



Atomically sharp interface enabled ultrahigh-speed non-volatile memory devices

Liangmei Wu^{1,2,3,8}, Aiwei Wang^{1,2,3,8}, Jinan Shi^{2,3,8}, Jiahao Yan^{1,2,3,8}, Zhang Zhou^{1,2,3}, Ce Bian^{1,2,3}, Jiajun Ma^{1,2,3}, Ruisong Ma^{1,2,3}, Hongtao Liu^{1,2,3}, Jiancui Chen^{1,2,3}, Yuan Huang¹, Wu Zhou^{1,2,3}, Lihong Bao^{1,2,3,4}✉, Min Ouyang⁵✉, Stephen J. Pennycook^{1,2,6}, Sokrates T. Pantelides^{1,2,7} and Hong-Jun Gao^{1,2,3,4}✉

The development of high-performance memory devices has played a key role in the innovation of modern electronics. Non-volatile memory devices have manifested high capacity and mechanical reliability as a mainstream technology; however, their performance has been hampered by low extinction ratio and slow operational speed. Despite substantial efforts to improve these characteristics, typical write times of hundreds of micro- or milliseconds remain a few orders of magnitude longer than that of their volatile counterparts. Here we demonstrate non-volatile, floating-gate memory devices based on van der Waals heterostructures with atomically sharp interfaces between different functional elements, achieving ultrahigh-speed programming/erasing operations in the range of nanoseconds with extinction ratio up to 10^{10} . This enhanced performance enables new device capabilities such as multi-bit storage, thus opening up applications in the realm of modern nanoelectronics and offering future fabrication guidelines for device scale up.

Memory technology has played a pivotal role in the development of the semiconductor industry over the past decades, driven by the explosive growth of massive data storage and desire of ultrafast data processing¹. Current bottleneck in the memory field includes operation speed, data retention, endurance and extinction ratio^{1,2}. In particular, while the scaling of devices continues, to meet the increasing demands for memory capacity, silicon-based technology will soon reach a critical limit. One of the key challenges is related to the unavoidable interfacial dangling bonds in ultrathin-body silicon, which causes substantial degradation in device performance³. It is thus an urgent need to seek atomically sharp interfaces and seamlessly integrate them into the device architecture. Among all candidates, emerging two-dimensional (2D) materials^{4–8} and their heterostructures^{9–13} represent ideal atomically flat in-plane surfaces potentially free from surface dangling bonds and are immune to short-channel effects that can allow effective electrostatic control and mechanical flexibility¹⁴. Indeed, a few examples employing 2D materials for flash memory devices have been recently attempted^{15–26} but with limited device performance. For example, a very long write time of the order of milliseconds to seconds was observed in 2D-materials-based floating-gate memory devices^{15–18}, while alternative semi-floating-gate configurations^{27,28} have shown improved write time (nanoseconds) but extremely short retention time in the range of seconds, making them unsuitable for long-term storage. Theoretically, an ideal floating-gate memory device based on planar layer materials should allow nanosecond-order operational time (Supplementary Figs. 1 and 2 and Supplementary Note 1), but the ultrahigh-speed floating-gate memory has not yet been reported so far.

Here we have demonstrated, for the first time, that ultrahigh-speed, non-volatile floating-gate memory devices can be achieved without the need to modify commercial device architecture by employing 2D van der Waals heterostructures with improved interfacial coupling and atomically sharp interfaces. In particular, ultrahigh-speed operation with nanosecond write and read times that is limited by instrumentation response, extremely high extinction ratio of 10^{10} and a retention time of 10 yr have been successfully achieved. Our result represents a leap forward to understand the current bottleneck of memory devices as well as future memory and storage technologies.

Figure 1a (top) shows a schematic of our floating-gate field-effect transistor (FGFET) device, consisting of a vertically stacked indium selenide (InSe)/hexagonal boron nitride (hBN)/multilayer graphene (MLG) van der Waals heterostructure on a $\text{SiO}_2/\text{p}^{++}\text{Si}$ substrate, where InSe, hBN, MLG, SiO_2 and p^{++}Si serve as the channel, tunnel barrier, floating gate, control-gate dielectric and control gate, respectively. Figure 1b shows a false-colour optical image of a typical device in which one MLG (marked by the blue dashed line) is directly placed on a 300 nm SiO_2/Si substrate, followed by sequential stacking of hBN (marked by the white dashed line) and InSe on top of the MLG. Source and drain electrodes are fabricated using standard electron-beam lithography followed by thermal evaporation. The detailed materials preparation and device fabrication processes are provided in Methods and Supplementary Fig. 3. The MLG floating gate can be charged/discharged by an electric field, and charges (ΔQ) can be retained in the MLG after removing the electric field (Fig. 1a, bottom), regulating the conductivity of the

¹Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China. ²University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, People's Republic of China. ³CAS Center for Excellence in Topological Quantum Computation, Chinese Academy of Sciences, Beijing, People's Republic of China. ⁴Songshan Lake Materials Laboratory, Dongguan, Guangdong, People's Republic of China. ⁵Department of Physics, University of Maryland, College Park, MD, USA. ⁶Department of Materials Science and Engineering & Centre for Advanced 2D Materials, National University of Singapore, Singapore, Singapore. ⁷Department of Physics and Astronomy & Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA. ⁸These authors contributed equally: Liangmei Wu, Aiwei Wang, Jinan Shi, Jiahao Yan. ✉e-mail: lhbao@iphy.ac.cn; mouyang@umd.edu; hjgao@iphy.ac.cn

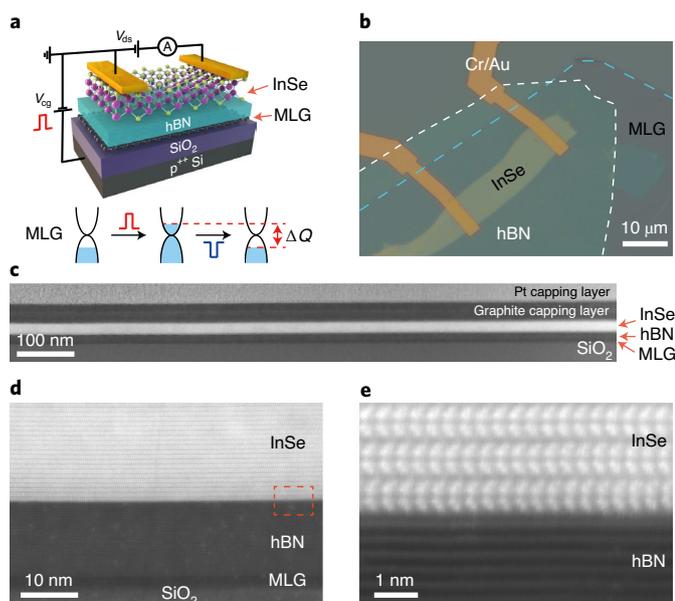


Fig. 1 | Atomically sharp interfaces embedded in non-volatile memory device with floating-gate configuration built upon van der Waals heterostructures. **a**, Schematic of a floating-gate memory device configuration and its operational principle. **b**, False-colour optical image of a memory device based on an InSe/hBN/MLG heterostructure placed on an SiO₂/Si substrate. **c**, Large-scale HAADF-STEM image of the InSe/hBN/MLG heterostructure on SiO₂ with capping layers of graphite and Pt, highlighting the achievable uniform, atomically sharp in-plane interfaces. **d,e**, Higher-resolution HAADF-STEM images acquired from the same heterostructure for different length scales, revealing clean atomic characteristics without defects or contamination among InSe, hBN and MLG. The image in **e** corresponds to the orange rectangular dotted area in **d**.

channel materials to fulfil non-volatile memory function²⁹. Our device fabrication is general and reproducible. For example, devices with the same configuration and structural quality have also been achieved by using other 2D layers as the channel material, including molybdenum disulfide (MoS₂).

We have employed an aberration-corrected scanning transmission electron microscope (STEM) to thoroughly characterize interfaces between different functional components in as-fabricated devices. Typical atomic-number-contrast (*Z*-contrast) high-angle annular dark-field (HAADF) STEM cross-sectional images acquired from the same device on different length scales are presented in Fig. 1c–e. Figure 1c highlights large-scale atomically sharp layered interfaces achievable in our device with uniformity, while Fig. 1d,e presents higher-resolution images to confirm interfacial quality at the atomic scale. More bright-field and HAADF-STEM images are presented in Supplementary Fig. 4. Our results have unambiguously revealed the existence of extremely uniform and clean atomically sharp interfaces between different functional layers in our devices, without any gap or observable defects/contamination (Fig. 1c–e). These features are the key for the exceptional device performance presented below.

Figure 2a shows the typical transfer curves of an InSe FGFET at room temperature. The data (drain–source current, I_{ds}) were acquired by sweeping the control-gate voltage (V_{cg}) from negative to positive values and then back to negative values, with floating MLG and a fixed drain–source bias (V_{ds}) of 0.05 V. A sizeable hysteresis is clearly observed in Fig. 2a. We have performed two control experiments by measuring the same FGFET devices with grounded

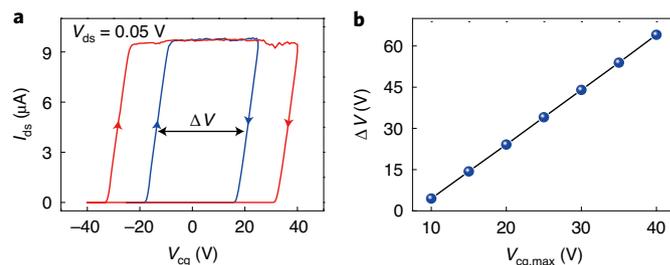


Fig. 2 | Two-dimensional InSe-based floating-gate memory device with large memory window. **a**, Typical dual-sweeping transfer curves with an FGFET configuration. Data were acquired by sweeping V_{cg} with floating MLG, while monitoring I_{ds} under a fixed V_{ds} of 0.05 V. The blue and red curves correspond to the maximum V_{cg} of 25 V and 40 V, respectively. Both transfer curves clearly reveal the existence of large ΔV . The arrows show the sweeping directions of V_{cg} . **b**, Linear dependence of ΔV on $V_{cg,max}$.

MLG (Supplementary Fig. 5a) and by fabricating independent devices without the MLG layer (Supplementary Fig. 5b). Both controls have shown negligible hysteresis in their transfer curves, indicating a low trapped charge density at the InSe/hBN interface. By comparing with the result in Fig. 2a, we conclude that the presence of the large voltage hysteresis can be attributed to the existence of the MLG floating gate. The appearance of substantial hysteresis signifies the storage capability of a memory device, and it is characterized by a memory window width (ΔV) that is defined as the shift in the threshold voltage in the dual-swept V_{cg} . In general, a larger memory window suggests that more charges can be stored in the MLG floating gate, resulting in a more reliable memory device performance (for example, a minimum width of 1.5 V is typically necessary to produce a reasonable on/off ratio for reliable memory application¹⁹). Figure 2b shows that ΔV proportionally increases with the maximum value of the control-gate voltage ($V_{cg,max}$), where the amount of charge stored in the MLG can be tuned by the applied V_{cg} . The complete set of data is shown in Supplementary Fig. 6. Here ΔV can be as high as 64 V when V_{cg} is swept between -40 V and $+40$ V. The corresponding charge density stored in the MLG floating gate can thus be estimated by $(\Delta V \times C_{CG-FG})/q$, where q is the electron charge of 1.6×10^{-19} C and C_{CG-FG} is the dielectric capacitance between the control gate and floating gate (300 nm SiO₂) of 1.15×10^{-8} F cm⁻², resulting in a value of 4.6×10^{12} cm⁻². Such a large charge density achievable in the MLG floating gate outperforms polycrystalline silicon in conventional flash memory, and this can facilitate easy detection in practical applications²⁹. There are a few advantages of utilizing MLG as the floating gate, including a significant reduction in floating-gate interference^{30,31} and effective suppression of ballistic currents across the floating gate due to low conductivity along the *c* axis. Furthermore, compared with single-layer graphene, MLG has a higher work function (4.6 eV) that is independent of the number of layers, enabling longer charge retention time as well as higher electronic density of states¹⁹. All these attributes of MLG have contributed to the observed large memory window in our devices.

We have proposed and defined, in Fig. 3a, the programming and erasing processes achievable in our FGFET devices for a non-volatile floating-gate memory: when a positive voltage pulse is applied to the control gate with the drain and source grounded and floating MLG (Fig. 3a(i)), electrons from the InSe channel tunnel through the hBN barrier and accumulate in the MLG, completing a program operation. The accumulated electrons in the MLG are well retained even after removing the external electric field due to a low return probability, resulting in a positive shift in the threshold voltage and thus low conductance of the InSe channel at $V_{cg}=0$ V. This state is

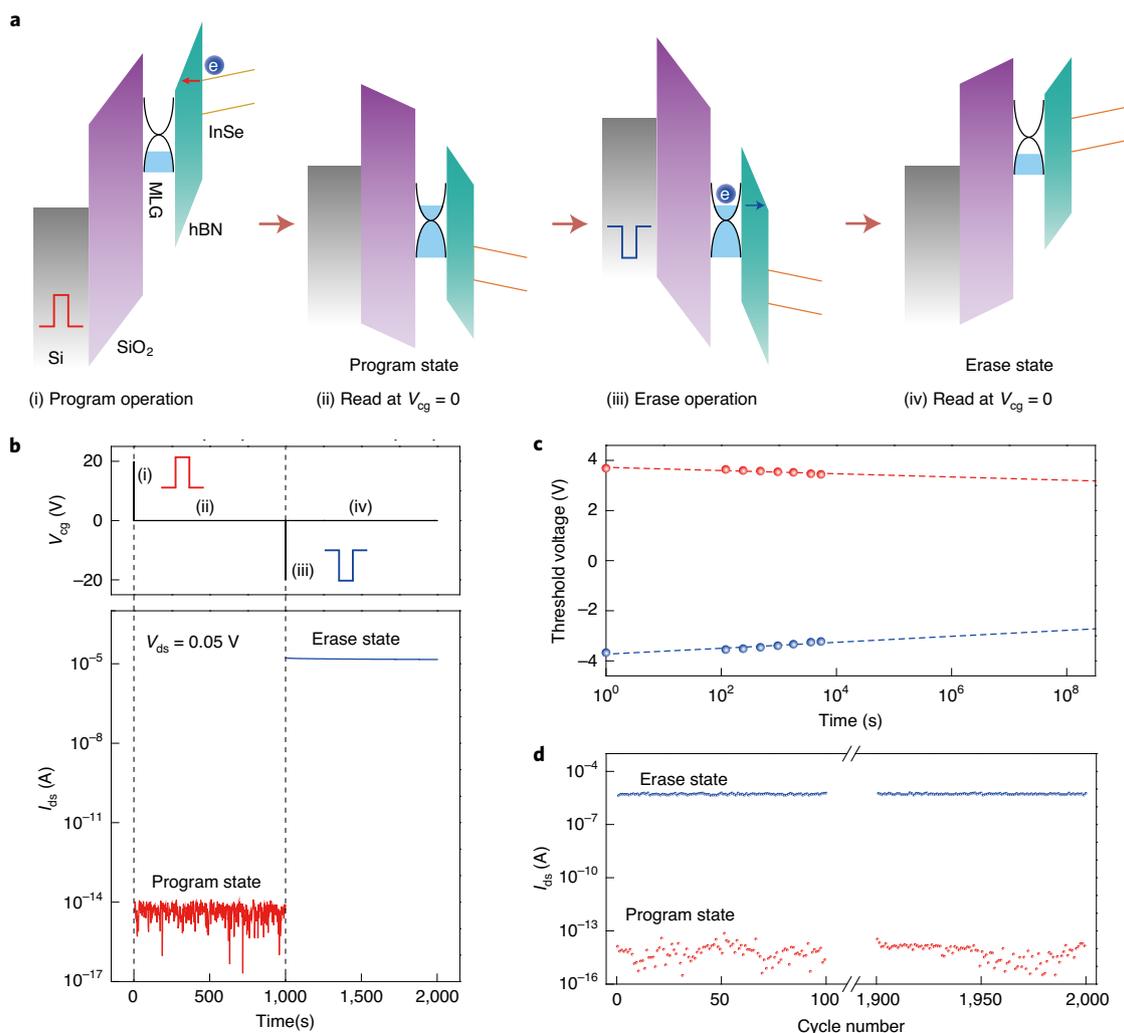


Fig. 3 | Programming and erasing non-volatile InSe floating-gate memory devices with large extinction ratio and robust performance. **a**, Schematic energy-band diagrams of a memory device for program, read and erase operations. **b**, Conductance of the InSe channel can be toggled between the low-conductance state (that is, program state) and high-conductance state (that is, erase state) by applying a single +17.7 V or -17.7 V voltage pulse with FWHM of 160 ns, respectively. **c**, Time variation of threshold voltage after +17.7/-17.7 V programming/erasing pulses with FWHM of 160 ns. **d**, Endurance performance test of the memory device executed with alternate voltage pulses (+17.7/-17.7 V programming/erasing pulses with FWHM of 160 ns).

defined as the ‘program state’ (Fig. 3a(ii)). On the contrary, when applying a negative voltage pulse to the control gate (Fig. 3a(iii)), the stored electrons can tunnel from the MLG floating gate back to the InSe channel, that is, the erase operation, leading to a negative shift in the threshold voltage after the external electric field is turned off. As a result, the InSe channel manifests high conductance at $V_{cg} = 0$ V, which is defined as the ‘erase state’ (Fig. 3a(iv)).

Figure 3b shows a typical realization of the proposed memory operations in our devices by using +17.7/-17.7 V voltage pulses with full-width at half-maximum (FWHM) of 160 ns for programming/erasing operations (the output voltage pulse waveforms are shown in Supplementary Fig. 7). When a positive square pulse is applied to the control gate, the InSe channel is driven to a low-conductance state, that is, the program state. This program state can be ‘erased’ by applying a negative voltage pulse to the control gate to recover the high conductance of the InSe channel, that is, the erase state. It is worth noting that the amplitude of the programming and erasing voltage pulses in Fig. 3b (17.7 V) is comparable to that of a commercial flash memory (~15 V) (ref. 1), whereas the extinction ratio between the erase and program states is up to 10^{10} —much higher than other floating-gate memory devices^{15–19} (Extended Data Fig. 1).

This high extinction ratio not only allows facile and reliable readout of the memory states to reduce the data sensing error but also opens up other attractive applications such as multi-bit storage in the same memory cell, which is challenging otherwise.

Our InSe floating-gate memory devices are very robust with high endurance and high reliability. We have extracted the threshold voltage of the InSe channel from transfer curves at different time intervals (Supplementary Fig. 8) and summarized the result in Fig. 3c. In addition, we also measured the current evolution of the program and erase states for an InSe floating-gate memory after 16.5 days, showing excellent retention performance (Supplementary Fig. 9). This result projects a 14.6% and 27.1% drop in the threshold voltage of the program (red dashed line) and erase (blue dashed line) states even after 10 yr, respectively, thus making long-term data retention become feasible at room temperature. The endurance test by repeatedly programming and erasing the same memory cell has also been performed; the results are presented in Fig. 3d. It can be clearly seen that both program and erase states remain almost unchanged with negligible degradation even after 2,000 cycles. The robust retention and endurance performance of our memory devices can be attributed

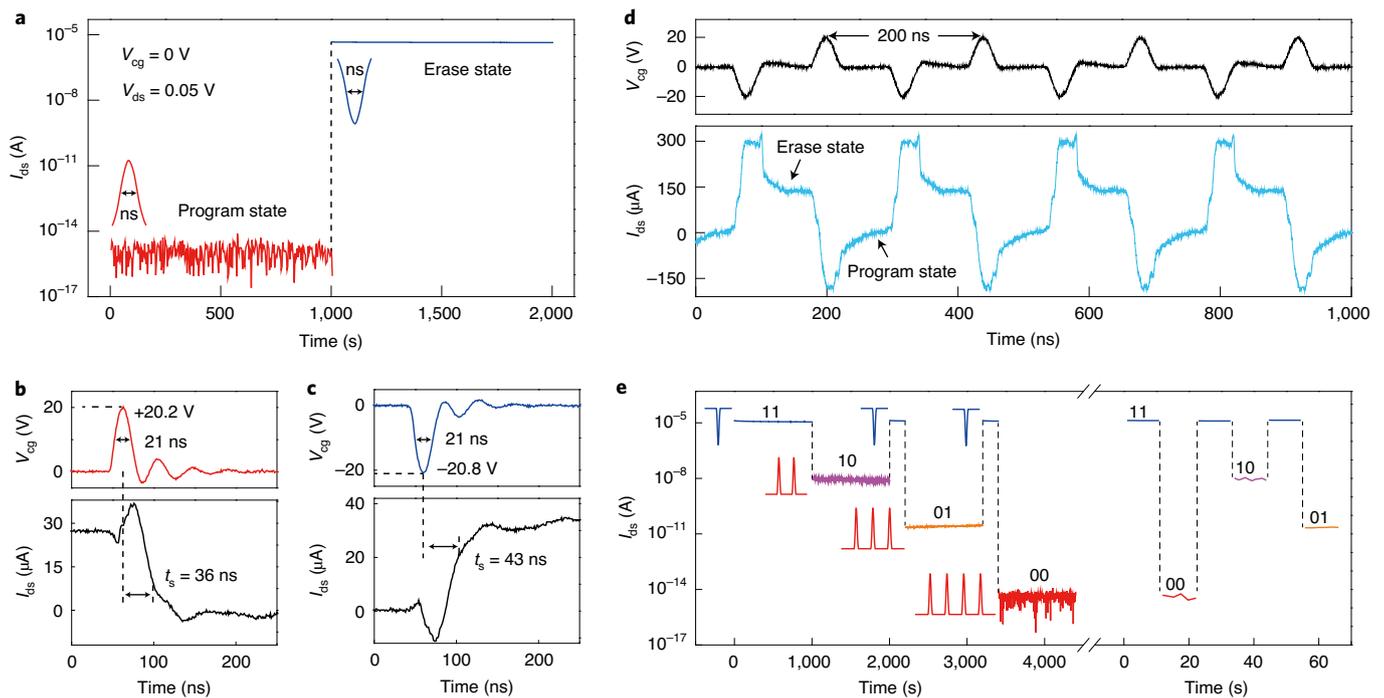


Fig. 4 | Ultrafast operation of memory cell and enabling the multi-bit storage paradigm. **a**, InSe memory cell with 10-nm-thick hBN layers can be successfully programmed/erased by a positive/negative voltage pulse with a nanosecond pulse width, while maintaining a high extinction ratio of 10^{10} . The amplitude of the voltage pulse is +20.2/-20.8 V. **b,c**, Ultrafast response of memory cells after programming (**b**) and erasing (**c**) pulses. **d**, Demonstration of ultrahigh-frequency operation by sequential programming and erasing pulses with ~100 ns intervals. **e**, Realization of reproducible multi-bit storage from devices with 12-nm-thick hBN layers by a simple combination of ultrafast pulse sequence, with an example of repeatable two-bit memory capacity demonstrated in the figure. By designing and applying sequential pulses for data programming, multiple-level states of the memory cell can be programmed with a discernible extinction ratio to achieve multi-bit storage. All programmed states can be consistently erased by a single-nanosecond negative voltage pulse. The voltage amplitude and FWHM of the programming and erasing pulses are (+20.2 V, 21 ns) and (-20.8 V, 21 ns), respectively.

to the defect-free hBN tunnelling barrier layer integrated in our devices^{18,32}.

Importantly, we have demonstrated that our InSe floating-gate memory devices feature ultrahigh-speed operation, which is not available in conventional flash memory devices. Figure 4a shows that the InSe floating-gate memory device can be successfully programmed/erased by an ultrashort +20.2/-20.8 V voltage pulse down to FWHM of 21 ns. The transient responses of memory devices by following ultrashort programming and erasing pulses are shown in Fig. 4b,c, respectively. Two important characteristics can be immediately identified under ultrafast operations: first, the data storage and erasure can still be consistently achieved with high extinction ratio up to 10^{10} even with ultrashort programming and erasing pulses; second, the memory states can respond instantaneously and settle without delay. We have defined the response time (t_s) of a device as the time required for it to reach $1/e$ value of its desired state after the peak pulse, and estimated a 36 ns and 43 ns response time for programming and erasing operations with an FWHM pulse width of 21 ns, respectively. It is worth noting that these response time estimations should be taken as the upper bound values and are currently limited by the voltage pulse source available in our work. Nevertheless, the observed ultrahigh-speed operation of floating-gate memory devices has approached the theoretical prediction of an ideal device possessing the same experimental device parameters (Supplementary Fig. 2), and it is ~5,000 times faster than that of commercially produced flash memory devices and comparable with that of commercially available volatile dynamic random access memory products³³. This is again consistent with our confirmation of an atomic sharp interfacial feature, as shown in Fig. 1c,d. Because of the extremely fast transient

response for both programming and erasing operations, it can allow a truly ultrahigh-speed memory device operation. Figure 4d shows one example of ultrahigh-frequency operation of a memory cell by sequential programming and erasing up to 5 MHz, which is again limited by instrumentation response.

To elucidate the underlying physics for the exceptional performance of our memory device and compare with an ideal model (Supplementary Note 1), we have performed a few control experiments. (1) To verify that the observed ultrafast operation indeed occurs in our memory devices and to exclude the possibility of direct tunnelling between the metal electrode and MLG floating gate, we have fabricated a device that can allow a side-by-side comparison of the writing and erasing performance by configuring the device in either the InSe/hBN/MLG or metal/hBN/MLG structure (Supplementary Fig. 10). We have confirmed that the metal/hBN/MLG configuration does not allow successful programming. Furthermore, an independent device that was fabricated by carefully eliminating overlapping between the top metal electrode and the MLG (Supplementary Fig. 11) successfully reproduced ultrafast device performance, as presented in Fig. 4a (Supplementary Note 3). These control experiments unambiguously confirm that the origin of the ultrafast memory operation of our devices is the InSe channel layer rather than other causes. (2) We have fabricated devices with different thicknesses of hBN tunnel barrier and studied their current-voltage relationships in the high-bias region; the data are summarized in Supplementary Fig. 12. A linear dependence of $\ln(1/V^2)$ on $1/V$ is clearly present in all our devices, suggesting that the Fowler-Nordheim tunnelling process is dominant in our devices, further substantiating the ideal model (Supplementary Note 4). This result is consistent with our observation of nanosecond

operation time, which is approaching the theoretical limit. Our mechanistic understanding of ultrahigh-speed device performance raises an important question about the generality of such memory devices as well as the criteria of choosing channel-layer materials that can offer unprecedented device performance. The Fowler–Nordheim process suggests that a reduction in the channel/hBN barrier height, which is determined by the work function of the channel semiconductor and hBN electron affinity (Supplementary Note 5), should result in a larger tunnelling current. Therefore, a low-work-function van der Waals semiconductor layer is preferred as a channel material to enable ultrafast performance in memory devices. Indeed, our fabrication of floating-gate memory devices with atomically sharp interfaces can be reproducibly extended to other 2D materials. An example is shown in Supplementary Fig. 13 using MoS₂ as the channel material (the work function of MoS₂ is comparable to that of InSe; also see Supplementary Note 5), highlighting the generality and reproducibility of the roles of atomically sharp interfaces in 2D memory technology.

The combination of high extinction ratio and ultrafast-operation characteristics achievable in our van der Waals memory devices immediately open up opportunities for novel device configurations and applications. For example, the current extinction ratio of 10¹⁰ should allow achievement of multi-bit memory operation to store more than one bit of information per cell for ultrahigh-density information storage. Nevertheless, in a commercial multi-level cell flash memory, multi-bit memory operation is often achieved by controlling the amplitude of V_{cg} to change the amount of charge stored in the floating gate, requiring extra control of the memory cells³⁴. Enabled by the ultrafast-operation capability (Fig. 4a), an alternative mechanism of multi-bit storage can be achieved by a judicious design of the pulse sequence to control the amount of charges stored in the MLG floating gate, offering more flexibility for data programming. Figure 4e exemplifies the realization of two-bit storage in our InSe-based memory cell. In this demonstration, we have utilized the same ultrashort pulses with FWHM of 21 ns as shown in Fig. 4b,c for programming and erasing. More specifically, by applying a pre-defined pulse sequence, we have shown that well-discernible multiple states of memory cells can be reproducibly achieved. For example, Fig. 4e shows that the memory cell can be programmed to different program states of (10), (01) and (00) with a sufficiently large extinction ratio and good retention performance after two, three and four positive voltage pulses (+20.2 V, 21 ns), respectively. It should be noted that in this demonstration, for every programming operation, an erase operation is applied to reset the memory state to (11), which has been consistently achieved by applying a single negative voltage pulse (−20.8 V, 21 ns). Furthermore, by optimizing our memory devices (for example, the thickness of the hBN tunnel-barrier layer), higher storage capacity such as triple-level cell and quadruple-level cell has also been demonstrated with reliable performance (Supplementary Figs. 14 and 15)^{35,36}. Such a multi-bit storage scheme can only be possible when both high extinction ratio and ultrafast operation become available, highlighting the uniqueness of our demonstrated memory device configuration and performance.

We have fabricated and exemplified emerging non-volatile floating-gate memory devices based on 2D functional layer materials, exhibiting device performance approaching the theoretical limit of an ideal memory structure and exciting ultrahigh-speed operation for the first time, and the key is our achieved uniform atomically sharp interfaces inherent in the as-fabricated memory devices, as highlighted in Fig. 1c. Our findings and device performances have important implications and open up exciting opportunities on several fronts. Our device configuration follows the commercial floating-gate memory architecture, but with a few orders of improvement in the device performance. As a result, it should facilitate the adoption of existing industry standards with

emerging technology. Although current device fabrication is based on the exfoliation method, our finding of the roles of an atomically sharp interface between functional layers in the performance limit of memory devices should offer a fabrication guideline for future integration with other techniques for device scale up^{37–41}. Moreover, the success of ultrahigh-speed nanosecond operation and response in our as-fabricated devices has solved the long-term challenge in the field of non-volatile memory technology; this is comparable to commercial volatile dynamic random access memory technology, representing a big leap forward in the present demand in data storage and data processing. In addition, such outstanding performance should allow many new unforeseen operational schemes, including multi-bit ultrahigh-density information storage demonstrated in the present work, which should be crucial for the recent development of big data science. All these together offer the ground for next-generation electronic devices based on emerging 2D materials with optimum device engineering.

Note added in proof: After submission of this paper, a manuscript reporting similar results was posted on arXiv⁴².

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41565-021-00904-5>.

Received: 16 August 2020; Accepted: 17 March 2021;

Published online: 03 May 2021

References

1. International Roadmap for Devices and Systems (IRDS) <https://irds.ieee.org/> (2017).
2. Hwang, C. S. Prospective of semiconductor memory devices: from memory system to materials. *Adv. Electron. Mater.* **1**, 1400056 (2015).
3. Chhowalla, M., Jena, D. & Zhang, H. Two-dimensional semiconductors for transistors. *Nat. Rev. Mater.* **1**, 16052 (2016).
4. Novoselov, K. S. et al. Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
5. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS₂ transistors. *Nat. Nanotechnol.* **6**, 147–150 (2011).
6. Li, L. et al. Black phosphorus field-effect transistors. *Nat. Nanotechnol.* **9**, 372–377 (2014).
7. Feng, W., Zheng, W., Cao, W. & Hu, P. Back gated multilayer InSe transistors with enhanced carrier mobilities via the suppression of carrier scattering from a dielectric interface. *Adv. Mater.* **26**, 6587–6593 (2014).
8. Wu, L. et al. InSe/hBN/graphite heterostructure for high-performance 2D electronics and flexible electronics. *Nano Res.* **13**, 1127–1132 (2020).
9. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499**, 419–425 (2013).
10. Liu, Y. et al. Van der Waals heterostructures and devices. *Nat. Rev. Mater.* **1**, 16042 (2016).
11. Novoselov, K. S., Mishchenko, A., Carvalho, A. & Castro Neto, A. H. 2D materials and van der Waals heterostructures. *Science* **353**, aac9439 (2016).
12. Haigh, S. J. et al. Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices. *Nat. Mater.* **11**, 764–767 (2012).
13. Kretinin, A. V. et al. Electronic properties of graphene encapsulated with different two-dimensional atomic crystals. *Nano Lett.* **14**, 3270–3276 (2014).
14. Fiori, G. et al. Electronics based on two-dimensional materials. *Nat. Nanotechnol.* **9**, 768–779 (2014).
15. Bertolazzi, S., Krasnozhan, D. & Kis, A. Nonvolatile memory cells based on MoS₂/graphene heterostructures. *ACS Nano* **7**, 3246–3252 (2013).
16. Choi, M. S. et al. Controlled charge trapping by molybdenum disulphide and graphene in ultrathin heterostructured memory devices. *Nat. Commun.* **4**, 1624 (2013).
17. Li, D. et al. Nonvolatile floating-gate memories based on stacked black phosphorus–boron nitride–MoS₂ heterostructures. *Adv. Funct. Mater.* **25**, 7360–7365 (2015).
18. Wang, S. et al. New floating gate memory with excellent retention characteristics. *Adv. Electron. Mater.* **5**, 1800726 (2019).
19. Hong, A. J. et al. Graphene flash memory. *ACS Nano* **5**, 7812–7817 (2011).

20. Lee, S. et al. Impact of gate work-function on memory characteristics in $\text{Al}_2\text{O}_3/\text{HfO}_x/\text{Al}_2\text{O}_3/\text{graphene}$ charge-trap memory devices. *Appl. Phys. Lett.* **100**, 023109 (2012).
21. Chen, M. et al. Multibit data storage states formed in plasma-treated MoS_2 transistors. *ACS Nano* **8**, 4023–4032 (2014).
22. Wang, J. et al. Floating gate memory-based monolayer MoS_2 transistor with metal nanocrystals embedded in the gate dielectrics. *Small* **11**, 208–213 (2015).
23. Zhang, E. et al. Tunable charge-trap memory based on few-layer MoS_2 . *ACS Nano* **9**, 612–619 (2015).
24. Feng, Q., Yan, F., Luo, W. & Wang, K. Charge trap memory based on few-layer black phosphorus. *Nanoscale* **8**, 2686–2692 (2016).
25. Lee, D. et al. Black phosphorus nonvolatile transistor memory. *Nanoscale* **8**, 9107–9112 (2016).
26. Liu, C. et al. Eliminating overerase behavior by designing energy band in high-speed charge-trap memory based on WSe_2 . *Small* **13**, 1604128 (2017).
27. Wang, P. F. et al. A semi-floating gate transistor for low-voltage ultrafast memory and sensing operation. *Science* **341**, 640–643 (2013).
28. Liu, C. et al. A semi-floating gate memory based on van der Waals heterostructures for quasi-non-volatile applications. *Nat. Nanotechnol.* **13**, 404–410 (2018).
29. Kahng, D. & Sze, S. M. A floating gate and its application to memory devices. *Bell Syst. Tech. J.* **46**, 1288–1295 (1967).
30. Lee, J.-D., Hur, S.-H. & Choi, J.-D. Effects of floating-gate interference on NAND flash memory cell operation. *IEEE Electron Device Lett.* **23**, 264–266 (2002).
31. Misra, A. et al. Multilayer graphene as charge storage layer in floating gate flash memory. In *2012 4th IEEE International Memory Workshop* 1–4 (2012).
32. Vu, Q. A. et al. Two-terminal floating-gate memory with van der Waals heterostructures for ultrahigh on/off ratio. *Nat. Commun.* **7**, 12725 (2016).
33. Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nat. Nanotechnol.* **8**, 13–24 (2013).
34. Cho, T. et al. A dual-mode NAND flash memory: 1-Gb multilevel and high-performance 512-Mb single-level modes. *IEEE J. Solid-State Circuits* **36**, 1700–1706 (2001).
35. Xiang, D. et al. Two-dimensional multibit optoelectronic memory with broadband spectrum distinction. *Nat. Commun.* **9**, 2966 (2018).
36. Tran, M. D. et al. Two-terminal multibit optical memory via van der Waals heterostructure. *Adv. Mater.* **31**, 1807075 (2019).
37. Kang, K. et al. Layer-by-layer assembly of two-dimensional materials into wafer-scale heterostructures. *Nature* **550**, 229–233 (2017).
38. Li, X. et al. Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science* **324**, 1312–1314 (2009).
39. Pan, Y. et al. Highly ordered, millimeter-scale, continuous, single-crystalline graphene monolayer formed on Ru (0001). *Adv. Mater.* **21**, 2777–2780 (2009).
40. Shi, Z. et al. Vapor-liquid-solid growth of large-area multilayer hexagonal boron nitride on dielectric substrates. *Nat. Commun.* **11**, 849 (2020).
41. Kang, K. et al. High-mobility three-atom-thick semiconducting films with wafer-scale homogeneity. *Nature* **520**, 656–660 (2015).
42. Liu, L., Ding, Y., Li, J., Liu, C. & Zhou, P. Ultrafast non-volatile flash memory based on van der Waals heterostructures. Preprint at <https://arxiv.org/abs/2009.01581> (2020).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Device fabrication and electrical characterization. InSe, MoS₂, hBN and graphite bulk crystals were purchased from 2D Semiconductors, SPI Supplies, HQ Graphene and NGS Naturgraphit, respectively. InSe/hBN/MLG and MoS₂/hBN/MLG heterostructures were prepared on a silicon wafer with 300 nm SiO₂ using mechanical exfoliation and dry-transfer approach^{43,44} carried out in an argon atmosphere glovebox with O₂ and H₂O concentrations below 0.5 ppm. Then Cr/Au drain and source electrodes were defined by standard electron-beam lithography, thermal evaporation and lift off. The floating-gate memory devices were protected by a layer of spin-coated poly(methylmethacrylate) at once after lift off⁴⁵. Electrical measurements were performed with a Keithley 4200 semiconductor characterization system (4200-SCS) and a home-made electric circuit in a vacuum probe station at room temperature. Memory characterization of the +17.7/-17.7 V voltage pulses with FWHM of 160 ns was performed in the 4200-SCS system equipped with 4225-PMU and 4225-RPM units. The ultrahigh-speed characterization of InSe and MoS₂ floating-gate memory devices was executed with the control-gate terminal connected to a home-made nanosecond voltage pulse signal, while the drain and source terminals were connected to the 4200-SCS source measure units. The grounds of the home-made electric circuit and 4200-SCS needed to be connected together. The response time tests were executed in atmospheric environment. The drain-source current is amplified with a FEMTO DHPCA-100 system (gain of 10⁴ V A⁻¹ and bandwidth of 14 MHz). The nanosecond voltage pulse and amplified drain-source current were recorded with a Tektronix MDO4104B-3 mixed domain oscilloscope (bandwidth of 1 GHz). The thicknesses of InSe, hBN and MLG flakes were determined by a Bruker Dimension Edge atomic force microscope after all tests.

STEM sample preparation and characterization. Before the transmission electron microscope (TEM) sample fabrication, the InSe/hBN/MLG heterostructure on an SiO₂/Si substrate was protected by covering it with three layers of graphite flakes (with total thickness reaching dozens of nanometres) by the dry-transfer approach. Then a standard cross-section TEM lift-out method was carried out with an FEI Helios NanoLab G3 CX focused-ion-beam microscope for high-resolution imaging. HAADF and bright-field imaging were performed in an aberration-corrected Nion HERMES-100 STEM at 60 kV acceleration voltage. Atomic-number-contrast images were obtained using an illumination semi-angle of 32 mrad; the collection semi-angle ranged from 75 to 210 mrad for HAADF imaging and was within 9 mrad for bright-field imaging.

Data availability

Source data are provided with this paper. All relevant data are available from the corresponding authors upon reasonable request.

References

43. Lee, G.-H. et al. Flexible and transparent MoS₂ field-effect transistors on hexagonal boron nitride-graphene heterostructures. *ACS Nano* **7**, 7931–7936 (2013).

44. Castellanos-Gomez, A. et al. Deterministic transfer of two-dimensional materials by all-dry viscoelastic stamping. *2D Mater.* **1**, 011002 (2014).

45. Wang, G. et al. Introduction of interfacial charges to black phosphorus for a family of planar devices. *Nano Lett.* **16**, 6870–6878 (2016).

Acknowledgements

We thank Y.-Y. Zhang, S. Du, G. Qian and Z. Zhu for helpful discussions, and H. Yang, J. Li, C. Gu and Q. Huan for assistance in device fabrication and measurement. This work was supported by the National Key Research & Development Projects of China (grant nos. 2016YFA0202300 and 2018FYA0305800), National Natural Science Foundation of China (grant no. 61888102), Strategic Priority Research Program of Chinese Academy of Sciences (CAS; grant nos. XDB30000000 and XDB28000000), Youth Innovation Promotion Association of CAS (Y201902) and Beijing Outstanding Young Scientist Program (BJJWZYJH01201914430039). M.O. acknowledges support from ONR (N000141712885) and NSF (DMR1608720). S.J.P. acknowledges support from the Ministry of Education, Singapore, under a Tier 2 grant (no. MOE2017-T2-2-139). A portion of the research was performed in the CAS Key Laboratory of Vacuum Physics.

Author contributions

H.-J.G. supervised the overall research. L.B., M.O. and H.-J.G. designed the experiments. L.W., A.W., J.Y. and L.B. fabricated the devices and carried out the electrical measurements. J.S., S.J.P. and W.Z. performed the STEM measurements. A.W. constructed the home-made electric circuit with ultrashort voltage pulse signals with FWHM of 21 ns. L.W., L.B., A.W., J.S., J.Y., W.Z., M.O., S.T.P. and H.-J.G. analysed the data. L.W., L.B., J.S., W.Z., S.J.P., S.T.P., M.O. and H.-J.G. wrote the paper. All the authors contributed to the preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41565-021-00904-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41565-021-00904-5>.

Correspondence and requests for materials should be addressed to L.B., M.O. or H.-J.G.

Peer review information *Nature Nanotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reference	Channel material	Structure	Write pulse	Extinction ratio
This work	InSe or MoS₂	Floating-gate memory	21 ns	10¹⁰
15	MoS ₂	Floating-gate memory	100 ms	10 ⁴
16	MoS ₂	Floating-gate memory	100 μs	10 ⁴
	graphene	Floating-gate memory	1 ms	2
17	Black phosphorus	Floating-gate memory	300 ms	50
18	MoS ₂	Floating-gate memory	100 ms	10 ⁶
19	silicon	Floating-gate memory (MLG as floating gate)	Not mentioned	Not mentioned

Extended Data Fig. 1 | Floating-gate memory devices based on 2D materials.