Published in final edited form as:
Nature. 2020 November 01; 587(7832): 72–77. doi:10.1038/s41586-020-2861-0.

Logic-in-Memory Based on an Atomically Thin Semiconductor

Guilherme Migliato Marega[1,2], **Yanfei Zhao**[1,2], **Ahmet Avsar**[1,2], **Zhenyu Wang**[1,2], **Mukesh Tripathi**[1,2], **Aleksandra Radenovic**[3], **Andras Kis**[1,2,*]

[1]Electrical Engineering Institute, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland [2]Institute of Materials Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland [3]Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

Abstract

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, brain-inspired in-memory computing uses the same basic device structure for logic operations and data storage[1–3], thus promising to reduce the energy cost of data-centric computing significantly[4]. While 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[5,6] such as semiconducting MoS2 could stand out as a promising candidate to face this obstacle thanks to their exceptional electrical and mechanical properties[7–9]. Here, we explore large-area grown MoS2 as an active channel material for developing logic-in-memory devices and circuits based on floating-gate field-effect transistors (FGFET). The conductance of our FGFETs can be precisely and continuously tuned, allowing us to use them as building blocks for reconfigurable logic circuits where 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 functionally complete sets of functions. Our findings highlight the potential of atomically thin semiconductors for the development of next-generation low-power electronics.

*Correspondence should be addressed to: Andras Kis, andras.kis@epfl.ch.

Contributions

AK initiated and supervised the work. GM performed the device fabrication with initial assistance of YZ. GM constructed the characterization setup and performed electrical measurements. YZ prepared the MOCVD grown MoS_2 monolayers. ZW performed Raman spectroscopy and growth of wafer-scale films, supervised by AR. MT performed HRTEM measurements and simulations. GM, AA and AK analyzed the data and wrote the manuscript with input of all authors.

Competing interests

The authors declare no competing financial interests.

Corresponding author

Correspondence and requests for materials should be addressed to andras.kis@epfl.ch.

Europe PMC Funders Author Manuscripts

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]. Since 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 standard processors from meeting optimal requirements for these applications[15]. Therefore, in the search of the most efficient solution, next-generation architectures have been an important subject of research[16–19]. Among them, in-memory computing, using the same basic device structure for logic operations and data storage[1,2], is presenting itself as an ideal hardware architecture for tackling portable data-intensive[20,21] and adaptative logic applications[22]. The success of this approach strongly depends 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 an appealing candidate material system for realizing scaled semiconducting devices and circuits[23] because of their atomic scale thickness, the absence of dangling bonds and enhanced electrostatic control[7]. Monolayer molybdenum disulphide (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 ($I_{ON}/I_{OFF} \sim 10^8$), reduced standby current even at nanometre-scale gate lengths[25] and a very low subthreshold slope (SS) (64 mV/dec) approaching the theoretical limit[7]. This makes it an appealing choice for both next-generation logic circuits[26,27] and memories in the form of floating-gate field-effect transistors (FGFET)[28–31], which is an attractive device for in-memory computing. Here, 2D materials can enable aggressive scaling beyond 12 nm and in the same time also increase device reliability thanks to the atomic scale thickness as well as reduced cell-to-cell interference between neighbouring thin film floating gates in FGFETs[32].

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[33–35]. Moreover, overcoming device-to-device variation and large-area integration at the system level remain crucial to realize large scale systems which could open the path for creating new, unexplored circuit functionalities.

Here, we show 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 floating-gate transistors (FGFETs) with a monolayer MoS_2 channel, allowing us to build simple logic-in-memory arrays[1]. Our MoS_2 is grown using a large-grain, large-area metal-organic chemical vapor deposition (MOCVD) process[36,37]. Figure 1a and b show 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 from the metal surface decreases the dielectric disorder in the interface between the top tunnel oxide and the 2D channel, improving performance and reliability[38]. Both blocking and tunnel oxides (30 nm and 7 nm thick, respectively) consist of ALD-deposited high-κ dielectric HfO_2 to achieve an effective modulation of the electric field within the semiconducting channel. Finally, the drain-source contacts are composed of a

Europe PMC Funders Author Manuscripts

Ti/Au (2 nm/100 nm) stack to obtain ohmic-like contacts with high charge carrier injection efficiency. Figure 1c shows the optical micrograph of a fabricated memory array. We note that all the device components are fabricated in an approach that is scalable, i.e. 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). For reading the memory state of the device, a constant voltage is applied to the gate ($V_{G,READ}$) while the drain-source conductance is measured. First, we perform the basic characterisation of our devices by sweeping the gate voltage in the ±12.5 V range under a constant 50 mV drain-source voltage (V_{DS}), Figure 2a. The total shift of the memory threshold voltage (V_{TH}) gives an estimated memory window of 10.6 V, taken for a 1 nA constant current. The linear behaviour of I_{DS} vs. V_{DS} traces (Figure 2b) indicates ohmic-like contacts. The same multilevel behaviour is illustrated in Figure 2c, in which we show the ability to set the channel conductance with the programming voltage (V_{PROG}). Prior to applying the observed multilevel behaviour of our memory to in-memory computing, we check retention times to verify if the programmed conductance values are stable over time. In Figure 2c, we show the evolution of the ON and OFF states of our memory as well as multiple intermediate states stable in a 1-hour time frame, Figure 2c. We project a ~10-year retention time for two-state operation (See Extended Data Figure 1). Other critical memory characterization concerning device variability and memory behaviour under different constraints is provided in Extended Data Figure 2 with band alignments shown on Extended Data Figure 3. In addition to programming the memory using the programming voltage (V_{PROG}), we can also fine-tune the conductance states to the desired level by applying short potentiative ($V_{G,PEAK}$ = -5 V) and depressive ($V_{G,PEAK}$ = 5 V) pulses with a 10 ms pulse width and 1 second 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 depressive stages for resetting the memory state. Results of the endurance test (Supplementary Note 2), shown on Extended Data Figure 4, demonstrate that our memories can sustain more than 10000 programming pulses with no performance loss.

Programmable Inverter

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

We take advantage of the fine control over the 2D material conductance states and tune the memory cell's threshold voltage by adding or removing charge carriers from the floating gate. This creates different electron transport regimes of the memory's operation in the inverter circuit. We limit the gate voltage during regular operation ($V_G = V_{IN}$) to a range

between 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_{OUT} and the corresponding logic state are defined by both the logic input and memory logic state $X^{(Q)}$. The relationship between them is shown in the tables in Figure 3a. As presented in Figure 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_{IN} in the 0−1 V range. The output then becomes independent of the input and the memory logic states are 0 (for $Q = 1$) or 1 (for $Q = 3$). For $Q = 2$, the amount of charge stored on the floating gate is insufficient for inhibiting the channel modulation and V_{TH} is tuned to be in the 0−1 V range. Here, the memory cell functions as a regular FET with a programmable threshold voltage. In this case, the memory state reflects the input logic state ($X^{(2)}$ = IN) and the circuit operates as an inverter.

The programmable shift in the threshold V_{TH} allows us to fine-tune the transfer curve of the inverter circuit, shown in Figure 3e for programming voltages in the 7.5 − 9 V range. The gain and noise margin of the circuit can also be configured for either a more precise (lower noise margin) or a more robust circuit (higher noise margin). Figure 3f show the evolution of both HIGH (NM_H) and LOW (NM_L) noise margin as a function of the programming voltage (V_{PROG}) (see Supplementary Note 3, Supplementary Table 1 and Extended Data Figure 5 for details). Time traces displayed on Figure 3g show that the different configurations of the circuit are stable and reproducible.

Logic-in-memory

This multitude of memory states (always ON/always OFF and a regular FET) opens a path for configuring 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 (2-input NOR and 3-input NOR) can be implemented using 2 or 3 devices and can have their functionality expanded up to 9 different Boolean functions, see Supplementary Note 4, Supplementary Tables 2 and 3 and Extended Data Figures 6 and 7. These new logic operations emerge as subsets of the main function. For instance, a 3-input NOR gate contains a 2-input NOR and NOT as subset operations. Hence, the footprint per functionality is greatly decreased as the circuit size grows.

We take advantage of this large number of available logic functions to propose in Figure 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 Figure 8 and Supplementary Note 5.). By combining two cells, we can perform more complex operations, such as the addition of two numbers using a half-adder, shown in Figure 4b (see Supplementary tables 4 and 5 for functions breakdown). This is made possible by adding polarity control in the input and output of the cell. With this new degree of freedom, one unit operates as an XOR logic gate producing as a result the binary SUM (S) and the second unit producing the logic NAND which after inversion by the output interface generates the

Europe PMC Funders Author Manuscripts

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[39] (see also Supplementary Note 6 and Supplementary Table 6), our approach allows cascading of 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 unit can be connected in parallel to execute more complex operations and the signal can be transferred to the next set of units, creating a field-programmable gate array-like structure.

To archive higher parallelism and more complex operation, the number of logic inputs can be further increased. As shown in Figure 4a, the concept of a 3-input cell increases the functionality that can be implemented compared to a 2-input structure. As a proof of concept, we show in Figure 4c a 3-input cell operating in one of its possible states, 3-input NAND, with the corresponding transfer curves for individual memory elements shown in Extended Data Figure 9.

In summary, we demonstrate 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 present a programmable logic circuit which operates directly in memory and does not require additional terminals for programming[40,41]. This direct integration of memory and logic can increase processing speed, opening the way to the realisation of energy-efficient circuits based on 2D materials for machine learning, the internet of things and non-volatile computing.

Methods

Material Synthesis

Single-crystal monolayer MoS_2 is grown in a home-built system using metal organic chemical vapor deposition (MOCVD) method. C-plane sapphire is used as the growth substrate and annealed at 1000 °C to achieve an atomically smooth surface, necessary for epitaxial growth[42]. Before growth, the substrate is spin coated with NaCl solution to suppress nucleation and promote the growth[36,37]. The two precursors $Mo(CO)_6$ and H_2S, with the flow rate ratio of 1:6028, are carried by Ar gas to the MOCVD chamber and undergo reaction at 820 °C for 30 min. $Mo(CO)_6$ is kept at 15 °C in a water bath and the valve is closed immediately after growth process, while H_2S continues flowing during the cooling process. Throughout the whole growth process, the furnace is kept at 850 mbar pressure. Raman spectroscopy confirms the monolayer nature of the grown material, Extended Data Figure 10, while TEM imaging also indicates the high quality of the material, Extended Data Figure 11.

Continuous, 2” wafer-scale monolayer MoS_2 film for complex circuits shown on Figure 4 was synthesized using metal-organic chemical vapor deposition method[37]. Similarly to the synthesis of single crystals, we anneal the sapphire wafers in air and coat them with a 0.2 $mol \cdot L^{-1}$ sodium chloride (NaCl) solution in de-ionized water. The growth process lasts for 30 min in a quartz tube at atmospheric pressure and a temperature of 870 °C. We use

molybdenum hexacarbonyl ($Mo(CO)_6$) and diethyl sulfide ($(C_2H_5)_2S$) as precursors. Argon/hydrogen mixture is used as a carrier gas, delivered with 210 sccm/4sccm flow rates respectively. Oxygen with a 1 sccm flow rate is separately introduced in the growth chamber, for the purpose of balancing the growth rate with the O_2 etching effect.

Sample transfer & TEM imaging

The sample was spinning-coated with PMMA A2 at a speed of 4000 r/min for 1 min and put on the hot plate at 85 °C for 10 min for drying. Afterward, PMMA film was detached with water tension and floated on the surface together with MoS_2 sample. Subsequently, PMMA film was fished by TEM grid and heated for 15 min on hot plate at 85 °C to improve the adhesion. To remove PMMA film, the sample was immersed with Acetone overnight and annealed at 250 °C in high vacuum for 6 h.

Atomic-resolution annular dark field scanning transmission electron microscopy (ADF-STEM) images were acquired with an aberration-corrected (with double Cs corrector) FEI Titan Themis TEM 60-300 kV, equipped with Schottky X-FEG electron source and a monochromator. Imaging was performed at a low acceleration voltage (80 kV). The electron probe semi convergence angle was set to 21.2 mrad and the typical beam current was 18 pA. Images were acquired with Gatan high annular angular dark field (HAADF) detector using 185 mm camera length which corresponds to a 49.5-198 mrad collection angle. To reduce the sample drift distortion, a short dwell time (8μs) with 512 × 512 pixels was used to capture the frames. Cross-section lamella was prepared by focused ion beam (Zeiss NVision 40). Transmission electron microscopy (TEM) cross-sectional imaging was performed with a FEI Talos F200S G2, using 80 kV acceleration voltage. Multislice STEM image simulations were performed using quantitative scanning transmission electron microscopy (QSTEM). The simulation parameters were chosen similar to the experimental conditions and higher order aberrations were reduce to zero.

Transfer Procedure

The MOCVD grown material is first spin-coated with PMMA A2 at 1500 rpm for 60 s. It is then dried in a vacuum atmosphere for 12 hours to dry the polymer material. After that, with the support of PDMS and Gelpak, the MoS_2 sample is detached from sapphire in DI water and transferred onto the patterned substrate. Finally, the sample is immersed in acetone and annealed at 250 °C in high vacuum to remove the polymer resist.

Memory Fabrication

First, 270 nm of SiO_2 is thermally grown using dry 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 HfO_2 blocking oxide was grown by thermal atomic layer deposition (ALD) using TMAH 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 HfO_2 tunnel barrier was grown. The MoS_2 is transferred on top of the tunnel barrier. For defining 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

Europe PMC Funders Author Manuscripts

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 density of 0.386 devices per 10 μm^2. Resulting FGFETs have a typical channel length of 1 μm and a width of 7.5 μm. A cross-sectional image TEM image of a representative device is shown on Extended Data Figure 12.

Logic-In-Memory Fabrication

First, 270 nm of SiO_2 is thermally grown using dry PECVD technique on a p-doped silicon wafer. The bottom gate contacts were patterned using the MLA150 Advanced Maskless Aligner and a 2 nm/40 nm Cr/Pt stack was deposited using e-beam evaporation. The 30 nm HfO_2 blocking oxide was grown by thermal atomic layer deposition (ALD) using TMAH 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 HfO_2 tunnel barrier was grown. Prepatterned pads are exposed by the Advanced Maskless Aligner and 2 nm/ 60nm Ti-Au are deposited by e-beam evaporation. The MoS_2 continuous film is transferred on top of the tunnel barrier. For defining 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 Characterisation

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 10pF load capacitor is used for simulating the input capacitance of a cascade of logical stages in both FGFET inverter and FGFET NOR time measurements.

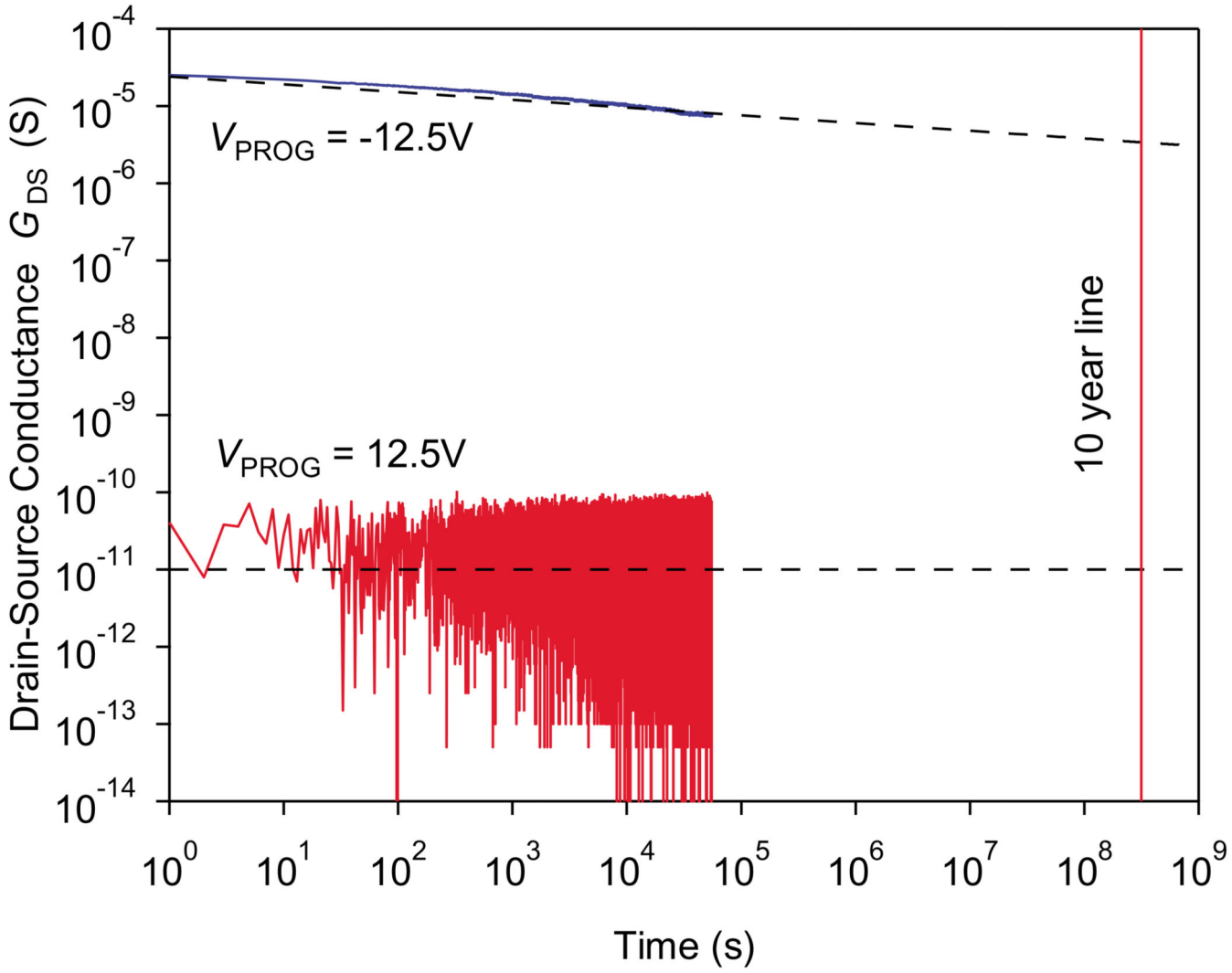
Logic-In-Memory

The logic measurements for a 2-input and 3-input unit cell and 3-input NOR where performed in air in a custom build programmer using NI ELVIS II Board I/O. More detailed explanation about the programmer can be seen in the Supplementary Note 9.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

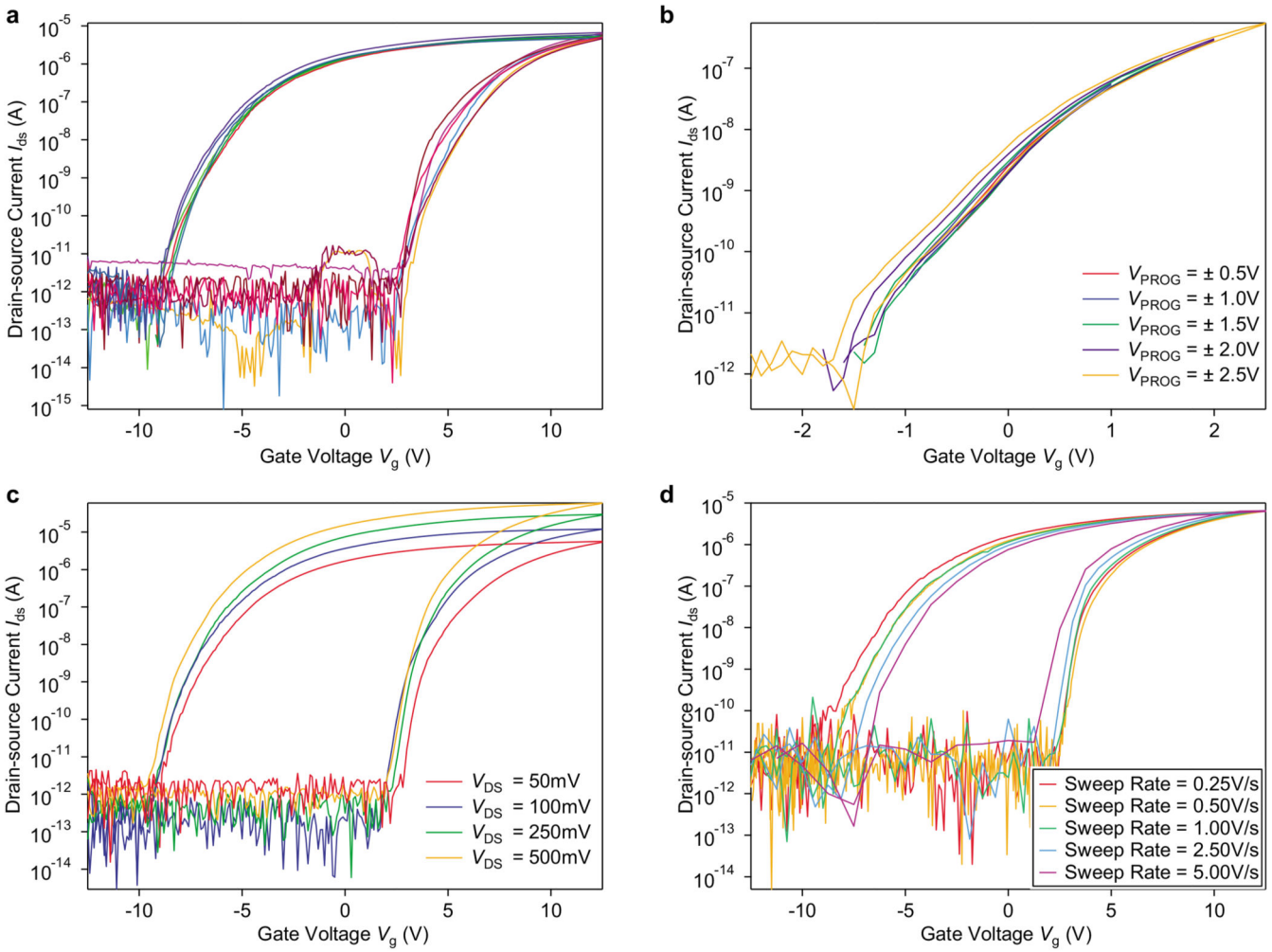
Extended Data



Extended Data Figure 1.

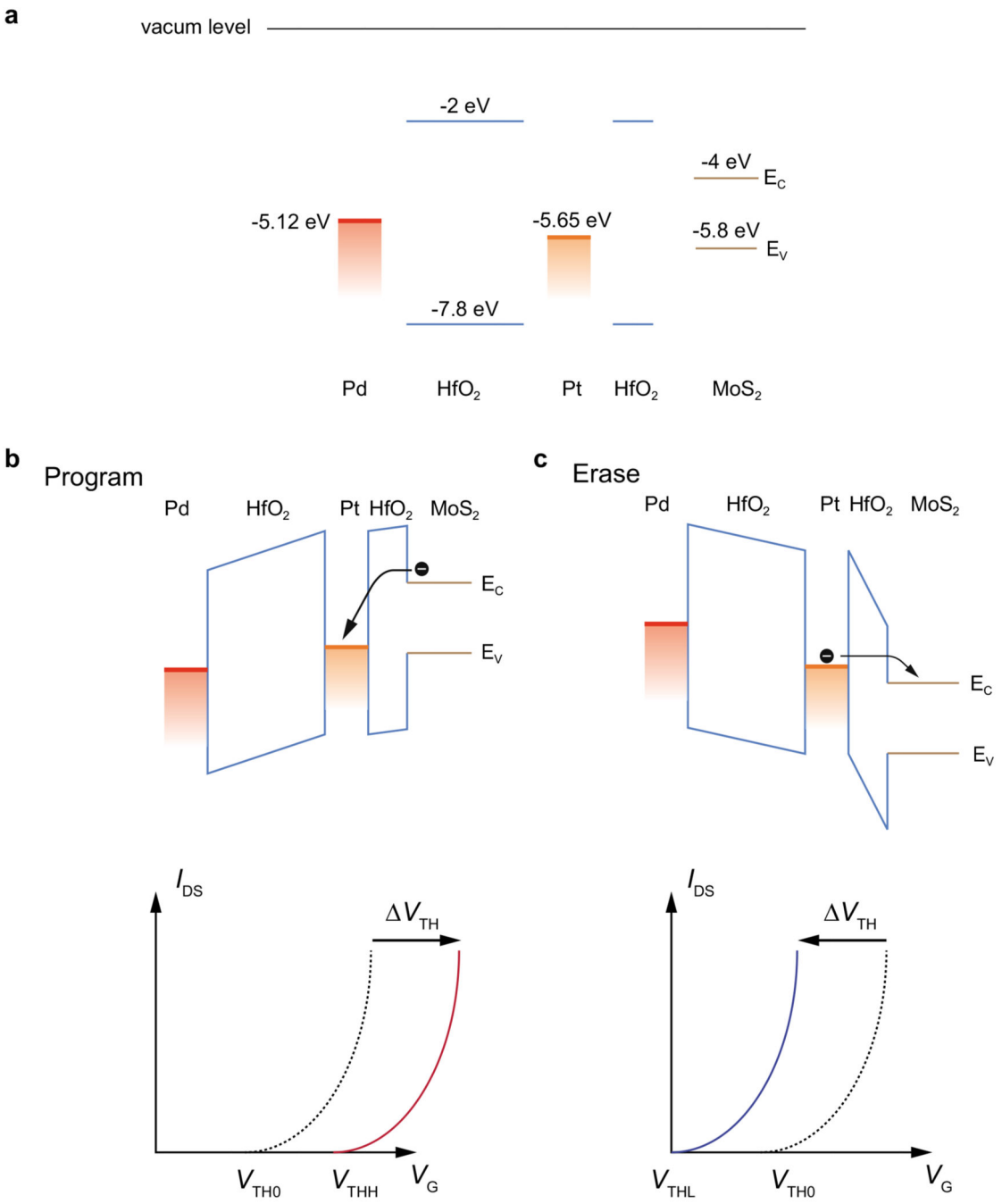
Two-State retention time. Conductance G_{DS} versus time. Blue curve – V_{PROG} = -12.5V. Red curve - V_{PROG} = +12.5V. Fitting of curve for predicting the trend of the decay using the following expression $f(x) = A \cdot x^{k}$. We expect that the device has a 10-year retention.

Europe PMC Funders Author Manuscripts



Extended Data Figure 2.

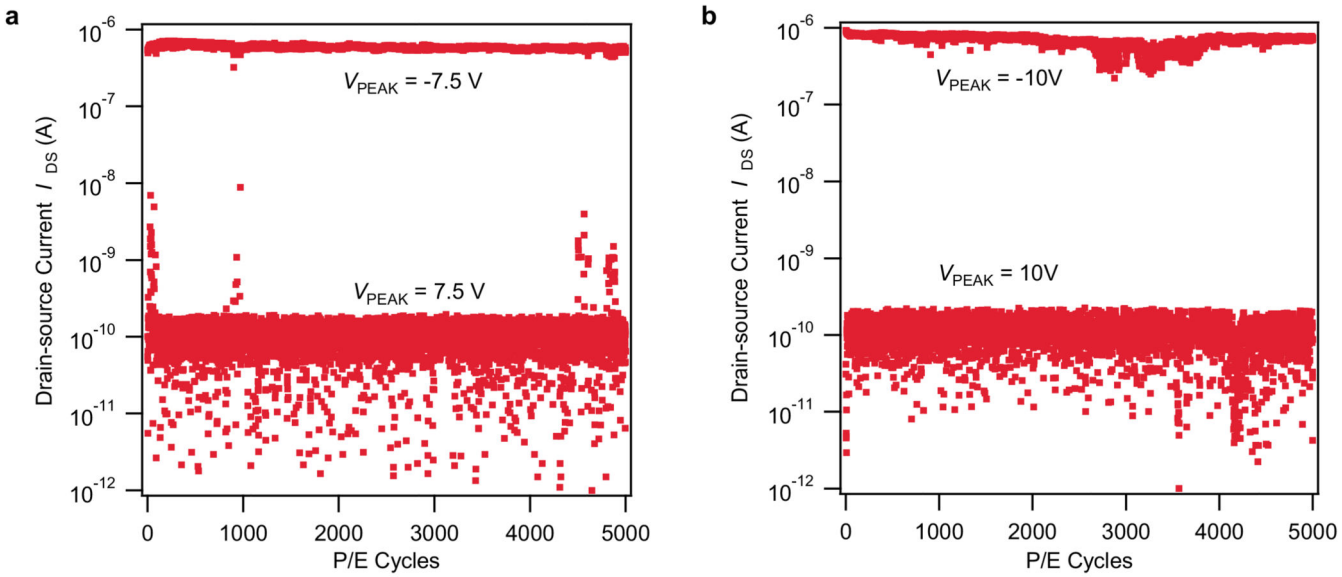
Additional characteristics of MoS_2 FGFETs. **a**, Device variability. I_{DS} versus V_G curves for 6 different devices on the same die. **b**, Fresh I_{DS} versus V_G. Curves with small $\pm V_{G,MAX}$ which is insufficient for inducing charge transfer into the floating gate memory. This shows the behavior of the FGFET in the initial state. **c**, I_{DS} versus V_G for different V_{DS}. Red curve - V_{DS} = 50mV. Blue curve - V_{DS} = 100mV. Green curve - V_{DS} = 250mV. Orange curve - V_{DS} = 500mV. The progressive increase of the current without decreasing 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 on the memory window in function of the sweep-rate. The decrease is most likely a result of a limit charge dynamics for charging and discharging the floating-gate.



Extended Data Figure 3.

Simplified band diagrams of MoS_2 FGFETs. **a**, Energy band diagrams of different materials comprising the FGFET before being brought into contact. **b**, Programming of the floating gate memory with electron injection into the floating-gate (application of positive gate voltage and positive shift in the threshold voltage). **c**, Erase operation with electron extraction from the floating-gate (application of negative gate voltage and negative shift in the threshold voltage).

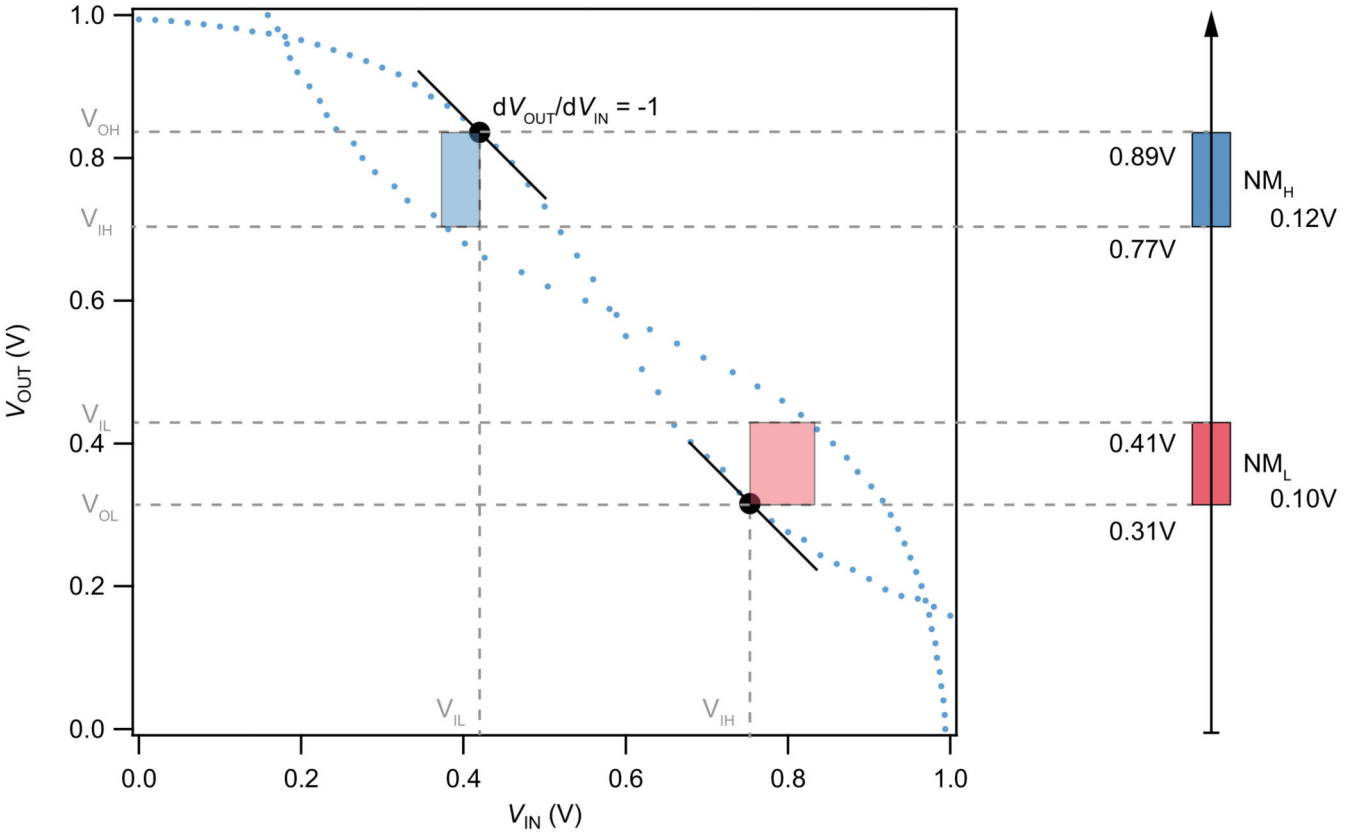
Europe PMC Funders Author Manuscripts



Extended Data Figure 4.

Floating gate endurance test. a, Each P/E Cycle is constituted of 100 ms positive 7.5 V pulse for the Erase operation and 100 ms negative -7.5V pulses for Programming. b, Each P/E Cycle is comprised of a 100 ms positive 10.0 V pulse for erasing and a 100 ms negative -10.0 V pulse for programming. Both measurements are taken using a constant V_{DS} = 50 mV.

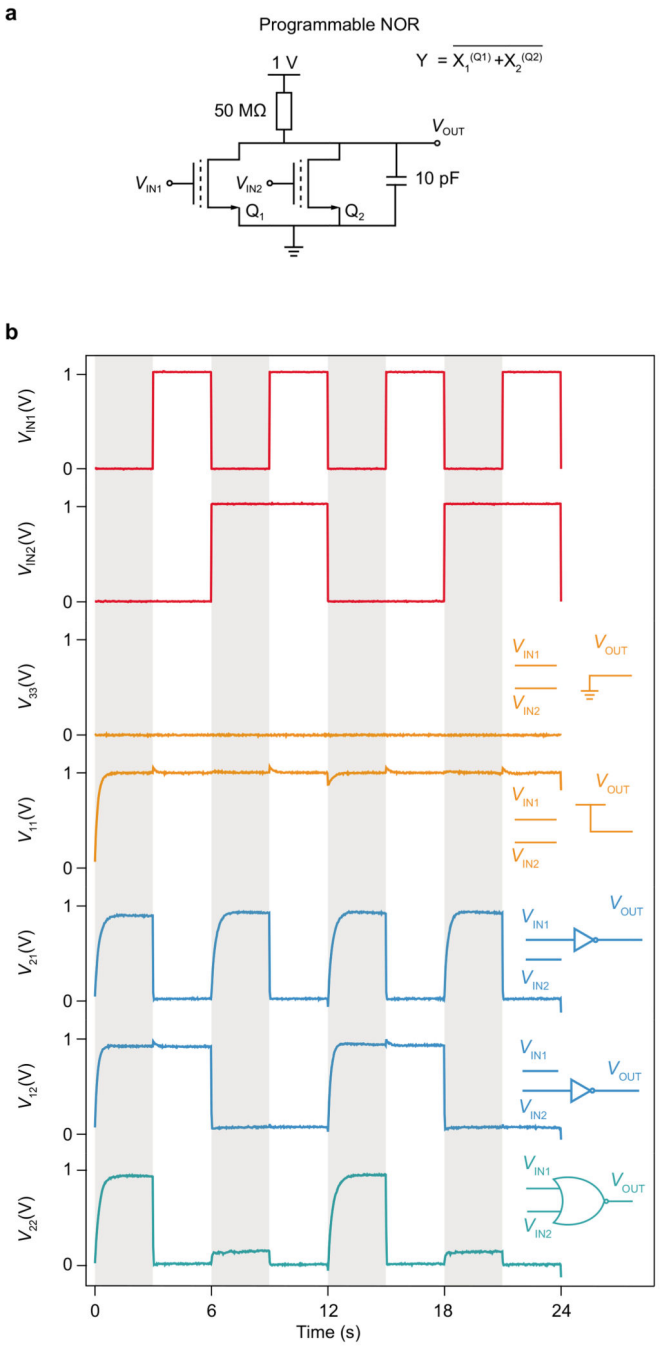
Europe PMC Funders Author Manuscripts



Extended Data Figure 5.

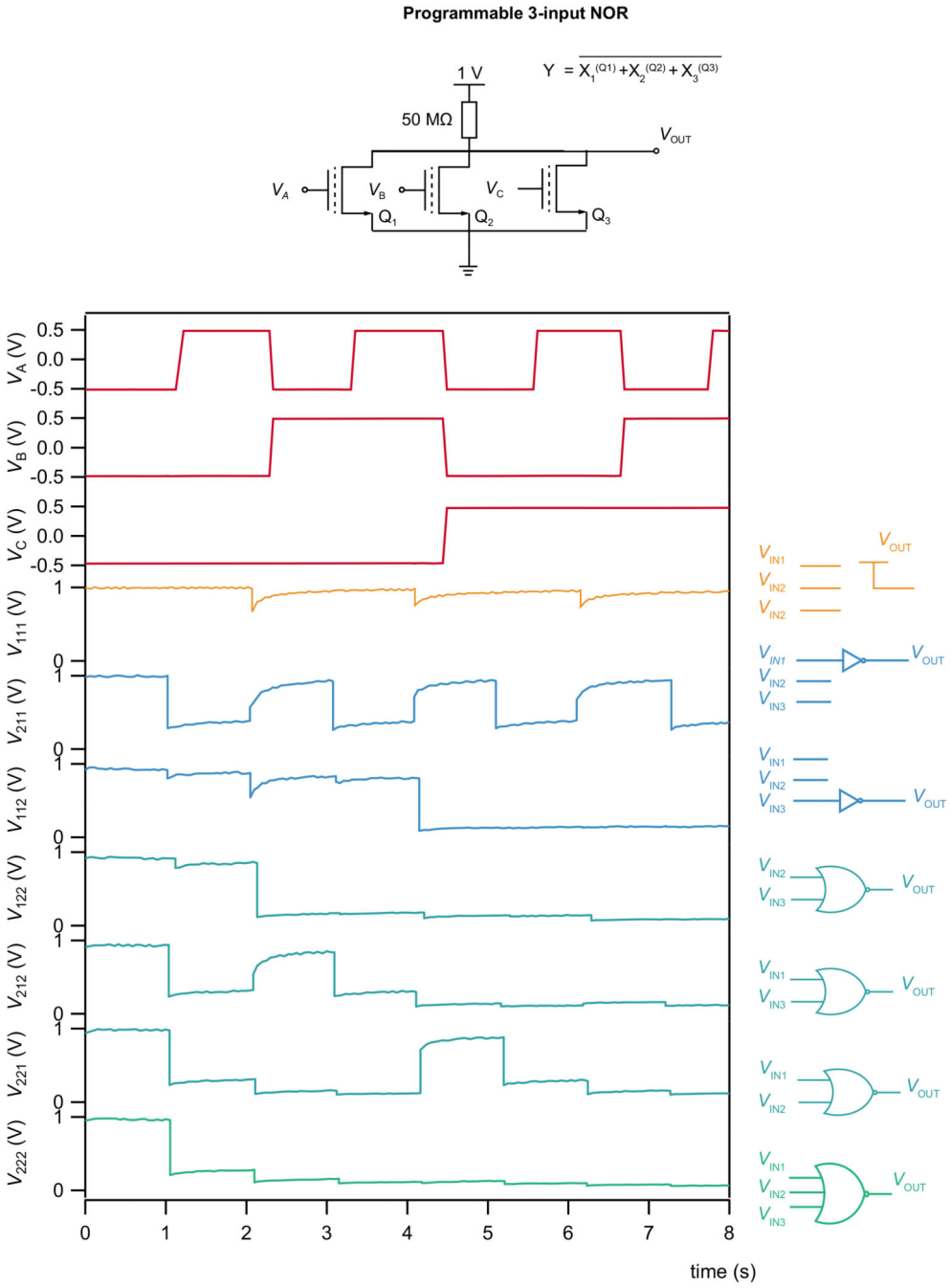
Example of the graphical estimation of the noise margin for the inverter programmed with V_{PROG} = 8.5 V. V_{IL} and V_{IH} are defined as the points where the gain of the transfer curve is unitary. V_{OL} and V_{OH} are the point where the curve is reflected from the transfer curve towards the V_{OUT} axis.

Europe PMC Funders Author Manuscripts



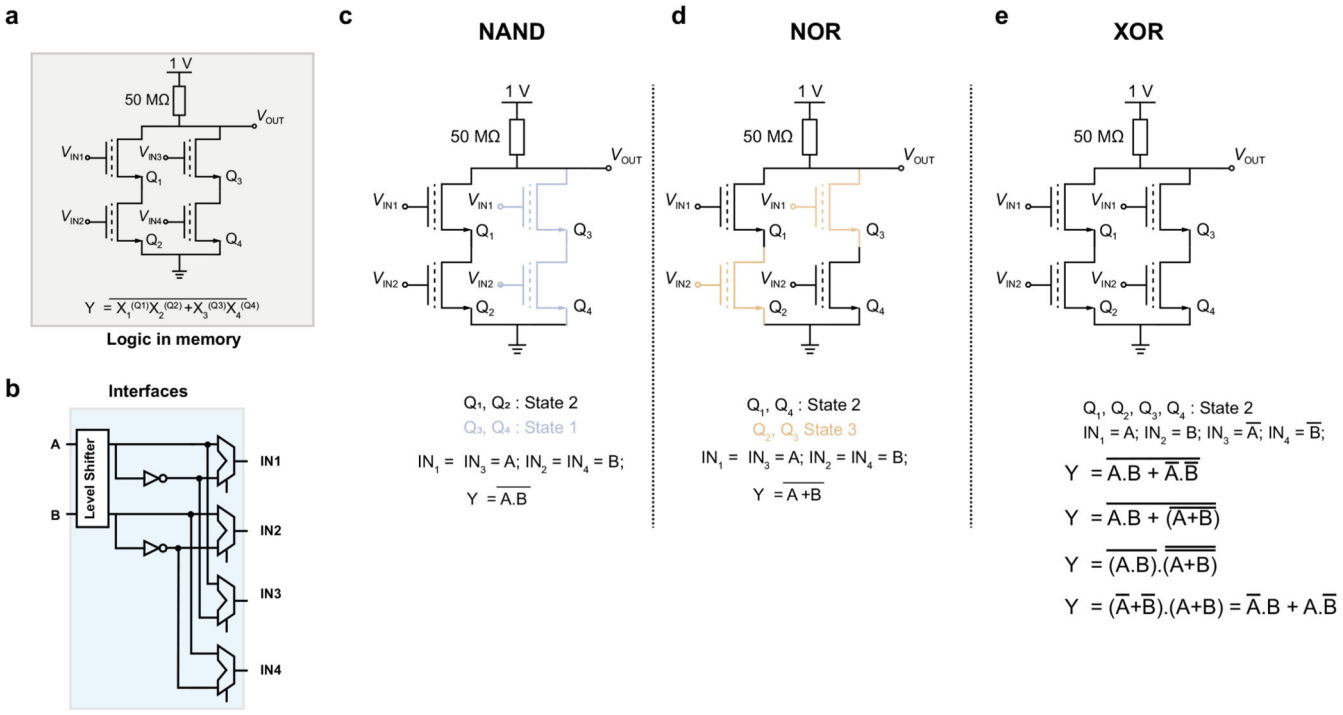
Extended Data Figure 6.

Circuit Schematic for a 2-Input NOR a; Circuit Schematic for a 2-Input NOR b; Logic over time for different programming states Q_1, Q_2. 33 – Constant LOW; 11 – Constant HIGH; 21 – Inverter A; 12 – Inverter B; 22 – NOR A,B.



Extended Data Figure 7.

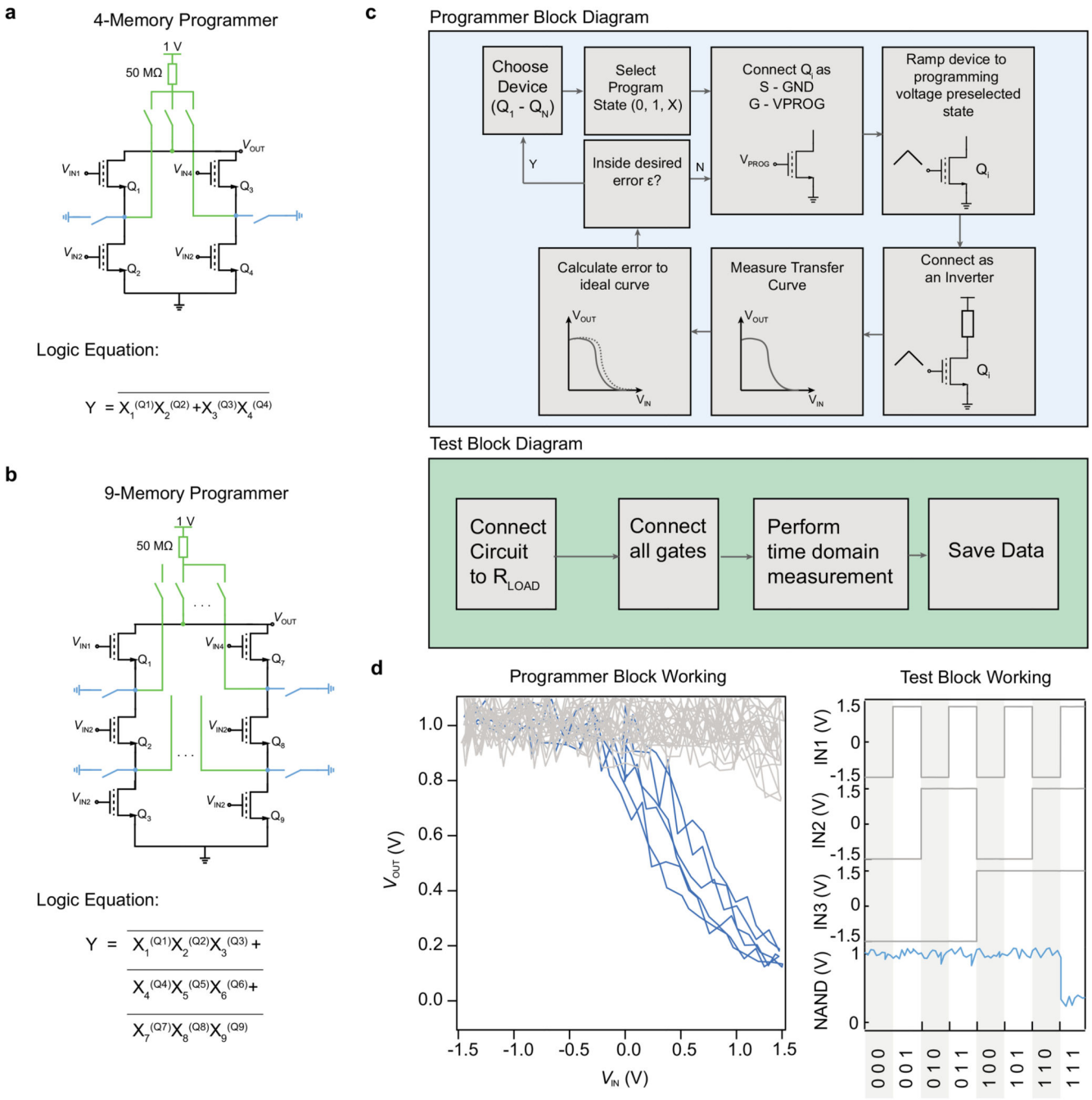
3-Input NOR a; Circuit Schematic for a 3-Input NOR b; Logic over time for different programming states Q_1, Q_2, Q_3. 111 – Constant HIGH; 211 – Inverter A; 112 – Inverter C; 122 – NOR B,C; 212 – NOR A,C; 221 – NOR A,B; 222 – NOR A,B,C



Extended Data Figure 8.

2-input logic-in-memory concept and interpretation a, 2-input schematic of the logic-in-memory concept; b, Interface model for input polarity control; c, NAND gate – $Q_{1\text{-}4}$ = 22 11; d, NOR gate – $Q_{1\text{-}4}$ = 23 32; e, XOR gate – $Q_{1\text{-}4}$ = 33 33, by applying the De Morgan's laws we derive the XOR canonical form.

Europe PMC Funders Author Manuscripts

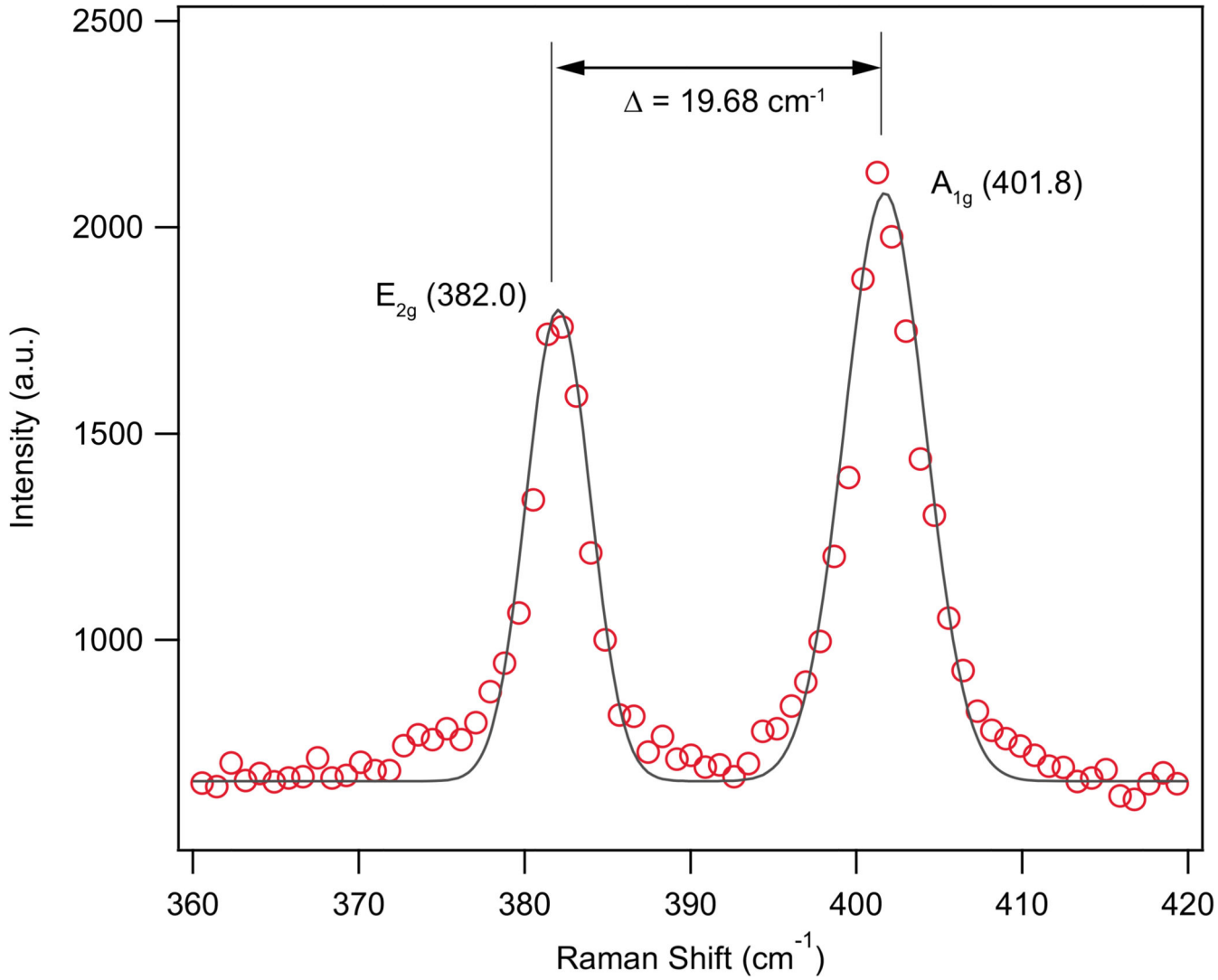


Extended Data Figure 9.

Hardware and software implementation of the logic-in-memory programmer a; Hardware implementation of the 4-memory programmer. b; Hardware implementation of the 9-memory programmer. c; Software working diagram of the programming and test blocks d; Example of programming and test using a 9-memory programmed into the following state $Q_{1\text{-}9}$ = 222 111 111 to perform a 3-Input NAND operation.

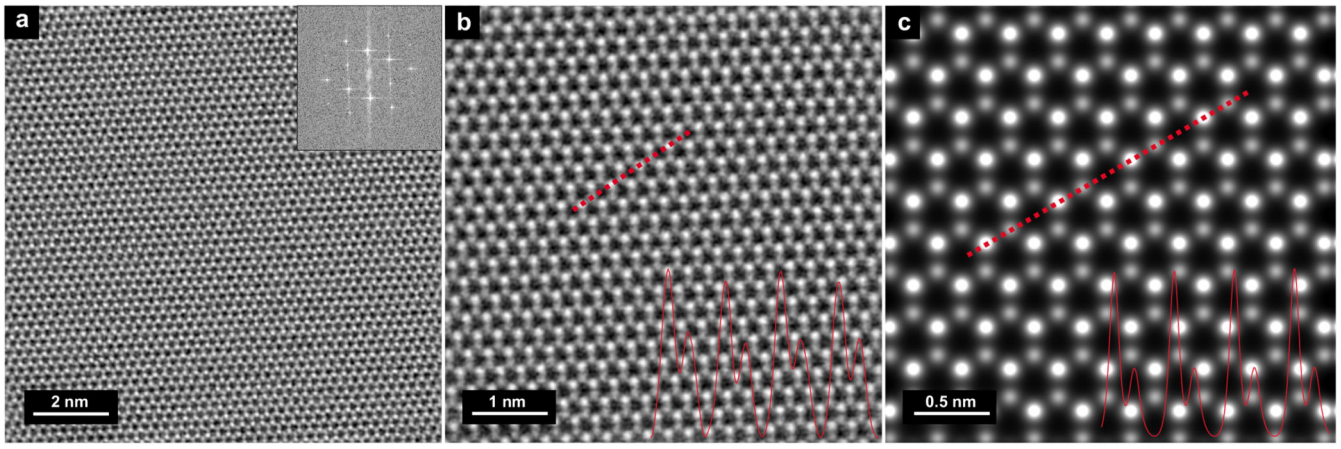
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



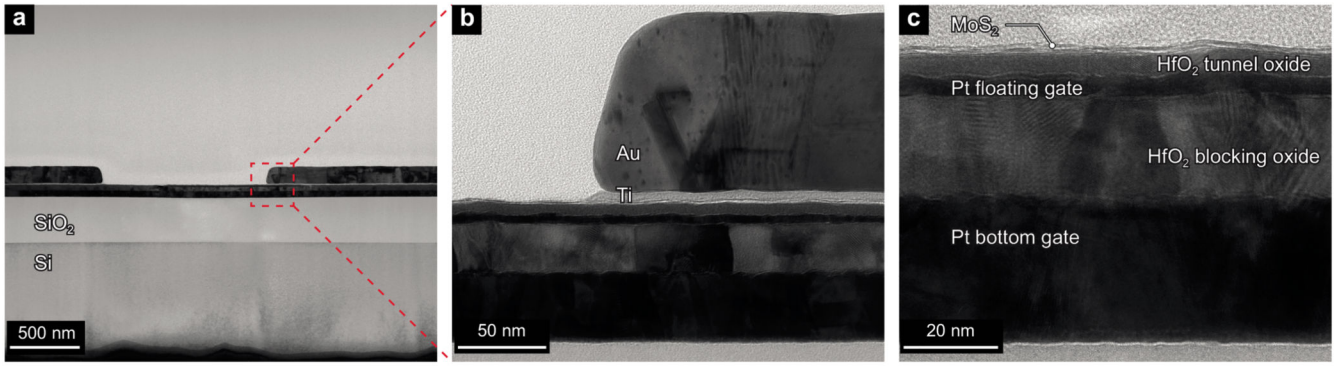
Extended Data Figure 10.

Raman characterization of monolayer MoS_2. Raman spectra of transferred MoS_2 from a single crystal from the same growth that the material used in the manuscript, using a 532 nm laser excitation and a 3000-line mm^{-1} grating. The observed difference between A_{1g} and E_{2g} Raman modes of MoS_2 is consistent with a monolayer.



Extended Data Figure 11.

ADF-STEM image of monolayer MoS_2. a, atomically-resolved STEM image shows the large field of view of the monolayer MoS_2. Inset: Fast Fourier transform (FFT) amplitude spectrum further shows the crystalline monolayer MoS_2 structure. b, the magnified filtered STEM image taken from a show the 2H crystal structure of monolayer MoS_2. C, STEM simulation image of monolayer MoS_2. The intensity line profiles along the dashed lines shows the peak positions of Mo-atoms and S-atoms in both panels b and c, respectively.



Extended Data Figure 12.

Floating-Gate transistor TEM cross-section a, wide-field view of the device b, magnified view of the contact area c, cross-section image of the gate stack consisting, from bottom to top: Pt bottom gate, HfO_2 blocking oxide, Pt floating-gate, HfO_2 tunnel oxide. MoS_2 2D channel is on top of the gate stack.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Z. Benes (CMI) for the help with electron-beam lithography. We would like to acknowledge support by the European Union's Horizon 2020 research and innovation programme under grant agreements No 829035 (QUEFORMAL) and No 785219 and 881603 (Graphene Flagship Core 2 and Core 3), Marie Curie-Sklodowska

COFUND (grant 665667), H2020 European Research Council (ERC, grant 682332) and support from the CCMX Materials Challenge grant “Large area growth of 2D materials for device integration”.

Data Availability

The data that support the findings of this study are available on Zenodo with the identifier(s) DOI: 10.5281/zenodo.4073060

References

1. Kautz WH. Cellular Logic-in-Memory Arrays. IEEE Trans Comput. 1969; C–18:719–727.
2. Stone HS. A Logic-in-Memory Computer. IEEE Trans Comput. 1970; C–19:73–78.
3. Gallo ML, et al. Mixed-precision in-memory computing. Nat Electron. 2018; 1:246–253.
4. Horowitz, M. 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC); IEEE; 2014. 10–14.
5. Wang QH, Kalantar-Zadeh K, Kis A, Coleman JN, Strano MS. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. Nat Nanotechnol. 2012; 7:699–712. [PubMed: 23132225]
6. Manzeli S, Ovchinnikov D, Pasquier D, Yazyev OV, Kis A. 2D transition metal dichalcogenides. Nat Rev Mater. 2017; 2:1733.
7. Radisavljevic B, Radenovic A, Brivio J, Giacometti V, Kis A. Single-layer MoS_2 transistors. Nat Nanotechnol. 2011; 6:147–150. [PubMed: 21278752]
8. Bertolazzi S, Brivio J, Kis A. Stretching and Breaking of Ultrathin MoS_2 . ACS Nano. 2011; 5:9703–9709. [PubMed: 22087740]
9. Iannaccone G, Bonaccorso F, Colombo L, Fiori G. Quantum engineering of transistors based on 2D materials heterostructures. Nat Nanotechnol. 2018; 13:183–191. [PubMed: 29511331]
10. Wu, B; Wan, A; Iandola, F; Jin, PH; Keutzer, K. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); IEEE; 2017. 446–454.
11. Graves, A; Mohamed, A; Hinton, G. Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013. 6645–6649.
12. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med. 2001; 23:89–109. [PubMed: 11470218]
13. Naffziger, S. High-Performance Processors in a Power-Limited World. 2006 Symposium on VLSI Circuits, 2006 Digest of Technical Papers; 2006. 93–97.
14. McKee, SA. Reflections on the memory wall. Proceedings of the first conference on computing frontiers on Computing frontiers - CF’04 162; ACM Press; 2004.
15. Xu X, et al. Scaling for edge inference of deep neural networks. Nat Electron. 2018; 1:216–222.
16. Arute F, et al. Quantum supremacy using a programmable superconducting processor. Nature. 2019; 574:505–510. [PubMed: 31645734]
17. Kwon J, et al. Three-dimensional monolithic integration in flexible printed organic transistors. Nat Commun. 2019; 10
18. Shulaker MM, et al. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. Nature. 2017; 547:74–78. [PubMed: 28682331]
19. Yu S. Neuro-inspired computing with emerging nonvolatile memorys. Proc IEEE. 2018; 106:260–285.
20. Chen W-H, et al. CMOS-integrated memristive non-volatile computing-in-memory for AI edge processors. Nat Electron. 2019; 2:420–428.
21. Li C, et al. Analogue signal and image processing with large memristor crossbars. Nat Electron. 2018; 1:52–59.
22. Fuller EJ, et al. Parallel programming of an ionic floating-gate memory array for scalable neuromorphic computing. Science. 2019; 364:570–574. [PubMed: 31023890]

23. Fiori G, et al. Electronics based on two-dimensional materials. Nat Nanotechnol. 2014; 9:768–779. [PubMed: 25286272]
24. Splendiani A, et al. Emerging Photoluminescence in Monolayer MoS_2 . Nano Lett. 2010; 10:1271–1275. [PubMed: 20229981]
25. Desai SB, et al. MoS_2 transistors with 1-nanometer gate lengths. Science. 2016; 354:99–102. [PubMed: 27846499]
26. Radisavljevic B, Whitwick MB, Kis A. Integrated circuits and logic operations based on single-layer MoS_2 . ACS Nano. 2011; 5:9934–9938. [PubMed: 22073905]
27. Wachter S, Polyushkin DK, Bethge O, Mueller T. A microprocessor based on a two-dimensional semiconductor. Nat Commun. 2017; 8:1–6. [PubMed: 28232747]
28. Bertolazzi S, Krasnozhon D, Kis A. Nonvolatile Memory Cells Based on MoS_2/Graphene Heterostructures. ACS Nano. 2013; 7:3246–3252. [PubMed: 23510133]
29. Li D, et al. Nonvolatile Floating-Gate Memories Based on Stacked Black Phosphorus-Boron Nitride-MoS_2 Heterostructures. Adv Funct Mater. 2015; 25:7360–7365.
30. Tan C, Liu Z, Huang W, Zhang H. Non-volatile resistive memory devices based on solution-processed ultrathin two-dimensional nanomaterials. Chem Soc Rev. 2015; 44:2615–2628. [PubMed: 25877687]
31. Wang J, et al. Floating Gate Memory-based Monolayer MoS_2 Transistor with Metal Nanocrystals Embedded in the Gate Dielectrics. Small. 2015; 11:208–213. [PubMed: 25115804]
32. Cao W, Kang J, Bertolazzi S, Kis A, Banerjee K. Can 2D-Nanocrystals Extend the Lifetime of Floating-Gate Transistor Based Nonvolatile Memory? IEEE Trans Electron Devices. 2014; 61:3456–3464.
33. Sebastian A, Pannone A, Subbulakshmi Radhakrishnan S, Das S. Gaussian synapses for probabilistic neural networks. Nat Commun. 2019; 10
34. Sivan M, et al. All WSe_2 1T1R resistive RAM cell for future monolithic 3D embedded memory integration. Nat Commun. 2019; 10
35. Sun L, et al. Self-selective van der Waals heterostructures for large scale memory array. Nat Commun. 2019; 10:1–7. [PubMed: 30602773]
36. Kim H, Ovchinnikov D, Deiana D, Unuchek D, Kis A. Suppressing Nucleation in Metal–Organic Chemical Vapor Deposition of MoS_2 Monolayers by Alkali Metal Halides. Nano Lett. 2017; doi: 10.1021/acs.nanolett.7b02311
37. Cun H, et al. Wafer-scale MOCVD growth of monolayer MoS_2 on sapphire and SiO_2 . Nano Res. 2019; 12:2646–2652.
38. Raja A, et al. Dielectric disorder in two-dimensional materials. Nat Nanotechnol. 2019; doi: 10.1038/s41565-019-0520-0
39. Papandroulidakis G, Vourkas I, Vasileiadis N, Sirakoulis G. Ch. Boolean Logic Operations and Computing Circuits Based on Memristors. IEEE Trans Circuits Syst II Express Briefs. 2014; 61:972–976.
40. Resta GV, et al. Polarity control in WSe_2 double-gate transistors. Sci Rep. 2016; 6
41. Resta GV, et al. Doping-Free Complementary Logic Gates Enabled by Two-Dimensional Polarity-Controllable Transistors. ACS Nano. 2018; 12:7039–7047. [PubMed: 29956911]
42. Dumcenco D, et al. Large-Area Epitaxial Monolayer MoS_2 . ACS Nano. 2015; 9:4611–4620. [PubMed: 25843548]

Europe PMC Funders Author Manuscripts

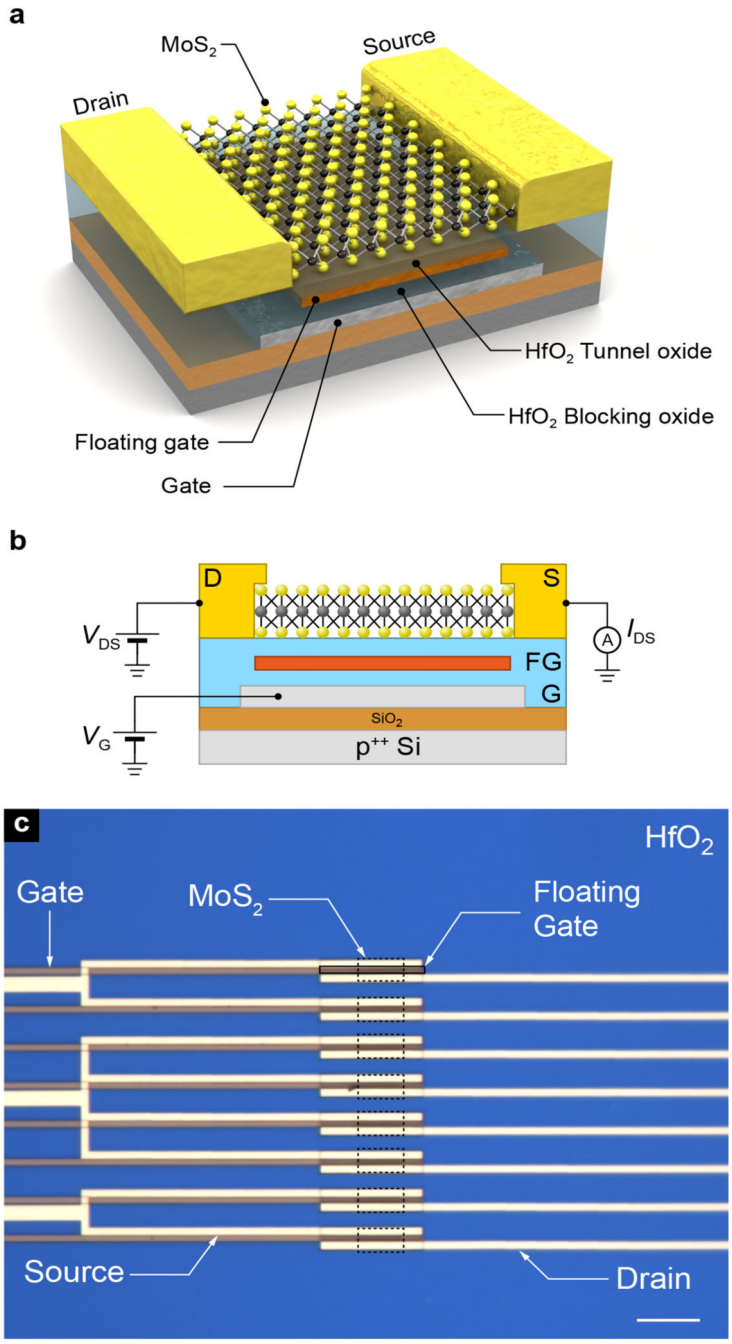


Figure 1. In-memory device structure.

a, Three-dimensional view of a floating-gate memory device based on MOCVD-grown monolayer MoS_2 with source (S) and drain contacts (D). The floating gate (FG) is separated from the MoS_2 channel by a 7 nm thick HfO_2 tunnel oxide layer and the bottom control gate (G) by a 30 nm thick HfO_2 blocking oxide layer. **b,** Schematic of the device. **c,** Optical image of the fabricated floating-gate memory array, comprised of 8 memory cells (scale bar: 10 μm).

Europe PMC Funders Author Manuscripts

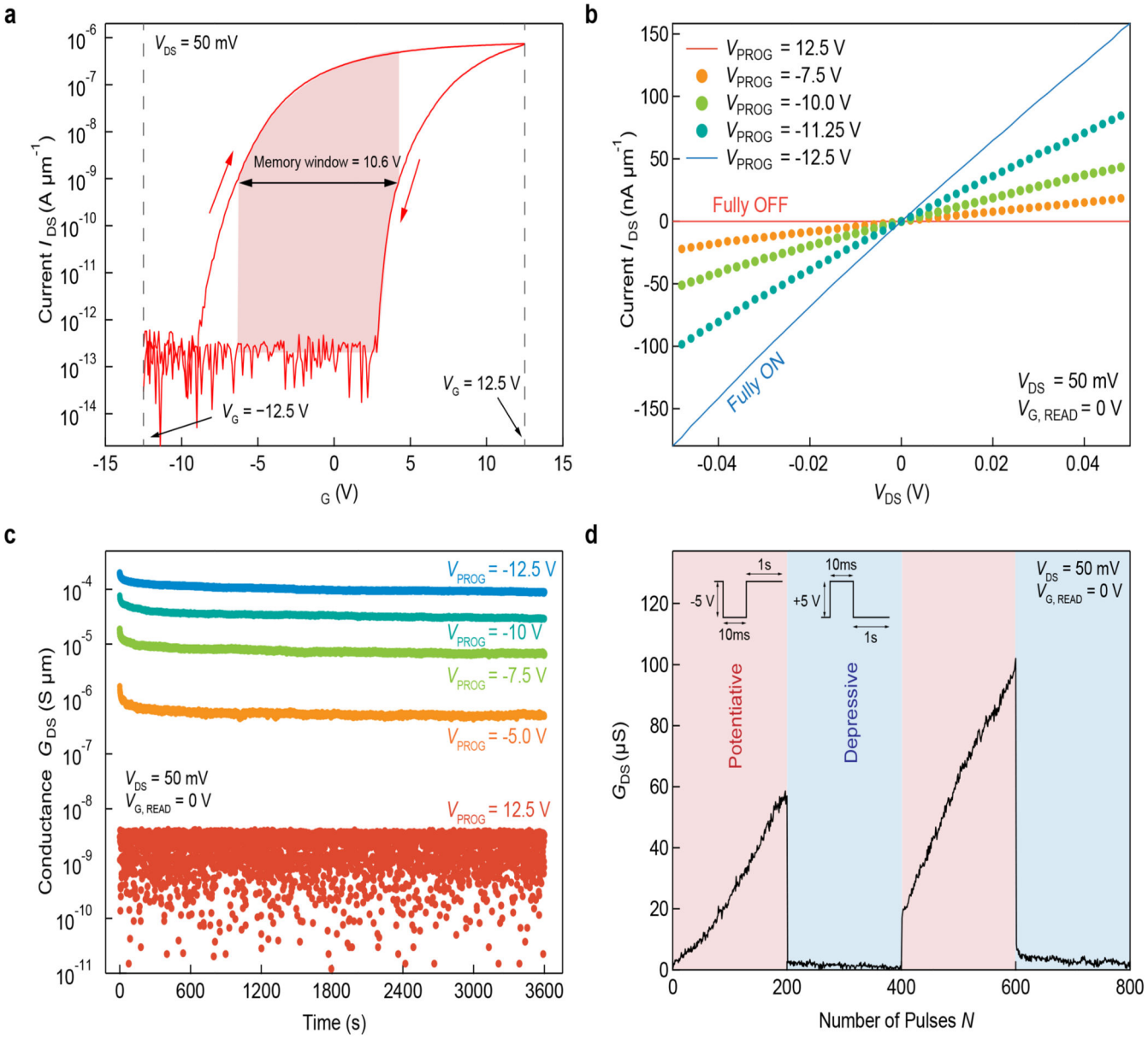


Figure 2. Characterization of non-volatile memories.

a, Transfer characteristic of the floating-gate transistor (I_{DS} vs. V_G) acquired for two different gate voltage sweep directions. The variation of the threshold voltage V_{TH}, the memory window, is estimated to be 10.6 V. **b,** Output characteristics (I_{DS} vs. V_{DS}) of the floating-gate transistor in the ON state, after having been programmed using different values of the programming voltage V_{PROG}. **c,** Time-dependence of the device conductance for different levels of programming voltage V_{PROG}. Retention measurements showing multilevel states of the normalized conductance G_{DS} versus time as a function of the programming voltage (V_{PROG} ranging from -12.5 V to 12.5 V). **d,** Demonstration of fine control over the conductance state of the memory using voltage pulses applied to the gate. Main plot shows the evolution of the memory device conductance G_{DS} as a function of the number of pulses. Potentiative pulses applied to the gate (-5 V amplitude, 10 ms duration

and 1s rest time) can be used to increase and fine-tune the conductance. Depressive pulses (+5 V amplitude, 10 ms duration and 1s 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 for programming the initial memory state into the floating gate.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

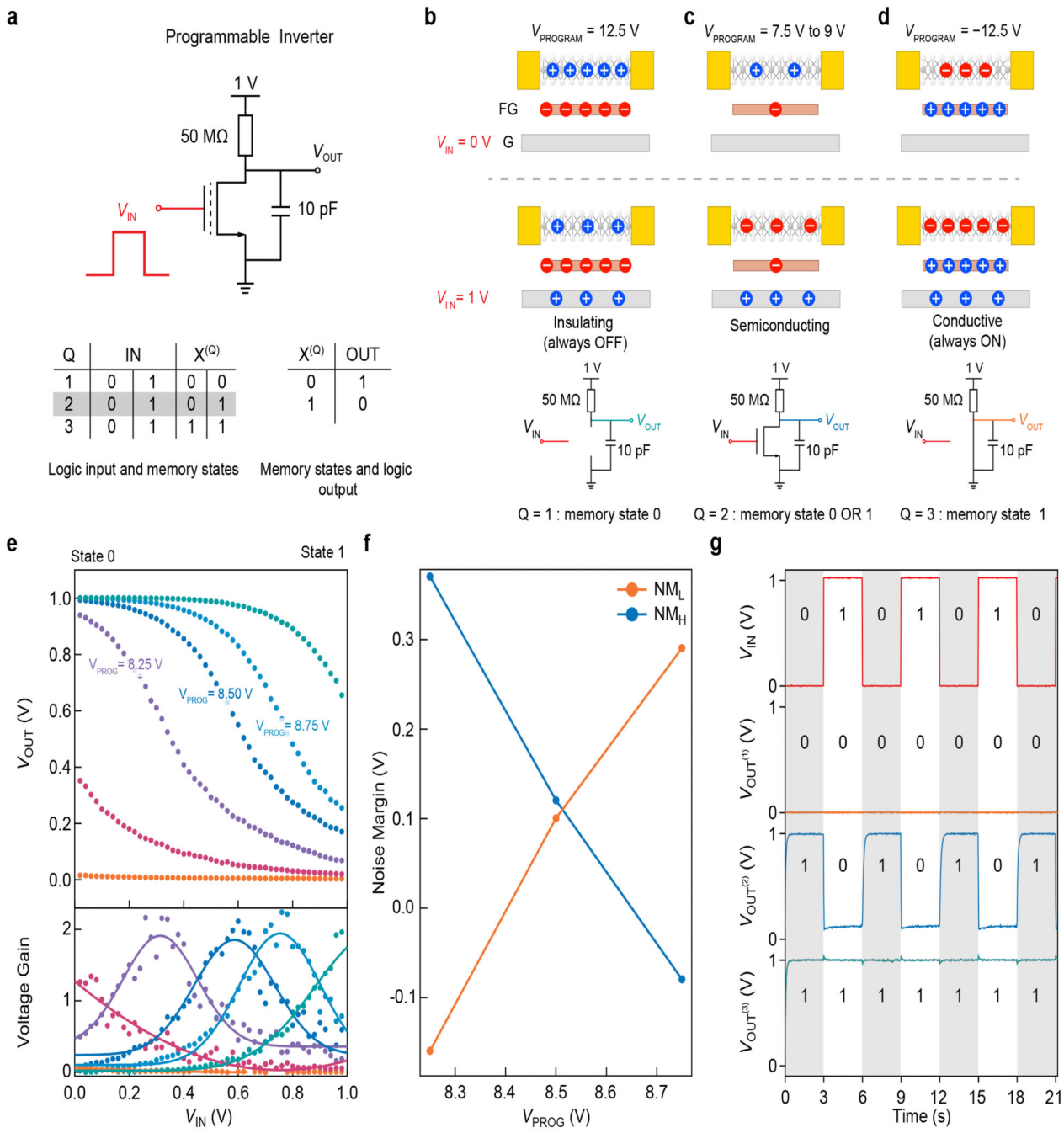


Figure 3. Programmable inverter based on a MoS_2 memory cell.

a, Schematic of the programmable inverter, different states of the device and the inverter truth table. **b-d**, Illustration of the discrete memory states used for programmable behavior and the corresponding circuit diagrams: **b**, Insulating state of the MoS_2 channel ($Q = 1$, memory state 0). **c**, semiconducting state with a continuously tunable conductance of the MoS_2 channel ($Q = 2$, memory states 0 or 1). **d**, Conductive state of the channel ($Q = 3$, memory state 1). **e**, Programmable output (V_{OUT}) curves of the inverter and the inverter voltage gain as a function of the input voltage (V_{IN}), for different programming conditions.

f, Evolution of the inverter noise margin (NM_L and NM_H) as a function of the programming voltage. **g**, Time traces showing stability of the output voltage for the three different configurations of the programmable inverter. Red: input voltage (V_{IN}), orange: constant 0 state, blue: inverter operation, green: constant 1 state.

Europe PMC Funders Author Manuscripts

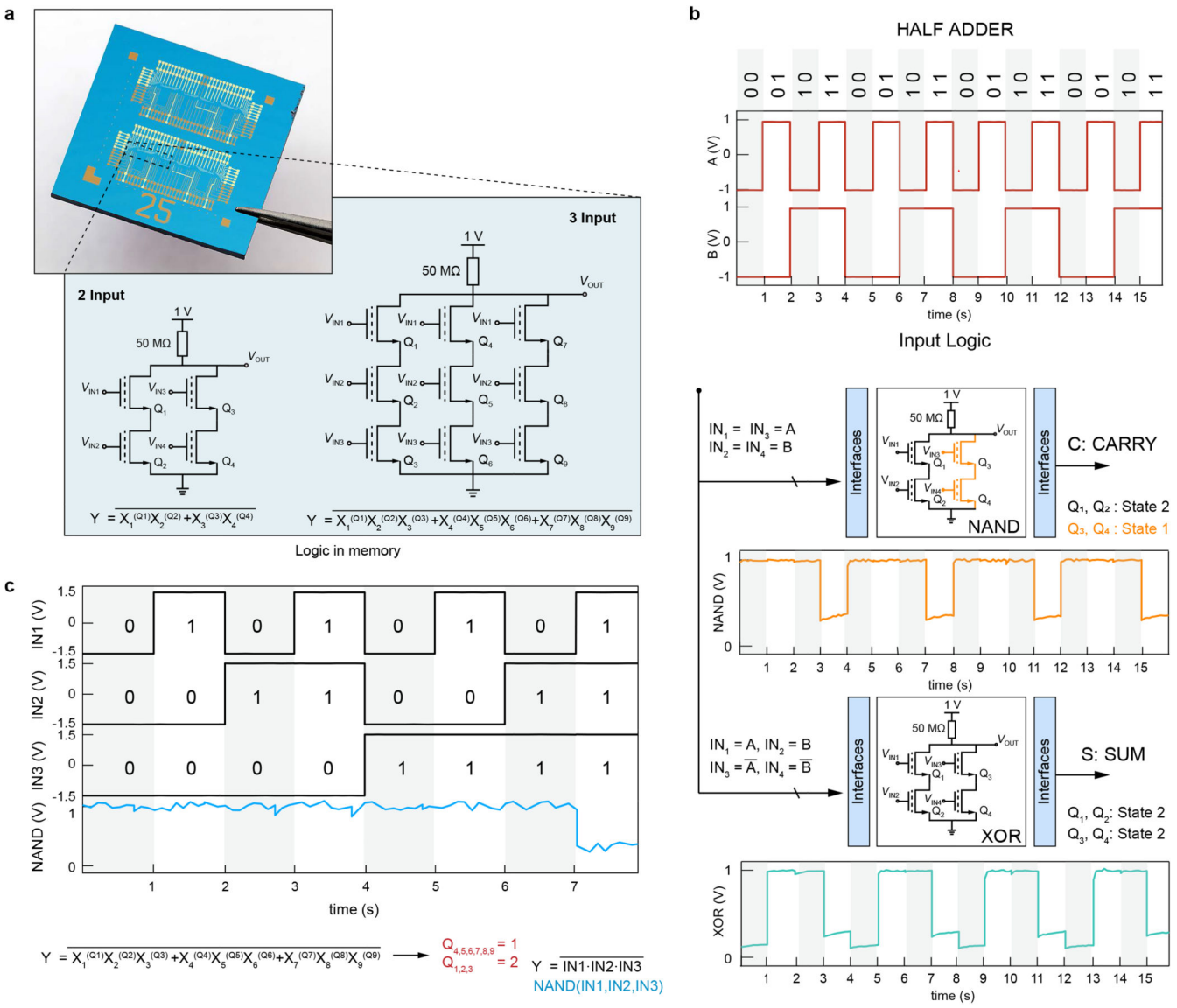


Figure 4. Logic-in-memory.

a, Photograph of a fabricated 12 × 12 mm die with logic-in-memory cell arrays and schematics of the 2-input and 3-input logic-in-memory cells **b**, System level operation of two 2-input cells to form a half-adder. XOR programmed as $Q_{1\text{-}4}$= 2 with inputs to memories Q_3 Q_4 inverted. NAND programmed as $Q_{1\text{-}2}$= 2, $Q_{3\text{-}4}$= 3 output is inverted to form AND logic. **c**, Time traces showing stability of the output voltage for the NAND operation of the 3-input unit cell. For this configuration, memories are programmed as following $Q_{1\text{-}3}$ = 2, $Q_{4\text{-}9}$ = 3. The transfer curves of each state can be seen in SI 15.

Europe PMC Funders Author Manuscripts