Reconfigurable and Ultrafast Non-volatile Floating-gate Memory based on van der Waals Heterostructures

Minh Chien Nguyen[1], Ngoc Thanh Duong[2], Hong Woon Yun[1], Van Dam Do[1], Van Tu Vu[1], Whan Kyun Kim[3], Vu Khac Dat[2], Huamin Li[4], and Woo Jong Yu[1,*]

[1]*Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea*
[2]*Department of Energy Science, Sungkyunkwan University, Suwon 16419, Republic of Korea*
[3]*Semiconductor R&D Center, Samsung Electronics, Hwaseong 18448, Republic of Korea*
[4]*Department of Electrical Engineering, University at Buffalo, The State University of New York, New York 14260, United States*
*E-mail: micco21@skku.edu

***Abstract*— We present the first experimental demonstration and performance analysis of reconfigurable, ultra-fast non-volatile floating gate memory (R-FGM) based on WSe_2/h-BN/graphene van der Waals (vdW) heterostructures. The device features an asymmetrically engineered WSe_2 channel with top (TE) and bottom (BE) contacts, where the shielding effect of BE enables n-type and p-type memory reconfiguration through electric field control via Si/SiO_2 gate. Benefiting from ultra-clean vdW interfaces, the R-FGM achieves a program/erase speed of 30 ns, excellent data retention up to 10^4 s, and endurance over 10^5 cycles while maintaining an on/off ratio exceeding 10^6 for both memory types. Furthermore, the device exhibits multi-conductance states with 7-bit operation, supporting long-term potentiation (LTP) and long-term depression (LTD) to emulate synaptic behavior. These results establish R-FGM as a promising candidate for next-generation non-volatile memory and neuromorphic computing applications.**

Keywords—floating-gate memory, 2D materials, van der Waals heterostructures, reconfigurable, ultrafast, neuromorphic computing

I. Introduction

Floating gate memory (FGM) is the dominant technology in the non-volatile memory market due to its high storage density, excellent stability, and low fabrication cost [1]. However, with the increasing demand for processing large volumes of data, the limitations of the traditional von Neumann computing architecture, where processing and memory units are physically separated have become a significant bottleneck for the semiconductor industry [2]. The low program speed of conventional silicon-based floating gate memory, typically in the microsecond range, restricts its suitability for next-generation neuromorphic computing, which requires high-speed data processing with low power consumption [2, 3].

Two-dimensional (2D) materials, with their atomic thickness, high carrier mobility, and excellent electrostatic control, hold great potential for ultra-scaled, ultra-fast devices [4]. These materials can significantly reduce the distance between memory and logic units to a few nanometers, thereby enabling faster data transfer [5]. Recent progress in 2D materials has enabled the realization of next-generation FGM using both two-terminal [6, 7] and three-terminal architectures [8, 9]. Notably, ultra-fast FGM devices based on 2D heterostructures such as MoS_2/h-BN/graphene [10], InSe/h-BN/graphene [11], and graphene-based hot-carrier injection systems [12] have demonstrated programming speeds in the nanosecond to sub-nanosecond range. However, these devices suffer from limitations, including fixed carrier type, which hinder their reliability, and versatility [13, 14]. In this regard, we present a novel R-FGM device based on WSe_2/h-BN/graphene vdW heterostructures. This device overcomes the limitations of conventional FG devices by enabling both n-type and p-type carrier reconfiguration while achieving an ultra-fast program/erase speed of 30 ns, excellent data retention up to 10^4 s, endurance of over 10^5 cycles, and multi-bit operation with 7-bit operation. These features offer enhanced performance, versatility, and significant potential for next-generation neuromorphic system applications.

II. Device Fabrication

The experimental configuration of the WSe_2/h-BN/graphene R-FGM is shown in Fig. 1, along with a summary of the fabrication process. A highly doped silicon wafer with a 300 nm SiO_2 layer is used as the substrate, which acts as the control gate and blocking oxide. Monolayer graphene and multilayer h-BN are exfoliated from bulk crystals and transferred onto the Si/SiO_2 substrate using a dry transfer method. The h-BN layer serves as the tunneling oxide, while graphene functions as the floating gate. The TE and BE electrodes are patterned using electron-beam lithography (EBL) and metalized by electron-beam evaporation of Ti/Au (5/50 nm). The TE acts as the source, and the BE serves as the drain, enabling asymmetric control of the WSe_2 ambipolar semiconducting channel.

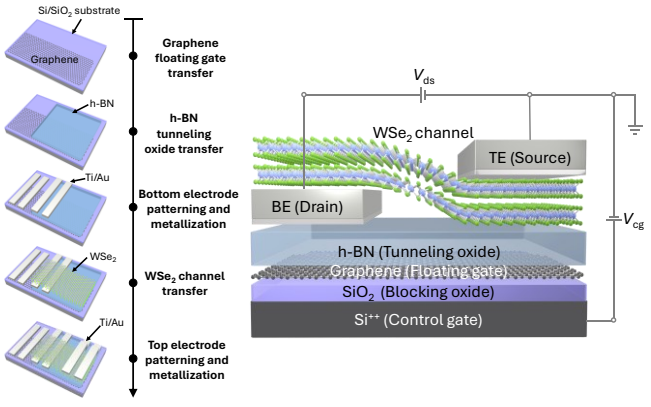

Fig. 1. Summary of the device fabrication process and 3D schematic of the R-FGM architecture.

The fabricated R-FGM device is shown in the optical microscope (OM) image in Fig. 2a, with graphene, h-BN, and WSe_2 flakes outlined by grey, blue, and green dashed lines, respectively. Raman spectra from the heterostructure

region, provided in Fig. 2b, confirm the characteristic features of graphene, h-BN, and WSe_2, validating the formation of the heterostructure. To further verify the atomic structure, scanning transmission electron microscopy (STEM) analysis was performed, as shown in Fig. 2c, revealing layer thicknesses of 8 layers WSe_2, 15 nm for h-BN, and a monolayer of graphene, which illustrates an ultra-clean interface. Detailed elemental mapping using energy dispersive X-ray spectroscopy (EDS) is also provided in Fig. 2d, confirming the spatial distribution of Se, W, B, N, and C within the heterostructure.

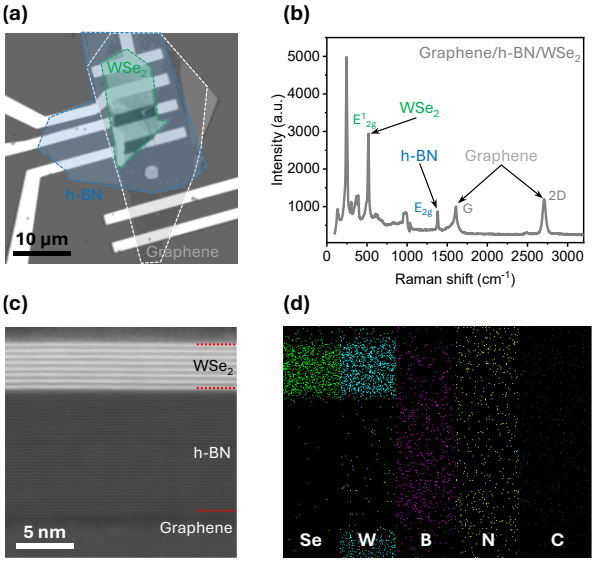

Fig. 2. (a) OM image of the fabricated R-FGM device. (b) Raman spectra confirming the heterostructure characteristics. (c) Cross-sectional STEM image showing the ultra-clean interface between graphene, h-BN, and WSe_2 layers. (d) Elemental mapping by STEM-EDS illustrating the spatial distribution of Se, W, B, N, and C in the heterostructure.

III. Results

A. *Ambipolar to reconfigurable unipolar transition of memory behavior in R-FGM*

To investigate the transition from ambipolar to reconfigurable unipolar memory behavior, we measured the transfer characteristics of two graphene/h-BN/WSe_2 floating gate devices: one with symmetric top contacts (Fig. 3a) and another with asymmetric TE and BE contacts (Fig. 3b). As shown in Fig. 3c, the ambipolar memory behavior of the FGM arises from the intrinsic energy band structure of ambipolar WSe_2, which facilitates balanced carrier injection at both the source and drain Schottky top contacts. The carrier type is modulated similarly to an ambipolar field effect transistor [15]. When a positive gate voltage is higher than flat-band voltage ($V_g > V_{FB}$) is applied, the energy band shifts downward, enabling electron tunneling from the source to the drain and activating n-type conduction. Conversely, when a negative gate voltage ($V_g < V_{FB}$) is applied, the energy band shifts upward, allowing hole tunneling from the drain to the source at both negative and positive V_{ds}. The hysteresis observed in the transfer curve is attributed to charge trapping in the graphene FG. Fig. 3d illustrates the reconfigurable unipolar memory behavior of the R-FGM with asymmetric contacts. In this configuration, the shielding effect of the BE blocks electric field modulation by the control gate voltage (V_{cg}), restricting energy band modulation and trapping effects to the WSe_2 channel above the BE region. When $V_{ds} > 0$ V, the transfer curve exhibits a clear n-type characteristic with pronounced hysteresis, a high on-state current of 10^{-5} A/μm, and an ultra-high on/off ratio exceeding 10^7. Conversely, when $V_{ds} < 0$ V, the transfer curve shows a p-type characteristic with significant hysteresis, a high on-state current of 10^{-6} A/μm, and an ultra-high on/off ratio greater than 10^6.

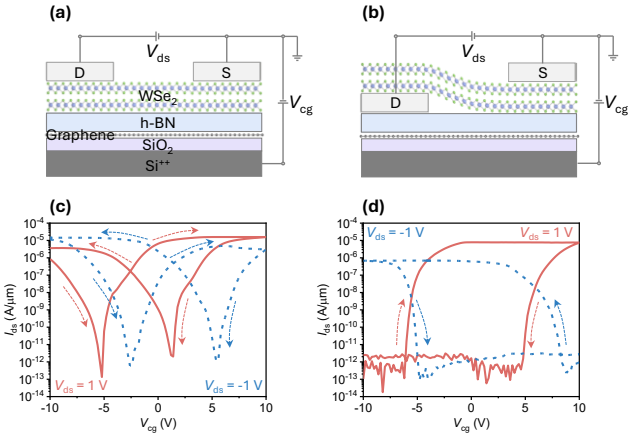

Fig. 3. (a) Schematic of the graphene/h-BN/WSe_2 FGM device with symmetric contacts. (b) Schematic of the device with asymmetric contacts. (c) Dual-sweep transfer characteristics at V_{ds} = 1 V and -1 V for the device with symmetric contacts. (d) Dual-sweep transfer characteristics at V_{ds} = 1 V and -1 V for the device with asymmetric contacts.

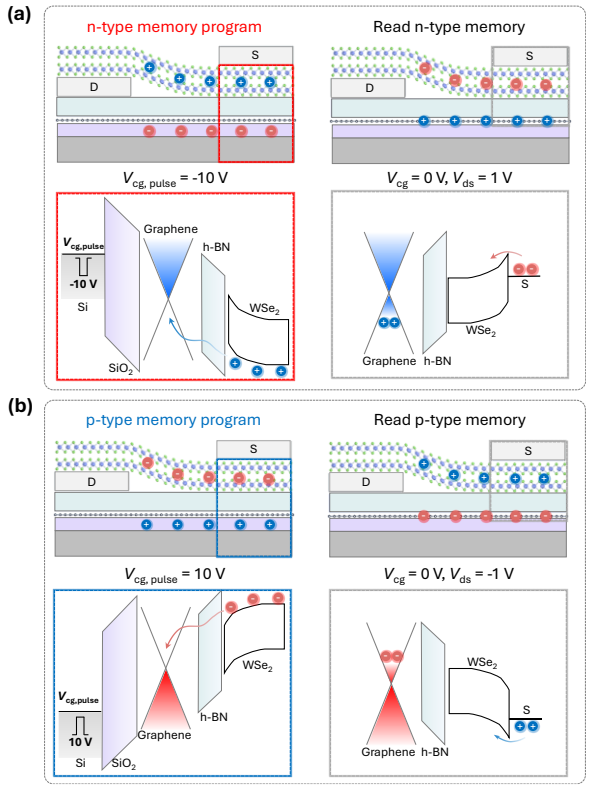

Fig. 4. Cross-sectional schematics and corresponding energy-band diagrams of the R-FGM device during program and read operations: (a) n-type memory, and (b) p-type memory.

The reconfigurable and non-volatile unipolar memory characteristics of the R-FGM device are explained by the energy band diagrams in Fig. 4. In n-type memory (Fig. 4a), applying a negative V_{cg} pulse attracts holes from the WSe_2

channel, which tunnel through the h-BN barrier and get trapped in the graphene, turning the channel into an electron-rich region. This state remains even after removing V_{cg}, enabling n-type conduction when a positive V_{ds} is applied. In p-type memory (Fig. 4b), a positive V_{cg} pulse attracts electrons from the WSe_2 channel, trapping them in the graphene and creating a hole-rich channel. This p-type memory state remains stable and can be read by applying a negative V_{ds}. The trapped charges in graphene FG ensure non-volatility, while the ability to switch between n-type and p-type memory makes the device highly reconfigurable.

B. High retention and enduance characteristics in R-FGM

We further investigated the memory performance of the R-FGM device. Fig. 5a and 5b present the transfer characteristics of n-type memory at V_{ds} = 1 V and p-type memory at V_{ds} = -1 V, confirming that the threshold voltage (V_{th}) can be modulated by the $V_{cg,\ pulse}$, a typical flash memory feature. As the V_{cg} range increases from 5 V to 40 V, the memory window expands to 63 V for n-type memory and 72 V for p-type memory, while maintaining high on/off current ratios of 10^7 and 10^6, respectively. The retention characteristics (Fig. 5c and 5d) were evaluated by applying a $V_{cg,\ pulse}$ = 10 V in 1 µs for the program state and a $V_{cg,\ pulse}$ = -10 V in 1 µs for the erase state. For n-type memory, retention was tested by reading at V_{ds} = 1 V, while for p-type memory, the program state used a -10 V pulse for 1 µs, followed by reading at V_{ds} = -1 V. Both memory types demonstrated excellent stability, maintaining an on/off ratio greater than 10^6 over an extended period (> 10^4 s). In the endurance test (Fig. 5e and 5f), the device exhibited outstanding durability, retaining stable memory performance over 10^5 program/erase cycles.

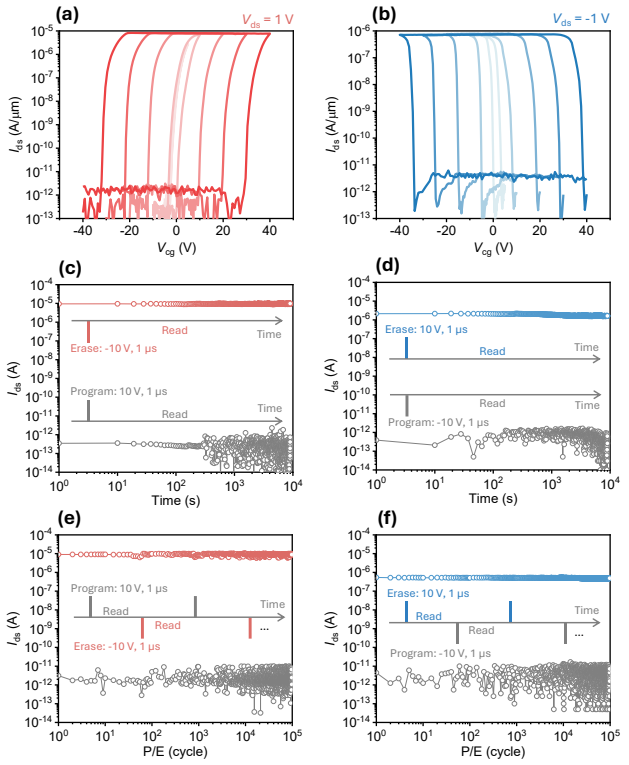

Fig. 5. Memory performances of R-FGM. (a) n-type memory with a memory window of 63 V at V_{ds} = 1 V and (b) p-type memory with a memory window of 72 V at V_{ds} = -1 V when Vbg sweeps from -40 V to 40 V. (c) Retention characteristics for n-type memory and (d) p-type memory, demonstrating stable data storage over 10^4 s. (e) Endurance performance of n-type memory and (f) p-type memory, showing robust cycling capability over 10^5 cycles. Insets show the pulse schemes used for measurements.

C. Ultra-fast swiching of the R-FGM

We have demonstrated that our R-FGM features ultrafast switching characteristics which cannot achieved in conventional flash memory. Fig. 6a shows the experimental set up for ultrafast transient *I-V* measurement A 120 MHz waveform generator in burst mode programs nanosecond pulse cycles, with I_{ds} pre-amplified by FEMTO DHPCA 100 at 10^7 V/A and captured by a mixed-domain oscilloscope in trigger mode, ensuring impedance matching with 50 Ω coaxial cables and terminations to minimize reflections above 1 MHz while still maintain the distinct data storage levels between program and erase states. Fig. 6b shows an example of the ultrafast switching operation of the R-FGM, demonstrating sequential programming and erasing up to 1 MHz, which is limited only by the instrumentation response. Fig. 6c and 6d illustrate the transient response speed of the device, with response times measured at 30 ns for erasing and 32 ns for programming operations, corresponding to a full width at half maximum (FWHM) of 22 ns. This speed is approximately 5000 times faster than that of commercial flash memory [2], attributed to the ultra-clean and flat interface in the WSe_2/h-BN/graphene heterostructure, as confirmed by the STEM data shown in Fig. 2c.

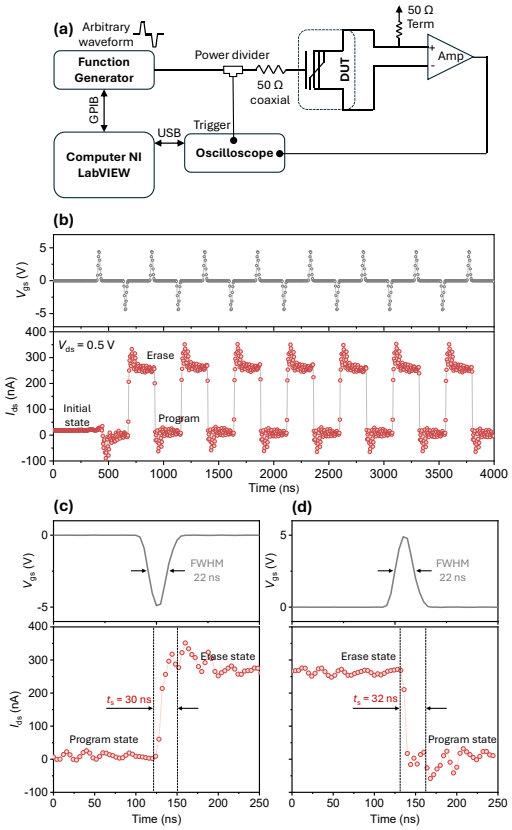

Fig. 6. (a) Experimetal set-up for ultrafast transient *I-V* measurement. (b) Demonstration of ultrafast switching by sequential program and erase pulse with 200 ns intervals. (c, d) Ultrafast response speed of the device after a FWHM of 22 ns pulses for erase and program.

D. Artificial synapse characteristics of R-NVM

Fig. 7a presents a schematic of the voltage spike-based artificial synapse device utilizing the R-FGM, inspired by the human biological synaptic system. In this device, V_{ds} from

the drain acts as the pre-synaptic terminal, while I_{ds}, serving as the post-synaptic terminal, simulates signal transmission from the pre-synaptic neuron to the post-synaptic neuron. When voltage spikes are applied to the V_{cg}, charges are stored in the graphene FG through the h-BN tunneling layer. These stored charges modulate the channel conductivity, transmitting the stimulus through the tunneling barrier, effectively adjusting the synaptic weight and enabling either LTP or LTD. Fig. 7b and 7c illustrate the continuous and repetitive modulation of channel conductance in the R-FGM-based analog synapse. By applying a series of identical pulses, the device demonstrates long-term synaptic plasticity, achieving 128 cycles of LTP and 128 cycles of LTD with finely tunable synaptic weights in both n-type and p-type memory modes. This behavior highlights ability of the R-FGM to mimic biological synaptic learning processes with high precision and stability.

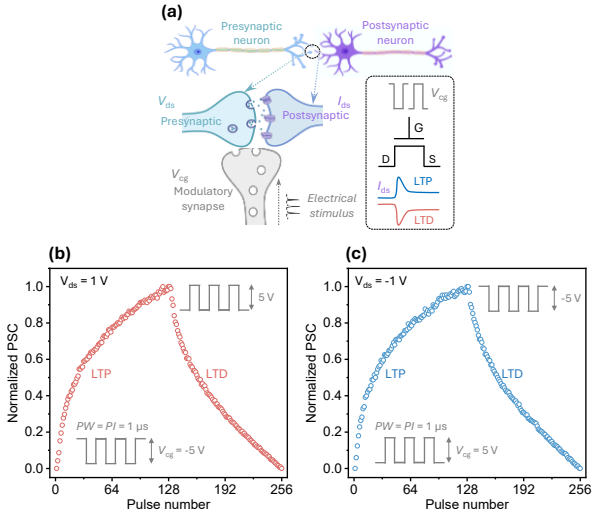

Fig. 7. (a) Schematic diagram of the R-FGM artificial synapse, where the drain, source, and control gate function as the pre-synaptic, post-synaptic, and modulatory synaptic terminals, respectively. Measured synaptic channel conductance at (b) V_{ds} = 1 V and (c) V_{ds} = -1 V as a function of identical pulse numbers, demonstrating LTP and LTD characteristics.

IV. Conclusion

We demonstrated the first high-performance R-FGM with ultra-fast switching and dual-mode memory behavior. The device can switch between n-type and p-type memory states, achieving response times of 30 ns for erase and 32 ns for program operations, which is 5000 times faster than conventional flash memory. It shows a large memory window of 63 V for n-type and 72 V for p-type, with an on/off ratio above 10^7 and stable retention for over 10^4 seconds. The device maintains excellent endurance for 10^5 cycles and benefits from the clean WSe_2/h-BN/graphene interface for stability and speed. Moreover, the R-FGM functions as an artificial synapse, mimicking biological synaptic plasticity with 128 cycles of LTP and LTD, making it a promising candidate for future high-speed memory and neuromorphic computing systems.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2021R1A2C2004027, NRF-2023K2A9A2A06059788, NRF-2020R1A5A1019649, RS-2024-00439520, RS-2024-00439092), ICT Creative Consilience program (IITP-2020-0-01821) the Competency Development Program for Industry Specialists of the Korean Ministry of Trade, Industry and Energy (MOTIE), operated by Korea Institute for Advancement of Technology (KIAT) (No. P0023704, Semiconductor Track Graduate School (SKKU)). Samsung Research Funding & Incubation Center of Samsung Electronics under Project No. SRF-MA170101.

References

[1] International Roadmap for Devices and Systems (IRDS) https://irds.ieee.org/ (2023).

[2] T. Kouchi, "13.5A 128Gb 1b/Cell 96-word-line-layer 3D flash memory to improve random read latency with t_{PROG} = 75 µs and t_R = 4 µs." 2020 IEEE International Solid-State Circuits Conference (ISSCC), 226–228 (IEEE, 2020).

[3] Hoi-Jun, Y. "Intelligence on Silicon: From Deep-Neural-Network Accelerators to Brain Mimicking AI-SoCs." 2019 IEEE International Solid - State Circuits Conference - (ISSCC) 20–26 (IEEE, 2019).

[4] Liu, Y., Weiss, N. O., Duan, X., Cheng, H. C., Huang, Y., and Duan, X., "Van der Waals heterostructures and devices." Nature Reviews Materials, vol. 1, no. 9, pp.1-17, 2016.

[5] Liu, C., Chen, H., Hou, X., Zhang, H., Han, J., Jiang, Y.G., Zeng, X., Zhang, D.W. and Zhou, P., "Small footprint transistor architecture for photoswitching logic and in situ memory." Nature Nanotechnology, vol. 14, no. 7, pp.662-667, 2016.

[6] Vu, Q. A., Shin, Y. S., Kim, Y. R., Nguyen, V. L., Kang, W. T., Kim, H., Luong, D. H., Lee, I. M., Lee, K., Ko, D. S., Heo, J., Park, S., Lee, Y. H., and Yu, W. J., "Two-terminal floating-gate memory with van der Waals heterostructures for ultrahigh on/off ratio." Nature Communications, vol. 7, no. 1, p.12725, 2016.

[7] Won, U. Y., An Vu, Q., Park, S. B., Park, M. H., Dam Do, V., Park, H. J., Yang, H., Lee, Y. H., and Yu, W. J., "Multi-neuron connection using multi-terminal floating–gate memristor for unsupervised learning.". Nature Communications, vol. 14, no. 1, p.3070, 2016.

[8] Choi, M. S., Lee, G. H., Yu, Y. J., Lee, D. Y., Lee, S. H., Kim, P., Hone, J., and Yoo, W. J., "Controlled charge trapping by molybdenum disulphide and graphene in ultrathin heterostructured memory devices." Nature Communications, vol. 4, no. 1, p.1624, 2013.

[9] Bertolazzi, S., Krasnozhon, D., and Kis, A., "Nonvolatile memory cells based on MoS2/graphene heterostructures." ACS Nano, vol. 7, no. 4, pp.3246-3252, 2013.

[10] Liu, L., Liu, C., Jiang, L., Li, J., Ding, Y., Wang, S., Jiang, Y.G., Sun, Y.B., Wang, J., Chen, S. and Zhang, D.W., "Ultrafast non-volatile flash memory based on van der Waals heterostructures." Nature Nanotechnology, vol. 16, no. 8, pp.874-881, 2021.

[11] Wu, L., Wang, A., Shi, J., Yan, J., Zhou, Z., Bian, C., Ma, J., Ma, R., Liu, H., Chen, J. and Huang, Y., "Atomically sharp interface enabled ultrahigh-speed non-volatile memory devices." Nature Nanotechnology, vol. 16, no. 8, pp.882-887, 2021.

[12] Xiang, Y., Wang, C., Liu, C., Wang, T., Jiang, Y., Wang, Y., Wang, S. and Zhou, P., "Subnanosecond flash memory enabled by 2D-enhanced hot-carrier injection." Nature, vol. 641, pp. 90-97, 2025.

[13] Sun, X., Zhu, C., Yi, J., Xiang, L., Ma, C., Liu, H., Zheng, B., Liu, Y., You, W., Zhang, W. and Liang, D., "Reconfigurable logic-in-memory architectures based on a two-dimensional van der Waals heterostructure device." Nature Electronics, vol. 5, no. 11, pp.752-760, 2022.

[14] Migliato Marega, G., Zhao, Y., Avsar, A., Wang, Z., Tripathi, M., Radenovic, A., and Kis, A., "Logic-in-memory based on an atomically thin semiconductor." Nature, vol. 578, no. 7832, pp.72-77, 2020.

[15] Zhou, Y., Tong, L., Chen, Z., Tao, L., Pang, Y., and Xu, J.B., "Contact-engineered reconfigurable two-dimensional Schottky junction field-effect transistor with low leakage currents." Nature Communications, vol. 14, no. 1, p.4270, 2023.