ARTICLES

Journal of Semiconductors
(2021) 42, 024101
doi: 10.1088/1674-4926/42/2/024101

Enhancement of refresh time in quasi-nonvolatile memory by the density of states engineering

Zhaowu Tang[1, ‡], Chunsen Liu[1, 2, ‡, †], Senfeng Zeng[1], Xiaohe Huang[1], Liwei Liu[1], Jiayi Li[1], Yugang Jiang[2], David Wei Zhang[1], and Peng Zhou[1, †]

[1]State Key Laboratory of ASIC and System, School of Microelectronics, Fudan University, Shanghai 200433, China

[2]School of Computer Science, Fudan University, Shanghai 200433, China

Abstract: The recently reported quasi-nonvolatile memory based on semi-floating gate architecture has attracted extensive attention thanks to its potential to bridge the large gap between volatile and nonvolatile memory. However, the further extension of the refresh time in quasi-nonvolatile memory is limited by the charge leakage through the p–n junction. Here, based on the density of states engineered van der Waals heterostructures, the leakage of electrons from the floating gate to the channel is greatly suppressed. As a result, the refresh time is effectively extended to more than 100 s, which is the longest among all previously reported quasi-nonvolatile memories. This work provides a new idea to enhance the refresh time of quasi-nonvolatile memory by the density of states engineering and demonstrates great application potential for high-speed and low-power memory technology.

Key words: quasi-nonvolatile memory; refresh time; density of states engineering

Citation: Z W Tang, C S Liu, S F Zeng, X H Huang, L W Liu, J Y Li, Y G Jiang, D W Zhang, and P Zhou, Enhancement of refresh time in quasi-nonvolatile memory by the density of states engineering[J]. *J. Semicond.*, 2021, 42(2), 024101. http://doi.org/10.1088/1674-4926/42/2/024101

1. Introduction

The existing commercial memory can be divided into two categories: volatile memory and nonvolatile memory. Volatile memory exhibits ultrafast accessing speed, but the data stored will disappear once the power is lost. Especially for dynamic random access memory (DRAM), frequent refresh is needed to retain the data. Nonvolatile memory can store data for up to 10 years without the power supply, but its accessing speed is much lower than volatile memory. To bridge the large timescale gap between volatile and non-volatile memory, the third type of memory called quasi-nonvolatile (QNV) memory has been proposed recently[1]. Benefit from its semi-floating gate (SFG) architecture and band-engineered van der Waals heterostructures, an ultrafast programming speed has been achieved and meanwhile its refresh time has been significantly enhanced to 10 s. Although the refresh time is much higher than DRAM (64 ms), it is not long enough to sufficiently reduce the power consumption due to frequent refresh. Therefore, it is necessary to propose new methods to further extend the refresh time of QNV memory to further reduce power consumption.

In consideration of that the limited refresh time of QNV memory is mainly caused by the leakage of high-energy electrons crossing the barrier between the channel and the floating gate through the p–n junction, a new approach to cut down the high-energy electrons needs to be adopted to extend the refresh time. Two dimensional (2D) Dirac material such as the monolayer graphene (Gr) has a linear band dispersion[2], and when suitably doped its electron density will decrease super-exponentially as energy increases toward the Dirac point thanks to the linear dependence of density of states (DOS) on energy, resulting in a more concentrated electron distribution around the Fermi level E_F[3]. It has been demonstrated in many works that inserting a properly doped Gr Dirac material between the source electrode and the channel can effectively cut off high-energy electrons injected into the channel and thus achieve more efficient switching[3–6]. By applying the Dirac material to QNV memory, the long thermal tail of high-energy electrons can be greatly cut down and thus suppressing the leakage of electrons from the floating gate to the channel.

In this work, we propose and fabricate a QNV SFG memory based on the DOS engineered van der Waals heterostructures. By utilizing the SFG architecture, the device exhibits the symmetric ultrafast programming and erasing speeds of 10 μs, which is 10^3 times faster than that of the previously reported 2D material based conventional flash memory devices[7–17]. Additionally, its refresh time is effectively extended to more than 100 s by DOS engineering, which is the longest among all previously reported QNV memory devices[1, 18–21]. Benefit from the significant enhancement of refresh time in QNV memory, considerable power consumption resulting from frequent refresh can be greatly reduced.

2. Experimental section

Firstly, the 30 nm Al_2O_3 layer was grown on highly p-doped Si substrates by atomic layer deposition as the back-gate dielectric. Then all of the 2D material including multilay-

Zhaowu Tang and Chunsen Liu contribute equally to this work.

Correspondence to: C S Liu, chunsen_liu@fudan.edu.cn; P Zhou, pengzhou@fudan.edu.cn

Received 8 JANUARY 2021; Revised 13 JANUARY 2021.

©2021 Chinese Institute of Electronics

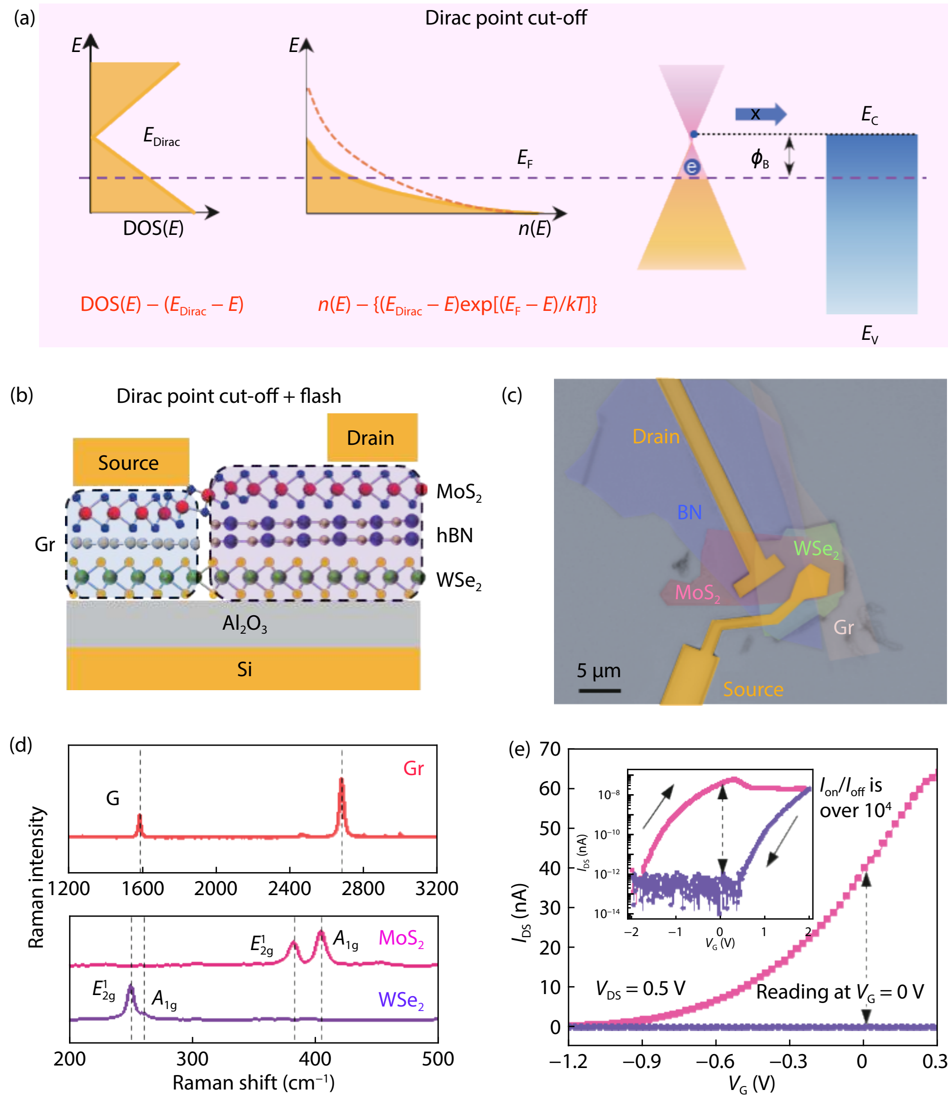


Fig. 1. (Color online) The schematic and characterization of the QNV memory. (a) The linearly varied density of states DOS(E), super-exponentially decreased electron density $n(E)$ with the energy of Gr Dirac material, and the cut-off of electron injection from Gr Dirac material to channel over the bulk barrier ϕ_B. The orange dashed line in $n(E)$ represents the Boltzmann distribution. (b) The schematic structure of the QNV memory in which MoS_2 serves as the channel, Gr-hBN servers as the semi-blocking layer, and WSe_2 serves as the floating gate. (c) The false-colored optical microscope image of the device. The scale bar is 5 μm. (d) The Raman spectra of monolayer Gr and multilayer MoS_2 and WSe_2 were measured by using a 532 nm solid-state laser. (e) The double sweep transfer characteristic of the QNV memory under V_{DS} = 0.5 V. A large I_{on}/I_{off} exceeding 10^4 is achieved when reading at V_G = 0 V. The inset is the full measurement range (V_G varying from −2 to 2 V) of transfer characteristics on a logarithmic scale.

er WSe_2, hBN, MoS_2, and monolayer Gr were mechanically exfoliated from bulk materials to polydimethylsiloxane (PDMS) and then were transferred in turn from the PDMS on the Al_2O_3/Si substrate to form the WSe_2/hBN/Gr/MoS_2 SFG van der Waals heterostructures. After that, Cr/Au (5 nm/30 nm) films were patterned by using electron-beam lithography and deposited by electron-beam evaporation as the source and drain electrodes. Finally, the annealing process was performed at 250 °C in a nitrogen atmosphere for 2 h to ensure good contacts between the films of the device. The corresponding optical microscope images of the device during the preparation process are shown in supplementary Fig. S1.

The Raman spectroscopy of the 2D material used in the device was measured in an SOL instrument by using a 532 nm solid-state laser. The measurement of the electrical performance of the device was performed in a probe station (Cascade Summit 11000 type) by using the Agilent B1500A semiconductor device parameter analyzer. The DC signals and the voltage pulses used in the test were generated by the source/monitor unit (SMU) and the semiconductor pulse generator unit (SPGU) of B1500A, respectively. Additionally, all the above measurements of the device were completed at room temperature in an air environment.

3. Results and discussion

As shown in Fig. 1(a), the p-type Gr Dirac material has a linearly decreased DOS with energy [DOS(E)–(E_{Dirac} − E)] and thus its electron density $n(E)$ above E_F super-exponentially decreases with energy increasing toward the Dirac point energy ($n(E)$–{(E_{Dirac} − E)exp[(E_F − E)/kT]}, where the k and T refer to the Boltzmann constant and temperature respectively), contributing to a more concentrated electron distribution

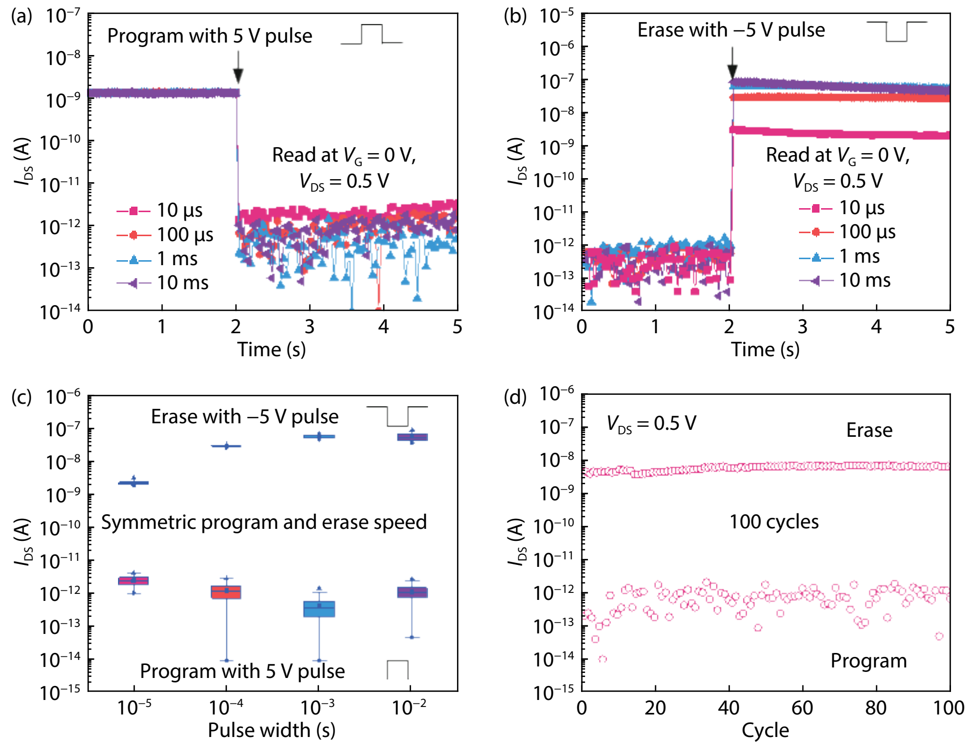


Fig. 2. (Color online) The symmetric ultrafast programming and erasing operation of the QNV memory. (a) The initial state is state-1 and the 5 V gate voltage pulse with different pulse widths is applied for the programming operation. After that, the reading operation is carried out at $V_G = 0$ V and $V_{DS} = 0.5$ V. (b) The initial state is state-0 after applying a positive gate voltage pulse for programming and the –5 V gate voltage pulse with different pulse width is applied for the erasing operation. After that the reading operation is carried out at $V_G = 0$ V and $V_{DS} = 0.5$ V. (c) The current state after the programming and erasing operations with different pulse width gate voltage pulse (the pulse amplitude is fixed 5 V). The data points of the current state are the mean value of the sampling points produced while monitoring the channel current 5 s after the programming/erasing operation. (d) The endurance of the device for 100 cycles of the programming/erasing pulse, demonstrating that no performance degradation was observed.

around E_F. When it is used as the electron source of an n-type channel, most electrons cannot cross the bulk barrier ϕ_B into the channel owing to the long thermal tail of high-energy electrons in Gr has been greatly cut down. This kind of Dirac point cut-off effect is suitable for SFG QNV memory to suppress the leakage of high-energy electrons crossing the bulk barrier ϕ_B between the channel and the floating gate so that its refresh time can be prolonged. Fig. 1(b) shows the schematic structure of the QNV memory where MoS_2 is used as the channel, Gr-hBN is used as the semi-blocking layer and WSe_2 is used as the floating gate. And the MoS_2 is n-type while the Gr is p-type (as shown in supplementary Fig. S2), which satisfies the design requirements of the Dirac point cut-off effect. Besides, the silicon substrate with 30 nm Al_2O_3 dielectric works as the global back gate and 5/30 nm Cr/Au film was patterned and deposited on the MoS_2 film as the source and drain electrode. Fig. 1(c) displays the optical microscope image of the device. The 2D crystals used in the device were characterized by Raman spectroscopy, as shown in Fig. 1(d). The Gr exhibits the G peak at ~1583.0 cm^{-1} and 2D peak at ~2673.9 cm^{-1} and the intensity ratio of 2D to G peak is ~2.56, indicating that it is a monolayer Gr[22, 23]. The Raman spectra of MoS_2 exhibits the $E^1{}_{2g}$ peak at ~381.2 cm^{-1} and the A_{1g} peak at ~403.8 cm^{-1}, and that of WSe_2 exhibits the $E^1{}_{2g}$ and A_{1g} peak at ~249.4 and ~259.3 cm^{-1} respectively, which are consistent with previous reports[24–26] and thus confirm the two kinds of 2D crystals are MoS_2 and WSe_2. The basic double sweep transfer characteristic of the QNV memory is displayed in Fig. 1(e). When sweeping the gate voltage (V_G) from –2 to 2 V at a fixed drain–source voltage (V_{DS}) of 0.5 V, a large memory window is acquired and a large current ratio between on state and off state (I_{on}/I_{off}) exceeding 10^4 is achieved when reading at $V_G = 0$ V.

The programming and erasing operation of the QNV memory are investigated by applying a series of gate voltage pulses with different pulse widths (10 μs, 100 μs, 1 ms, 10 ms), as presented in Fig. 2. The memory was set to state-0 and state-1 by the positive and negative gate voltage pulse respectively, and the state-0 and state-1 correspond to a high resistance state of around 500 GΩ and a relatively low resistance state of less than 500 MΩ respectively. As shown in Fig. 2(a), the initial state of the device is state-1 with a drain-source current (I_{DS}) of around 1 nA. A series of 5 V gate voltage pulses were applied to perform the programming state-0 operation, and then V_G and V_{DS} were set to 0 and 0.5 V respectively to perform the reading operation. An output current of around 1 pA was observed, indicating that the device has been successfully set to state-0 after the 5 V programming pulse with the minimum pulse width of 10 μs. Additionally, the erasing operation of the device is presented in Fig. 2(b). Its initial state has been pre-set to state-0 after a 5 V gate voltage pulse of 100 μs, and next a series of –5 V gate voltage pulses were applied to perform the erasing operation. When reading at $V_G = 0$ V, $V_{DS} = 0.5$ V, a high output cur-

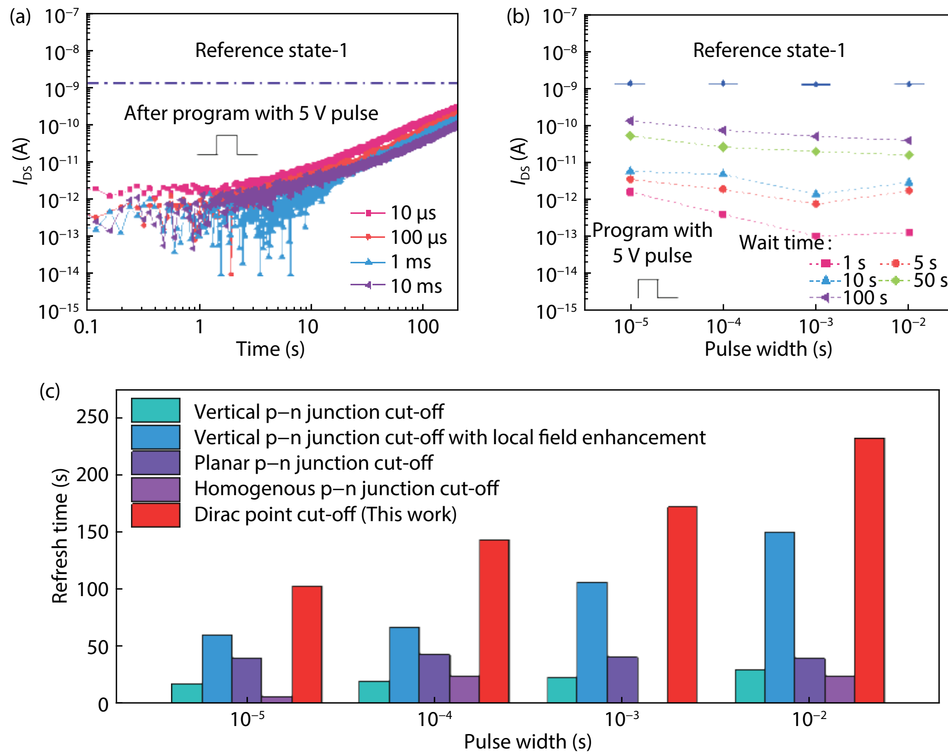


Fig. 3. (Color online) The data retention ability of the QNV memory. (a) The retention characteristic of the memory after the programming operation with different gate pulse widths (the pulse amplitude is fixed 5 V). (b) The relationship between the state-0 output current and program pulse width for different wait times. (c) The comparison of the refresh time of our device with the other QNV memory devices[1, 18–20] under different pulse widths.

rent over 1 nA was measured, indicating that the device can also be successfully erased to state-1 after the –5 V erasing pulse with the minimum pulse width of 10 μs. Fig. 2(c) summarizes the current state after the programming and erasing operations under different gate voltage pulse widths from 10 μs to 10 ms. It can be seen that all the programming and erasing operations can effectively set the device to state-0 and state-1 even if the gate voltage pulse width is reduced to as short as 10 μs, which indicates that the QNV memory can achieve the symmetric ultrafast programming and erasing speeds of 10 μs. What's more, the endurance of the device is presented in Fig. 2(d). After 100 cycles of the programming/erasing operations, the device can still work normally without significant performance degradation.

Fig. 3 shows the data retention ability of the QNV memory after the programming operation with different pulse widths (10 μs, 100 μs, 1 ms, 10 ms). As shown in Fig. 3(a), due to the unavoidable leakage of electrons stored in the WSe_2 floating gate through the semi-blocking layer, the channel current gradually increases with time after the gate voltage pulse is removed. And the reference state-1 refers to the initial state of the device before the programming operation. Fig. 3(b) displays the relationship between the state-0 channel current and program pulse width for different wait times (the time after the programming operation). It can be seen that the longer wait time results in the larger detected output channel current due to the gradual leakage of the storage charge. When the state-0 current is too close to the initial reference state-1, a refresh operation is needed. As shown in Fig. 3(b), even after a long wait time of 100 s, the state-0 current is still sufficiently distinguished with reference state-1 (the current ratio between state-1 and state-0 is over 10), which indicates that the refresh time of the QNV memory is more than 100 s. To more intuitively reflect the advantages of our device in refresh time, Fig. 3(c) shows the comparison of refresh time between our device and all other reported QNV memory devices[1, 18–21] under different programming pulse widths. It can be easily concluded that the data retention ability of our device is much better than other devices. Even when compared with the QNV memory with enhanced refresh time through local field modulation[18], there is still a great refresh time increment of about 54%–113% in our device.

Since our QNV memory combines the Dirac point cut-off effect and the conventional flash memory technology, its mechanism can be perspicuously illustrated by the band diagrams and charges transport in the QNV memory, as shown in Fig. 4. The Gr/hBN composite semi-blocking layer determines that there are two charge transport paths in our device: one is the flash memory path in which charges tunnel through the hBN layer, and the other one is the transparent path in which charges cross through the Gr layer. For the programming operation, a 5 V voltage pulse is applied on the back gate and the electrons in the MoS_2 channel cross through the hBN or Gr layer into the WSe_2 floating gate, as shown in Fig. 4(a). Since the Gr path is transparent with no barrier, the electrons can be easily and quickly injected into the floating gate, which results in an ultrafast programming speed on the timescale of 10 μs. When the read-1 operation is performed at V_G = 0 V and V_{DS} = 0.5 V [Fig. 4(b)], the Dirac point cut-off effect between the n-type MoS_2 and p-type Gr [see Fig. 1(a)] and the hBN blocking layer keep the electrons

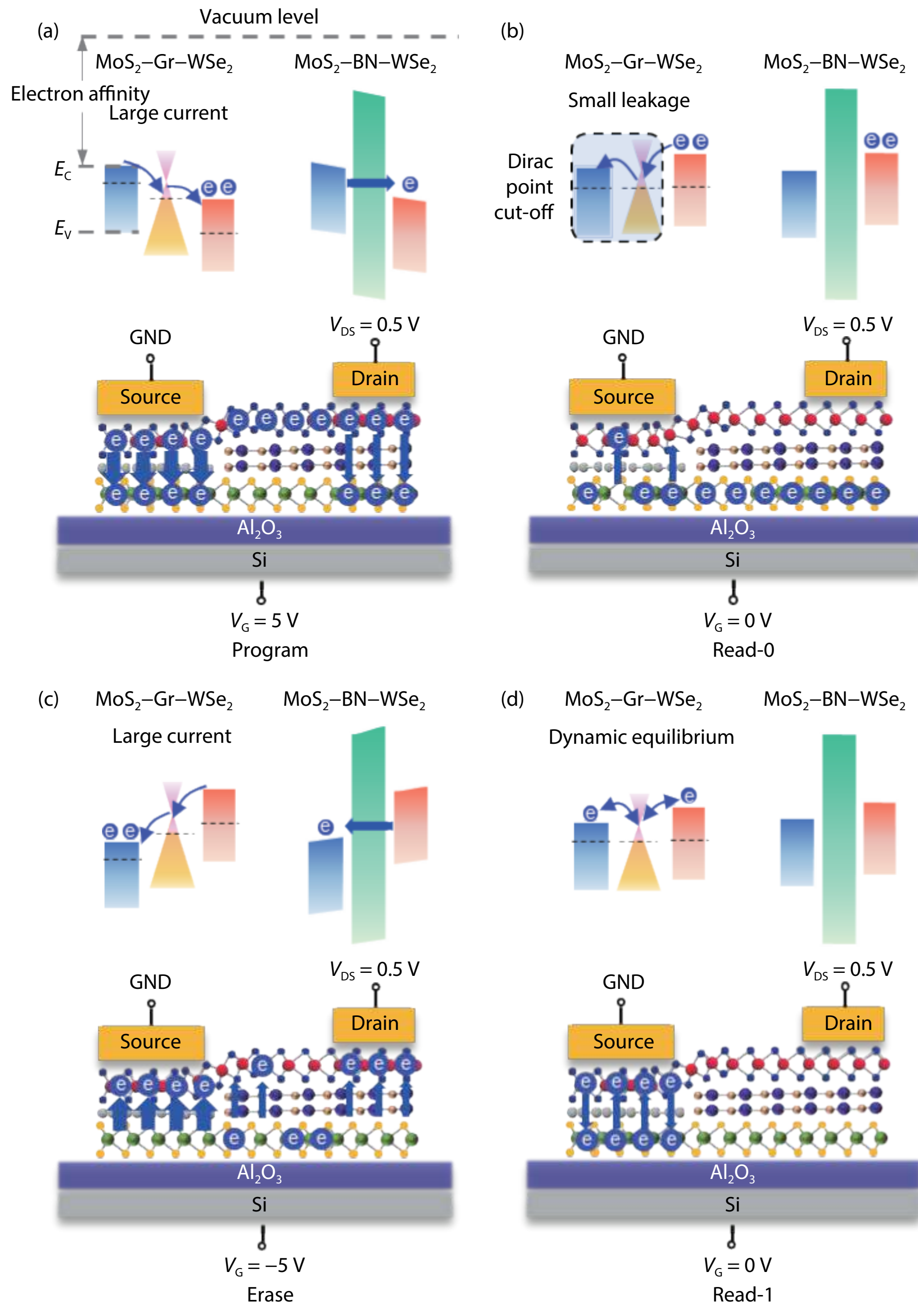


Fig. 4. (Color online) The schematic illustrations of the band diagrams and charges transport in the QNV memory at different operation modes: (a) program, (b) read-0, (c) erase, and (d) read-1. The blue balls and arrows represent the electrons and electron flow direction, respectively. The approximate electron affinities of MoS_2 and WSe_2 are 4.0–4.2, 3.5–4.0 eV respectively, and the work function of monolayer Gr is ~4.3 eV.

in the WSe_2 floating gate from leaking to MoS_2 channel. However, it should be noted that the thermal diffusion of high-energy electrons cannot be completely cut off, which results in non-permanent electron storage. For the erasing operation [Fig. 4(c)] with a –5 V gate voltage pulse, the electrons stored in the floating gate are injected back to the channel in the same way and thus also resulting in a symmetric ultrafast erasing speed. When the device is in the initial state-1 after the erasing operation, the electrons maintain thermal equilibrium in the MoS_2/Gr/WSe_2 region, and thus the charge transfer between the channel and the floating gate is a dynamic equilibrium process, as shown in Fig. 4(d).

4. Conclusion

In summary, based on the SFG architecture and the DOS engineered van der Waals heterostructures, an ultrafast 2D QNV memory with a long refresh time is demonstrated. Inheriting the superiority of the SFG memory in the accessing speed, the device exhibits the symmetric ultrafast programming and erasing speeds of 10 μs. More importantly, the leakage of electrons from the floating gate to the channel is greatly suppressed by the Dirac point cut-off effect between the n-type MoS_2 and p-type Gr, contributing to a long refresh time of more than 100 s that exceeds all previously reported QNV memory devices. Benefit from this significant enhancement of refresh time, considerable power consumption resulting from frequent refresh can be greatly reduced. Therefore, the novel structure in this work provides a reliable way to develop the next-generation memory device with high-speed and low-power.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61925402, 61851402 and 61734003), Science and Technology Commission of Shanghai Municipality (19JC1416600), National Key Research and Development Program (2017YFB0405600), Shanghai Education Development Foundation and Shanghai Municipal Education Commission Shuguang Program (18SG01), China Postdoctoral Science Foundation (2019M661358, 2019TQ0065).

Appendix A. Supplementary material

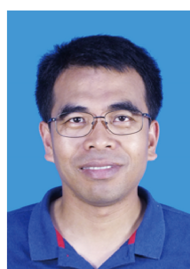
Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1088/1674-4926/42/2/024101.

References

[1] Liu C, Yan X, Song X, et al. A semi-floating gate memory based on van der Waals heterostructures for quasi-non-volatile applications. Nat Nanotechnol, 2018, 13(5), 404
[2] Novoselov K S, Geim A K, Morozov S V, et al. Electric field effect in atomically thin carbon films. Science, 2004, 306(5696), 666
[3] Qiu C, Liu F, Xu L, et al. Dirac-source field-effect transistors as energy-efficient, high-performance electronic switches. Science, 2018, 361(6400), 387
[4] Liu F, Qiu C, Zhang Z, et al. Dirac electrons at the source: breaking the 60-mV/decade switching limit. IEEE Trans Electron Devices, 2018, 65(7), 2736
[5] Liu F, Qiu C, Zhang Z, et al. First principles simulation of energy efficient switching by source density of states engineering. IEEE International Electron Devices Meeting (IEDM), 2018, 33
[6] Lyu J, Pei J, Guo Y, et al. A new opportunity for 2D van der Waals heterostructures: making steep-slope transistors. Adv Mater, 2020, 32(2), 1906000
[7] Yu W J, Chae S H, Lee S Y, et al. Ultra-transparent, flexible single-walled carbon nanotube non-volatile memory device with an oxygen-decorated graphene electrode. Adv Mater, 2011, 23(16), 1889
[8] Lee S, Song E B, Kim S, et al. Impact of gate work-function on memory characteristics in $Al_2O_3/HfO_x/Al_2O_3$/graphene charge-trap memory devices. Appl Phys Lett, 2012, 100(2), 023109
[9] Kim S M, Song E B, Lee S, et al. Transparent and flexible graphene charge-trap memory. ACS Nano, 2012, 6(9), 7879
[10] Bertolazzi S, Krasnozhon D, Kis A. Nonvolatile memory cells based on MoS_2/graphene heterostructures. ACS Nano, 2013, 7(4), 3246
[11] Choi M S, Lee G H, Yu Y J, et al. Controlled charge trapping by molybdenum disulphide and graphene in ultrathin heterostructured memory devices. Nat Commun, 2013, 4(1), 1624
[12] Zhang E, Wang W, Zhang C, et al. Tunable charge-trap memory based on few-layer MoS_2. ACS Nano, 2015, 9(1), 612
[13] Vu Q A, Shin Y S, Kim Y R, et al. Two-terminal floating-gate memory with van der Waals heterostructures for ultrahigh on/off ratio. Nat Commun, 2016, 7(1), 12725
[14] He G, Ramamoorthy H, Kwan C P, et al. Thermally assisted nonvolatile memory in monolayer MoS_2 transistors. Nano Lett, 2016, 16(10), 6445
[15] Lee Y T, Kwon H, Kim J S, et al. Nonvolatile ferroelectric memory circuit using black phosphorus nanosheet-based field-effect transistors with P(VDF-TrFE) polymer. ACS Nano, 2015, 9(10), 10394
[16] Hou X, Zhang H, Liu C, et al. Charge-trap memory based on hybrid 0D quantum dot-2D WSe_2 structure. Small, 2018, 14(20), 1800319
[17] Mishra A, Janardanan A, Khare M, et al. Reduced multilayer graphene oxide floating gate flash memory with large memory window and robust retention characteristics. IEEE Electron Device Lett, 2013, 34(9), 1136
[18] Ding Y, Liu L, Li J, et al. A semi-floating memory with 535% enhancement of refresh time by local field modulation. Adv Funct Mater, 2020, 30(15), 1908089
[19] Li J, Liu L, Chen X, et al. Symmetric ultrafast writing and erasing speeds in quasi-nonvolatile memory via van der Waals heterostructures. Adv Mater, 2019, 31(11), 1808035
[20] Wu G, Tian B, Liu L, et al. Programmable transition metal dichalcogenide homojunctions controlled by nonvolatile ferroelectric domains. Nat Electron, 2020, 3(1), 43
[21] Liu C, Zhou P. Memory Devices Based on Van der Waals Heterostructures. ACS Mater Lett, 2020, 2(9), 1101
[22] Liu J, Huang Z, Lai F, et al. Controllable growth of the graphene from millimeter-sized monolayer to multilayer on Cu by chemical vapor deposition. Nanoscale Res Lett, 2015, 10(1), 455
[23] Dai G P, Wu M H, Taylor D K, et al. Hybrid 3D graphene and aligned carbon nanofiber array architectures. RSC Adv, 2012, 2(24), 8965
[24] Cao Y, Wang Z, Bian Q, et al. Phonon modes and photonic excitation transitions of MoS_2 induced by top-deposited graphene revealed by Raman spectroscopy and photoluminescence. Appl Phys Lett, 2019, 114(13), 133103
[25] Mouri S, Miyauchi Y, Matsuda K. Tunable photoluminescence of monolayer MoS_2 via chemical doping. Nano Lett, 2013, 13(12), 5944
[26] Ahn G H, Amani M, Rasool H, et al. Strain-engineered growth of two-dimensional materials. Nat Commun, 2017, 8(1), 608

Zhaowu Tang was born in Hunan Province, China, in 1998. He received a B.E. degree in materials science and engineering from Northwestern Polytechnical University, China, in 2019. He is a currently graduate student of School of Microelectronics in Fudan University, China. His research interests include 2D materials and their application in logic devices and memory devices. Mr. Tang's awards and honors include Fudan University Outstanding Academic Scholarship and Huawei Scholarship.

Chunsen Liu received the B.S. degree in School of Electronic Science and Technology from Jilin University, Jilin, China, in 2015 and the Ph.D. degree in school of microelectronics from Fudan University, Shanghai, China, in 2019. He is currently working in computer science at Fudan University. His research interest includes the innovation of new logic and memory devices, and exploration of new electronic system architecture. He has first-authored several journal papers in Nature Nanotechnology, Small, etc.



Peng Zhou received his bachelor and Ph.D. degree in physics from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He is currently a full professor on novel electronic devices and process in School of Microelectronics, Fudan University. He has authored or co-authored more than 100 journal papers and conference presentations.