Applied Physics Letters

From memristive devices to neuromorphic systems

Cite as: Appl. Phys. Lett. **122**, 110501 (2023); https://doi.org/10.1063/5.0133044
Submitted: 01 November 2022 • Accepted: 21 February 2023 • Published Online: 13 March 2023

Yi Huang, Fatemeh Kiani, Fan Ye, et al.

AIP Publishing

Applied Physics Letters PERSPECTIVE scitation.org/journal/apl

From memristive devices to neuromorphic systems

Cite as: Appl. Phys. Lett. **122**, 110501 (2023); doi: 10.1063/5.0133044
Submitted: 1 November 2022 · Accepted: 21 February 2023 ·
Published Online: 13 March 2023

Yi Huang, Fatemeh Kiani, Fan Ye, and Qiangfei Xia[a)]

AFFILIATIONS

Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA

[a)]Author to whom correspondence should be addressed: qxia@umass.edu

ABSTRACT

Progress in hardware and algorithms for artificial intelligence (AI) has ushered in large machine learning models and various applications impacting our everyday lives. However, today's AI, mainly artificial neural networks, still cannot compete with human brains because of two major issues: the high energy consumption of the hardware running AI models and the lack of ability to generalize knowledge and self-adapt to changes. Neuromorphic systems built upon emerging devices, for instance, memristors, provide a promising path to address these issues. Although innovative memristor devices and circuit designs have been proposed for neuromorphic computing and applied to different proof-of-concept applications, there is still a long way to go to build large-scale low-power memristor-based neuromorphic systems that can bridge the gap between AI and biological brains. This Perspective summarizes the progress and challenges from memristor devices to neuromorphic systems and proposes possible directions for neuromorphic system implementation based on memristive devices.

INTRODUCTION

Advances in machine learning algorithms and hardware have enabled impressive performance in various artificial intelligence (AI) applications, such as computer vision, language processing, game playing, etc.[1] However, training these models requires tremendous computational resources and energy. As reported in a survey paper, the estimated CO_2 emissions from training a big transformer are five times that of an average U.S. car in its lifetime.[2] Also, most AI models only carry out specific tasks in specific contexts, far from natural intelligent systems, which can generalize knowledge according to different goals in various circumstances.[3] Thus, energy-efficient hardware with generality, adaptability, and flexibility has been prioritized in developing next-generation intelligent systems. Neuromorphic systems modeling biological brains in very-large-scale integration (VLSI) circuits are promising candidates for future AI regarding energy efficiency and learning ability.[4] Digital neuromorphic chips implementing spiking neural networks (SNNs) with high energy efficiency, including TrueNorth,[5] Loihi,[6,7] and Tianjic,[8,9] have been fabricated and applied to real-time object detection, multimodal tracking, robotics, etc. However, they are all based on digital computing and store data in digital memories, which are different from biological brains with spike-encoded and event-driven representations. Moreover, thousands of transistors are required to model the functionality of merely one biological neuron in digital chips,[6] limiting the intelligent systems' area and energy efficiencies. As possible solutions for these problems, memristors whose resistance can be dynamically reconfigured not only enabled the analog in-memory computing to improve power consumption, latency, and area of neuromorphic chips but also modeled biological synapses and neurons in a single device for spike-encoded neural networks. Both nonvolatile memristors, whose conductance retains after removing electrical bias, and volatile memristors, whose conductance relaxes back to OFF states upon removing the bias after ON switching, have been demonstrated in the hardware implementations of artificial neural networks (ANNs) and SNNs. Nonvolatile memristors are mostly used as in-memory computing components in crossbar arrays to accelerate the vector-matrix multiplications (VMMs) in ANNs, while volatile memristors are mainly used to emulate the dynamic behaviors of synapses and neurons in SNNs, which mimic the physics of the human brain and neural system. The breakthroughs in memristor devices laid a solid foundation for neuromorphic systems in analog in-memory computing and brain dynamics modeling.[10] However, the development of peripheral circuits, architectures, models, etc., to provide a route from memristor devices to neuromorphic systems is still in its infancy.

In this Perspective, we briefly summarize recent progress achieved from memristive devices to experimental neuromorphic hardware, evaluate the challenges from data conversions to learning and communications, and propose research opportunities from circuit

innovations to architecture designs to ultimately build a neuromorphic system with comparable energy efficiency and learning ability to biological brains.

PROGRESS: FROM MEMRISTIVE DEVICES TO NEUROMORPHIC HARDWARE

Inspired by biological brains where synapses and neurons work as both processing and memory units for analog signals, analog in-memory computing has been implemented using memristors for traditional ANNs. To conduct VMM in parallel, the most common and computationally expensive operation in ANNs, memristor devices are integrated into crossbar arrays. Memristor crossbar arrays use physical laws, Ohm's law for multiplications and Kirchhoff's current law for summations, to perform VMM within one step, resulting in orders of magnitude higher computing throughput. Also, analog data from sensors can be directly processed,[11–13] and synaptic weights are stored in memristors where VMM performs, significantly improving the power consumption and speed compared to frequent and inefficient data movement as in traditional von Neumann architectures.

The fundamental research is on memristor devices and crossbar arrays for VMM with lower power consumption and higher throughput. Various memristive devices have been developed for crossbar arrays with desirable properties for representation capability, switching speed and energy, reliability, and device density.[14] However, from memristive devices to crossbar arrays, sneak path current is the first issue to be addressed. Active memristor arrays with one transistor and one memristor (1T1R) structure in each cross point have dominated because the transistors can be used as selectors to limit sneak path currents and control the currents through memristors when programming.[12] Two-terminal selectors could potentially increase the packing density, but research and development in this direction are left behind, mainly because of the demanding requirements on the device properties. While some research engineered memristors to achieve high uniformity and linearity to build passive crossbar arrays with only one memristor in each crosspoint,[15] the scalability of passive memristor arrays remains a challenge for system-level designs because of the leakage current and IR drop in large arrays. The largest passive memristor array reported is 64×64 with discrete peripherals,[15] which is not comparable to the 1024×512 1T1R arrays used in an integrated chip.[16] For further area and throughput improvement of memristor crossbar arrays and reconfigurable and hierarchical neuromorphic systems, structure innovations arise in three-dimensional (3D) arrays. 3D crossbar arrays were fabricated with self-rectifying memristors,[17] which are CMOS-compatible. Moreover, 3D arrays were designed compactly to support complex ANNs, such as convolutional neural networks (CNN)[18] and capsule networks.[19]

Advances in memristor crossbar arrays highly improved the area, throughput, and energy efficiency of VMM. However, the main barriers to energy-efficient memristor array-based ANN hardware systems are other resource-hungry operations other than VMMs, such as analog/digital conversions performed in analog-to-digital converters (ADCs) and digital-to-analog converters (DACs). These peripheral circuits are needed at the current stage when digital processors are still required. As shown in Fig. 1, most memristor-based ANN systems require additional digital computing units to complete the pre- and post-processing, and input and output peripheral circuitry to complete signal conversions other than the analog VMM cores realized by 2D or 3D memristor crossbar arrays. As reported in a CNN system based on memristor arrays, the area and power consumption of the peripheral circuits account for 98.3% and 93.9% of those of the whole chip, respectively.[20] The numbers do not even include the area and power consumption of digital processors computing the activation functions, weight gradients, and conductance update values. Different approaches have been proposed to address the area and power issues caused by digital peripherals. The first is a different conversion scheme, like encoding input information in the time domain to complete the VMM in the charge domain.[21] This method can reduce the number of DACs required for each input of memristor arrays and simplify the analog to digital conversions after VMM. In the meantime, energy-efficient DAC and ADC designs were also proposed to accommodate the VMM in the charge domain.[22–24] The second way is to implement calculations or communications in the analog domain, eliminating analog/digital conversions and reducing off-chip communications. Previous work realized transfer and subtraction functions using analog amplifiers in the analog domain for hidden neurons of memristor arrays for a small neural network application.[25] Recent research developed analog-based hidden neurons to implement Rectified Linear Unit (ReLU) activation functions, taking the analog output currents from the previous layer and outputting voltages for the next layer in multilayer networks.[25] These methods removed the need for analog-to-digital conversions and vice versa between subsequent layers with acceptable signal degradations. Another work used analog communications in a reservoir computing system, in which information was moved from the reservoir layer to the readout layer as analog voltages.[26] Also, recent work demonstrated to use analog light communication in optoelectronic device arrays to facilitate chip-to-chip communications of memristor arrays for highly parallel data processing.[11]

Based on the memristor array and peripheral innovations, there have been system designs to implement different types of ANNs, like multilayer perceptron,[27,28] CNN,[20] long short-term memory (LSTM),[29] and reinforcement learning.[30] However, besides generalizing ANN algorithms, system designs are supposed to facilitate fully hardware ANNs and compensate for the accuracy degradation caused by hardware nonidealities. Most early works only implemented forward propagation for inference of neural networks since they only focused on validating the device performance. Recent research starts to merge the backward propagation in the same memristor array by adding neuron circuits to both rows and columns of the memristor arrays[22] or sharing neuron circuits of subarrays.[24] For the learning of ANNs, which corresponds to synaptic weight changes and memristor programming, the synaptic weights were trained in software and programmed to memristor arrays in most existing ANN-based systems.[29] However, some system progress proposed self-adaptive *in situ* training to reduce latency and power consumption.[31,32] Calibration algorithms for *in situ* training were also proposed and implemented in hardware to improve the training accuracy.[33] With system designs realizing more functions of ANNs in the analog domain, the performance degradation caused by inevitable nonidealities of memristor devices and peripheral circuits is boosted. To compensate for these circuit nonidealities in memristor-based systems, various hardware-algorithm co-optimization methods were proposed to achieve comparable accuracy to software implementations. These system optimization approaches include training the networks with noise from

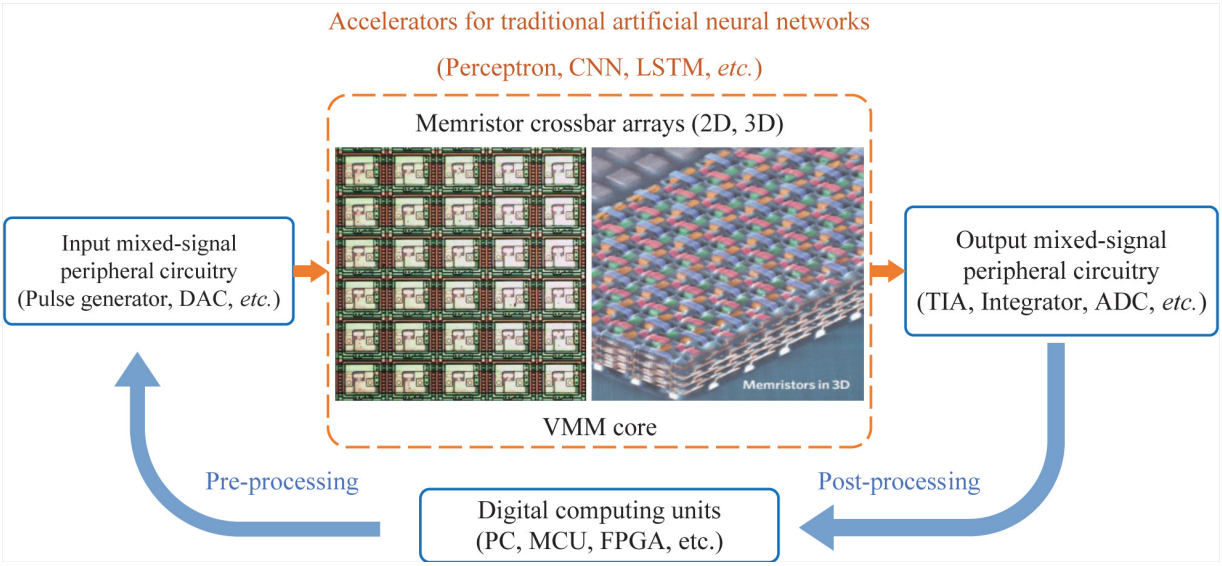

FIG. 1. The general implementation of traditional ANN accelerators based on memristor crossbar arrays.[18,30] [Reproduced with permission from Lin *et al.*, Nat. Electron. **3**, 225 (2020). Copyright 2020 Springer Nature Limited and reproduced with permission from Wang, *et al.*, Nat. Electron. **2**, 115 (2019). Copyright 2019 Springer Nature Limited.]

measurement data,[34] calibrating peripheral circuits with different models,[24] fine-tuning the trained weights with actual outputs from hardware systems,[35] etc.

While the bottom-to-top progress from memristor arrays to system designs has led to more efficient and flexible compute-in-memory accelerators compared to traditional von Neumann architectures,[24,36–38] the energy efficiency and learning ability of these chips are still limited because traditional ANNs only model a neuron as a weighted sum of all its synapses with a nonlinear function.[39] The next generation of neuromorphic systems attempt to model biological brains, where neural information is encoded in spike trains and active patterns in memristor-based hardware to achieve similar performances of brains in terms of energy efficiency, speed, and learning ability. The brain-inspired approach has the potential to harness the randomness and dynamics of memristor devices to realize robust and efficient learning in spike-based networks.[40] Also, the training of most brain-inspired neural networks does not become dependent on gradient-based algorithms that need high-precision computing, saving a huge amount of computing resource. Based on the internal dynamics of memristor devices and memristor arrays, brain-inspired systems have modeled the neural dynamics and network topology of biological brains at multiple levels.[41–46]

Different from device requirements (non-volatility, linearity, stable switching, etc.) for synaptic weights in traditional ANNs, the key requirements for memristor devices modeling biological synapses are similar dynamical behaviors for spike-related plasticity, which is important in communication and learning of brains. Memristors with diffusive dynamics were first developed to model the short-term and long-term spike-timing-dependent plasticity (STDP) of synapses,[47] and second-order memristors were proposed to model the triplet-STDP in the following work.[48] Moreover, memristor devices were used to mimic heterosynaptic plasticity[49,50] as well as homosynaptic plasticity,[51,52] providing synaptic connections at different levels. In addition to synaptic plasticity, memristor devices have also been used to emulate the rich dynamics of neurons composed of soma, axon, and dendrites.[10] One popular research focus is the hardware implementation of mathematic neuron models based on memristors, such as the Hodgkin–Huxley[42,53] and leaky integrate and fire neuron circuits[41,54,55] widely used in SNNs. Also, more neuronal sub-structures like dendrites and soma are modeled in memristor-based circuits to fully emulate the biological neurons and benefit the spatial-temporal data processing.[50–52] Another rising research direction is the modeling of emerging neuron models like hierarchical temporal memory[56,57] and afferent nerves that convert analog signals from sensors into spikes for neuronal processing.[58,59]

Since the development of volatile memristor devices has provided various options for basic components of brain-inspired systems, the primary focus from device innovations to neuromorphic systems is on incorporating the plasticity mechanisms and neuronal dynamics modeled by memristors in the learning of brain-inspired hardware systems for multimodal data processing.[46] One example is the fully memristive neural networks integrating neurons based on diffusive memristors emulating STDP and synapses based on nonvolatile memristors to realize unsupervised CNNs.[60] Another utilization of memristor dynamics is reservoir computing, which maps input signals into higher dimensional computational spaces. In memristor-based reservoir computing, the internal dynamics of diffusive memristors is applied to the reservoir layer, and the programmable resistance of drift memristors is used in the trainable readout layer. The reservoir computing hardware based on memristors has exhibited high performance in temporal data processing like classification tasks,[44] second-order nonlinear tasks,[61] and spoken-digit recognitions.[62] However, hardware implementation of brain-inspired networks based on dynamic memristors is still in the proof-of-concept phase. Most existing system-level works mainly use simulations based on measurement data from memristor devices to validate the efficiency of the emulated biological dynamics in temporal applications. Examples shown in Fig. 2 are diffusive and drift

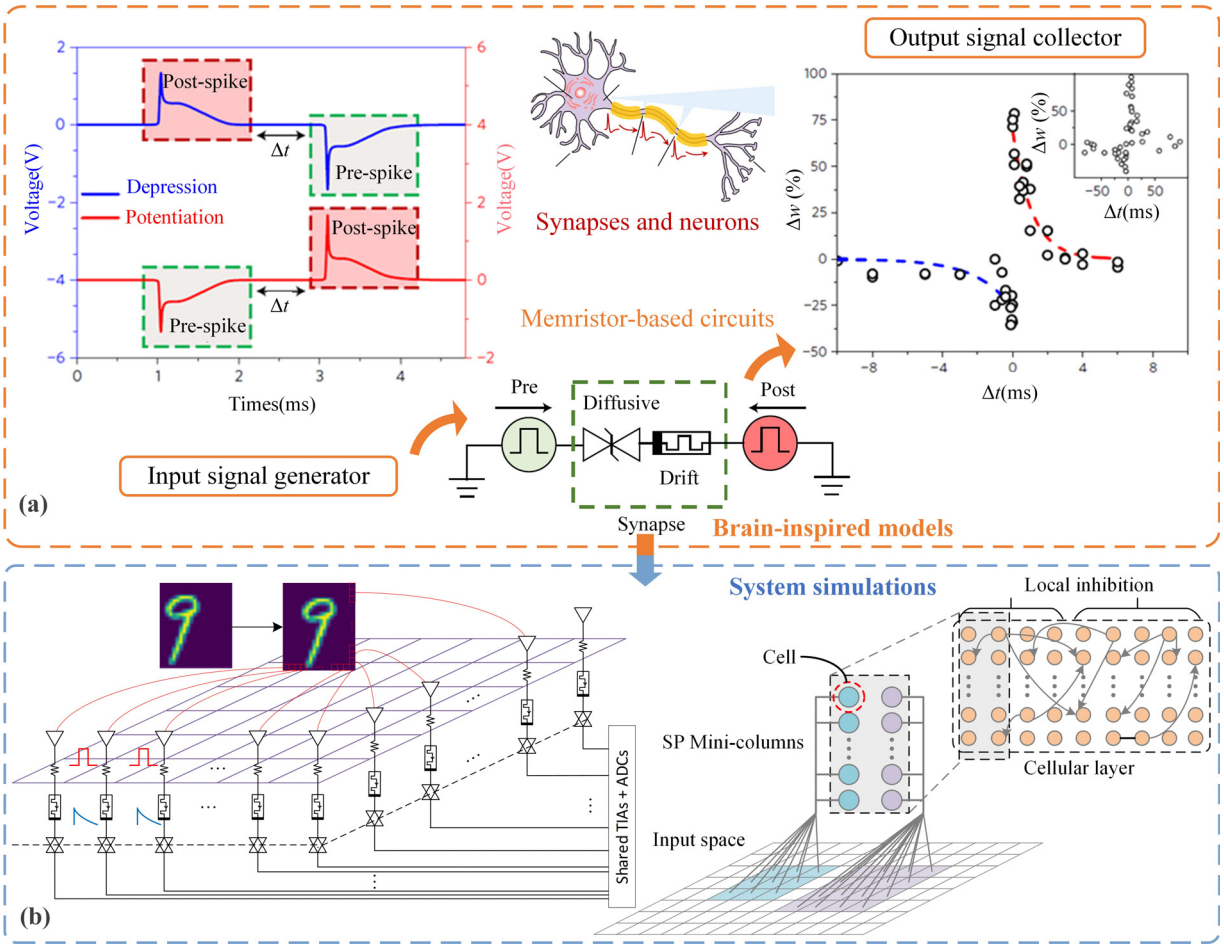

FIG. 2. Examples of emulations of biological synapses and neurons in memristor devices and simulations of brain-inspired systems. (a) Experimental input spike signals (left), biological and memristor-based synapses and neurons (middle), and their responses to spike signals[47] (right). (b) The system simulations of the hierarchy of event-based time surfaces (HOTS) using diffusive memristors[46] (left) and hierarchical temporal memory (HTM) using drift memristors[57] (right). [Reproduced with permission from Z. Wang, *et al.*, Nat. Mater. 16, 101 (2017). Copyright 2017 Springer Nature Limited and F. Ye, *et al.*, Advanced Materials 2204778 (2022). Copyright 2022 John Wiley & Sons, Inc. All rights reserved., and X. Liu, *et al.*, IEEE Trans. Syst. Man Cybern. Syst. 52, 1901 (2022). Copyright 2022 IEEE - All rights reserved.]

memristors fabricated and experimentally tested for synaptic circuits[47] and then used in system simulations based on their dynamics for hierarchical algorithms of event-based time-surfaces[46] and temporal memory network.[57] Moreover, triplet-STDP was tested in memristive devices and simulated to implement the spike-rate-dependent learning rule in orientation pattern selection.[63] Long-term and short-term memory with tunable dynamics and homeostatic plasticity were realized in phase-change memtransistive devices and simulated in networks processing sequential image recognition and solving combinatorial optimization problems.[64] Neuronal and synaptic plasticity were integrated into one memristive cell and simulated for feedback learning like the retraining process in biological systems.[65]

CHALLENGES: FROM DATA CONVERSIONS TO LEARNING AND COMMUNICATIONS

With the progress in both volatile and nonvolatile memristor devices, there are numerous circuit designs for neuromorphic systems and proof-of-concept system simulations for various applications.[66] However, there is still a huge gap between memristor-based circuits and a neuromorphic system that can be used in real-scenario applications with comparable efficiency and learning ability of biological brains.

The first challenge is the efficiency of analog-digital conversions. As shown in Fig. 1, memristor-based hardware for traditional ANNs are mainly inference-only accelerators, in which only VMM is completed in memristor arrays with high throughput and energy efficiency in the inference phase. Other critical operations, including activation functions, error backward propagation, and gradient calculations, are conducted in peripheral circuits and digital computing platforms. Even in the latest work implementing the backpropagation of ANNs, the network training and circuit calibrations are assisted by a digital processor and a field-programmable gate array (FPGA).[24] Therefore, analog-digital conversions and additional memory to store the intermediate data are still required in most mixed-signal systems for

traditional ANNs. Eliminating the analog-digital conversions between the intermediate layers of ANNs for both inference and training phases will be a significant improvement in future system designs. Furthermore, neuromorphic systems based on brain-inspired networks can also reduce data conversions because the inputs and outputs of these networks are encoded in analog signals. Since digital computing performs better in specific tasks like high-precision computing and data storage,[67] and many existing learning models are designed for digital processing units, a mixed-signal system with both digital and analog components is essential to neuromorphic systems. Implementations of more neural network functions in the analog domain and designs of next-generation conversion circuits for inevitable analog-digital conversions will be critical to improve energy and area efficiency.

The second challenge is using memristor dynamics to emulate brain dynamics to boost the learning ability of future neuromorphic systems. Since most memristor-based ANNs are hardware implementations of algorithms used to run in high-precision digital platforms, their learning aims to update the synaptic weights or conductance of memristor devices using stochastic gradient descent and error back-propagation algorithms. The precision required by the training algorithms to achieve high performance contradicts the relatively small number of memristor states and the nonidealities of memristor-based circuits, which are inevitable in analog hardware. Brain-inspired systems have the potential to use the internal stochasticity and dynamics of memristor devices in their training process. However, most synaptic and neuronal dynamics were emulated only in memristor devices and tested in certain conditions. As shown in Fig. 2, the test signals are usually generated from input signal generators and applied to memristor devices, while the responses are recorded and analyzed by output signal collectors. The signal generators and collectors are either semiconductor device analyzers or traditional digital computing units like processors or FPGAs. Experimental data from these tests are used to create memristor models and further used in network-level simulations to verify the efficiency of memristor dynamics in the learning of SNN. Therefore, research on peripheral circuits and learning models that can unleash the full benefits of memristor dynamics in brain-inspired learning of hardware systems is required.

The third challenge lies in the architectures and interfaces supporting data movement and communications inside and outside hierarchical neuromorphic systems. For ANNs based on memristor arrays, especially extensive 3D arrays, the parallel inputs and outputs result in tremendous data from sensors or to the post-processing modules. Also, multiple memristor arrays distributed across the chip were designed for large ANN models as computing cores.[20,24] However, the interfaces for data movement and communications in different cores are still traditional digital-based methods, which hinder the efficiency of parallel analog computing in memristor arrays. On the other hand, with the development of brain-inspired hardware based on memristors, future neuromorphic systems will integrate different network structures for spatial and temporal data processing. Implementing versatile neural networks in one neuromorphic system requires efficient analog data modulation and transmission, for example, the spike-encoded analog signals for SNNs to voltage-encoded signals for ANNs and vice versa. As shown in Fig. 3, the advancements in memristor-based neuromorphic research are either from the device and circuit level or neural network algorithms and applications. There are a few works on system-level implementations in architectures, interfaces, and corresponding models to incorporate the hardware and software breakthroughs to build a neuromorphic hardware system targeting various real-scenario applications.

OPPORTUNITIES: FROM CIRCUIT INNOVATIONS TO ARCHITECTURE DESIGNS

Different from conventional designs for pure-digital intelligent systems, neuromorphic systems are highly dependent on interdisciplinary knowledge from devices to algorithms. To address the discussed issues and bridge the gap shown in Fig. 3, we propose three underexplored directions and multidisciplinary opportunities that can help achieve the envisioned energy efficiency and general intelligence of neuromorphic systems. Moreover, a possible architecture for future neuromorphic systems is proposed in Fig. 4.

The first opportunity is mixed-signal in-memory and analog computing designs based on memristors. The focus of traditional mixed-signal designs is mainly on the performance of ADCs and DACs. However, with the memristor-based neuromorphic systems processing analog information encoded in voltages, currents, and optical signals, various conversion schemes are required to accommodate the mixed-signal circuits based on memristors working as both computing and memory components. For the digital-to-analog conversion, since the memristor-based circuits can directly process analog signals, more research efforts should be put into the modulation and merging of signals from different sensors for the processing cores. Several works include converting analog inputs into spikes for SNNs[58,59] and multimodal signal integration.[68–70] However, they are only designed for

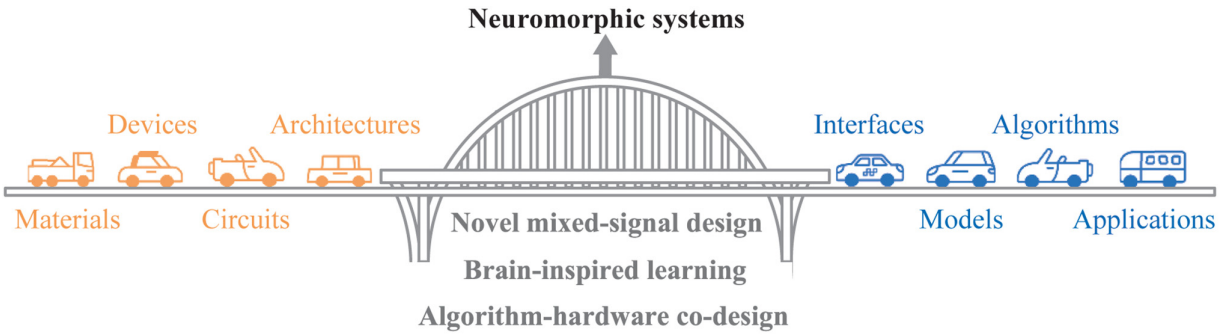

FIG. 3. Three research directions proposed to bridge the memristor-based hardware and brain-inspired models toward energy-efficient and intelligent neuromorphic systems.

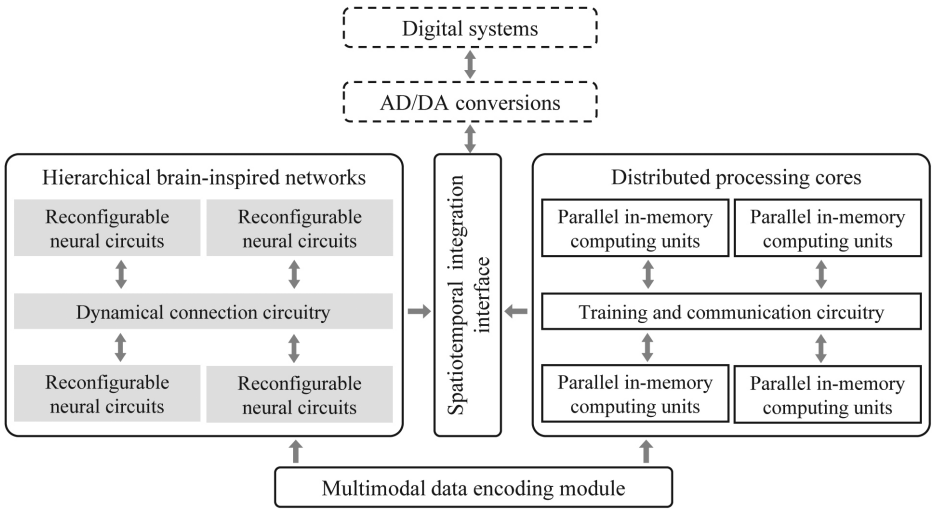

FIG. 4. A possible architecture for neuromorphic systems including multimodal data encoding module for sensory data processing, hierarchical brain-inspired networks for event-driven applications, distributed processing cores for computing-intensive applications, interfaces for spatiotemporal integration, and optional analog-digital conversions for applications requiring digital processing.

specific networks and applications, and system-level implementation is still lacking. For the analog-to-digital conversion, the basic research should be directed to analog-friendly hardware to eliminate the conversions in the neuromorphic systems and further reduce memory and communications in the digital domain. For required conversions to interact with digital computing systems, besides the high precision and high-speed traditional ADCs target, future conversion schemes should be able to convert various analog signals other than voltages and accommodate the functions of the brain-inspired models. Existing works start to explore memristor-based ADCs with high performance[71,72] and ADCs integrated with the computing circuits.[22,24] However, none are designed for brain-inspired and spike-encoded systems, which will be the central parts of future neuromorphic systems.

The second direction is the learning model and hardware to harness the internal dynamics of memristors to simplify the training process and boost learning abilities. Since biological brains do not precisely calculate loss functions, and gradients for learning and high-precision computing in the analog domain are difficult, the learning of future neuromorphic systems will not only rely on gradient-based algorithms, which have dominated the ANNs and have even been modified for SNNs. Many learning mechanisms in biological brains, such as homosynaptic and heterosynaptic plasticity, local and global inhibitions, spatial and temporal information integration, and hierarchical structures and communications, have not been fully explored to facilitate the learning of neuromorphic hardware. Learning models inspired by the memory and learning systems of biological brains are expected to be designed to integrate rich learning mechanisms into the training of neuromorphic systems based on memristors. Hardware at different levels should also be developed to facilitate these learning models in the hierarchical architecture of future neuromorphic systems.

The third outlook is on the interfaces and architectures for memory and communications in hierarchical and distributed neuromorphic systems. In recent system works,[20,22,24] analog data were still converted to digital values inside each computing core and transmitted with traditional interfaces. Unlike traditional digital signal transmission, information will be encoded and stored in various formats in future neuromorphic systems, for example, information is encoded in voltage amplitude,[73] pulse width,[22] and spikes,[74] weights are stored as temporary and long-term conductance,[38] chip-to-chip analog signals are transmitted as light,[11,75] etc. Interfaces and architectures are required to coordinate the data movement within local networks for specific tasks and global networks for various applications. Exploration of spike-based and event-driven communications inspired by biological brains is one valuable direction for signal transmissions. Moreover, the structural and functional connectivity for different regions of brain networks may provide some insights into the hardware designs for communications and models for data management. At the system architecture level, research efforts should also be directed to hierarchical architectures inspired by brain topology to integrate various neural networks specialized in different tasks to complete the implementation of neuromorphic systems.

Based on the proposed directions, a potential architecture of future neuromorphic systems taking advantage of both brain-inspired computing and parallel ANNs is shown in Fig. 4. With breakthroughs in dynamical volatile memristors, neural circuits are expected to benefit from brain-inspired learning and communications with reconfigurability and dynamical connections. Based on high performance nonvolatile memristors, parallel in-memory computing units are expected to achieve high accuracy in analog inference and training assisted by the training and communication circuitry. At the system level, hierarchical brain-inspired networks with local and global activation and inhibition will be built from the neural circuits for event-driven applications. Distributed processing cores with configurable resources will be built from the in-memory computing units for computing-intensive applications. For the system integration, multimodal data encoding modules are critical components to fuse multimodal sensory data, while future analog interfaces and learning models are required to complete the spatiotemporal integration for information from both subsystems and generate processed results or final decisions. For applications requiring digital peripherals, efficient analog-to-digital (AD) and digital-to-analog (DA) schemas will be designed to integrate the analog neuromorphic hardware with essential digital systems.

In summary, advances in recent memristor-related research from nanomaterials to spatiotemporal applications have demonstrated the efficiency and learning ability of intelligent hardware based on memristor devices. We summarized progress from memristor devices to preliminary system designs for ANN-based accelerators and brain-inspired neuromorphic systems. We also discussed three major challenges and proposed possible directions to embrace existing achievements to build efficient and intelligent neuromorphic systems based on memristors.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding support from the National Science Foundation (NSF) (Nos. ECCS-2023752 and CCF-2133475).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yi Huang: Conceptualization (equal); Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal). **Fatemeh Kiani:** Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal). **Fan Ye:** Formal analysis (equal); Writing – original draft (equal); Writing – review & editing (equal). **Qiangfei Xia:** Conceptualization (equal); Funding acquisition (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1]Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
[2]E. Strubell, A. Ganesh, and A. McCallum, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, Stroudsburg, PA, 2019), pp. 3645–3650.
[3]B. Goertzel, J. Artif. General Intell. **5**, 1–46 (2014).
[4]Y. Zhang, P. Qu, Y. Ji, W. Zhang, G. Gao, G. Wang, S. Song, G. Li, W. Chen, W. Zheng, F. Chen, J. Pei, R. Zhao, M. Zhao, and L. Shi, Nature **586**, 378 (2020).
[5]P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, Science **345**, 668–2014 (2014).
[6]M. Davies, N. Srinivasa, T. H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. H. Weng, A. Wild, Y. Yang, and H. Wang, IEEE Micro **38**, 82 (2018).
[7]G. Orchard, E. P. Frady, D. B. D. Rubin, S. Sanborn, S. B. Shrestha, F. T. Sommer, and M. Davies, in *IEEE Workshop on Signal Processing Systems (SiPS)* (IEEE, 2021), pp. 254–259.
[8]J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, and L. Shi, Nature **572**, 106 (2019).
[9]L. Deng, G. Wang, G. Li, S. Li, L. Liang, M. Zhu, Y. Wu, Z. Yang, Z. Zou, J. Pei, Z. Wu, X. Hu, Y. Ding, W. He, Y. Xie, and L. Shi, IEEE J. Solid-State Circuits **55**, 2228 (2020).
[10]J. Tang, F. Yuan, X. Shen, Z. Wang, M. Rao, Y. He, Y. Sun, X. Li, W. Zhang, Y. Li, B. Gao, H. Qian, G. Bi, S. Song, J. J. Yang, and H. Wu, Adv. Mater. **31**, 1902761 (2019).
[11]C. Choi, H. Kim, J.-H. Kang, M.-K. Song, H. Yeon, C. S. Chang, J. M. Suh, J. Shin, K. Lu, B.-I. Park, Y. Kim, H. E. Lee, D. Lee, J. Lee, I. Jang, S. Pang, K. Ryu, S.-H. Bae, Y. Nie, H. S. Kum, M.-C. Park, S. Lee, H.-J. Kim, H. Wu, P. Lin, and J. Kim, Nat. Electron. **5**, 386 (2022).
[12]Z. Zhao, J. Tang, J. Yuan, Y. Li, Y. Dai, J. Yao, Q. Zhang, S. Ding, T. Li, R. Zhang, Y. Zheng, Z. Zhang, S. Qiu, Q. Li, B. Gao, N. Deng, H. Qian, F. Xing, Z. You, and H. Wu, ACS Nano **16**, 16784 (2022).
[13]Z. Zhang, X. Zhao, X. Zhang, X. Hou, X. Ma, S. Tang, Y. Zhang, G. Xu, Q. Liu, and S. Long, Nat. Commun. **13**, 6590 (2022).
[14]Z. Wang, H. Wu, G. W. Burr, C. S. Hwang, K. L. Wang, Q. Xia, and J. J. Yang, Nat. Rev. Mater. **5**, 173 (2020).
[15]H. Kim, M. R. Mahmoodi, H. Nili, and D. B. Strukov, Nat. Commun. **12**, 5198 (2021).
[16]J. M. Hung, C. X. Xue, H. Y. Kao, Y. H. Huang, F. C. Chang, S. P. Huang, T. W. Liu, C. J. Jhang, C. I. Su, W. S. Khwa, C. C. Lo, R. S. Liu, C. C. Hsieh, K. T. Tang, M. S. Ho, C. C. Chou, Y. der Chih, T. Y. J. Chang, and M. F. Chang, Nat. Electron. **4**, 921 (2021).
[17]C. Li, L. Han, H. Jiang, M.-H. Jang, P. Lin, Q. Wu, M. Barnell, J. J. Yang, H. L. Xin, and Q. Xia, Nat. Commun. **8**, 15666 (2017).
[18]P. Lin, C. Li, Z. Wang, Y. Li, H. Jiang, W. Song, M. Rao, Y. Zhuo, N. K. Upadhyay, M. Barnell, Q. Wu, J. J. Yang, and Q. Xia, Nat. Electron. **3**, 225 (2020).
[19]Y. Huang, R. Hu, and Z. Zeng, in *8th International Conference on Information Science and Technology (ICIST 2018)* (IEEE, 2018), pp. 170–175.
[20]P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, Nature **577**, 641 (2020).
[21]P. M. Sheridan, F. Cai, C. Du, W. Ma, Z. Zhang, and W. D. Lu, Nat Nanotechnol **12**, 784 (2017).
[22]F. Cai, J. M. Correll, S. H. Lee, Y. Lim, V. Bothra, Z. Zhang, M. P. Flynn, and W. D. Lu, Nat. Electron. **2**, 290 (2019).
[23]M. Bavandpour, S. Sahay, M. R. Mahmoodi, and D. Strukov, IEEE Trans. VLSI. Syst. **28**, 823 (2020).
[24]W. Wan, R. Kubendran, C. Schaefer, S. B. Eryilmaz, W. Zhang, D. Wu, S. Deiss, P. Raina, H. Qian, B. Gao, S. Joshi, H. Wu, H.-S. P. Wong, and G. Cauwenberghs, Nature **608**, 504 (2022).
[25]F. Kiani, J. Yin, Z. Wang, J. J. Yang, and Q. Xia, Sci. Adv. **7**, 4801 (2021).
[26]Y. Zhong, J. Tang, X. Li, X. Liang, Z. Liu, Y. Li, Y. Xi, P. Yao, Z. Hao, B. Gao, H. Qian, and H. Wu, Nat. Electron. **5**, 672 (2022).
[27]M. Hu, J. P. Strachan, Z. Li, E. M. Grafals, N. Davila, C. Graves, S. Lam, N. Ge, R. S. Williams, J. J. Yang, H. P. Labs, and R. S. Williams, in IEEE Design Automation Conference, 2016.
[28]P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H. P. Wong, and H. Qian, Nat. Commun. **8**, 15199 (2017).
[29]C. Li, Z. Wang, M. Rao, D. Belkin, W. Song, H. Jiang, P. Yan, Y. Li, P. Lin, M. Hu, N. Ge, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, Nat. Mach. Intell. **1**, 49 (2019).
[30]Z. Wang, C. Li, W. Song, M. Rao, D. Belkin, Y. Li, P. Yan, H. Jiang, P. Lin, M. Hu, J. P. Strachan, N. Ge, M. Barnell, Q. Wu, A. G. Barto, Q. Qiu, R. S. Williams, Q. Xia, and J. J. Yang, Nat. Electron. **2**, 115 (2019).
[31]C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, Nat. Commun. **9**, 2385 (2018).
[32]F. Alibart, E. Zamanidoost, and D. B. Strukov, Nat. Commun. **4**, 2072 (2013).
[33]B. Gao, Y. Zhou, Q. Zhang, S. Zhang, P. Yao, Y. Xi, Q. Liu, M. Zhao, W. Zhang, Z. Liu, X. Li, J. Tang, H. Qian, and H. Wu, Nat. Commun. **13**, 2026 (2022).
[34]I. Chakraborty, D. Roy, and K. Roy, IEEE Trans. Emerging Top. Comput. Intell. **2**, 335 (2018).
[35]Z. Wang, C. Li, P. Lin, M. Rao, Y. Nie, W. Song, Q. Qiu, Y. Li, P. Yan, J. P. Strachan, N. Ge, N. Mcdonald, Q. Wu, M. Hu, H. Wu, R. S. Williams, Q. Xia, and J. J. Yang, Nat. Mach. Intell. **1**, 434 (2019).
[36]Q. Xia and J. J. Yang, Nat. Mater. **18**, 309 (2019).

[37]S. Jung, H. Lee, S. Myung, H. Kim, S. K. Yoon, S. W. Kwon, Y. Ju, M. Kim, W. Yi, S. Han, B. Kwon, B. Seo, K. Lee, G. H. Koh, K. Lee, Y. Song, C. Choi, D. Ham, and S. J. Kim, Nature **601**, 211 (2022).
[38]S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, G. W. Burr, C. Nolfo, S. Sidler, B. Killeen, C. Cheng, Y. Jaoudi, M. Giordano, M. Bodini, and C. P. Nathan, Nature **558**, 60 (2018).
[39]Y. Cui, S. Ahmad, and J. Hawkins, Front. Comput. Neurosci. **11**, 111 (2017).
[40]A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, and A. J. Kenyon, Adv. Intell. Syst. **2**, 2000085 (2020).
[41]J. Lin, W. Ye, X. Zhang, Q. Lian, S. Wu, T. Guo, and H. Chen, IEEE Electron Device Lett. **43**, 1231 (2022).
[42]H. M. Huang, R. Yang, Z. H. Tan, H. K. He, W. Zhou, J. Xiong, and X. Guo, Adv. Mater. **31**, 1803849 (2019).
[43]X. Mou, J. Tang, Y. Lyu, Q. Zhang, S. Yang, F. Xu, W. Liu, M. Xu, Y. Zhou, W. Sun, Y. Zhong, B. Gao, P. Yu, H. Qian, and H. Wu, Sci. Adv. **7**, 648 (2021).
[44]R. Midya, Z. Wang, S. Asapu, X. Zhang, M. Rao, W. Song, Y. Zhuo, N. Upadhyay, Q. Xia, and J. J. Yang, Adv. Intell. Syst. **1**, 1900084 (2019).
[45]K. S. Woo, J. Kim, J. Han, W. Kim, Y. H. Jang, and C. S. Hwang, Nat. Commun. **13**, 5762 (2022).
[46]F. Ye, F. Kiani, Y. Huang, and Q. Xia, Adv. Mater. **34**, 2204778 (2022).
[47]Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G. L. Li, H. L. Xin, R. S. Williams, Q. Xia, and J. J. Yang, Nat. Mater. **16**, 101 (2017).
[48]R. Yang, H. M. Huang, Q. H. Hong, X. B. Yin, Z. H. Tan, T. Shi, Y. X. Zhou, X. S. Miao, X. P. Wang, S. B. Mi, C. L. Jia, and X. Guo, Adv. Funct. Mater. **28**, 1704455 (2018).
[49]J. H. Nam, S. Oh, H. Y. Jang, O. Kwon, H. Park, W. Park, J. Kwon, Y. Kim, and B. Cho, Adv. Funct. Mater. **31**, 2104174 (2021).
[50]J. Qin, B. Zhu, C. Wang, C. Zhu, R. Sun, L. Zhen, Y. Chai, and C. Xu, Adv. Electron. Mater. 2200721 (2022).
[51]X. B. Yan, J. H. Zhao, S. Liu, Z. Y. Zhou, Q. Liu, J. S. Chen, and X. Y. Liu, Adv. Funct. Mater. **28**, 1705320 (2018).
[52]Y. Fu, Y. Zhou, X. Huang, B. Dong, F. Zhuge, Y. Li, Y. He, Y. Chai, and X. Miao, Adv. Funct. Mater. **32**, 2111996 (2022).
[53]L. Chua, V. Sbitnev, and H. Kim, Int. J. Bifurcation Chaos **22**, 1230011 (2012).
[54]Z. Wang, M. Rao, J. W. Han, J. Zhang, P. Lin, Y. Li, C. Li, W. Song, S. Asapu, R. Midya, Y. Zhuo, H. Jiang, J. H. Yoon, N. K. Upadhyay, S. Joshi, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, Q. Qiu, R. S. Williams, Q. Xia, and J. J. Yang, Nat. Commun. **9**, 3208 (2018).
[55]D. Lim, K. Cho, and S. Kim, IEEE Electron Device Lett. **42**, 649 (2021).
[56]D. Fan, M. Sharad, A. Sengupta, and K. Roy, IEEE Trans. Neural Networks Learn. Syst. **27**, 1907 (2016).
[57]X. Liu, Y. Huang, Z. Zeng, and D. C. Wunsch, IEEE Trans. Syst. Man, Cybern.-Syst **52**, 1901 (2022).
[58]H. Tan, Q. Tao, I. Pande, S. Majumdar, F. Liu, Y. Zhou, P. O. Å. Persson, J. Rosen, and S. van Dijken, Nat. Commun. **11**, 1369 (2020).
[59]C. Wan, G. Chen, Y. Fu, M. Wang, N. Matsuhisa, S. Pan, L. Pan, H. Yang, Q. Wan, L. Zhu, and X. Chen, Adv. Mater. **30**, 1801291 (2018).
[60]Z. Wang, S. Joshi, S. Savel'Ev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, and J. J. Yang, Nat. Electron. **1**, 137 (2018).
[61]C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, Nat. Commun. **8**, 2204 (2017).
[62]Y. Zhong, J. Tang, X. Li, B. Gao, H. Qian, and H. Wu, Nat. Commun. **12**, 408 (2021).
[63]Z. Wang, T. Zeng, Y. Ren, Y. Lin, H. Xu, X. Zhao, Y. Liu, and D. Ielmini, Nat. Commun. **11**, 1510 (2020).
[64]S. G. Sarwat, B. Kersting, T. Moraitis, V. P. Jonnalagadda, and A. Sebastian, Nat. Nanotechnol. **17**, 507 (2022).
[65]S. H. Sung, T. J. Kim, H. Shin, T. H. Im, and K. J. Lee, Nat. Commun. **13**, 2811 (2022).
[66]M. Lanza, A. Sebastian, W. D. Lu, M. le Gallo, M. F. Chang, D. Akinwande, F. M. Puglisi, H. N. Alshareef, M. Liu, and J. B. Roldan, Science **376**(2022), eabj9979.
[67]A. Mehonic and A. J. Kenyon, Nature **604**, 255 (2022).
[68]H. Tan, Y. Zhou, Q. Tao, J. Rosen, and S. van Dijken, Nat. Commun. **12**, 1120 (2021).
[69]R. Yuan, Q. Duan, P. J. Tiw, G. Li, Z. Xiao, Z. Jing, K. Yang, C. Liu, C. Ge, R. Huang, and Y. Yang, Nat. Commun. **13**, 3973 (2022).
[70]J. Zhu, X. Zhang, R. Wang, M. Wang, P. Chen, L. Cheng, Z. Wu, Y. Wang, Q. Liu, and M. Liu, Adv. Mater. **34**, 2200481 (2022).
[71]W. Cao, X. He, A. Chakrabarti, and X. Zhang, IEEE Trans. Comput-Aided. Des. Integr. Circuits Syst. **39**, 1841 (2020).
[72]W. Cao, L. Ke, A. Chakrabarti, and X. Zhang, IEEE Trans. Comput-Aided. Des. Integr. Circuits Syst. **40**, 808 (2021).
[73]C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Dávila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, and Q. Xia, Nat. Electron. **1**, 52 (2017).
[74]T. Wang, J. Meng, X. Zhou, Y. Liu, Z. He, Q. Han, Q. Li, J. Yu, Z. Li, Y. Liu, H. Zhu, Q. Sun, D. W. Zhang, P. Chen, H. Peng, and L. Chen, Nat. Commun. **13**, 7432 (2022).
[75]D. Lee, M. Park, Y. Baek, B. Bae, J. Heo, and K. Lee, Nat. Commun. **13**, 5223 (2022).