NEUROMORPHIC CHIPS

2D materials–based homogeneous transistor-memory architecture for neuromorphic hardware

Lei Tong1, Zhuiyi Peng2, Runfeng Lin3, Zheng Li4, Yilun Wang1, Xinyu Huang1, Kan-Hao Xue1, Hangyu Xu2, Feng Liu2, Hui Xia2, Peng Wang2, Mingsheng Xu4, Wei Xiong1, Weida Hu2, Jianbin Xu2, Xiniang Zhang1, Lei Ye1*, Xiangshui Miao1

In neuromorphic hardware, peripheral circuits and memories based on heterogeneous devices are generally physically separated. Thus, exploration of homogeneous devices for these components is key for improving module integration and resistance matching. Inspired by the ferroelectric proximity effect on two-dimensional (2D) materials, we present a tungsten diselenide–on–lithium niobate cascaded architecture as a basic device that functions as a nonlinear transistor, assisting the design of operational amplifiers for analog signal processing (ASP). This device also functions as a nonvolatile memory cell, achieving memory operating (MO) functionality. On the basis of this homogeneous architecture, we also investigated an ASP-MO integrated system for binary classification and the design of ternary content-addressable memory for potential use in neuromorphic hardware.

Recently proposed diverse neuron-inspired hardware based on various emerging nanomaterials has effectively advanced neural networks (1–6), especially for two-dimensional (2D) materials. 2D materials can provide a platform to develop transistor architectures for memory operating (MO)—including field-effect transistors (FETs), tunneling transistors, junction transistors, ferroelectric (FE) transistors, and ferromagnetic transistors—owing to their rich electrostatic control capabilities (1, 7). 2D materials–based FE FETs, FE bipolar junction transistors (BJTs), and FE tunneling transistors exhibit large on-off resistance ratios, fast operation, low power consumption, nonvolatile electronic control, and weight updates under reversible polarization (1) because of the strong proximal coupling of FE materials with 2D materials. Therefore, these 2D materials–based FE proximal coupled devices are being intensively investigated for neuromorphic computing (1, 4), in which they are used as memories by dynamically modulating the FE polarization to program the conductivity of superjacent 2D channels (8, 9). Achieving computing tasks requires these memories to be integrated with peripheral circuits, because analog signal processing (ASP) is essential before and after MO (2). However, peripheral circuits are generally based on complementary metal-oxide semiconductor (CMOS) transistors, and thus the heterogeneous architectures between memory cells and peripheral circuits lead to their physical separation, making it necessary to consider module integration compatibility issues for chip design (2). In addition, an emerging challenge regards how to achieve efficient resistance matching between heterogeneous device architectures as device dimensions are scaling down, which may hinder the pursuit of higher performance and energy efficiency (10). Therefore, it is crucial to explore the integration between ASP and MO.

Designing an ASP-MO integrated system with a homogeneous device architecture for peripheral circuits and memory cells offers the potential to realize the above-mentioned issues, which can also be realized by the mechanism of 2D-FE proximal coupling. On the one hand, FE polarization proximity-induced nonvolatile electronic gating in 2D materials enables the design of nonlinear transistors, including p-n diodes and BJTs (8, 11–16). On the other hand, FE polarization can modulate the built-in potential in BJTs (17), enabling nonvolatile memory functionalities with an improved on-off resistance ratio. Moreover, the reconfigurable FE polarization domains can seamlessly manipulate arrayed doping domains in 2D materials, showing potential for the fabrication of massive cascaded devices with enhanced compactness. Therefore, a homogeneous, 2D material–based FE proximal coupled BJT architecture is proposed to design peripheral circuits for ASP, as well as nonvolatile memory cells for MO, enabling development of an ASP-MO integrated system.

In this work, seamlessly arrayed periodically polarized LiNbO3 (LNO) domains formed a grating-like structure (fig. S1), which effectively tailored the WSe2 channels into seamlessly arrayed junctions. ASP, MO, and their corresponding cascade were investigated to demonstrate the success of an integrated system based on the same device architecture, which included a WSe2 channel crossing three FE domains. Our operational amplifier (OPAMP) was designed for ASP (18), and memory cells with encoded synapse weights were cascaded with the OPAMP to demonstrate the applicability in binary classification (19). Furthermore, ternary content-addressable memory (TCAM) with a two-transistor-two-resistor (2T2R) configuration was designed with the homogeneous transistor-memory architecture, yielding a ratio of 898.4 between the high-resistance state (HRS) and the low-resistance state (LRS). Such an integrated system architecture could provide a feasible approach to solve the heterogeneous issue and improve neuromorphic applications.

Few-layer WSe2 flakes were exfoliated and transferred onto LNO to demonstrate the reconfigurable electronic functionalities under FE proximal coupling (WSe2 characterization and optical images are shown in fig. S1). The basic device functioned as a nonlinear transistor when the domain polarization state was fixed under zero gate voltage ($V_G$) (Fig. 1A), as the FE proximal coupling induces carrier doping in the WSe2 (Fig. 1B). This doping mechanism was indicated by the Kelvin probe force microscopy (KPFM) mapping shown in fig. SIC, where the higher (or lower) surface potential induced by the polarization-down ($P_u$) [or polarization-up ($P_d$)] domain was in accordance with the p-doping (or n-doping) nature (8). Transfer curves for the intrinsic WSe2 and FE-doped WSe2 FETs are shown in fig. SII. The neutral point shifts were 3.8 and −6.2 V for the $P_d$ and $P_u$ domains, respectively, corresponding to a hole doping density of $2.07 \times 10^{12}$ cm$^{-2}$ and an electron doping density of $3.37 \times 10^{12}$ cm$^{-2}$. This doping character induced a built-in potential of −0.43 eV and a depletion width (w) of 48.25 nm (fig. SIM) (20). For a basic n-p-n BJT on the $P_d$-$P_u$-$P_d$ domains, current amplifications were measured under the base, collector, and emitter voltage control. The common-base configuration in Fig. 1E showed an average gain of $a = \beta = 0.979$ ($I_c$, collector current; $I_e$, emitter current) for the active region, and the common-emitter configuration in Fig. 1F showed a maximum gain of $\beta = 5.7 = 11.2$ ($I_e$, base current), offering the capability to design analog circuits. More details about signal rectification, amplification, and performance uniformity are shown in figs. S2 to S5.

The basic device could also operate as a nonvolatile memory (Fig. 1C), the mechanism of which differs from those of conventional FE-FET and MemFlash (1, 4, 21). The FE...
The polarization state was fixed to be $P_u$ for the collector and emitter, and the HRS and LRS were dominated by the built-in potential \( (17) \), which was controlled by \( V_g \) for the base. In the potentiation (or depression) process, a positive \( V_g \) from 6 to 9 V (or a negative \( V_g \) from −6 to −9 V) changed the FE polarization state for the base to $P_u$ (or $P_d$), resulting in a low built-in potential with enhanced channel conductance (or a high built-in potential with reduced channel conductance) (Fig. 1D) \((21, 22)\). The mechanism was indicated from Tong et al., Science 373, 1353–1358 (2021) 17 September 2021.
the modulated current rectifications under the collector-emitter voltage ($V_{ce}$) sweep for the potentiation (Fig. 1G) and depression (Fig. 1H) processes, respectively ($V_{be}$) = 0 V; $V_g$ step of 0.5 V]. This rectification behavior can provide a high resistance ratio between the HRS and LRS. In contrast to the conventional FE-FET structure with the WSe$_2$ channel placed on a single FE domain (fig. S6), the resistance for the FE polarization state of P$_{u}$P$_{d}$P$_{u}$ and P$_{u}$P$_{u}$P$_{u}$ showed an average resistance ratio of ~10$^6$ under driving voltages of $V_{ce}$ = −3 V, $V_{be}$ = 0 V, and $V_g$ = 0 V (Fig. 1I). Moreover, as an artificial synapse, the conductance update for the potentiation or depression process with a $V_g$ step of 0.1 V under $V_{ce}$ = 1 V and $V_{be}$ = 0 V (Fig. 1J) was encoded as a synapse weight (fig. S7A), and the on-off resistance ratio was ~10$^3$. The stable memory performance is shown in fig. S7 for measurements of multiple cycles, and the memory performances were compared with those from other studies of diverse memory architectures (table S1). The few-layer WSe$_2$
film grown by chemical vapor deposition was then transferred and patterned on LNO to fabricate an ASP-MO integrated system.

The OPAMP was designed by cascading multiple basic devices for ASP (Fig. 2, A to C) (18), with the input electrodes designated X1 and X2. The dc driving voltages were collector voltage ($V_{cc}$) = emitter voltage ($V_{ee}$) = 6 V, the resistances were $R_1 = R_2 = 10$ megohms, the capacitance ($C$) was 1 μF, and the input bias current was varied between 30 and 50 nA to adjust the performance. The OPAMP characteristics were discussed and compared with those of previously reported CMOS-based devices (table S2). We applied an ac input signal $V_i = 0.2\sin200\pi \times t$ (t, time) (Fig. 2D) at only one input port, and the other port was grounded. The output signal $V_o$ was inverted when the signal was input at port X1 (Fig. 2E), the output $V_o^+$ was noninverted when the signal was input at port X2 (Fig. 2F), and the

---

**Fig. 3. ASP-MO integration for binary classification.**

(A) Class 1 with letter patterns H, U, S, and T and class 0 with number patterns 2, 0, 2, and 1. (B) Schematic of the classification with the sigmoid activation function. (C) Schematic of the circuit diagram, where the memory cells were integrated with the TIA and VC. (D) Output character of the VC (solid red curve), which was similar to the sigmoid function (dashed blue curve). (E to G) Training dataset label (E), predicted class in the simulation (F), and predicted class in the experiment (G). (H to J) Average Score(V) (H), accuracy (I), and cost (J) for the training process in the simulation and experiment.
voltage gain was related to the ratio \( R_f/R_o \). For the addition operation, port X2 was grounded, and both input signals were cascaded to port X1, with a resistance \( R_o \) for negative feedback. The two ac input signals \( V_1 = 0.2 \sin(200\pi t) \) and \( V_2 = 0.3 \sin(200\pi t) \) and the output signal \( V_o \) are shown in Fig. 2G. Under the condition of \( R_f = R_k = R_o = 2 \) megohms, the output signal was \(-V_o = V_1^1 + V_1^2\). When a capacitance was used for negative feedback, the OPAMP achieved the integral calculation with a time constant of \( \tau = R_o C = 2 s \) (\( R_o = 2 \) megohms, \( C = 1 \mu F \)). The square input signal \( V_1 \) and the corresponding integrated output signal \( V_o = -\frac{V_1}{\tau} t \) are shown in Fig. 2H. The triangle-shaped input signal and the approximately sine wave–shaped integrated output signal are shown in Fig. 2I.

On the basis of the reconfigurable functionalities, the ASP-MO integrated system was used to demonstrate the proof of principle for binary classification (23, 24). The input three pixel–by–three pixel patterns for the letters H, U, S, and T were labeled as class 1, and those for the numbers 2, 0, and 1 were labeled as class 0 (Fig. 3A shows eight randomly selected patterns; and the entire training dataset is shown in fig. S8). The calculation schematic and circuit diagram are shown in Fig. 3, B and C, respectively. The pixels were coded into nine input voltages \( (V_{ee}) \) to calculate the weighted average current (Eq. 1) in nine memory cells. The weighted average current was converted into a voltage \( \text{Score}(V) \) in the transimpedance amplifier (TIA) (19, 25), with a resistance \( (R_o) \) of 10 megohms. The nonlinear output voltage (Eq. 2) was processed by the voltage comparator (VC) with a reference voltage \( (V_{\text{ref}}) \) of 1.5 V, which was similar to the sigmoid function (Fig. 3D).

\[
\text{Score}(V) = w_0 + w_1 V_1^1 + w_2 V_2^2 + w_3 V_3^3 + \ldots + w_j V_j^j
\]

\[
V_{\text{out}} = \begin{cases} V_{ee}, & \text{Score}(V) > V_{\text{ref}} \\ V_{ce}, & \text{Score}(V) < V_{\text{ref}} \end{cases}
\]

The drive voltages were \( V_{ee} \) (6 V) and \( V_{ce} \) (0 V), and the conductance was updated after calculating the cross-entropy cost for each batch, as summarized in fig. S9, C and D. The system was trained for 30 epochs, after which the
ductance for the address read at Xr and thus (HRS). The matched state exhibited a low conductance, defined when R1 was in the HRS and R2 were in the HRS (Fig. 4, E and F), with an average conductivity of ~0.4 nS (Fig. 4I). The cumulative probability was analyzed in fig. S10.

In addition to binary classification and TCAM design, the basic device should be suitable for various applications, including digital computing (fig. S11) and artificial neural systems with optical sensing abilities (fig. S12) (28, 29). The basic device should also be applicable for the design of analog-to-digital converters, digital-to-analog converters, and analog filters (2). The device size can be reduced by scaling down the polarization domain size, which will improve the current gain and integration density. A lower coercive voltage would be helpful for achieving a lower power consumption, which can be attained by reducing the LNO thickness or using other FE materials. Finally, inspired by conventional CMOS-based chips, we have proposed a neuromorphic ASP-MO 3D stacking system derived from 2D integration (fig. S13). The main challenge lies in the growth of wafer-scale, high-quality 2D materials, and recent works have achieved promising breakthroughs to overcome this challenge (30). Thus, this homogeneous transistor-memory architecture will help to promote analogous neuromorphic systems on-chip.

REFERENCES AND NOTES
1. C. Liu et al., Nat. Nanotechnol. 15, 545–557 (2020).
2. Q. Xia, J. J. Yang, Nat. Mater. 18, 309–323 (2019).
3. L. Danail and Nat. Electron. 2, 596–605 (2019).
4. D. Ielmini, H. P. Wong, Nat. Electron. 1, 333–343 (2018).
5. W. Zhang et al., Nat. Electron. 3, 371–382 (2020).
6. A. Sebastian, M. Le Gallo, R. Khaddar-Aljameh, E. Eleftheriou, Nat. Nanotechnol. 15, 529–544 (2020).
7. M. L. B. Uthama et al., Nat. Electron. 2, 60–65 (2019).
8. G. Wu et al., Nat. Electron. 3, 43–50 (2020).
9. X. Wang et al., Nat. Commun. 11, 1109 (2021).
10. P. Kapur, J. P. McVittie, K. C. Saraswat, IEEE Trans. Electron Dev. 49, 590–597 (2002).
11. L. Lv et al., Nat. Commun. 10, 3331 (2019).
12. C. Bauern et al., Nat. Commun. 6, 6136 (2015).
13. J.-W. Chen et al., Nat. Commun. 9, 3143 (2018).
14. H. Lu et al., Nat. Commun. 5, 5518 (2014).
15. N. Park et al., ACS Nano 9, 10729–10736 (2015).
16. H. Mulasosmanovic, E. T. Breyer, T. Mikoajick, S. Slesazeck, Nat. Electron. 3, 391–397 (2020).
17. Z. Xiao, J. Song, D. K. Ferry, S. Ducharme, X. Hong, Phys. Rev. Lett. 118, 236801 (2017).
18. D. K. Polyushkin et al., Nat. Electron. 3, 486–491 (2020).
19. L. Meniel et al., Nature 579, 62–66 (2020).
20. B. Wen et al., ACS Nano 13, 5335–5343 (2019).
21. M. K. Kim, J.-S. Lee, Nano Lett. 19, 2044–2050 (2019).
22. S. Boynt et al., Nat. Commun. 8, 14736 (2017).
23. S. Ambrogio et al., Nature 558, 60–67 (2018).
24. S. Seo et al., Nat. Commun. 11, 3936 (2020).
25. C. Li et al., Nat. Electron. 1, 52–59 (2018).
26. K. Ni et al., Nat. Electron. 2, 521–529 (2019).
27. R. Yang et al., Nat. Electron. 2, 108–114 (2019).
28. L. Tong et al., Nat. Commun. 11, 2308 (2020).
29. C. Y. Wang et al., Sci. Adv. 6, eaba6173 (2020).
30. X. Xu et al., Science 372, 195–200 (2021).
31. L. Tong et al., 2D materials-based homogeneous transistor-memory architecture for neuromorphic hardware, version 1.0. Zenodo (2023), https://doi.org/10.5281/zenodo.5163973.

ACKNOWLEDGMENTS
Funding: This work was supported by the National Natural Science Foundation of China (grants 61974050, 51872257, 61927817, 61725055, 11774036, 61905266, 11991063, and 61774067), the National Key Research and Development Program of China (grants 2019YFB2205100 and 2019YFB2203102), the Priority Research Program of Chinese Academy of Sciences (grants XDB22030400 and XD40300200), the Fundamental Research Funds of Jiangsu Natural Science Foundation (grant S2019190082) and the Fundamental Research Funds of Zhejiang Province (grant LR20F040001).

Author contributions: L.T., X.M., W.H., and L.Y. conceived the idea and designed the experiments. L.T., Z.P., and R.L. performed device fabrication and characterization. X.Z., J.X., and X.Z. helped with device fabrication. L.T., Z.P., and R.L. conducted cumulative probability analysis. H.Xia conducted the KPFM measurement. F.L. conducted optoelectronic measurements. K.-H.X. conducted binary logic function measurements. K.-H.X. and L.Y. wrote the manuscript. All authors provided critical comments and analyses.

Competing interests: The authors declare no conflicts of interest.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. All findings of this study are archived on Zenodo (31).

SUPPLEMENTARY MATERIALS
https://science.org/doi/10.1126/science.abg3161
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 and S2
References (S2–S4)

24 December 2020; resubmitted 5 June 2021
Accepted 6 August 2021
Published online 19 August 2021
10.1126/science.abg3161
2D materials–based homogeneous transistor-memory architecture for neuromorphic hardware
Lei Tong Zhuirui Peng Runfeng Lin Zheng Li Yilun Wang Xinyu Huang Kan-Hao Xue Hangyu Xu Feng Liu Hui Xia Peng Wang Mingsheng Xu Wei Xiong Weida Hu Jianbin Xu Xinliang Zhang Lei Ye Xiangshui Miao

Science, 373 (6561), • DOI: 10.1126/science.abg3161

Memory and logic in the same device
Future artificial intelligence applications and data-intensive computations require the development of neuromorphic systems beyond traditional heterogeneous device architectures. Physical separation between a peripheral signal-processing unit and a memory-operating unit is one of the main bottlenecks of heterogeneous architectures, blocking further improvements in efficient resistance matching, energy consumption, and integration compatibility. Tong et al. present a transistor-memory architecture based on a homogeneous tungsten selenide-on-lithium niobate device array (see the Perspective by Rao and Tao). Analog peripheral signal preprocessing and nonvolatile memory were possible within the same device structure, promising diverse neuromorphic functionalities and offering potential improvements in neuromorphic systems on-chip. —YS

View the article online
https://www.science.org/doi/10.1126/science.abg3161
Permissions
https://www.science.org/help/reprints-and-permissions