The growing importance of applications based on machine learning is driving the need to develop dedicated, energy-efficient electronic hardware. Compared with von Neumann architectures, which have separate processing and storage units, brain-inspired in-memory computing uses the same basic device structure for logic operations and data storage\(^1\text{-}^3\), thus promising to reduce the energy cost of data-centred computing substantially\(^4\). Although there is ample research focused on exploring new device architectures, the engineering of material platforms suitable for such device designs remains a challenge. Two-dimensional materials\(^5\text{-}^6\) such as semiconducting molybdenum disulphide, MoS\(_2\), could be promising candidates for such platforms thanks to their exceptional electrical and mechanical properties\(^7\text{-}^9\). Here we report our exploration of large-area MoS\(_2\) as an active channel material for developing logic-in-memory devices and circuits based on floating-gate field-effect transistors (FGFETs). The conductance of our FGFETs can be precisely and continuously tuned, allowing us to use them as building blocks for reconfigurable logic circuits in which logic operations can be directly performed using the memory elements. After demonstrating a programmable NOR gate, we show that this design can be simply extended to implement more complex programmable logic and a functionally complete set of operations. Our findings highlight the potential of atomically thin semiconductors for the development of next-generation low-power electronics.

Emerging data-intensive applications in fields including machine learning and the Internet of Things require highly energy-efficient hardware for operations such as autonomous driving\(^10\), speech recognition\(^11\) and disease diagnosis\(^12\). Because these specific applications require both high-performance and energy-efficient computation, the power\(^13\) and memory\(^14\) constraints imposed by von Neumann computers, with separate processing and storage units, limit the ability of standard processors to meet optimal requirements for these applications\(^15\). Next-generation architectures have therefore been an important subject of research\(^16\text{-}^20\). Among them, in-memory computing, using the same basic device structure for logic operations and data storage\(^21\), is presenting itself as an ideal hardware architecture for tackling portable data-intensive\(^20\text{-}^21\) and adaptive logic applications\(^22\). The success of this approach depends strongly on identifying an ideal material system capable of harnessing the full potential of this architecture. Two-dimensional (2D) transition metal dichalcogenides (TMDs) have been considered as a candidate material system for realizing scaled semiconducting devices and circuits\(^23\) because of their atom-scale thickness, the absence of dangling bonds and enhanced electrostatic control\(^2\). Monolayer MoS\(_2\), in particular possesses a sizeable direct bandgap\(^24\), enabling a strong modulation of the semiconducting channel with a high ON/OFF current ratio \(I_{\text{on}}/I_{\text{off}} = 10^8\), reduced standby current even at nanometre-scale gate lengths\(^25\) and a subthreshold slope approaching the theoretical limit\(^2\). This makes it an appealing choice for both next-generation logic circuits\(^26\text{-}^28\) and memories in the form of FGFETs\(^29\text{-}^33\), which are attractive devices for in-memory computing. In this context, 2D materials can enable aggressive scaling below 12 nm and at the same time also increase device reliability, thanks to the atom-scale thickness and reduced cell-to-cell interference between neighbouring thin-film floating gates in the FGFETs\(^33\).

2D materials therefore combine advantages for realizing both logic and memory. Their applications in neuromorphic computing are, however, rare and have been limited to single devices\(^33\text{-}^34\). Moreover, overcoming device-to-device variation and large-area integration at the system level remain crucial to the realization of large-scale systems that could open the path to creating new, unexplored circuit functionalities. Here we demonstrate the integration of MoS\(_2\) memories into a subsystem for in-memory computing, and demonstrate reprogrammable logic operations. The basic building block of our circuits are FGFETs with a monolayer MoS\(_2\) channel, allowing us to build simple logic-in-memory arrays\(^2\). Our MoS\(_2\) is grown using a large-grain, large-area metal–organic chemical vapour deposition (MOCVD) process\(^36\text{-}^37\). Figure 1a, b shows the floating-gate memory structure used throughout this work and its side-view schematic. Our device has a local Cr/Pd (2 nm/80 nm) bottom gate and a thin-film Pt floating gate (5 nm thickness), which results in a continuous and smooth surface. The reduced roughness...
source and drain contacts. The floating gate is separated from the MoS2 floating-gate memory device based on MOCVD-grown monolayer MoS2 with a 7-nm-thick, respectively) consist of high-k dielectric HfO2 deposited by atomic-layer deposition (ALD) to achieve effective modulation of the electric field within the semiconducting channel. Finally, the contacts to the drain and source are composed of a Ti/Au (2 nm/100 nm) stack to obtain ohmic-like contacts with high charge carrier injection efficiency. Figure 1c shows an optical micrograph of a fabricated memory array.

We note that all the device components are fabricated in an approach that is scalable, that is, no exfoliated materials were used.

**Floating-gate memory**

The FGFET memory behaviour manifests itself in a shift of the transistor threshold voltage controlled by the amount of charge stored in the charge trap layer (see Supplementary Note 1 for details). To read the memory state of the device, a constant voltage is applied to the gate (\(V_{\text{G,READ}}\)) while the drain–source conductance is measured. We perform the basic characterization of our devices by sweeping the gate voltage in the range ±12.5 V under a constant 50-mV drain–source voltage (\(V_{\text{DS}}\)) Fig. 2a. The total shift of the memory threshold voltage (\(V_{\text{TH}}\)) gives an estimated memory window of 10.6 V, taken for a 1-nA constant current. The linear behaviour of the drain–source current (\(I_{\text{DS}}\)) versus \(V_{\text{DS}}\) traces (Fig. 2b) indicates ohmic-like contacts. The same multilevel behaviour is illustrated in Fig. 2c, in which we show the ability to set the channel conductance with the programming voltage (\(V_{\text{PROG}}\)). Before applying the observed multilevel behaviour of our memory to in-memory computing, we check retention times to verify that the programmed conductance values are stable over time. In Fig. 2c, we show the evolution of the ON and OFF states of our memory as well as multiple intermediate states stable in a 1-h time frame. Fig. 2c. We project a retention time of about 10 years for two-state operation (see Extended Data Fig. 1). Other critical memory characterization concerning device variability and memory behaviour under different constraints is provided in Extended Data Fig. 2 with band alignments shown in Extended Data Fig. 3.

In addition to programming the memory using the programming voltage (\(V_{\text{PROG}}\)), we can also fine-tune the conductance states to the desired level by applying short potentiative (\(V_{\text{G,PEAK}} = 5 \text{ V}\)) and depressive (\(V_{\text{G,PEAK}} = -5 \text{ V}\)) pulses with a 10-ms pulse width and 1-s rest time, allowing a finer control over the device conductance. Figure 2d shows the linear evolution of the conductance values for potentiation (which can be used to rapidly set the desired conductance value) and for depressive stages (to reset the memory state). Results of the endurance test (Supplementary Note 2), shown in Extended Data Fig. 4, demonstrate that our memories can sustain more than 10,000 programming pulses with no performance loss.

**Programmable inverter**

As shown in Fig. 3a, using FGFETs as the basic building blocks instead of normal FETs brings us the capability of programming the threshold voltage, giving an additional degree of freedom for applications in both digital and analogue circuits. The gate terminal can then be used for both setting the state of the memory using a programming voltage \(V_{\text{PROG}}\) and as a terminal for applying the input voltage (\(V_{\text{IN}}\)) during logic operations.

We take advantage of the fine control over the conductance states of the 2D material, and tune the memory cell’s threshold voltage by adding or removing charge carriers from the floating gate. This enables different electron transport regimes to be accessed when the memory is operated in the inverter circuit. We limit the gate voltage during regular operation (\(V_{\text{G}} = V_{\text{IN}}\)) to a range of 0–1 V, corresponding to logic ‘0’ and ‘1’. With this, we can avoid programming currents and preserve the pre-programmed memory state \(Q\). The output voltage \(V_{\text{OUT}}\) and the corresponding logic state are defined by both the logic input and the memory logic state \(X\). The relationship between them is shown in the tables in Fig. 3a. As presented in Fig. 3b–d, we can differentiate between three distinct and discrete states of the memory device according to how efficiently the gate electrode is screened by the charges present in the floating gate. For states \(Q = 1\) and 3, the charges present in the floating gate strongly dope the FGFET channel which remains in the OFF (\(Q = 1\)) or ON (\(Q = 3\)) states for all values of \(V_{\text{IN}}\) in the 0–1 V range. The output then becomes independent of the input and the memory logic
Provision, for $V$ values of the programming voltage, the FGFET in the ON state, after having been programmed using different windows, is estimated to be 10.6 V.

Directions (red arrows). The variation of the threshold voltage $V_{TH}$, the memory window, is tuned to be in the 0–1 V range. Here, the memory cell functions as a normal FET with a programmable threshold voltage.

The programmable shift in the threshold $V_{TH}$ allows us to fine-tune the conductance. Depressive pulses (+5 V amplitude, 10 ms duration and 1 s rest time) can be used to increase and fine-tune the conductance. Potentiative pulses applied to the gate (=5 V amplitude, 10 ms duration and 1 s rest time) can reset the device. The conductance measurement is performed at the end of the rest time period. Before the first pulse, the memory was initialized with a reset voltage $V_{PROG}$ = 12.5 V.

Deceptive pulses (+5 V amplitude, 10 ms duration and 1 s rest time) can be used to increase and fine-tune the conductance. Depressive pulses (+5 V amplitude, 10 ms duration and 1 s rest time) can reset the device. The conductance measurement is performed at the end of the rest time period. Before the first pulse, the memory was initialized with a reset voltage $V_{PROG}$ = 12.5 V.

Logic-in-memory

This multitude of memory states (always ON/always OFF and a normal FET) opens a way to configure memory arrays as a large set of distinct logic circuits. When multiple FGFETs are assembled into a logic gate, the number of possible functions grows exponentially with the number of devices (see Supplementary Note 4). To demonstrate this principle, we show that simple logic gates (two-input NOR and three-input NOR) can be implemented using two or three devices, and can have their functionality expanded up to nine different Boolean functions (see Supplementary Note 4, Supplementary Tables 2, 3 and Extended Data Figs. 6, 7). These new logic operations emerge as subsets of the main function. For instance, a three-input NOR gate contains a two-input NOR and NOT as subset operations. Hence, the circuit area per functionality is greatly decreased as the circuit size grows.

We take advantage of this large number of available logic functions to propose in Fig. 4a a two-input logic-in-memory unit cell capable of acting as a universal logic gate, performing any logic operation from a complete set of two-input logic operations (Extended Data Figs. 8 and Supplementary Note 5). By combining two cells, we can perform more complex operations, such as the addition of two numbers using the number of possible functions grows exponentially with the number of devices (see Supplementary Note 4). To demonstrate this principle, we show that simple logic gates (two-input NOR and three-input NOR) can be implemented using two or three devices, and can have their functionality expanded up to nine different Boolean functions (see Supplementary Note 4, Supplementary Tables 2, 3 and Extended Data Figs. 6, 7). These new logic operations emerge as subsets of the main function. For instance, a three-input NOR gate contains a two-input NOR and NOT as subset operations. Hence, the circuit area per functionality is greatly decreased as the circuit size grows.

We take advantage of this large number of available logic functions to propose in Fig. 4a a two-input logic-in-memory unit cell capable of acting as a universal logic gate, performing any logic operation from a complete set of two-input logic operations (Extended Data Figs. 8 and Supplementary Note 5). By combining two cells, we can perform more complex operations, such as the addition of two numbers using the number of possible functions grows exponentially with the number of devices (see Supplementary Note 4). To demonstrate this principle, we show that simple logic gates (two-input NOR and three-input NOR) can be implemented using two or three devices, and can have their functionality expanded up to nine different Boolean functions (see Supplementary Note 4, Supplementary Tables 2, 3 and Extended Data Figs. 6, 7). These new logic operations emerge as subsets of the main function. For instance, a three-input NOR gate contains a two-input NOR and NOT as subset operations. Hence, the circuit area per functionality is greatly decreased as the circuit size grows.
the CARRY ($C$) value. Because the half-adder is a basic building block in modern processors, this shows that logic-in-memory based on 2D materials could be extended to complex computational accelerators. In contrast to current logic-in-memory circuits (see also Supplementary Note 6 and Supplementary Table 6), our approach allows cascading different cells without the need for complex current–voltage conversion circuits. This eliminates the extra power consumption and enables the creation of more complex circuits similar to modern CMOS digital processors. Logic-in-memory units can be connected in parallel to execute more complex operations, and the signal can be transferred to the next set of units, creating a structure like a field-programmable gate array.

To archive higher parallelism and more complex operation, the number of logic inputs can be further increased. As shown in Fig. 4a, the concept of a three-input cell increases the functionality that can be implemented compared to a two-input structure. As a proof of concept, we show in Fig. 4c a three-input cell operating in one of its possible states, three-input NAND, with the corresponding transfer curves for individual memory elements shown in Extended Data Fig. 9.

We have demonstrated here reprogrammable logic devices for in-memory processing architectures based on monolayer MoS$_2$. By employing an innovative way of realizing a universal logic gate based on logic-in-memory, we have produced a programmable logic circuit that operates directly in memory and does not require additional terminals.
for programming. This direct integration of memory and logic can increase processing speed, opening the way to the realization of energy-efficient circuits based on 2D materials for machine learning, the Internet of Things and non-volatile computing.

**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/s41586-020-2861-0.

Fig. 4 | Logic-in-memory. a, Top left, photograph of a fabricated 12 mm × 12 mm die with logic-in-memory cell arrays. The boxed area contains two-input and three-input logic-in-memory cells, and their schematics are shown below with the output logic function. b, System level operation of two-input cells (‘Input logic’, bottom) to form a half-adder (top). XOR is programmed as Q1 = 2 with inputs to memories Q7 and Q2 inverted. NAND is programmed as Q2 = 2, Q4 = 1, and output is inverted to form the AND logic operation. c, Time traces showing stability of the output voltage for the NAND operation of the three-input unit cell. For this configuration, memories are programmed as follows: Q3 = 2, Q4 = 1. The transfer curves of each state can be seen in Extended Data Fig 9.

1. Kautz, W. H. Cellular logic-in-memory arrays. *IEEE Trans. Comput.* **C-18**, 719–727 (1969).
2. Stone, H. S. A logic-in-memory computer. *IEEE Trans. Comput.* **C-19**, 73–78 (1970).
3. Le Gall, M. et al. Mixed-precision in-memory computing. *Nat. Electron.* **1**, 246–253 (2018).
4. Horowitz, M. Computing’s energy problem (and what we can do about it). In 2014 IEEE Intl Solid-State Circuits Conf. Digest of Technical Papers (ISSCC) 10–14 (IEEE, 2014).
5. Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**, 699–712 (2012).
6. Manzel, S., Ochrimko, D., Pasquier, D., Yazyev, O. V. & Kis, A. 2D transition metal dichalcogenides. *Nat. Rev. Mater.* **2**, 17033 (2017).
7. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer MoS2 transistors. *Nat. Nanotechnol.* **6**, 147–150 (2011).
8. Bertolazzi, S., Brivio, J. & Kis, A. Stacking and strain effects in MoS2 transistors. *ACS Nano* **5**, 9703–9709 (2011).
9. Iannaccone, G., Bonaccorso, F., Colombo, L. & Fiori, G. Quantum engineering of 2D materials heterostructures. *Nat. Rev. Mater.* **3**, 183–191 (2018); correction 15, 520 (2018).
10. Wu, B., Wan, A., Landola, J., Jin, P. H. & Keutzer, K. SqueezeDet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In 2017 IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW) 446–454 (IEEE, 2017).
11. Graves, A., Mohamed, A. & Hinton, G. Speech recognition with deep recurrent neural networks. In 2013 IEEE Intl Conf. Acoustics, Speech and Signal Processing (2013).
12. Kononenko, I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif. Intell. Med. 23, 89–109 (2001).
13. Naffziger, S. High-performance processors in a power-limited world. In 2006 Symp. VLSI Circuits. 2006. Digest of Technical Papers 93–97 (IEEE, 2006).
14. McKee, S. A. Reflections on the memory wall. In Proc. First Conf. Computing Frontiers CF’04 162 (ACM Press, 2004).
15. Xu, X. et al. Scaling for edge inference of deep neural networks. Nat. Electron. 1, 216–222 (2018).
16. Azuete, F. et al. Quantum supremacy using a programmable superconducting processor. Nature 574, 505–510 (2019).
17. Kwon, J. et al. Three-dimensional monolithic integration in flexible printed organic transistors. Nat. Commun. 10, 54 (2019).
18. Shulaker, M. M. et al. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. Nature 547, 74–78 (2017).
19. Yu, S. Neuro-inspired computing with emerging nonvolatile memories. Proc. IEEE 106, 260–285 (2018).
20. Chen, W.-H. et al. CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors. Nat. Electron. 2, 420–428 (2019).
21. Li, C. et al. Analogue signal and image processing with large memristor crossbars. Nat. Electron. 1, 52–59 (2018).
22. Fuller, E. J. et al. Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. Science 364, 570–574 (2019).
23. Fiori, G. et al. Electronics based on two-dimensional materials. Nat. Nanotechnol. 9, 768–779 (2014), erratum 9, 1063 (2014).
24. Splendiani, A. et al. Emerging photoluminescence in monolayer MoS2. Nano Lett. 10, 1271–1275 (2010).
25. Desai, S. B. et al. MoS2 transistors with 1-nanometer gate lengths. Science 354, 99–102 (2016).
26. Raja, A. et al. Dielectric disorder in two-dimensional materials. Nat. Nanotechnol. 14, 832–837 (2019).
27. Papandroulakis, G., Vourkas, I., Vasiliadis, N. & Sirakoulis, G. Ch. Boolean logic operations and computing circuits based on memristors. IEEE Trans. Circuits Syst. I 161, 972–976 (2014).
28. Resta, G. V. et al. Polarity control in WSe2 double-gate transistors. Sci. Rep. 6, 29448 (2016).
29. Resta, G. V. et al. Doping-free complementary logic gates enabled by two-dimensional polarity-controllable transistors. ACS Nano 12, 7039–7047 (2018).

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 2020
**Article**

**Methods**

**Material synthesis**

Single-crystal monolayer MoS2 was grown in a home-built system using the metal–organic chemical vapour deposition (MOCVD) method. C-plane sapphire was used as the growth substrate and annealed at 1,000 °C to achieve an atomically smooth surface, necessary for epitaxial growth. Before growth, the substrate is spin-coated with NaCl solution to suppress nucleation and promote growth. The two precursors (molybdenum hexacarbonyl, Mo(CO)6, and hydrogen sulfide, H2S), with a flow ratio of 1:6,028, were carried by Ar gas to the MOCVD chamber and underwent reaction at 820 °C for 30 min. Mo(CO)6 was kept at 15 °C in a water bath and the valve was closed immediately after the growth process, while H2S continued flowing during the cooling process. Throughout the whole growth process, the furnace was kept at 850 mbar pressure. Raman spectroscopy confirmed the monolayer nature of the grown material (Extended Data Fig. 10), and transmission electron microscopy (TEM) imaging also indicated the high quality of the material (Extended Data Fig. 11).

Continuous, 2-inch wafer-scale monolayer MoS2 film for complex circuits shown in Fig. 4 was synthesized using MOCVD. Similarly to the synthesis of single crystals, we anneal the sapphire wafers in air and coat them with a 0.2 mol l⁻¹ sodium chloride (NaCl) solution in deionized water. The growth process lasts for 30 min in a quartz tube at atmospheric pressure and a temperature of 870 °C. We use Mo(CO)6 and diethyl sulphide ((C2H5)2S) as precursors. An argon/hydrogen mixture is used as a carrier gas, delivered with flow rates of 210 cm³ STP min⁻¹/4 cm³ STP min⁻¹. Oxygen with a flow rate of 1 cm³ STP min⁻¹ is separately introduced into the growth chamber, for the purpose of balancing the growth rate with the O2 etching effect.

**Sample transfer and TEM imaging**

The sample was spin-coated with PMMA (poly(methyl methacrylate)) A2 at a speed of 4,000 rpm for 1 min and put on a hot plate at 85 °C for 10 min for drying. Afterwards, the PMMA film was detached with water tension and floated on the water’s surface together with the MoS2 sample. Subsequently, the PMMA film was fished out using a TEM grid and annealed at 15 °C in a water bath and the valve was closed immediately after the growth process, while H2S continued flowing during the cooling process. Throughout the whole growth process, the furnace was kept at 850 mbar pressure. Raman spectroscopy confirmed the monolayer nature of the grown material (Extended Data Fig. 10), and transmission electron microscopy (TEM) imaging also indicated the high quality of the material (Extended Data Fig. 11).

**Transfer procedure**

The MOCVD-grown material is first spin-coated with PMMA A2 at 1,500 rpm for 60 s. It is then dried in vacuum for 12 h. After that, with the support of PDMS (polydimethylsiloxane) and Gel-pak elastomer film, the MoS2 sample is detached from the sapphire in deionized water and transferred onto the patterned substrate. Finally, the sample is immersed in acetone, and subsequently annealed at 250 °C in high vacuum to remove the polymer resist.

**Memory fabrication**

A 270-nm thick SiO2 layer is thermally grown using a dry plasma-enhanced chemical vapour deposition (PECVD) technique on a p-doped silicon wafer. The bottom gate contacts were patterned using e-beam lithography (EBL) and a 2 nm/80 nm Cr/Pd stack was deposited using e-beam evaporation. The 30-nm HfO2 blocking oxide was grown by thermal atomic layer deposition (ALD) using TEMAH (tetramethylammonium hydroxide) and water as precursors. The floating gate was patterned similarly by EBL and a 5-nm thick Pt layer was deposited using e-beam evaporation. Using the same process as described earlier, a 7-nm thick HfO2 tunnel barrier was grown. The MoS2 is transferred on top of the tunnel barrier. To define the active region, PMMA polymer was used and patterned by EBL. The exposed area was then etched by oxygen plasma. Finally, drain-source contacts were patterned by EBL and a 2 nm/100 nm thick Ti/Au stack was deposited using e-beam evaporation. Each die has 8 devices with a density of 0.386 devices per 10 µm². Resulting FGFETs have a typical channel length of 1 µm and a width of 7.5 µm. A cross-sectional TEM image of a representative device is shown in Extended Data Fig. 12.

**Logic-in-memory fabrication**

A 270-nm thick SiO2 layer is thermally grown using a dry PECVD technique on a p-doped silicon wafer. The bottom gate contacts were patterned using an MLA150 Advanced Maskless Aligner and a 2 nm/40 nm Cr/Pt stack was deposited using e-beam evaporation. The 30-nm HfO2 blocking oxide was grown by thermal ALD using TEMAH and water as precursors. The floating gate was patterned by EBL and a 5-nm thick Pt layer was deposited using e-beam evaporation. Using the same process as described earlier, a 7-nm thick HfO2 tunnel barrier was grown. Prepatterned pads were exposed using the Advanced Maskless Aligner and 2 nm/60 nm Ti-Au were deposited using e-beam evaporation. The MoS2 continuous film was transferred on top of the tunnel barrier. To define the active region, PMMA polymer was used and patterned by EBL. The exposed area was then etched by oxygen plasma. Finally, drain-source contacts were patterned by EBL and a 2 nm/100 nm thick Ti/Au stack was deposited using e-beam evaporation. Resulting FGFETs have a typical channel length of 1 µm and a width of 12.5 µm.

**Memory characterization**

Memory characterization is performed in high vacuum after in situ annealing at 120 °C. I–V curve acquisition and pulse programming are performed using an Agilent E5270B mainframe with E5287A-ATO and E5281B-FG modules. A 10 pF load capacitor is used for simulating the input capacitance of a cascade of logical stages in both FGFET inverter and FGFET NOR time measurements.

**Logic-in-memory**

The logic measurements of a two-input and a three-input unit cell, and a three-input NOR, were performed in air in a custom-built programmer using NI ELVIS II Board I/O. A detailed description of the programmer is given in Supplementary Note 9.

**Data availability**

The data that support the findings of this study are available at http://doi.org/10.5281/zenodo.4073060.

42. Dumcenco, D. et al. Large-area epitaxial monolayer MoS2, ACS Nano 9, 4611–4620 (2015).
research and innovation programme under grant agreements 829035 (QUEFORMAL), 785219 and 881603 (Graphene Flagship Core 2 and Core 3), from the Marie Curie-Sklodowska COFUND (665667), from the H2020 European Research Council (ERC, grant 682332) as well as from the CCMX Materials Challenge grant ‘Large area growth of 2D materials for device integration’. Device preparation was carried out in part in the EPFL Centre of MicroNanotechnology (CMI). TEM imaging was carried out in the EPFL Interdisciplinary Centre for Electron Microscopy (CIME).

**Author contributions** A.K. initiated and supervised the work. G.M.M. performed the device fabrication with initial assistance of Y.Z. G.M.M. constructed the characterization setup and performed electrical measurements. Y.Z. prepared the MOCVD grown MoS2 monolayers. Z.W. performed Raman spectroscopy and growth of wafer-scale films, supervised by A.R. M.T. performed HRTEM measurements and simulations. G.M.M., A.A. and A.K. analysed the data and wrote the manuscript with input from all authors.

**Competing interests** The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-020-2861-0.

**Correspondence and requests for materials** should be addressed to A.K.

**Peer review information** Nature thanks Devin Verreck, Takhee Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at http://www.nature.com/reprints.
Extended Data Fig. 1 | Two-state retention time. The drain–source conductance $G_{DS}$ is shown versus time. Blue curve, $V_{\text{PROG}} = -12.5\,\text{V}$; red curve, $V_{\text{PROG}} = +12.5\,\text{V}$. To predict the trend of the decay, we fit both curves using the following expression, $f(x) = Ax^k$ (dashed black lines). We expect that the device has a 10-year retention.
**Extended Data Fig. 2 | Additional characteristics of MoS₂ FGFETs.**

**a**, Device variability. $I_{DS}$ versus $V_G$ curves for six different devices on the same die.

**b**, Fresh $I_{DS}$ versus $V_G$ curves, corresponding to the first $V_G$ sweep carried out on these devices. Maximal gate voltage $\pm V_{G,\text{MAX}}$ (corresponding to $V_{\text{PROG}}$) is insufficient for inducing charge transfer into the floating gate memory. This shows the behaviour of the FGFET in the initial state.

**c**, $I_{DS}$ versus $V_G$ for different values of $V_{DS}$ (red curve, 50 mV; blue curve, 100 mV; green curve, 250 mV; orange curve, 500 mV). The progressive increase of the current without a decrease in the memory window demonstrates that the memory effect is not due to capacitive charges in the contacts.

**d**, $I_{DS}$ versus $V_G$ for different sweep rates. The decrease of the memory window is a function of the sweep rate. The decrease is most probably a result of charge dynamics limiting the charging and discharging of the floating gate.
Extended Data Fig. 3 | Simplified band diagrams of MoS₂ FGFETs. 

a, Energy band diagrams of different materials comprising the FGFET before being brought into contact. $E_C$ and $E_V$ are the positions of the bottom of the conduction band and the top of the valence band, respectively.

b, Programming of the floating-gate memory by electron injection into the floating gate with the application of a positive gate voltage (upper panel). Lower panel, accompanying positive shift in the threshold voltage.

c, Erase operation with electron extraction from the floating gate under the application of a negative gate voltage (upper panel). Lower panel, accompanying negative shift in the threshold voltage.
Extended Data Fig. 4 | Floating-gate endurance test. $I_{DS}$ is shown as a function of the number of program/erase (P/E) cycles. **a**, Each P/E cycle consists of a 100-ms $+7.5$ V pulse for the erase operation, and a 100-ms $-7.5$ V pulse for the program operation. **b**, As **a** but with a $+10.0$ V pulse for erasing and a $-10.0$ V pulse for programming. Both measurements are taken using a constant $V_{DS} = 50$ mV and on the same device.
Extended Data Fig. 5 | Example of the graphical estimation of the noise margin for the inverter programmed with $V_{\text{prog}} = 8.5$ V. Output voltage $V_{\text{OUT}}$ as a function of input voltage $V_{\text{IN}}$ (blue dots), and its mirror reflection (red dots). $V_{\text{OH}}$ and $V_{\text{OL}}$ are defined as the points where the slope of the transfer curve ($V_{\text{OUT}}$ as a function of $V_{\text{IN}}$, blue dots) is equal to $-1$, whereas $V_{\text{IL}}$ and $V_{\text{IH}}$ are the corresponding values of $V_{\text{IN}}$. 

$0.10 V_{\text{IL}}$ 

$0.12 V_{\text{IL}}$ 

$0.31 V_{\text{OL}}$ 

$0.41 V_{\text{OL}}$ 

$0.77 V_{\text{OH}}$ 

$0.89 V_{\text{OH}}$ 

$V_{\text{IL}}$ 

$V_{\text{IH}}$ 

$dV_{\text{OUT}}/dV_{\text{IN}} = -1$
Extended Data Fig. 6 | Circuit schematic and logic for a two-input NOR.
a. Circuit schematic for a two-input NOR. b. Logic over time for different programming states $Q_1, Q_2$. Red: time traces of input voltages $V_{IN1}$ and $V_{IN2}$. Orange: output curves for $Q_{1,2} = 33$, constant LOW and $Q_{1,2} = 11$, constant HIGH. Blue: output curves for $Q_{1,2} = 21$, inverter A (IN1); $Q_{1,2} = 12$, inverter B (IN2). Green: output curve for $Q_{1,2} = 22$, NOR A, B. (Here and in Extended Data Figs. 7, 8, we denote (for example) programming state ‘$Q_1 = 3$, $Q_2 = 3$’ by ‘$Q_{1,2} = 33$’.)
Extended Data Fig. 7 | Three-input NOR. a, Circuit schematic for a three-input NOR. b, Logic over time for different programming states $Q_1$, $Q_2$, $Q_3$, $Q_{1-3} = 111$, constant HIGH; $Q_{1-3} = 211$, inverter A (IN1); $Q_{1-3} = 112$, inverter C (IN2); $Q_{1-3} = 122$, NOR B; $Q_{1-3} = 212$, NOR A; $Q_{1-3} = 221$, NOR A, B; $Q_{1-3} = 222$, NOR A, B, C.
Extended Data Fig. 8 | Two-input logic-in-memory concept and interpretation. a, Two-input schematic of the logic-in-memory concept. b, Interface model for input polarity control. c, NAND gate, $Q_{1-4} = 2211$; d, NOR gate, $Q_{1-4} = 2332$; e, XOR gate, $Q_{1-4} = 2222$. We derive the XOR canonical form by applying De Morgan’s laws.
Extended Data Fig. 9 | Hardware and software implementation of the logic-in-memory programmer. a, b, Hardware implementation of the four-memory programmer (a) and of the nine-memory programmer (b). c, Software working diagram of the programming (top) and test (bottom) blocks. d, Example of programming (left) and test (right) blocks working, using a nine-memory programmed into the following state $Q_{1-9} = 222111111$ to perform a three-input NAND operation.
Extended Data Fig. 10 | Raman characterization of monolayer MoS$_2$. Raman spectrum of transferred MoS$_2$ from a single crystal (which also provided the material used in this paper) using 532-nm laser excitation and a 3,000 lines mm$^{-1}$ grating. The observed wavenumber difference between the $A_{1g}$ and $E_{2g}$ Raman modes of MoS$_2$ is consistent with a monolayer. Black line is a fit to the data points (red circles).
Extended Data Fig. 11 | ADF-STEM images of monolayer MoS$_2$. 

**a**, Atomically resolved STEM image showing a large region of monolayer MoS$_2$. Inset, fast Fourier transform (FFT) amplitude spectrum further shows the crystalline monolayer MoS$_2$ structure. 

**b**, A magnified filtered STEM image taken from **a** shows the 2H crystal structure of monolayer MoS$_2$. 

**c**, STEM simulation image of monolayer MoS$_2$. The intensity line profiles at bottom right of **b** and **c** are taken along the dashed lines in those images, and show the peak positions of Mo atoms and S atoms.
Extended Data Fig. 12 | FGFET TEM cross-section. a, Wide-field view of the device fabricated using the logic-in-memory process. b, Magnified view of the contact area boxed in a. c, Cross-section image of the gate stack consisting of (from bottom to top) Pt bottom gate, HfO₂ blocking oxide, Pt floating gate, HfO₂ tunnel oxide. The MoS₂ 2D channel is on top of the gate stack.