nature

https://doi.org/10.1038/s41586-023-05973-1

Accelerated Article Preview

Hybrid 2D/CMOS microchips for memristive applications

Received: 4 March 2022

Accepted: 17 March 2023

Accelerated Article Preview

Cite this article as: Zhu, K. et al. Hybrid 2D/CMOS microchips for memristive applications. *Nature* https://doi.org/10.1038/s41586-023-05973-1 (2023)

Kaichen Zhu, Sebastian Pazos, Fernando Aguirre, Yaqing Shen, Yue Yuan, Wenwen Zheng, Osamah Alharbi, Marco A. Villena, Bin Fang, Xinyi Li, Alessandro Milozzi, Matteo Farronato, Miguel Muñoz-Rojo, Tao Wang, Ren Li, Hossein Fariborzi, Juan B. Roldan, Guenther Benstetter, Xixiang Zhang, Husam Alshareef, Tibor Grasser, Huaqiang Wu, Daniele Ielmini & Mario Lanza

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. Nature is providing this early version of the typeset paper as a service to our authors and readers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

ACCELERATED ARTICLE PREVIEW

Hybrid 2D/CMOS microchips for memristive applications

Kaichen Zhu[1]†, Sebastian Pazos[1]†, Fernando Aguirre[1], Yaqing Shen[1], Yue Yuan[1], Wenwen Zheng[1], Osamah Alharbi[1], Marco A. Villena[1], Bin Fang[1], Xinyi Li[2], Alessandro Milozzi[3], Matteo Farronato[3], Miguel Muñoz-Rojo[4-5], Tao Wang[6], Ren Li[7], Hossein Fariborzi[7], Juan B. Roldan[8], Guenther Benstetter[9], Xixiang Zhang[1], Husam Alshareef[1], Tibor Grasser[10], Huaqiang Wu[2], Daniele Ielmini[3], Mario Lanza[1]*

[1] Materials Science and Engineering Program, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[2] Institute of Microelectronics, Tsinghua University, 100084, Beijing, China
[3] Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy
[4] Department of Thermal and Fluid Engineering, Faculty of Engineering Technology, University of Twente, Enschede, Netherlands
[5] Instituto de Micro y Nanotecnología, IMN-CNM, CSIC (CEI UAM+CSIC), Madrid, Spain.
[6] Institute of Functional Nano & Soft Materials, Collaborative Innovation Center of Suzhou Nanoscience and Technology, Soochow University, 199 Ren-Ai Road, Suzhou 215123, China
[7] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia
[8] Departamento de Electrónica y Tecnología de Computadores, Facultad de Ciencias, Universidad de Granada, Avenida Fuentenueva s/n, 18071 Granada, Spain
[9] Department of Electrical Engineering and Media Technology, Deggendorf Institute of Technology, Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany
[10] Institute for Microelectronics, TU Wien, Gusshausstrasse 27-29, 1040 Vienna, Austria

† These authors contributed equally
* Corresponding author email: mario.lanza@kaust.edu.sa

ACCELERATED ARTICLE PREVIEW

Abstract

Exploiting the excellent electronic properties of two-dimensional (2D) materials to fabricate advanced electronic circuits is a major goal for the semiconductors industry[1-2]. However, most studies in this field have been limited to the fabrication and characterization of isolated large (>1μm^2) devices on unfunctional SiO_2/Si substrates. Some studies integrated monolayer graphene on silicon microchips as large-area (>500μm^2) interconnection[3] and as channel of large transistors (~16.5μm^2)[4-5], but in all cases the integration density was low, no computation was demonstrated, and manipulating monolayer 2D materials was challenging because native pinholes and cracks during transfer increase variability and reduce yield. Here we present the fabrication of high-integration-density 2D/CMOS hybrid microchips for memristive applications — CMOS stands for complementary metal oxide semiconductor. We transfer a sheet of multilayer hexagonal boron nitride (h-BN) onto the back-end-of-line (BEOL) interconnections of silicon microchips containing CMOS transistors of the 180nm node, and finalize the circuits by patterning the top electrodes and interconnections. The CMOS transistors provide outstanding control over the currents across the h-BN memristors, which allows us to achieve endurances of ~5 million cycles in memristors as small as ~0.053μm^2. We demonstrate in-memory computation by constructing logic gates, and measure spike-timing dependent plasticity (STDP) signals that are suitable for the implementation of spiking neural networks (SNN). The high performance and the relatively-high technology readiness level achieved represent a significant advance towards the integration of 2D materials in microelectronic products and memristive applications.

Keywords: microchip; integrated circuit; back-end-of-line; two-dimensional materials; memristor.

MAIN TEXT

Our 2cm×2cm silicon microchips have been designed via Synopsys software and fabricated in a 200mm silicon wafer in an industrial clean room using a 180nm CMOS technology node (Figure 1a and Extended Data Figure 1). The circuits fabricated in this study consist of 5×5 crossbar arrays of one-transistor-one-memristor cells (1T1M, Figure 1b-c and Supplementary Figure 1), although some standalone memristors and CMOS transistors were fabricated as reference (Supplementary Figure 2). The microchips have been designed to integrate the memristors into the BEOL interconnections, i.e., they have been terminated at the latest metallization layer (fourth in our wafer) and have been left without passivation. Hence, silicon oxide naturally grows on the wafers when they are extracted from the industrial clean room (Figure 1d), which can be easily etched away to expose the tungsten vias (Figure 1e and Supplementary Figure 3). Then, a ~18-layers-thick sheet of h-BN (i.e., ~6nm), grown on a Cu substrate via chemical vapour deposition (CVD), was transferred on the microchips (Figure 1f) using a low-temperature process (see Methods). Finally, the h-BN on the contact pads was etched, and top electrodes made of different materials (i.e., Au/Ti, Au or Ag) were patterned and deposited on the h-BN to finalize the circuits (see Figure 1g).

As the tungsten vias of the fourth metallization layer have a diameter of ~260nm (Figure 1h, Supplementary Figure 3), the lateral size of the resulting h-BN memristors is, at most, $0.053\mu m^2$. Figure 1h shows a high-angle annular dark-field (HAADF) cross-sectional scanning transmission electron microscope (STEM) image of a 1T1M cell (with top Au/Ti electrode) in the crossbar array (see also Supplementary Figures 4-5). The correct layered structure of the h-BN stack is confirmed before and after transfer via cross-sectional transmission electron microscopy (TEM, see the inset in Figure 1h and Extended Data Figure 2). Nano-chemical analyses via electron energy loss spectroscopy demonstrate the correct composition of the h-BN sheet (see Extended Data Figure 3). The optical microscope images (Figure 1c) reveal that the h-BN sheet does not crack during the transfer; this is an important advantage of using ~6-nm-thick 2D layered materials, and it remarkably increases the yield of the devices and circuits compared to counterparts using monolayer 2D materials[6].

Electronic memory

When we apply sequences of ramped voltage stresses (RVS) to several standalone $\sim 0.053\mu m^2$ Au/Ti/h-BN/W structures, most (~90%) of them show erratic current fluctuations and no resistive switching (RS) is observed (see Figure 2a); the currents don't reach linear regime (i.e., dielectric breakdown) even if we apply 11V. This is striking because most (>75%) h-BN devices with larger areas ($25\mu m^2$) show dielectric breakdown voltages (V_{DB}) between 3 and 11V followed by filamentary non-volatile bipolar RS[6]. The reason should be the lower probability to find clusters of defects in small devices, which remarkably increases V_{DB}[7]. Few (~10%) $\sim 0.053\mu m^2$ Au/Ti/h-BN/W structures show V_{DB} between ~2.5 and 4V followed by filamentary non-volatile bipolar RS (if a current limitation ≥1mA is applied, see Supplementary Figure 6). However, the endurance is only ~100 cycles, in great part due to the poor controllability of the current across the memristor and the overshoot during the dielectric breakdown[6].

Quite to the contrary, the CMOS transistor in the 1T1M cell can precisely control the current across the h-BN memristor and avoid the current overshoot, which results in outstanding performance. First, we obtain the output characteristic of one standalone CMOS transistor by applying a constant voltage to the gate (V_G) and a RVS to the drain (V_{DS}), and measuring the drain-to-source current (I_{DS}); the CMOS transistor works correctly as expected (see Figure 2b). And second, we measure the 1T1M cell by applying RVS at the top Au/Ti electrode of the memristor while keeping the source terminal of the transistor grounded and simultaneously applying a constant V_G. When a sequence of RVS is applied to the top electrode of the Au/Ti/h-BN/W structure using V_G=1.1V, most 1T1M cells exhibit non-volatile bipolar RS (see Figure 2c and Extended Data Figure 4). The high state resistances (R_{HRS}~20GΩ and R_{LRS}~100KΩ) —beneficial to reduce power consumption—, the non-linearity of the currents in both states, and the progressive state transitions indicate that the RS is non-filamentary[8]. However, we do see an activation process, as the first RVS slightly increases the conductance of the devices (i.e., softly degrades the h-BN stack, see blue line in Figure 2c and Extended Data Figure 4). In the first microchip that we fabricated this stable non-filamentary bipolar RS regime was observed in 32 out of 40 cells (yield 80%), and in the last one it was observed in 25 out of 25 devices (yield 100%). On the contrary, standalone $\sim 0.053\mu m^2$ Au/Ti/h-BN/W structures and 1T1M cells without h-BN never exhibited this behaviour; this

ACCELERATED ARTICLE PREVIEW

confirms that the RS is produced by the h-BN stack and that the CMOS transistor is key to control its soft-degradation. Note that in the standalone Au/Ti/h-BN/W structures the current is limited using the semiconductor parameter analyser, which activation time is long (~70µs) and parasitic capacitance is high (~300pF, related to the cables)[9]; on the contrary, in the 1T1M cell the series transistor acts as an instantaneous current limitation (it cannot drive more current than that allowed by the size of its channel) and the parasitic capacitance is much lower (~50pF, internal connections in the microchip), which reduces the duration of the switching transient and undesired currents across the Au/Ti/h-BN/W structure[10]. The values of R_{HRS} and R_{LRS} are stable over time, and multiple stable conductance levels can be programmed either by adjusting V_G during the set process (which fixes R_{LRS}) and/or by adjusting the end voltage of the negative RVS (which fixes R_{HRS}, see Extended Data Figure 5).

The most surprising observation however relates to the endurance, which readily reaches 2.5 million cycles (see Figure 2d-f) when applying sequences of pulsed voltage stresses (PVS). Under this type of stress, the values of R_{HRS}, R_{LRS} and R_{LRS}/R_{HRS} can be accurately controlled in three different ways: by tuning the duration of the write pulse, by tuning the amplitude of the write pulse, and by tuning the amplitude of the erase pulse (see Figure 2d-f and Extended data Figure 6). This endurance is very high considering the small size of the memristors (see Supplementary Note 2), and similar to that of commercial metal-oxide-based resistive random access memories (0.5 million cycles)[11] and phase-change memories (10 million cycles)[12-13]. However, the switching time of the 1T1M cells using top Au/Ti electrodes is rather long (t_{SET} ~232µs and t_{RESET} ~783ns, see Extended Data Figure 7).

The properties of the 1T1M cells can be adjusted using different top electrodes (see Extended Data Figure 8). When Au electrodes are employed, the devices show reliable switching at lower state resistances (see Figure 2g), as well as shorter switching time (t) and lower switching energy (E); and when Ag electrodes are employed these values can be pushed down to t_{SET}=680ns, t_{RESET}=60ns, E_{SET}=21.11pJ and E_{RESET}=1.41pJ (see Figure 2h). The reasons behind these observations are the lack of an interfacial TiO_X layer and the higher conductivity and diffusivity of Au^{X+} and Ag^{X+} ions[14] (see Supplementary Note 1). The performance observed in h-BN/CMOS 1T1M cells using Au/Ti electrodes may allow them to cover niche applications between NAND Flash and DRAM within the memory hierarchy (e.g., persistent memory), and when using Au or Ag electrodes their performance may be valid for low-power application-specific integrated circuits (ASIC) within the internet-of-things[15] (see Supplementary Figure 12).

Data computation

Based on the above measured performance metrics, the hybrid 2D/CMOS 1T1M cells show good potential for data computation. The high R_{HRS}/R_{LRS} ratio and the stability of the resistive states over time allows us to implement in-memory computing operations taking advantage of the internal connections of our 5×5 crossbar array of 2D/CMOS 1T1M cells. As a proof of concept, we realized “or” and “implication” operations (see Extended Data Figure 9), although more sophisticated operations could be easily realized by modifying the interconnections between the devices via custom design.

Furthermore, the 1T1M cells with Au/Ti/h-BN/W memristors exhibit spike-timing dependent plasticity (STDP) when applying pairs of PVS displaced in time at the input and output (see Figure 2i). This non-volatile RS performance is very attractive to construct electronic synapses for spiking neural networks (SNNs)[16], which consume less energy than traditional deep neural networks[17].

While implementing via hardware a reliable 2D-materials-based memristive SNN capable to compete with state-of-the-art developments[18-19] is at this moment not achievable due to the lower maturity of these materials, we do analyse the performance of a SNN made of memristors that exhibit STDP characteristics like those in Figure 2i (see Supplementary Note 3). First we fit the measured STDP data from Figure 2i, including the device-to-device variability, using an exponential decaying model to implement the learning rule (see Supplementary Figure 13). Second, we simulate a SNN to demonstrate the unsupervised learning capability (Figure 3a), and benchmark it by classifying the images from the Modified National Institute of Standards and Technology (MNIST) database of handwritten digits[20-21] (see Methods). The SNN has 784 input neurons, an excitatory layer of 400 neurons, and an inhibitory layer of 400 neurons, plus a decision block that determines which is the most probable digit (0-9) represented by the input pattern. We trained the SNN with the complete MNIST dataset and evaluated the accuracy every 1000 images. Figures 3b-d show the main three figures-of-merit for this type of SNNs (i.e., evolution of the synaptic weights with the number of training images, confusion matrix of the

ACCELERATED ARTICLE PREVIEW

network, and the training accuracy versus number of training images) and all of them indicate an excellent performance. To account for the device variability, we considered a Monte Carlo simulation with 50 iterations that randomizes the exponential fitting of the STDP plot and the initial value of the synapses, and the deviations observed in the accuracy are very low (<5%, see Figure 3d and Supplementary Figure 14). The best average accuracy reaches ~90%, which is a very high value considering the simplicity of the SNN and the unsupervised training protocol (see Supplementary Table 4).

We also propose a CMOS circuit for the hardware implementation of an electronic neuron based on our h-BN memristors (Figure 3e), which is capable of accounting for the adaptative firing threshold and the refractory period after firing (see the pre/post-synaptic traces and the evolution of the membrane potential, simulated via SPICE, in Figure 3f-g).

Discussion

Very few commercial electronic products today already include 2D materials, and the ones that do (sensors[22], specialty cameras[23]) use very low integration density (>100μm^2/device) —because in larger devices the local defects in the 2D material are not so detrimental. Our hybrid 2D/CMOS microchips are still far from being ready for production, but we can safely claim that our work represents the highest performance and technology readiness level ever achieved in high-integration-density 2D-materials-based electronic devices/circuits. The electrical characteristics of the h-BN memristors connected to a CMOS transistor are by orders of magnitude superior to those of standalone h-BN memristors[6,24-27] and h-BN memristors connected to 2D-materials-based transistors[28-29].

The voltages needed to switch our devices (from ±1.4V to ±5V) are quite low compared to other prototypes in the field of 2D materials (even >20V)[30-32], but still higher than that used at the 180nm CMOS node. Nevertheless, this is not an impediment for the development of this technology, as there are many commercial microchips that operate at much higher voltages; that is the case of all Flash memories[33] (state-of-the-art 3D-NAND Flash memories are programmed at ~20V)[34] and all bipolar-CMOS microchips for automotive applications (which require up to 40V)[35]. Strategies to fabricate wafers with devices that operate at different voltages are widespread[36], and many companies[37-38] offer versions of their 180nm CMOS technology that operate at high voltages >18V. Note that prototype memristive devices developed by companies also operate at ±5V[39].

We finally wish to remark that, at a first glance, the use of Au and Ag electrodes may not appear ideal because they are categorized as contaminant in front-end-of-line (FEOL) processes. However, our h-BN memristors are integrated in the latest metallic layer of the BEOL interconnections (see Figure 1h), where Au pads, liners and wires are usually employed[40] (see Extended Data Figure 10). The semiconductor industry has also developed ferroelectric memories with high content of Iridium[41] (a contaminant material forbidden in FEOL processes), and companies working in the field of 2D materials use Au electrodes in their studies and FEOL prototypes (see Supplementary Table 5). Hence, the use of Au, Au/Ti or Ag electrodes in our hybrid 2D/CMOS microchips for memristive applications does not prevent their adoption by the industry.

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/XXXXXXXXXXXXX

References

[1] Akinwande, D. et al. Graphene and two-dimensional materials for silicon technology. *Nature* **573**, 507–518 (2019).

[2] The International Roadmap for Devices and Systems. 2021 Edition (accessed 04 February 2023); https://irds.ieee.org/editions/2021.

[3] Goossens, S. et al. Broadband image sensor array based on graphene–CMOS integration. *Nat. Photon.* **11,** 366–371 (2017).

[4] Lin, Y. *et al*. Wafer-scale graphene integrated circuit. *Science* **332**, 1294–1297 (2011).

ACCELERATED ARTICLE PREVIEW

[5] Han, S., Garcia, A. V., Oida, S., Jenkins, K. A. & Haensch, W. Graphene radio frequency receiver integrated circuit. *Nat. Commun*. **5**, 3086 (2014).
[6] Shen, Y. et al, Variability and yield in h-BN-based memristive circuits: the role of each type of defect. *Adv. Mater.* **33**, 2103656 (2021).
[7] Chen, A. Forming voltage scaling of resistive switching memories. *71st Device Research Conference*. 181–182 (2013).
[8] Sawa. A, Resistive switching in transition metal oxides, *Mater. Today* **11**, 28–36 (2008).
[9] Website of Tektronix. 2600B Source Measure Units (SMU) Instruments (accessed 04 February 2023); https://www.tek.com/en/datasheet/2600b-source-measure-units-smu-instruments.
[10] Chen, A. Current overshoot during set and reset operations of resistive switching memories, *2012 IEEE International Reliability Physics Symposium (IRPS)* MY.2.1–MY.2.4 (IEEE, 2012)
[11] Website of Fujitsu. Non-volatile memory with very small operating current ReRAM (accessed 04 February 2023); https://www.fujitsu.com/jp/group/fsm/en/products/reram/.
[12] Website of Anandtech. Analyzing Intel-Micron 3D XPoint: The Next Generation Non-Volatile Memory (accessed 04 February 2023); https://www.anandtech.com/show/9470/intel-and-micron-announce-3d-xpoint-nonvolatile-memory-technology-1000x-higher-performance-endurance-than-nand.
[13] Kau, D. C. et al. Stackable cross point phase change memory, *2009 IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2009).
[14] Yang, Y. et al. Electrochemical dynamics of nanoscale metallic inclusions in dielectrics. *Nat. Commun.* **5,** 4232 (2014).
[15] Lanza, M. et al. Memristive technologies for data storage, computation, encryption, and radio-frequency communication. *Science* **376**, eabj9979 (2022).
[16] Yang, J., Strukov, D. & Stewart, D. Memristive devices for computing. *Nat. Nanotech.* **8**, 13–24 (2013).
[17] Merolla, P. A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
[18] Li, C. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. *Nat. Commun.* **9**, 2385 (2018).
[19] Ishii, M. et al. On-chip trainable 1.4M 6T2R PCM synaptic array with 1.6K stochastic LIF neurons for spiking RBM. *2019 IEEE International Electron Devices Meeting (IEDM)* 14.2.1–14.2.4 (IEEE, 2019).
[20] Diehl, P.U., Cook, M. Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Front. Comput. Neurosc.* **9**, 1662–5188 (2015).
[21] LeCun, Y., Cortes, C., Burges, C. MNIST handwritten digit database. ATT Labs. 2010 (accessed 04 February 2023); http://yann.lecun.com/exdb/mnist/.
[22] Website of Graphenea (accessed 04 February 2023);); https://www.graphenea.com/.
[23] Website of Graphene Flagship, news related to products released by the company Emberion, (accessed 04 February 2023); https://graphene-flagship.eu/graphene/news/emberion-launches-hyperspectral-array-sensor-at-laser-world-of-photonics/.
[24] Kim, M. et al. Analogue switches made from boron nitride monolayers for application in 5G and terahertz communication systems. *Nat. Electron.* **3**, 479–485 (2020).
[25] Shi, Y. et al. Electronic synapses made of layered two-dimensional materials. *Nat. Electron.* **1,** 458–465 (2018).
[26] Qian, K. et al. Hexagonal boron nitride thin film for flexible resistive memory applications. *Adv. Funct. Mater.* **26**, 2176–2184 (2016).
[27] Zhuang, P. et al. Nonpolar resistive switching of multilayer-hBN-based memories. *Adv. Electron. Mater.* **6**, 1900979 (2020).
[28] Yang, R. et al. Ternary content-addressable memory with MoS_2 transistors for massively parallel data search. *Nat. Electron.* **2,** 108–114 (2019).
[29] Wang, C. et al. 3D monolithic stacked 1T1R cells using monolayer MoS_2 FET and hBN RRAM fabricated at low (150°C) temperature. *2018 IEEE International Electron Devices Meeting (IEDM)* 22.5.1-22.5.4 (IEEE, 2018).
[30] Wu, F. et al. Vertical MoS2 transistors with sub-1-nm gate Lengths. *Nature* **603**, 259–264 (2022).
[31] Liu, L. et al. Uniform nucleation and epitaxy of bilayer molybdenum disulfide on sapphire. *Nature* **605**, 69–75 (2022).

ACCELERATED ARTICLE PREVIEW

[32] Marega, G. et al. Logic-in-memory based on an atomically thin semiconductor. *Nature* **587**, 72–77 (2020).
[33] Goda, A. 3-D NAND Technology Achievements and Future Scaling Perspectives. *2020 IEEE Transactions on Electron Devices* **67**, 1373–1381 (2020).
[34] Lin, W. L. et al. Grain boundary trap-induced current transient in a 3-D NAND flash cell string, *IEEE Trans. Electron Devices* **66**, 1734–1740 (2019).
[35] Automotive Standard Analog Robustness and Performance. ST Microelectronics. Brochure number BR2206AUTOICS (accessed 04 February 2023); https://www.st.com/en/automotive-analog-and-power.html#documentation.
[36] Niimi, H., Khamankar, R. & Alshareef, H. N. Method for non-thermally nitride gate formation for high voltage devices. *US patent* 6,730,566 B2 (2004).
[37] Mixed-Signal/CMOS – Tower Semiconductor. (n.d.) (accessed 04 February 2023); https://towersemi.com/technology/mixed-signal-cmos/.
[38] Minixhofer, R. et al. A 120V 180nm high voltage CMOS smart power technology for system-on-chip integration. *22nd International Symposium on Power Semiconductor Devices & IC's (ISPSD)*, 75–78 (2010).
[39] Saitoh, M. et al. HfO_2-based FeFET and FTJ for Ferroelectric-Memory Centric 3D LSI towards low-power and high-density storage and AI applications, *2020 IEEE International Electron Devices Meeting (IEDM)*, 375–378 (IEEE 2020).
[40] Goodman, P. Current and Future Uses of Gold in Electronics. *Gold Bulletin* **35**, 21–26 (2002).
[41] Summerfelt S. R. et al. Ferroelectric capacitor stack etch cleaning methods, *US Patent* 7,220,600 B2 (2007).

Methods

Microchip fabrication: The metal-oxide-semiconductor field-effect-transistor circuits were fabricated in a standard CMOS foundry. The size of the wafers is 200mm, and the technology node is 180nm. Each wafer contains 60 chips with a size of 2cm×2cm, and each of them contains different circuits, including the unfinished memristors, finished transistors, and unfinished 5×5 crossbar arrays of 1T1M cells. As we receive the silicon wafer without a passivation layer, first we cut it to separate the microchips, and then etched the native oxide by immersing them in a diluted hydrofluoric acid solution (10:1) for 1 min to etch the native oxide (SiO_2). This step is to expose the conductive tungsten vias to make a good electrical contact with 2D materials. Second, we transfer a sheet of multilayer h-BN (previously grown by CVD on a Cu foil) using a wet transfer method. One layer of PMMA with thickness around 300 nm was spin coated on the cut h-BN. The PMMA/h-BN/Cu sample was deposited on a $FeCl_3$ solution (0.1 g/mL) to etch the Cu substrate and, once the Cu disappeared, the resulting PMMA/h-BN sample was washed in diluted HCl solution (1 mol/L for 1 min) and deionized water (for 1 hour). The PMMA/h-BN sample was picked up using the native-oxide-free CMOS microchip and dried naturally in the dry box. The PMMA was then removed by immersing the sample in acetone for 24 hours. Third, we used photolithography (mask aligner from SUSS MicroTec, model MJB4) to expose the h-BN on top of the metallic pad. Then we used a dry etching method with Ar/O_2 plasma (Plasma Cleaner from PVA TePla America Inc, Model IoN 40) to etch the h-BN (300W for 10 minutes) and expose the pads. Finally, we used photolithography, electron beam evaporation (Kurt J. Lesker, model PVD75) and lift-off process (rinse in acetone for 1 minute) to pattern and deposit the top electrodes and/or drain electrodes (3nm Ti 40nm Au without breaking the vacuum, or 50nm Au or 50nm Ag). The process was simple and reproducible, although we believe it could be considerably improved if optimized methods in an industrial clean room are used. Ideally, the h-BN should be grown in large CVD systems (e.g., Aixtron[42]) and transferred on the wafers before cutting them in multiple microchips, using methods like laser debonding[43]. Note that even large companies are still using small (3 inch) tube furnaces to grow the h-BN for their prototypes[44]. We also confirm that getting the microchips finalized and etch the passivation film before transferring the h-BN works well.

Device characterization: The morphology of the devices was investigated by an optical microscope (DM 4000M, Leica), AFM (Dimension Icon, Bruker), and TEM (Titan Themis, FEI). The thin lamellae for TEM inspection were prepared using a scanning electron microscope provided with focused ion beam (Helios G4 UX, Thermo Fisher Scientific). The electrical characterization was performed by using two probe stations (both M150, Cascade) connected to different semiconductor

ACCELERATED ARTICLE PREVIEW

parameter analysers: a Keithley 4200 and a Keysight B1500A. All the I-V curves under DC voltages were collected using the Keithley 4200 in the ramped voltage sweep (RVS) mode, for which 3 source-measure-unit (SMU) are needed for drain, source and gate. Also, all the I-V curves under pulse mode were collected by Keysight B1500A with two Waveform Generator/Fast Measurement Units (WGFMU) connected to the drain and source. An Agilent E3631A DC Power Supply was used to apply constant voltage stress on the transistor gate as gate voltage for pulse measurement. All the endurance plots were collected using the recommended method described in reference 45.

SNN simulation: The SNN architecture[20] has been developed using Brian2[46], an SNN simulator written in Python. The learning process is based on the empirical measurement of STDP made in the 1T1M cells combining an Au/Ti/h-BN/W memristor on a CMOS transistor. We considered the variability of the network by running a Monte Carlo engine. We benchmark the accuracy of the SNN during image classification[20] of the Modified National Institute of Standards and Technology (MNIST) dataset of handwritten digits[21] under an unsupervised learning scheme. We propose a circuit-level model for the neuron-synapse-neuron system, as well as its implementation in SPICE. A deep description of the SNN and its performance is given in Supplementary Note 3.

Data availability

The data needed to evaluate the conclusions in this work are publicly available online at https://doi.org/10.5281/zenodo.7607096. The datasets that we used for benchmarking are publicly available in reference 21. The training methods are provided in reference 20.

Code availability

The simulator Brian2 used here is publicly available in reference 46. The codes used for the simulations described in Methods are publicly available online at https://doi.org/10.5281/zenodo.7607096.

Methods references

[42] Gomollon-Bel, F. Cost effective, large scale graphene with AIXTRON systems. Graphene Flagship. https://graphene-flagship.eu/graphene/news/cost-effective-large-scale-graphene-with-aixtron-systems/.

[43] Phommahaxay, A. et al. The growing application field of laser debonding: from advanced packaging to future nanoelectronics. *2019 International Wafer Level Packaging Conference (IWLPC)*, 1–8 (2019).

[44] Chen, T. A. et al. Wafer-scale single-crystal hexagonal boron nitride monolayers on Cu (111). *Nature* **579**, 219–223 (2020).

[45] Lanza, M. et al. Standards for the characterization of endurance in resistive switching devices. *ACS Nano* **15**, 17214–17231 (2021).

[46] Stimberg, M., Brette, R. & Goodman, D. F. Brian 2, an intuitive and efficient neural simulator. *eLife* **8**, e47314 (2019).

Acknowledgement

This work has been supported by the Ministry of Science and Technology of China (grants no. 2019YFE0124200, 2018YFE0100800), the National Natural Science Foundation of China (grant no. 61874075) and the Baseline funding scheme of the King Abdullah University of Science and Technology.

Authors contribution section

KZ, SP and ML designed the experiments. XL and HW provided the CMOS wafers. KZ, YS, YY, WZ and OA fabricated the hybrid 2D/CMOS microchips. KZ and SP characterized the microchips. KZ measured the logic gates. RL, HF and TW assisted in electrical characterization. FA did the simulation of the neural network. MAV, BF, AM and MF assisted in the fabrication, characterization and simulation

ACCELERATED ARTICLE PREVIEW

of structures included in intermediate version of this manuscript that were finally removed. MMR, JBR, XZ, HA, TG and DI gave technical recommendations on fabrication, characterization and/or simulations. ML wrote the manuscript, which was revised by all the authors.

Competing interests. The authors declare no competing interests.

Captions

Figure 1 | Fabrication of hybrid 2D/CMOS memristive microchips. a, Photograph of the 2cm×2cm microchips containing the CMOS circuitry. **b-c**, Optical microscope images of a part of the microchip containing a 5×5 crossbar array of 1T1M cells, as-received and after fabrication (respectively). The size of the squared pads is 50μm×50μm. **d-f**, Topographic maps collected with atomic force microscopy of the vias in the 5×5 crossbar arrays on the wafers as-received, after native oxide etching, and after the transfer of the h-BN sheet (respectively). **g**, Optical microscope image of a finished 5×5 crossbar array of 1T1M, i.e., after h-BN transfer and top electrodes deposition. **h**, High-angle annular dark-field cross-sectional scanning transmission electron microscope image of a 1T1M cell in the crossbar array. The inset, which is 20nm×16nm, shows a cross-sectional transmission electron microscopy image of the Au/Ti/h-BN/W memristor on the via; the correct layered structure of h-BN can be seen.

Figure 2 | Electrical characterization of h-BN/CMOS based 1T1M cells. a, Electrical characterization of ~0.053μm^2 Au/Ti/h-BN/W structures, showing erratic current fluctuations and no stable resistive switching (each colour line corresponds to one RVS with two polarities). **b**, Typical output characteristic of all standalone CMOS transistor. **c**, Typical non-volatile bipolar RS measured in most 1T1M cells with an Au/Ti/h-BN/W memristor and a CMOS transistor (when applying V_G=1.1V). **d**,**f**, Endurance plots of 1T1M cell showing around 1.4 million cycles and 1 million cycles for write pulse durations of 0.1 and 1ms (respectively). **e,** R_{LRS} and R_{HRS} in a 1T1M cell when applying pulsed voltage stresses of different duration. **g**, Endurance plot showing non-volatile bipolar RS at V_G=1V for 1T1M cells using Au/h-BN/W memristors. All the endurance tests have been conducted following the recommended characterization process described in reference 45. **h**, Voltage and current versus time in a 1T1M cell with Ag/h-BN/W memristor, showing a low switching energy. **i**, Spike-timing dependent plasticity (STDP) characteristic of the 1T1M cell with Au/Ti/h-BN/W memristor. Before STDP characterization, the devices are always tuned to the same initial conductance (lower box charts, which relate to the right Y axis).

Figure 3 | Implementation of a spiking neural network using CMOS/h-BN based 1T1M cells. a, Structure of the considered SNN. Each MNIST image is reshaped as a 784×1 column vector, and the intensity of the pixels is encoded in terms of the firing frequency of the input neurons. The only trainable synapses are those connecting the input layer with the excitatory layer, and they are modelled with the STDP characteristic of the CMOS/h-BN based 1T1M cells. The learning is unsupervised, and the neurons are labelled only after the training. These label-neuron assignments are then feed to the decision block altogether with the firing patterns of the neurons, to infer the class of the image presented in the input. **b**, Evolution of the synaptic connections between the input and excitatory layers during training for the case of 400 excitatory/inhibitory neurons. The red square identifies 784 synapses arranged in a 28×28 representation. **c,** Confusion matrix indicating the classification accuracy for each class from the dataset. **d,** Classification accuracy as a function of the number of presented training images for the neural network comprising 400 excitatory/inhibitory neurons. The error bars show the standard deviation for 50 Monte Carlo simulation runs for every accuracy point. **e,** Circuit schematic of the proposed neuron-synapse-neuron block combining h-BN based 1T1M cells and CMOS circuitry. The colours indicate the complete neuron (grey surrounding box), the core block (light-blue box) and the individual building blocks (light-red boxes). **f**, SPICE simulation of the pre- and post- synaptic signals applied to the CMOS/h-BN based 1T1M. **g,** SPICE simulation of the neuron's membrane potential. The firing events progressively separate from each other due to the adaptative firing threshold.

Extended Data Figure 1 | As-received silicon wafers. Photo of the 200mm wafer received from the foundry containing all the 2cm×2cm microchips, made of CMOS circuits.

Extended Data Figure 2 | Morphological analysis of the h-BN. Cross-sectional TEM images of the as-grown multilayer h-BN sheet on Cu. The layered structure is evident, although it contains local defects (i.e., lattice distortions) produced during the CVD process – they are not related to the FIB process because our

ACCELERATED ARTICLE PREVIEW

exfoliated samples never show these features. Those native defects are necessary for the observation of memristive effect. The scale bars (from top to down) 6 nm, 6 nm, 4 nm and 3 nm.

Extended Data Figure 3 | Nano-chemical characterization of the h-BN stack. Electron energy loss spectroscopy signal showing the correct structure of the h-BN stack on the CMOS microchip. The image also shows that the titanium electrode can absorb a remarkable amount of oxygen. This oxygen peak is not observed when using gold or silver electrodes.

Extended Data Figure 4 | Electrical characterization of multiple h-BN memristors. Current versus voltage plots collected for different 1T1M cells using Au/Ti/h-BN/W memristors with a lateral size of 0.053μm^2. The plots have been drawn with 75% transparency to emphasize the general trend. Each plot contains 50 cycles or more. The first ramped voltage stress is shown in blue colour; it is normally more insulating than the rest, indicating the need of an activation step that softly degrades the h-BN and produces RS. However, no filament is completely formed (see Supplementary Note 1).

Extended Data Figure 5 | Multilevel RS in 1T1M cells. Current versus time plots for a 1T1M cell when using different end voltage of the reset RVS. Multiple stable conductance levels can be programmed. The composition of the top electrode is Au/Ti. The read voltage applied is 0.1 V. We measured the retention of several states up to ~7 hours, and the rest up to 200 s due to laboratory availability. The devices show no sign of conductance degradation for any measured state after ~7 hours, although typical drift (<5%) is detected, which is acceptable. According to the Semiconductor Research Corporation [V. Zhirnov, Decadal Plan for Semiconductors: New Trajectories for Memory and Storage, presented in the 2022 Non-Volatile Memory Technology Symposium 2022, December 7th-11, Stanford, USA], the unsaved bytes turnover and lifetime is currently less than 15 hours, meaning that no retention time >10 years will be needed for multiple memory applications. Similarly, the requirements in terms of retention time for memristive electronic synapses is much more relaxed than in most memories, and correct functioning with short retention times of few hours has been claimed [see Kuzum, D., Yu, S., & Wong, H. P. Synaptic electronics: materials, devices and applications. Nanotechnology 24, 382001 (2013)].

Extended Data Figure 6 | Adjustment of R_{LRS}/R_{HRS} in 1T1M cells with Au/Ti top electrodes. Values of R_{HRS} and R_{LRS} depending on the pulse amplitude of write and erase operation. The value of the R_{LRS}/R_{HRS} window can also be adjusted in this way.

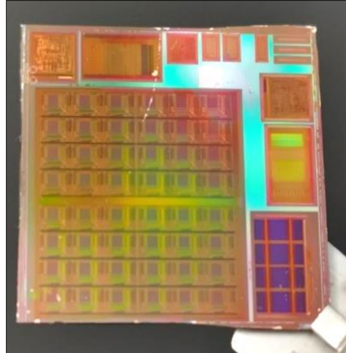
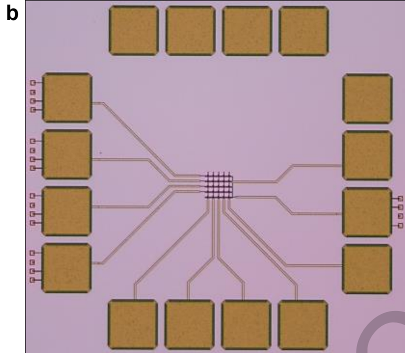
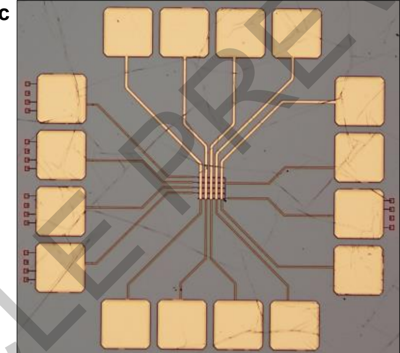
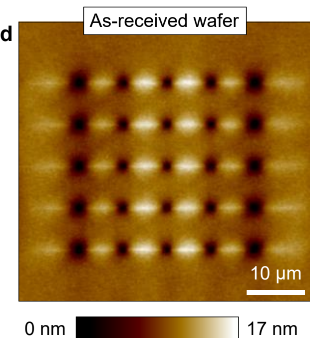
Extended Data Figure 7 | Switching time of 1T1M cells using top Au/Ti electrodes. Voltage and current versus time plots showing the non-volatile bipolar RS with high temporal resolution. The switching time for the set transition is 232μs, and the one for the reset transition is 783ns.

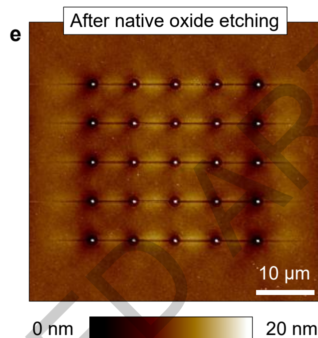
Extended Data Figure 8 | Non-volatile bipolar RS in the 1T1M cells depending on the chemical composition of the top electrode. Voltage and current versus time plots showing reproducible non-volatile bipolar RS in devices with Au/Ti, Au and Ag top electrode. The operation voltages and switching times are different, providing enough flexibility to operate in different regimes.

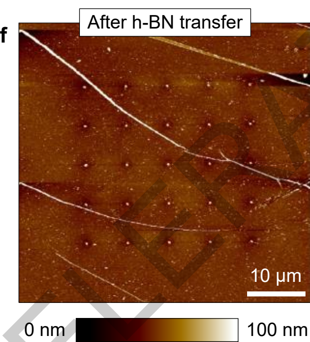
Extended Data Figure 9 | In-memory computation with crossbar arrays of h-BN memristors. a,**b,** Schematic and current versus time plot for "or" and "implication" logic operations (respectively) made with the crossbar array of 1T1M cells in the hybrid 2D/CMOS microchip. The value of the output current changes depending on the conductance of the memristors M1 and M2, which is indicated as "00", "01", "10" and "11" on each segment of the plot ("0" and "1" means low and high conductance, respectively).

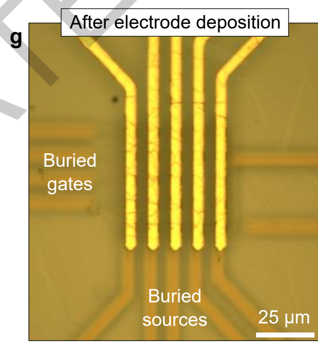
Extended Data Figure 10 | Non-exhaustive list of materials for microchip manufacturing. Schematic representation of a microchip; the position of our memristor is indicated with a green star. Our memristors are integrated in the last metallization layer of the BEOL, where Au liners are often employed, and very far from the transistors in the FEOL. Apart from this, Au wire bonding to the socket is also often employed. Hence, the use of Au electrodes of the memristors is not a concern for this type of devices/circuits. STI means shallow trench isolation.

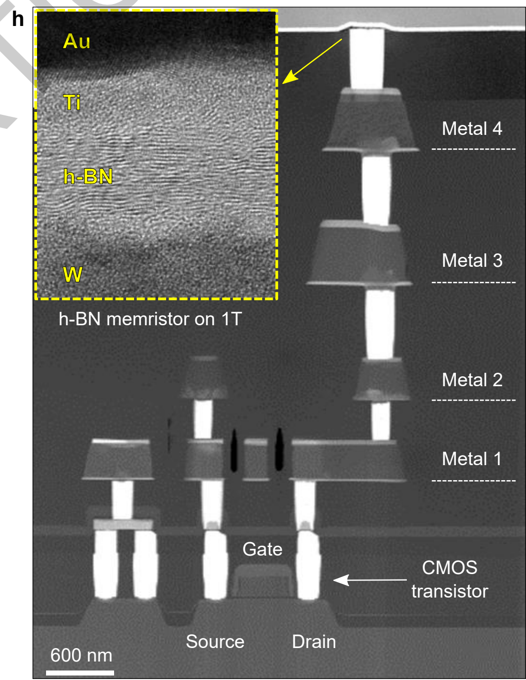
ACCELERATED ARTICLE PREVIEW

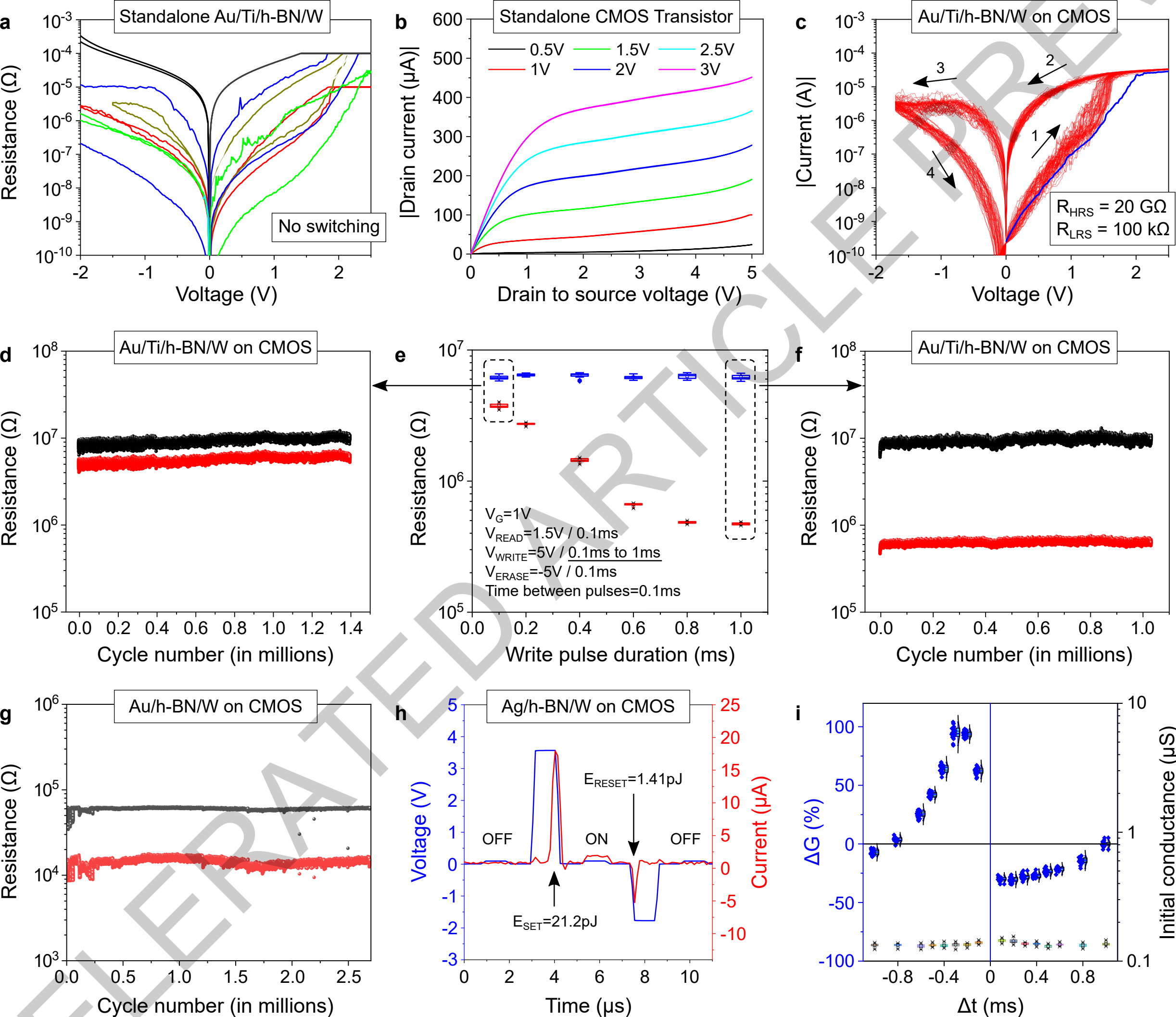




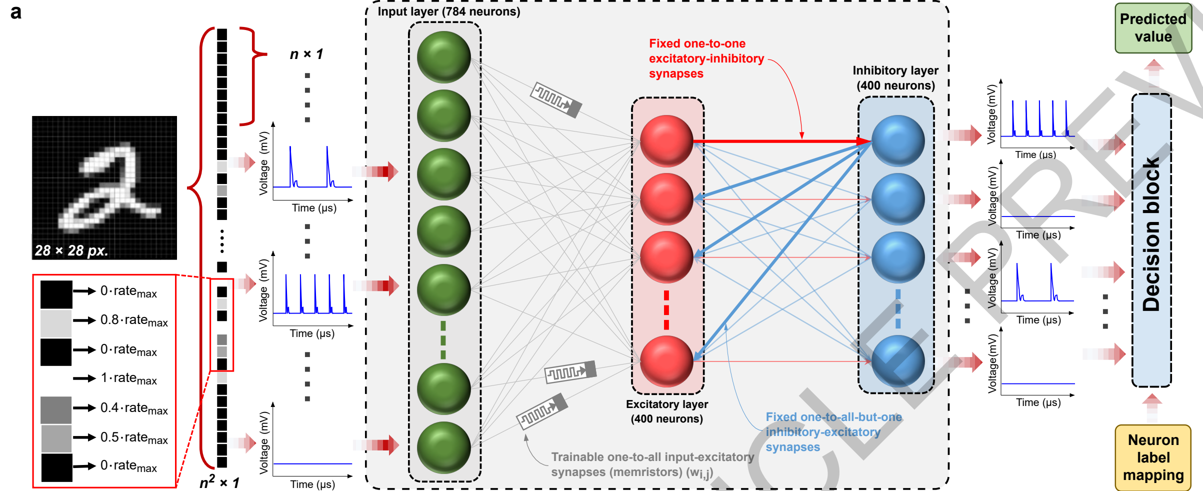


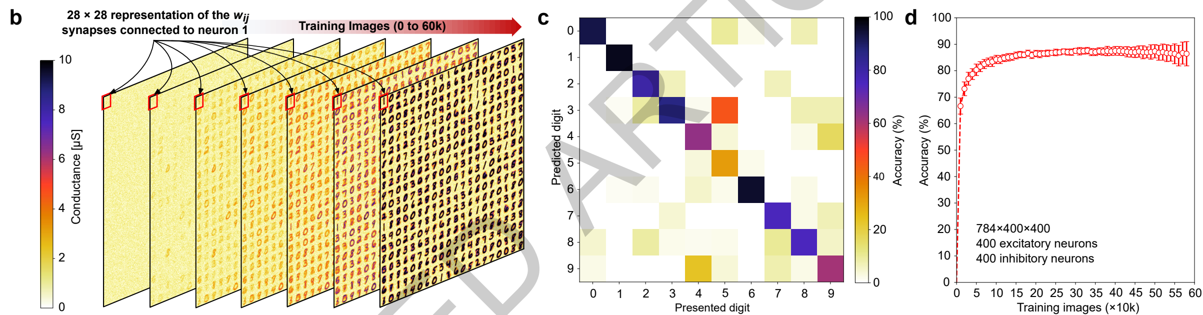


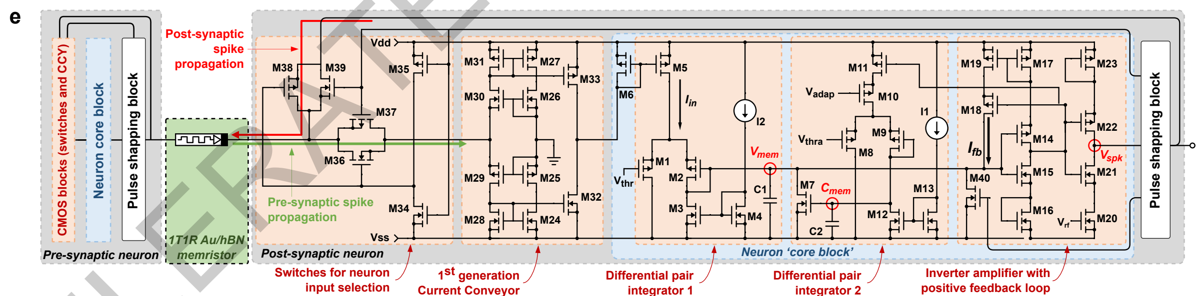
ACCELERATED ARTICLE PREVIEW

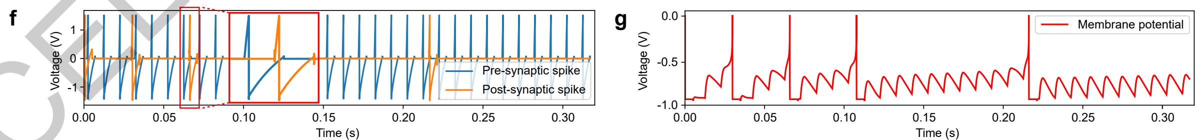


ACCELERATED ARTICLE PREVIEW

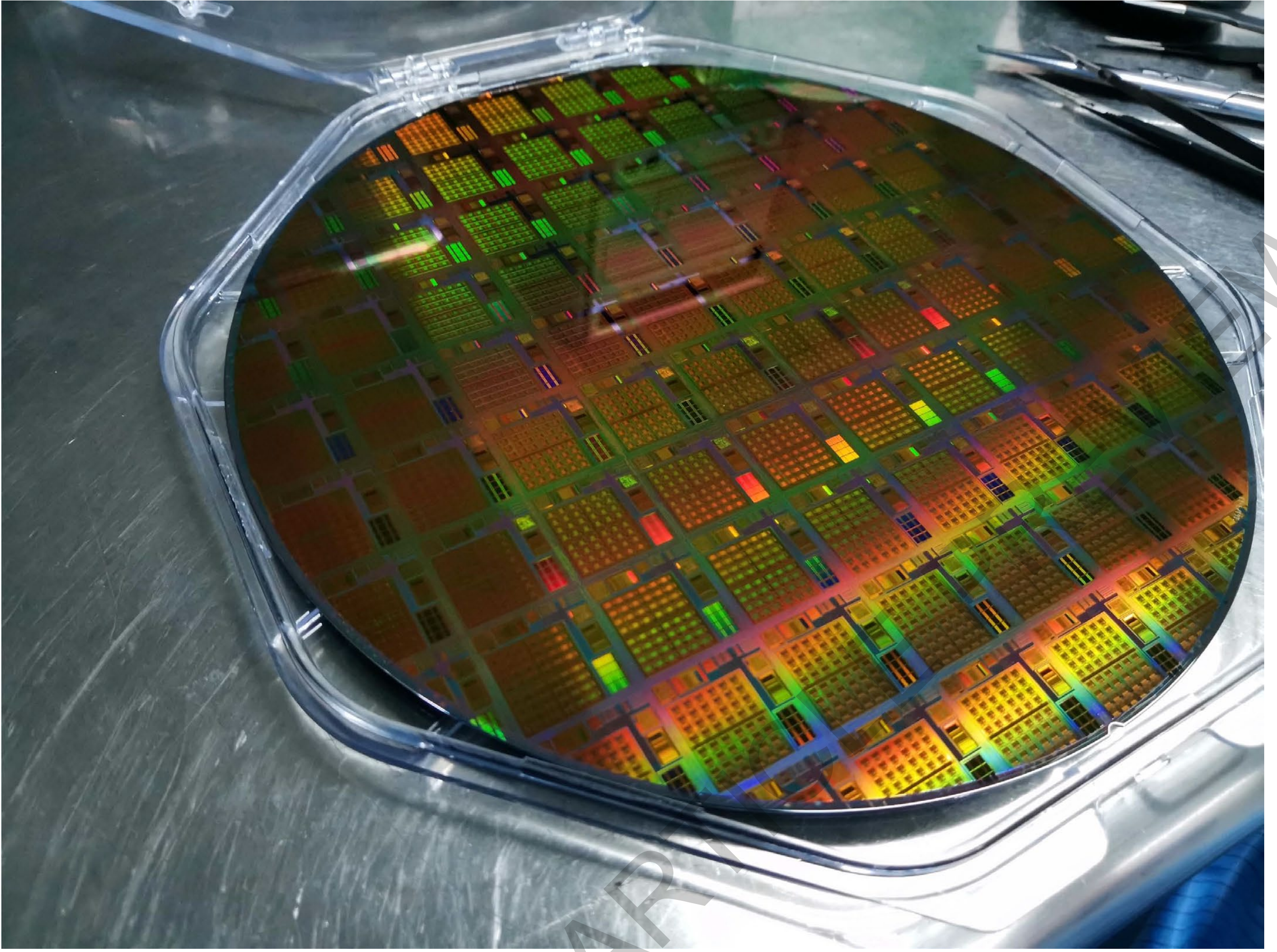






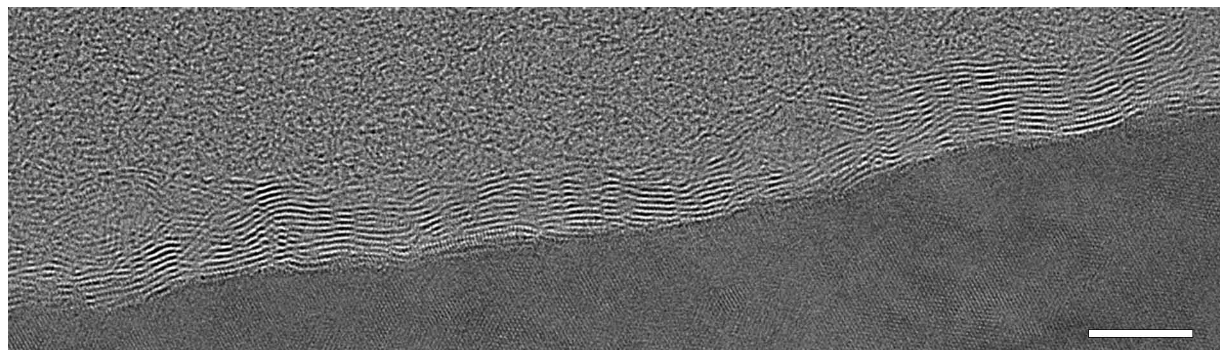
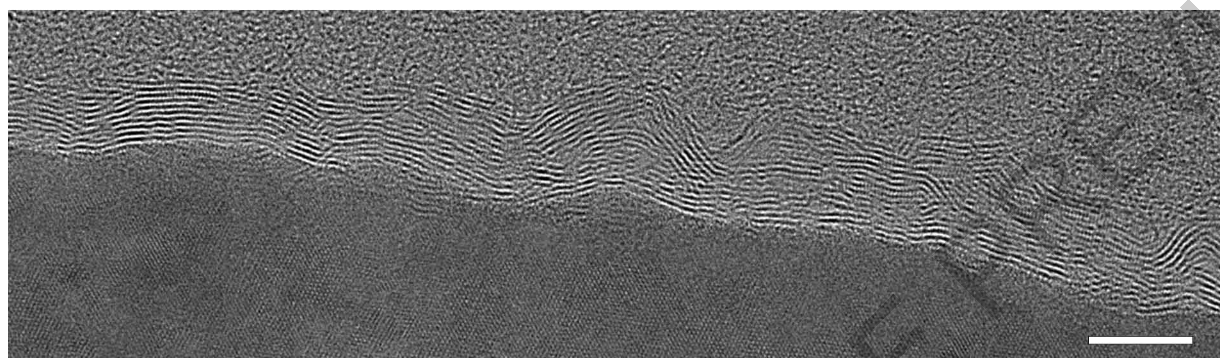
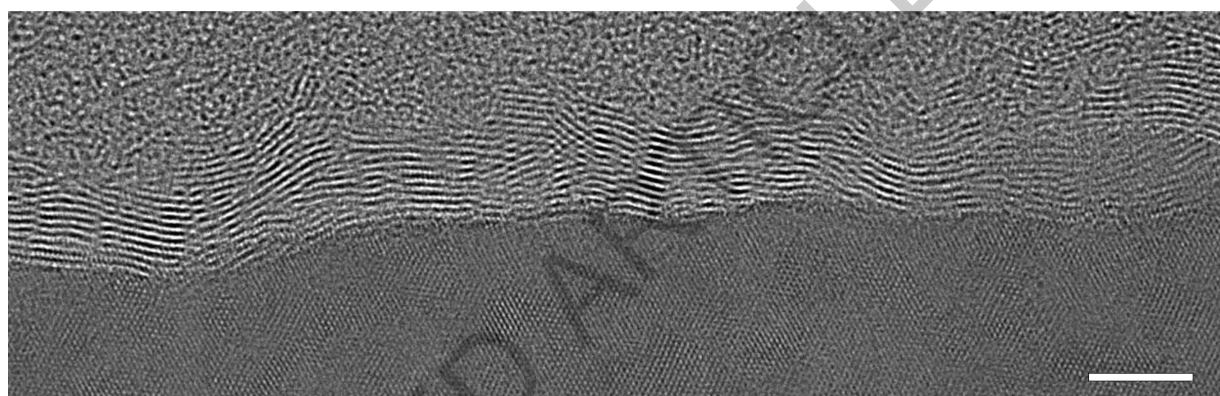
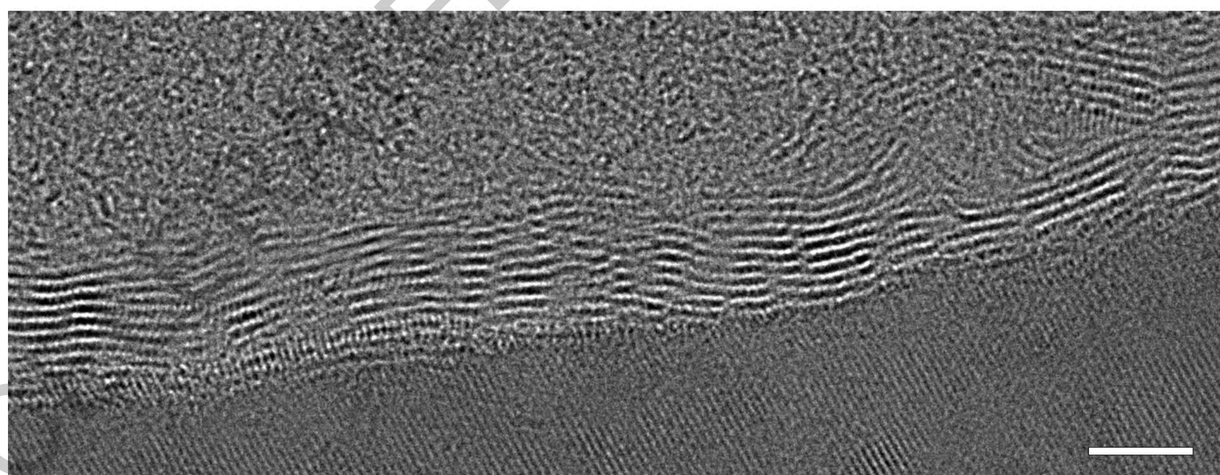


ACCELERATED ARTICLE PREVIEW



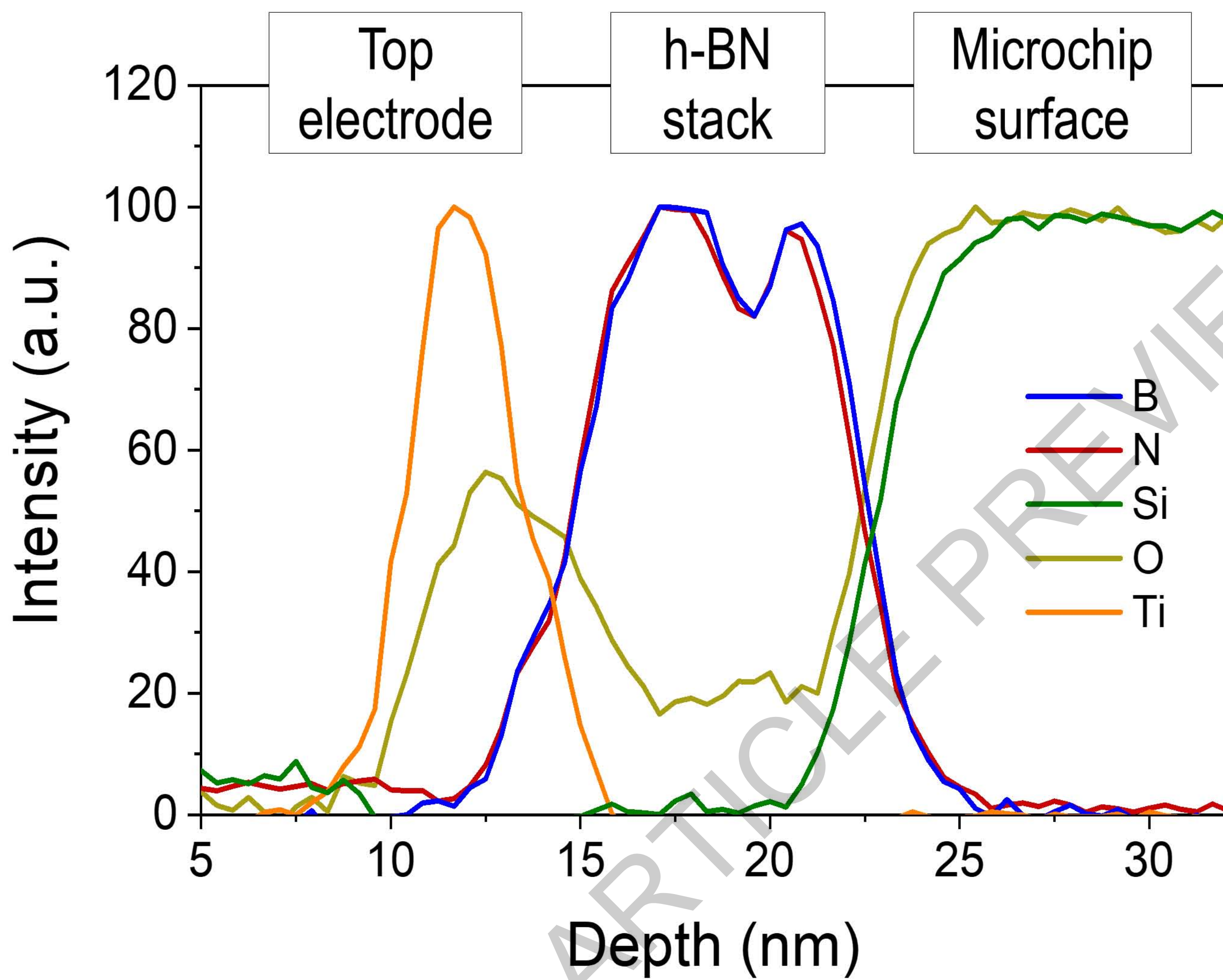
Extended Data Fig. 1

ACCELERATED ARTICLE PREVIEW

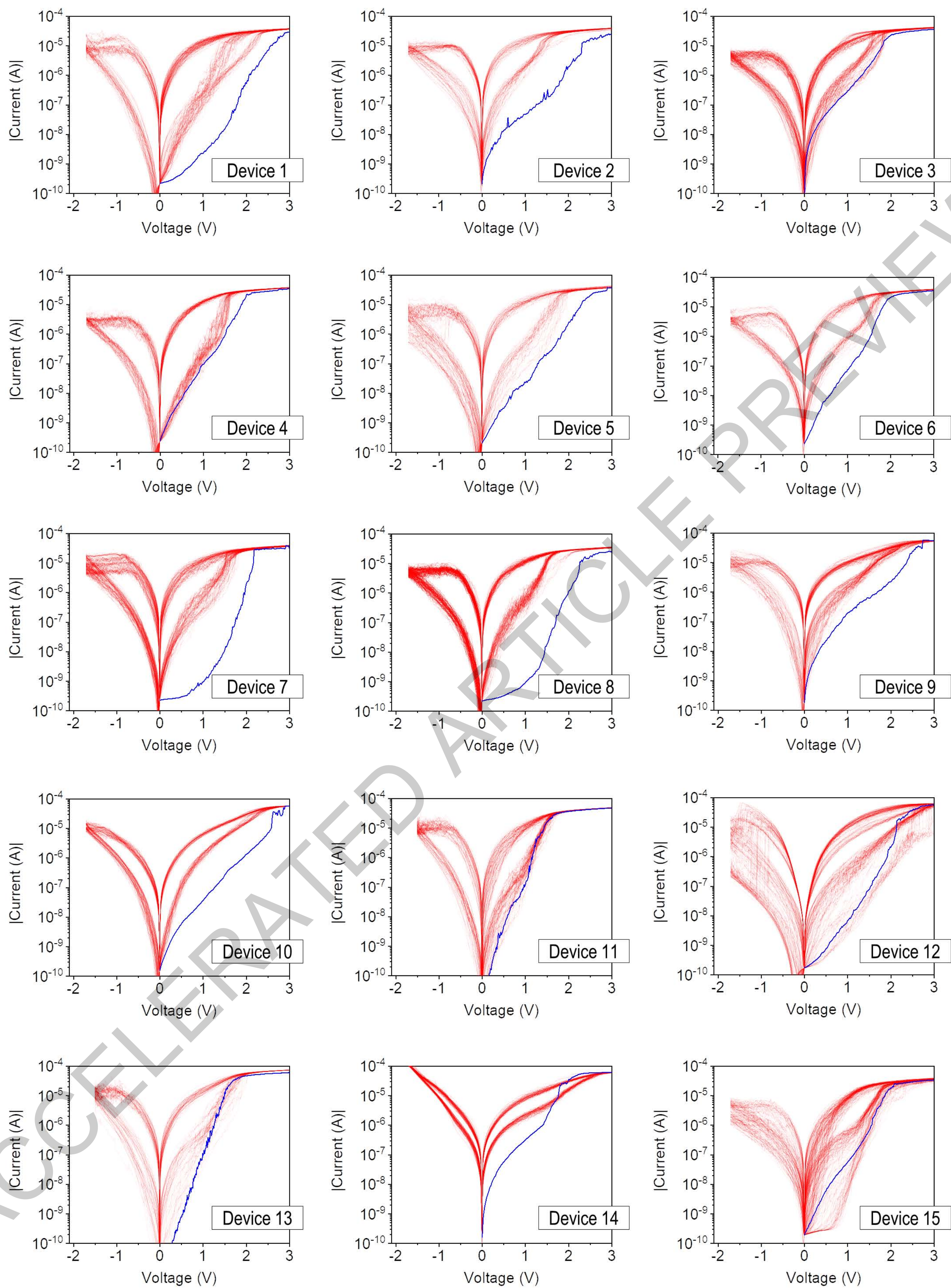
Extended Data Fig. 2

ACCEPTED ARTICLE PREVIEW



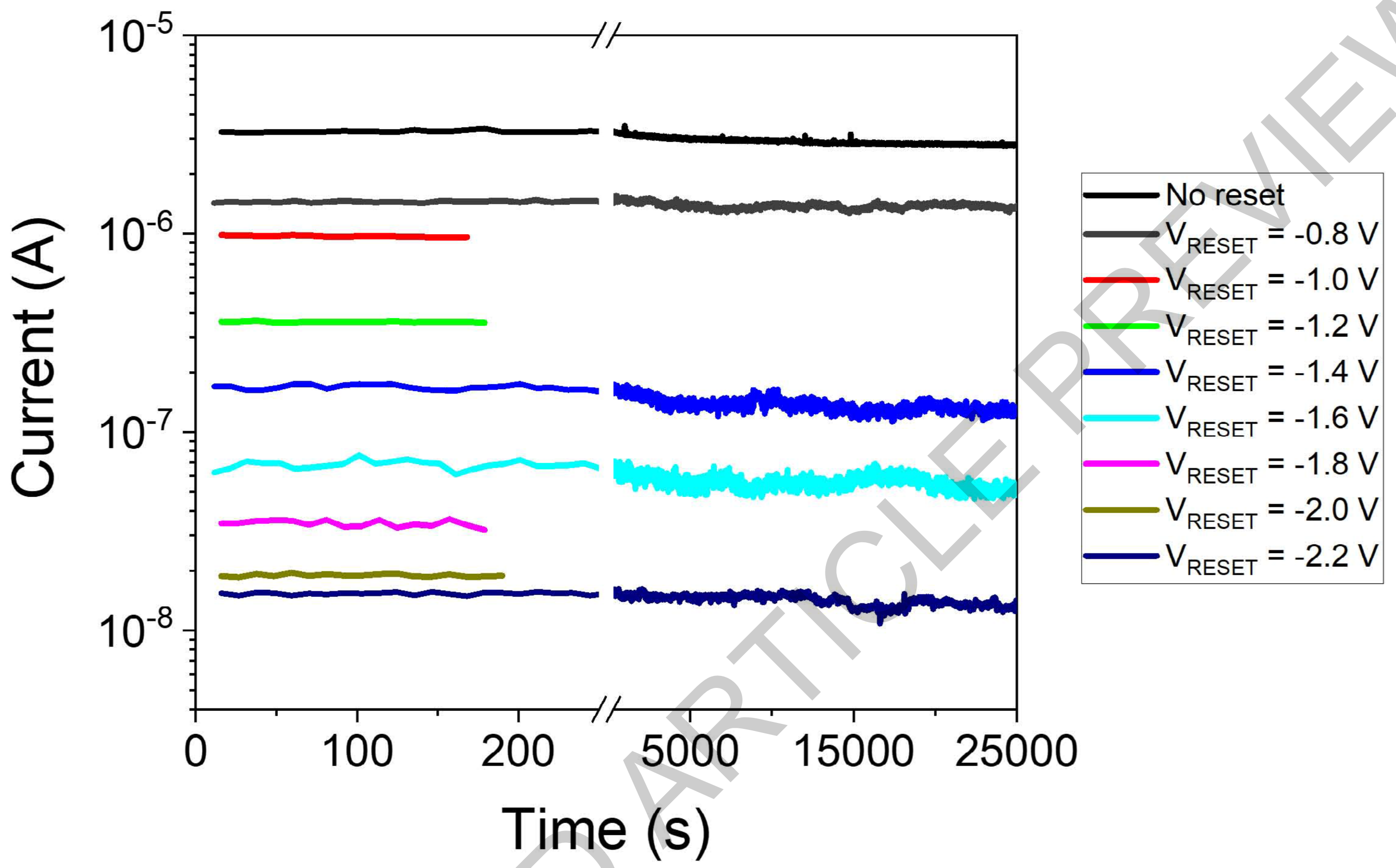
Extended Data Fig. 3

ACCELERATED ARTICLE PREVIEW



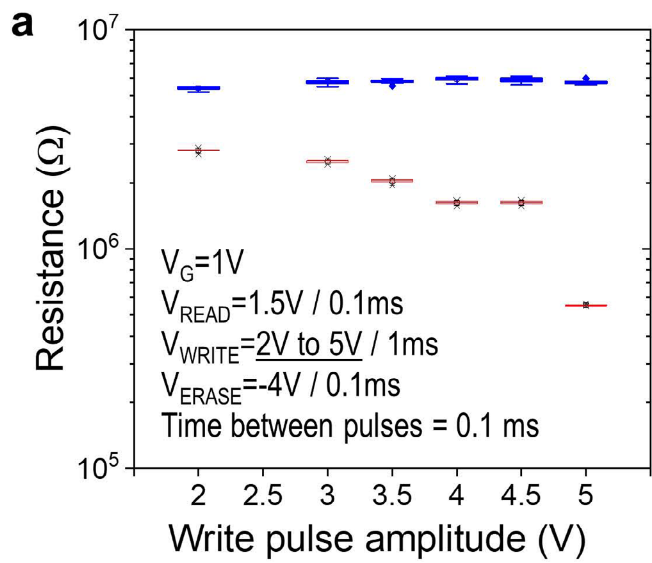
ACCELERATED ARTICLE PREVIEW

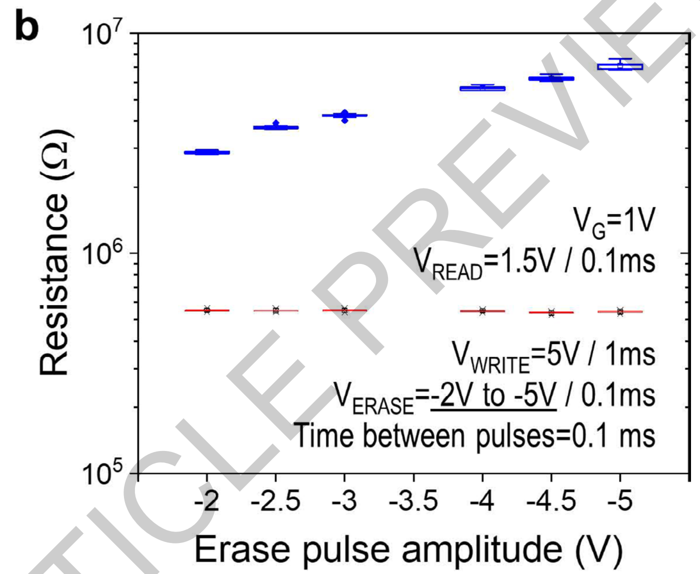
Extended Data Fig. 4



Extended Data Fig. 5

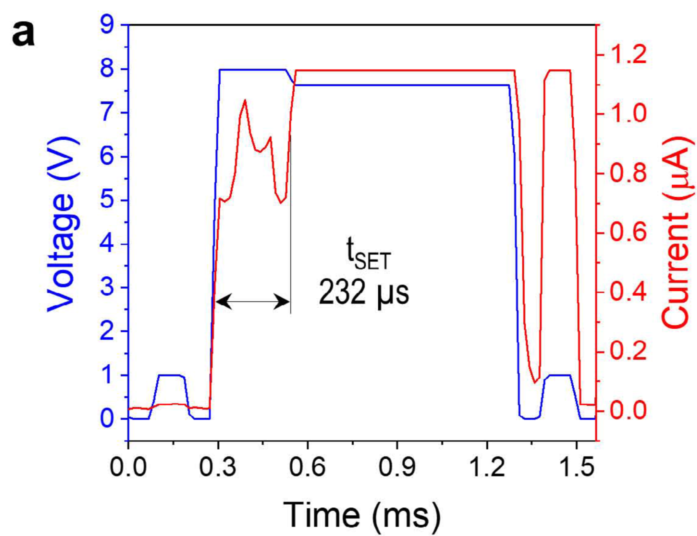
ACCELERATED ARTICLE PREVIEW

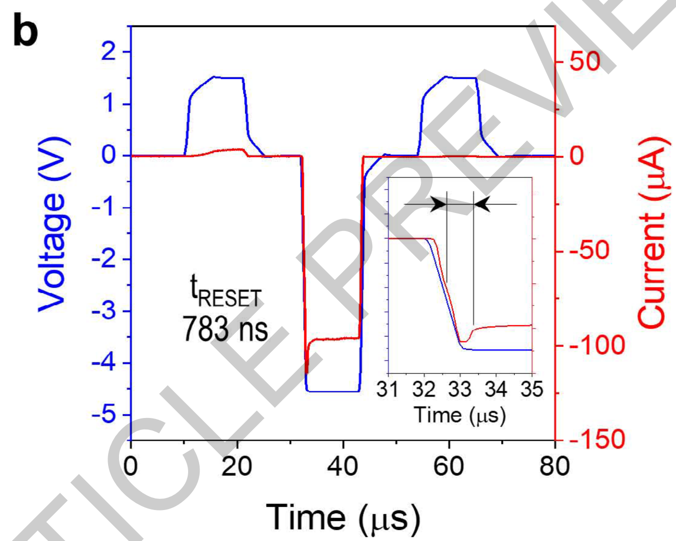




Extended Data Fig. 6

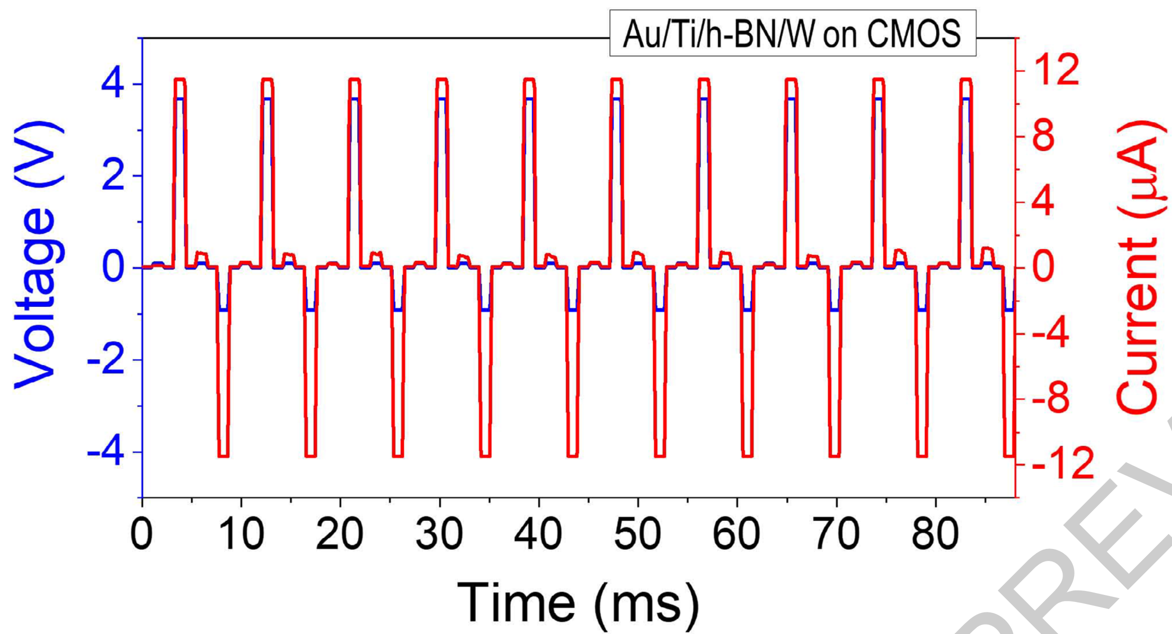
ACCELERATED ARTICLE PREVIEW

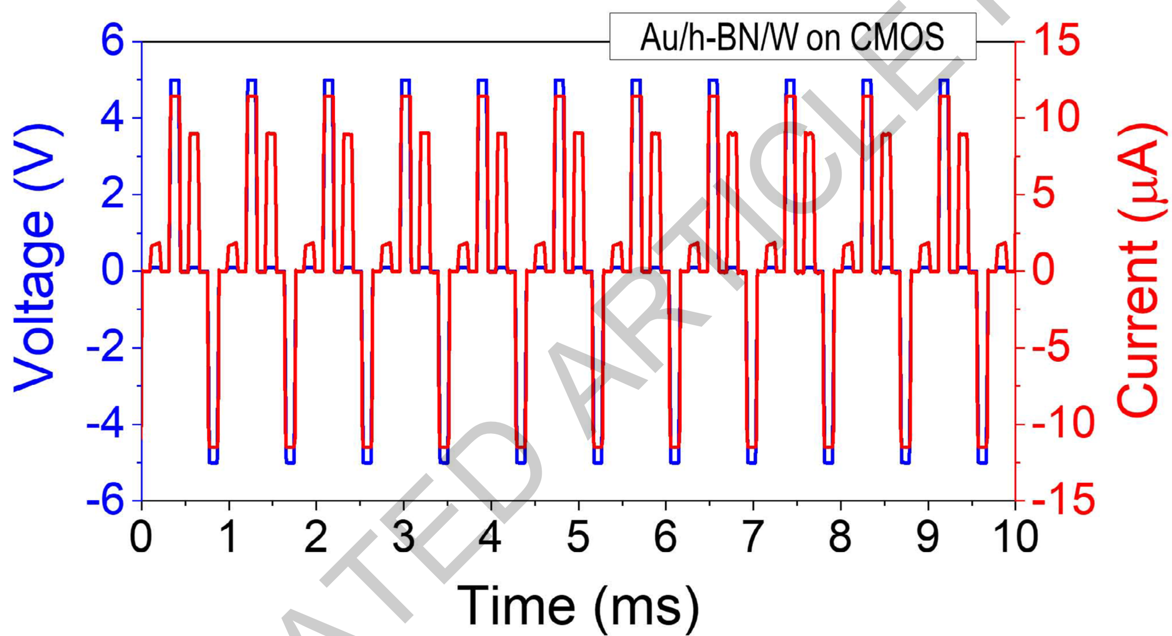


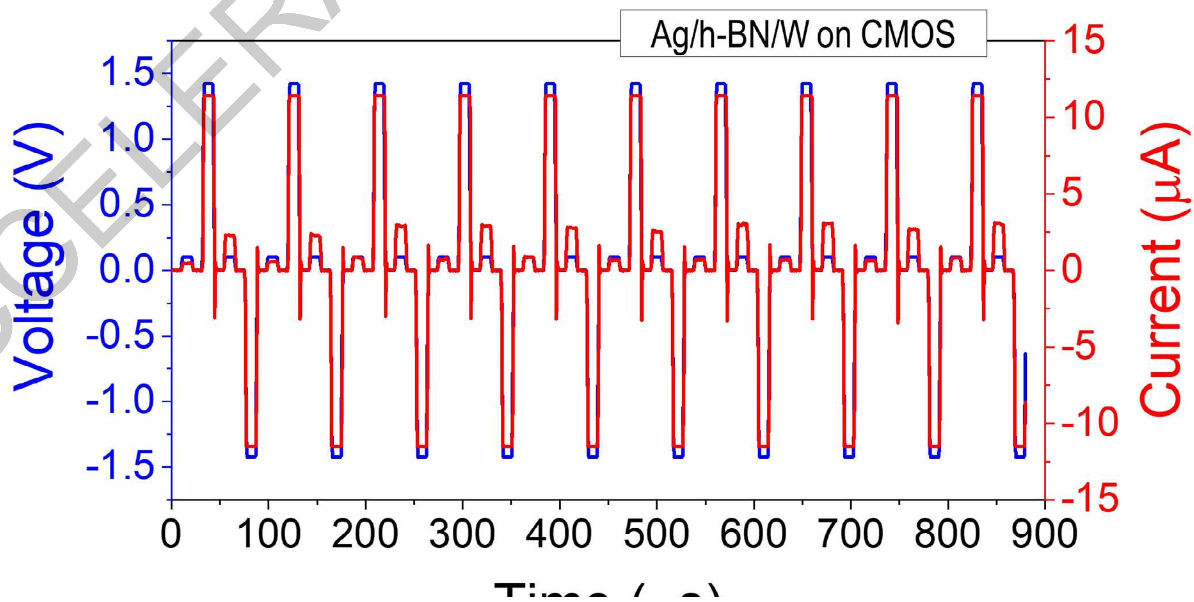


Extended Data Fig. 7

ACCELERATED ARTICLE PREVIEW

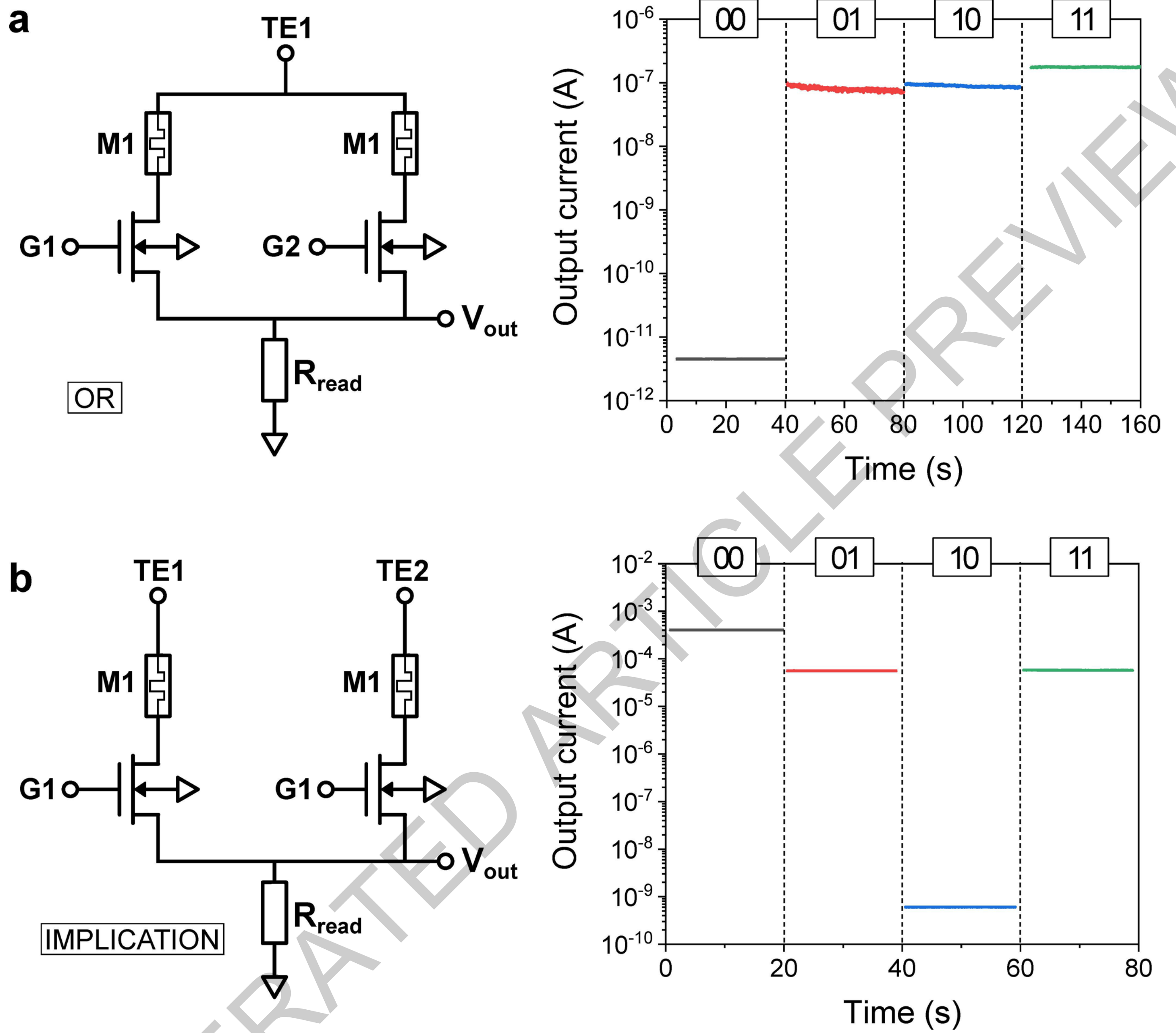






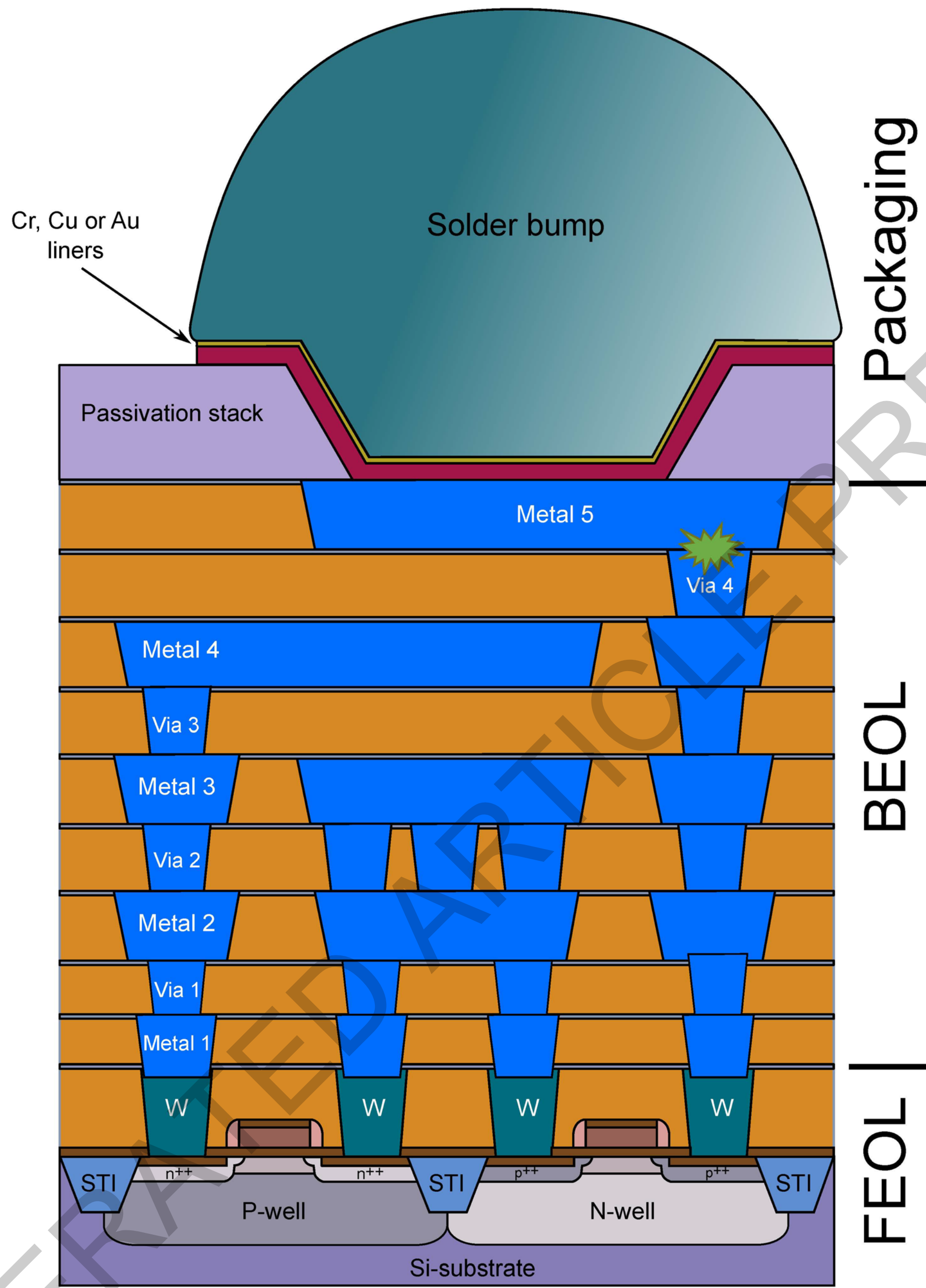
Extended Data Fig. 8

ACCELERATED ARTICLE PREVIEW



Extended Data Fig. 9

ACCELERATED ARTICLE PREVIEW



Extended Data Fig. 10

ACCELERATED ARTICLE PREVIEW