically a bulk metal-oxide film, such as SiO₂ or HfO₂, in a metal-insulator-metal (MIM) stacking structure (Chang et al., 2016; Lin et al., 2017). By applying external electrical bias, repeatable resistive switching behavior between high resistance state (HRS) and low resistance state (LRS) can be observed in the device and can be subsequently sustained without power supply (Wong et al., 2012). In recent years, two-dimensional (2D) materials have been investigated and applied in the development of the next generation nanoelectronics, optoelectronics, flexible electronics, and biosensors. Various 2D materials, including graphene oxide, solution-processed transitional metal dichalcogenides (TMDs), degraded black phosphorus and multilayer hexagonal boron nitride (h-BN), have been reported to exhibit NVRS phenomenon and applied in RRAM devices (Bai et al., 2015; Han et al., 2017; Ge et al., 2018; Wu et al., 2019). It has been reported in our previous work that monolayer chemical vapor deposition (CVD)-grown rhenium diselenide (ReSe₂) shows intrinsic NVRS phenomenon in a vertical MIM configuration (Wu et al., 2021). The structure holds promise for practical integration with smaller die space and 3D stacking capability.

Analog resistive switching has been realized in metal-oxide-based as well as 2D-material-based RRAM devices in the early work (Shen et al., 2020; Cao et al., 2021). Compared with the conventional binary switching with only two stable resistance states, RRAM devices with analog switching property can potentially mimic the function of biological synapse, making them feasible to be implemented in brain-inspired neuromorphic computing systems (Li et al., 2017; Zhang et al., 2019; Wang et al., 2021). When designing circuits with RRAM devices, Verilog-A provides a good platform for compact analog modeling as a de facto standard language widely used in semiconductor industry (Jiang et al., 2014; Mcandrew et al., 2015). By fitting the experimental data into functions of resistance and voltage, resistive switching behaviors can be described by a well-designed Verilog-A model.

In this work, RRAM devices in vertical MIM structure based on monolayer ReSe₂ were fabricated. Raman and photoluminescence (PL) spectra have been conducted to characterize the 2D material used in the devices. DC and pulse measurements are used to demonstrate the electrical properties of the ReSe₂ RRAM devices including resistive switching, retention, endurance, and LTP/LTD behaviors. A Verilog-A model has been established based on the experimental data. To study the interference to a neuromorphic computing system brought by the ReSe₂ RRAM device, an artificial neural network (ANN) is simulated for pattern recognition using Modified National Institute of Standards and Technology (MNIST) dataset.

**EXPERIMENTS**

**Device Fabrication**
The monolayer CVD-grown ReSe₂ films on mica substrate were produced by Sixcarbon Technology. All the fabricated devices in this work make use of the crossbar MIM structure. The 2D RRAM devices are fabricated with continuous CVD-grown monolayer ReSe₂ film sandwiched between Au top and bottom electrodes in the vertical MIM structure. The active device area is defined as the overlapped region between the crossbar electrodes. Figures 1A,B show the schematic and top-view optical image of the ReSe₂ RRAM devices, respectively. The active layer thickness is below 1 nm for the monolayer ReSe₂. Gold, which is an inert metal, is used as the electrode material in this work to avoid the possible formation of metal oxides at the metal-ReSe₂ interface. The bottom electrodes (BE) were patterned by e-beam lithography and deposited by e-beam evaporation (2 nm Cr/80 nm Au) on a 285 nm SiO₂/Si substrate. Polydimethylsiloxane (PDMS) pick-and-place water assisted transfer method was used to transfer the monolayer ReSe₂ film from mica to the target substrate with BE (Ma et al., 2017). In this method, the ReSe₂ film contacts with the PDMS stamp conformally. Then, the mica-ReSe₂-PDMS system was soaked into deionized water for 30 min. During this process, water can diffuse into the interface between ReSe₂ and mica due to the hydrophilic surface of mica. The ReSe₂-PDMS film was separated from the mica substrate and brought into contact with the target substrate. The PDMS stamp was peeled off subsequently and the ReSe₂ film was left on the target substrate with BE. After the completion of transfer, top electrodes (TE) were patterned and deposited with the same method as BE. The overlapped region between TE and BE is defined as the device area with a typical size of 1 × 1 and 2 × 2 μm² in this work.

**Material and Electrical Characterization**
Raman spectroscopy and photoluminescence (PL) were performed on a Renishaw in-Via system with 532 nm wavelength source to evaluate the ReSe₂ film before transferring. The DC and pulse characteristics of the devices were taken on a Cascade probe station under ambient conditions. An Agilent B1500A semiconductor parameter analyzer was used for DC and pulse measurements. During the DC measurements, electrical bias was applied on TE and BE was grounded. By involving one additional probe for pulses on TE, voltage pulses were injected to the device for pulse measurements.

**Verilog-A Model**
To describe the analog switching behavior of ReSe₂ RRAM devices, a compact Verilog-A model was constructed based on the data from DC and pulse measurements. By fitting the analog switching data into functions of resistance and voltage, the model was then used to simulate analog resistive switching and LTP/LTD behaviors when implemented in HSPICE.

**Artificial Neural Network Simulation**
A fully connected multi-layer perceptron ANN model was designed for supervised learning with the MNIST database. To demonstrate the potential application of the ReSe₂ RRAM devices, LTP/LTD characteristics were implemented in the stochastic gradient descent (SGD) optimizer of the ANN model.
RESULTS AND DISCUSSIONS

Material characterizations have been conducted on the samples to confirm the material used in the device. Raman spectra of the monolayer ReSe$_2$ films (Figure 2A) has been used to confirm the quality of material. The two characteristic peaks for ReSe$_2$, were observed at $\sim$190 cm$^{-1}$ and $\sim$140 cm$^{-1}$, corresponding to the $A_g$-like mode and $E_g$-like mode of 2H ReSe$_2$. The peak distance of $\sim$50 cm$^{-1}$ is in a good agreement with the reported value, suggesting a high-quality uniform monolayer crystalline film (Jiang et al., 2018; Apte et al., 2019). Figure 2B shows the PL of the monolayer ReSe$_2$. The PL peak position for the ReSe$_2$ is located at $\sim$1.5 eV, consistent with the reported optical band gap (Hong M. et al., 2018; Qiu et al., 2019).

DC electrical measurements were performed on as-fabricated crossbar MIM RRAM devices and revealed nonvolatile resistive...
switching in monolayer ReSe$_2$ active layers. Figure 3A shows a typical I-V curve of resistive switching. The device generally starts from a HRS. Without any forming process, the current shows a sudden increase as voltage sweeps to $\sim$3 V, indicating a transition from HRS to LRS (SET). For bipolar operation, the device will start a multi-step analog transition to HRS from $-0.5$ V (RESET) when a reverse voltage sweep is applied, which shows the potential for neuromorphic applications. A compliance current of 1 mA is applied at SET process to prevent irreversible breakdown, while no compliance current is applied at RESET process. The grey curves in Figure 1A illustrates a low cycle-to-cycle variation. After each SET and RESET measurement process, a read process is applied to acquire the resistance of the device. The LRS of the device is around 50 $\Omega$, while the HRS can reach up to over $1 \times 10^6$ $\Omega$, highlighting a promising ON/OFF ratio of over 4 orders of magnitude. Manual endurance data shown in Figure 3B demonstrates over 200 DC cycles, which is similar to the endurance data of metal oxide-based RRAM devices. Retention of nonvolatile states tested up to 6 h (Figure 3C) without obvious resistance change, which is sufficient for certain neuromorphic applications involving short term plasticity.

Based on our previous studies on 2D TMD-based RRAM devices, a “conductive-point” model was proposed to illustrate the resistive switching mechanism (Wu et al., 2020; Wu et al., 2021). When applying external voltage bias, metal ions/atoms from TE can be dissociated triggered by the electrical field and adsorbed into the chalcogen defects in the ReSe$_2$ layer. As electrons transport through the metallic conductive point at the vacancy site, the device would go into a lower resistance state. During the RESET process, the Joule heating effect induced by the high current in the device would move the metal ions/atoms out of the vacancies, resulting in a transition to a higher resistance state. Atomic resolution STM measurement with the evidence of Au moving in and out from the vacancy site after SET and RESET process supports the model. Figure 4 shows the schematics of the conductive-point model with metal adsorption. After the DC characterization, variable pulses are applied to the ReSe$_2$ devices to access its capability of emulating LTP and LTD. In this work, the selected pulse setup for SET is 100 ns in width and 3.2 V in height after a series of tuning. While for RESET, the pulse setup is 100 ns in width and $-0.8$ V in height. A DC read process was applied after each pulse to acquire the resistance state. By applying the optimum SET/RESET pulses, the conductance of the ReSe$_2$ device shows gradual increase/decrease. Figure 5 presents the LTP/LTD. The device exhibits a resistive transition from HRS to LRS after receiving 25 SET pulses (Figure 5A). While after 40 RESET pulses, the conductance gradually decreases from $\sim$2.4 mS to 1.8 mS (Figure 5B). Based on the conductive-point model, Au ions/atoms would be gradually driven into the active layer, leading to a multi-step conductance increase after a series of pulses. While in the pulse RESET process, the dissociation of Au ions/atoms from the active layer is relatively slower than DC RESET due to the low amount of power injected into the device by RESET pulses, resulting in a decrease of conductance after a larger number of pulses. In this way, we can mimic the synaptic responses of depression and potentiation by changing the polarization. As is shown in Table 1, among RRAM devices based on different types of low-dimensional materials including graphene oxide, hBN, metal dichalcogenides and quantum dots, ReSe$_2$ RRAM demonstrated in this work is highlighted by its small pulse width, and relatively high ON/OFF ratio, which shows the potential in the large scale applications of neuromorphic computing system.

With the acquisition of DC and pulse experimental data, a compact Verilog-A model is constructed. Based on the I-V characteristics from DC measurements, the LRS can be fitted with Ohmic conduction model, which illustrates the linear relation between voltage and current. For the analog RESET
process, resistance states are fitted by Schottky emission model. This is consistent with our previous work. The following simplifications and functions are used in the construction of the Verilog-A model:

- The resistance states are continuous as the maximum voltage ($V_{\text{max}}$) in the RESET process increases.
- The resistance of LRS is controlled by a metallic Ohmic conduction function as is shown in Function 1, where the parameter $A_1$ reflects the current and $B_1$ is affected by the intrinsic resistance of measurement tools.
  
  \[
  R = A_1 V + B_1. \tag{1}
  \]

- The resistance of HRS is derived from the Schottky emission model:

  \[
  J \propto A^* T^2 \exp \left[ -\frac{q (\phi_B - \sqrt{qE})}{4\pi \varepsilon_r \varepsilon_0 kT} \right]. \tag{2}
  \]
  
  \[
  R = \frac{V}{A_2 \cdot \exp \left( B_2 \cdot \sqrt{V} + C_2 \right)}. \tag{3}
  \]

- The parameter $A_2$, $B_2$ and $C_2$ in function (3) is defined by the maximum voltage ($V_{\text{max}}$) in the RESET process.
- The resistance of the RRAM device under LTP/LTD follows an exponential degradation function:

  \[
  R = R_{\text{init}} \alpha^n. \tag{4}
  \]

Here, $J$ is the current density, $A^*$ is the effective Richardson constant, $T$ is the absolute temperature, $q$ is the elementary charge, $\phi_B$ is the Schottky barrier height, $E$ is the electrical field across the dielectric, $k$ is the Boltzmann’s constant, $\varepsilon_r$ is the optical dielectric constant, $R_{\text{init}}$ is the initial resistance when LTP/LTD begins, $\alpha$ is the degradation rate during LTP/LTD, and $n$ is the number of pulses. The $\alpha$ is $\sim 0.85$ for LTP and $\sim 1.01$ for LTD, which shows good linearity. The fitting is performed by Python 3.7 with the SciPy library. The continuous curves in Figures 6B,C illustrate a good consistency between the experimental and simulation results.

By implementing the LRS and HRS function fitted by data extracted from 20 resistive switching cycles, an analog Verilog-A compact model for the ReSe$_2$ RRAM device is constructed. Figures 6A,B displays the HSPICE simulation result based on the Verilog-A model when applying DC sweep. The results obtained from the simulation indicate that the resistance switching occurs when the input voltage sweep across threshold voltage for SET and RESET, which are 3 V and $-0.5$ V, respectively. Further function implementation has been conducted based on the basic analog resistive switching model. The degradation rate fitted by function (4) enables the model to have LTP/LTD behavior when a series of pulse with the same height and width is input into the RRAM device in a HSPICE simulation. Figures 6C,D shows the simulated LTP/LTD results. The grey plots are the resistance output from the simulation. The blue and red curves are the calculated conductance states after each pulse, which shows similar behavior as the experimental results.
FIGURE 6 | Simulation results of ReSe₂ RRAM Verilog-A compact model using HSPICE. Simulated I-V curve of (A) RESET and (B) SET under DC sweep. No current compliance was applied to the simulation, causing an increasing current response after passing the threshold voltage of SET. Simulated (C) long-term potentiation, and (D) long term depression.

FIGURE 7 | Application of ReSe₂ RRAM for handwriting number recognition with MNIST dataset. (A) Schematic illustration of the ANN structure designed for the simulation. Comparison of (B) accuracy and (C) loss between models with standard and ReSe₂ RRAM-based SGD optimizers.
To demonstrate the potential of our synaptic ReSe₂ RRAM devices, a training simulation based on ANN is performed. A fully connected two-layer perceptron neural network for supervised learning with the MNIST dataset for handwriting number recognition. As is shown in Figure 7A, the two-layer perceptron neural network consists of 784 neurons on the input layer, 200 neurons on the hidden layer, and 10 output neurons. The 784 input neurons correspond to 28 x 28 MNIST image data, and the 10 output neurons correspond to 10 types of numbers (0−9).

In the simulation learning process, a customized SGD optimizer is designed to demonstrate the interference brought by the implementation of ReSe₂ RRAM devices. After each epoch of training, the optimizer will calculate the gradient of the current weights. After that, the weights will be updated following Eq. 5.

\[ W_{\text{new}} = W - lr \times g. \] (5)

Here \( W_{\text{new}} \) is the updated weight, \( W \) is the weight after one epoch of training, \( lr \) is learning rate, and \( g \) is the gradient. When the gradient is positive, the weight will decrease, which goes into the process of LTD. While the gradient is negative, the weight will go through the process of LTP. Taking the effect of LTP/LTD into consideration, Eq. 5 will be modified as:

\[ W_{\text{new}} = W - lr \times g \times |k_{\text{LTD}}|, \quad g > 0. \] (6)
\[ W_{\text{new}} = W - lr \times g \times |k_{\text{LTP}}|, \quad g < 0. \] (7)

Here we consider the LTP/LTD processes are linear. The \( k_{\text{LTD}} \) and \( k_{\text{LTP}} \) are the slope of the conductance curve under a linear fitting, which follows Eq. 8. Here, \( G_{\text{fin}} \) is the normalized final conductance, \( G_{\text{init}} \) is the normalized initial conductance, and \( P \) is the number of pulses.

\[ k = \frac{G_{\text{fin}} - G_{\text{init}}}{P}. \] (8)

The ANN with customized SDG optimizer is trained for 50 epochs. An ideal ANN with no customization is trained under the same conditions as a benchmark. After 50 epochs of training, the ANN based on ReSe₂ RRAM devices achieved high recognition accuracy of 91.3%, which is comparable to the 95.71% accuracy of the ideal ANN model (Figure 7B). The loss of each model shows that neither of them exhibits over-fitting issue, which is desirable in the training process (Figure 7C). Note that the recognition accuracy of the ideal model used in this work is not as high as that in previously reported works. The main reason is that the model is not optimized for the specific application of handwriting recognition with MNIST dataset. By comparing the model with default parameters with the model implementing customized SDG optimizer, we can quickly assess the performance of the ANN using ReSe₂ RRAM devices.

**CONCLUSION**

In summary, vertical crossbar MIM structure RRAM devices based on 2D ReSe₂ have been found to exhibit excellent characteristics, e.g., stable multiple-step RESET when different levels of DC bias are applied. Furthermore, with the analog switching characteristic of the ReSe₂, essential LTP and LTD behaviors are successfully mimicked. By implementing the physical properties, a Verilog-A compact model is designed to provide a standard method for future design of neuromorphic computing systems. In addition, the demonstration of a high handwriting recognition accuracy with ANN simulation shows the potential of the ReSe₂ RRAM devices. The findings from our work provide additional insights into the application of 2D ReSe₂ RRAM devices in the construction of brain-inspired neuromorphic computing systems.

**DATA AVAILABILITY STATEMENT**

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

**AUTHOR CONTRIBUTIONS**

YH and YG contributed equally to this work. The electrical characterization, model design and data process are done by YH. The device fabrication and material characterization are done by YG. XhW and RG provided instructions on measurement and fabrication. Y-FC gave instructions on the development of Verilog-A and ANN model. XyW wrote the SGD optimizer codes for the ANN model. JZ wrote part of the Verilog-A code.

**FUNDING**

This work was supported in part by the National Science Foundation (NSF) Grant #1809017. The authors acknowledge use of Texas Nanofabrication Facilities supported by the NSF NNCl award #1542159.

**ACKNOWLEDGMENTS**

The authors appreciate Jo Wozniak and Anne Bowen of Texas Advanced Computing Centre (TACC) for 3D renderings.

**REFERENCES**

Amirsoleimani, A., Alibart, F., Yon, V., Xu, J., Pazhouhandeh, M. R., Ecolfey, S., et al. (2020). In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal-Oxide-Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives. Adv. Intell. Syst. 2, 2000115. doi:10.1002/aisy.202000115

Apte, A., Krishnamoorthy, A., Hachtel, J. A., Susarla, S., Yoon, J., Sassi, L. M., et al. (2019). Two-Dimensional Lateral Epitaxy of 2H (MoSe2)-1T (ReSe2) Phases. Nano Lett. 19, 6338–6345. doi:10.1021/ acs.nanolett.9b02476
