

ROADMAP • OPEN ACCESS

2022 roadmap on neuromorphic devices and applications research in China

To cite this article: Qing Wan *et al* 2022 *Neuromorph. Comput. Eng.* **2** 042501

View the [article online](#) for updates and enhancements.

You may also like

- [Hadamard product-based in-memory computing design for floating point neural network training](#)
Anjunyi Fan, Yihan Fu, Yaoyu Tao et al.
- [General spiking neural network framework for the learning trajectory from a noisy mmWave radar](#)
Xin Liu, Mingyu Yan, Lei Deng et al.
- [Designing polar textures with ultrafast neuromorphic features from atomistic simulations](#)
Sergey Prosandeev, Sergei Prokhorenko, Yousra Nahas et al.



ROADMAP

OPEN ACCESS

RECEIVED

15 November 2021

REVISED

17 June 2022

ACCEPTED FOR PUBLICATION

20 June 2022

PUBLISHED

8 December 2022

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



2022 roadmap on neuromorphic devices and applications research in China

Qing Wan^{1,23,*} , Changjin Wan^{1,23,*} , Huaqiang Wu^{2,23} , Yuchao Yang^{3,23} , Xiaohe Huang⁴ , Peng Zhou⁴ , Lin Chen⁴ , Tian-Yu Wang⁴ , Yi Li^{5,6} , Kan-Hao Xue^{5,6} , Yu-Hui He^{5,6} , Xiang-Shui Miao^{5,6} , Xi Li⁷ , Chenchen Xie⁷ , Houpeng Chen⁷ , Zhitang Song⁷ , Hong Wang⁸ , Yue Hao⁸ , Junyao Zhang⁹ , Jia Huang⁹ , Zheng Yu Ren^{10,11,12} , Li Qiang Zhu^{10,11} , Jianyu Du¹³ , Chen Ge¹³ , Yang Liu¹⁴ , Guanglong Ding¹⁵ , Ye Zhou¹⁵ , Su-Ting Han¹⁴ , Guosheng Wang¹⁶ , Xiao Yu¹⁶ , Bing Chen¹⁷ , Zhufei Chu¹⁸ , Lunyao Wang¹⁸ , Yinshui Xia¹⁸ , Chen Mu⁴ , Feng Lin⁴ , Chixiao Chen⁴ , Bojun Cheng¹⁹ , Yannan Xing²⁰ , Weitao Zeng²⁰ , Hong Chen² , Lei Yu²¹ , Giacomo Indiveri^{19,22} and Ning Qiao^{19,20,22}

- ¹ School of Electronic Science and Engineering, Nanjing University, Nanjing, People's Republic of China
- ² Tsinghua University, Beijing, People's Republic of China
- ³ Peking University, Beijing, People's Republic of China
- ⁴ State Key Laboratory of ASIC and System, School of Microelectronics, Fudan University, Shanghai 200433, People's Republic of China
- ⁵ School of Integrated Circuits, School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, People's Republic of China
- ⁶ Hubei Yangtze Memory Laboratories, Wuhan 430205, People's Republic of China
- ⁷ Shanghai Key Laboratory of Nanofabrication Technology for Memory, Shanghai Institute of Micro-System and Information Technology, Chinese Academy of Sciences, People's Republic of China
- ⁸ Key Laboratory of Wide Band Gap Semiconductor Technology, School of Microelectronics, Xidian University, Xi'an 710071, People's Republic of China
- ⁹ Interdisciplinary Materials Research Center, School of Materials Science and Engineering, Frontiers Science Center for Intelligent Autonomous Systems, Tongji University, Shanghai, 201804, People's Republic of China
- ¹⁰ School of Physical Science and Technology, Ningbo University, Ningbo 315211, Zhejiang, People's Republic of China
- ¹¹ Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, Zhejiang, People's Republic of China
- ¹² School of Physical Science and Technology, ShanghaiTech University, Shanghai 201210, People's Republic of China
- ¹³ Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
- ¹⁴ Institute of Microscale Optoelectronics, Shenzhen University, Shenzhen 518060, People's Republic of China
- ¹⁵ Institute for Advanced Study, Shenzhen University, Shenzhen 518060, People's Republic of China
- ¹⁶ Research Center for Intelligent Chips, Zhejiang Lab, Hangzhou 311121, People's Republic of China
- ¹⁷ School of Micro-Nano Electronics, Zhejiang University, Hangzhou 310058, People's Republic of China
- ¹⁸ Faculty of Electrical Engineering and Computer Science (EECS), Ningbo University, Ningbo, People's Republic of China
- ¹⁹ SynSense AG, Thurgauerstrasse 40, 8050 Zurich, Switzerland
- ²⁰ Chengdu SynSense Tech. Co. Ltd., 1577 Tianfu Road, Chengdu, Sichuan, People's Republic of China
- ²¹ School of Electronic and Information, Wuhan University, 299 Bayi Road, Wuhan, People's Republic of China
- ²² University of Zurich, Rämistrasse 71, 8006, Zurich, Switzerland
- * Authors to whom any correspondence should be addressed.
- ²³ Guest editors of the roadmap.

E-mail: wanqing@nju.edu.cn and cjwan@nju.edu.cn

Keywords: neuromorphic computing, memristors, in-memory computing, 2D memristors, neuromorphic devices, China

Abstract

The data throughput in the von Neumann architecture-based computing system is limited by its separated processing and memory structure, and the mismatching speed between the two units. As a result, it is quite difficult to improve the energy efficiency in conventional computing system, especially for dealing with unstructured data. Meanwhile, artificial intelligence and robotics nowadays still behave poorly in autonomy, creativity, and sociality, which has been considered as the unimaginable computational requirement for sensorimotor skills. These two plights have urged the imitation and replication of the biological systems in terms of computing, sensing, and even motoring. Hence, the so-called neuromorphic system has drawn worldwide attention in recent decade, which is aimed at addressing the aforementioned needs from the mimicking of neural

system. The recent developments on emerging memory devices, nanotechnologies, and materials science have provided an unprecedented opportunity for this aim.

This roadmap profiles the potential trend in building neuromorphic systems from the view of Chinese scientists. The content of this roadmap will cover some core topics from multidisciplinary researchers, including electronics, computer science, materials, physics, and so on. The perspectives and challenges are also discussed partly, which may serve as guidance for the pursuing of high energy-efficient and/or bio-realistic systems that can compute, sense, and act as our human beings. This will also give birth to some excited paradigm breakers and advanced technologies in a wide spectrum of areas.

Contents

Introduction	2
1. From memristors to neuromorphic computing	4
2. Two-dimensional two-terminal memristors	7
3. Metal-oxide memristors for IMC	9
4. Neuromorphic devices and applications based on phase-change random-access memory	12
5. Physically transient memristors for neuromorphic computing	14
6. Recent advances in transistor-based organic synapses in China: from device to system	16
7. Brief discussion of oxide neuromorphic transistors in China	20
8. Neuromorphic devices based on functional oxides	23
9. Development of 2D-based neuromorphic transistor in China	26
10. Neuromorphic transistor for bionic perception	28
11. Neuromorphic computing going efficient	31
12. Automated synthesis and mapping	34
13. Survey on machine learning SoC: from near-memory computing to IMC	37
14. Market-oriented neuromorphic SoC solution	40
Acknowledgments	41
Data availability statement	42
References	42

Introduction

Changjin Wan and Qing Wan
Nanjing University, People's Republic of China
E-mail: cjwan@nju.edu.cn and wanqing@nju.edu.cn

The recent surge in artificial intelligence (AI) drives the exploration of machine learning hardware, especially the development of fundamental building blocks, namely, neuromorphic devices. China is becoming one of the most thriving regions of this research field. The general aim of the *roadmap on neuromorphic devices and applications research in China* is to provide an overview of the different fields of research and progress that are related to neuromorphic devices and applications, evaluate the possible technical routes and challenges, and give guidelines and perspectives that would shed light on future development. The roadmap addresses the following topics:

- Materials for neuromorphic devices
- Applications and tools for neuromorphic devices and systems
- Perspectives for neuromorphic computing

Materials for neuromorphic devices. The development of neuromorphic devices is still dependent on what materials can render. Even the scaling of state-of-the-art (SOTA) complementary metal-oxide-semiconductor (CMOS) technologies benefits from the utilizing of advanced materials. For example, the success of fin-shaped field-effect transistor (FinFET) was partly attributed to the implementation of high- κ dielectric materials. In view of transistor-type neuromorphic devices, proper semiconducting materials would bring about low off current, which could improve the energy efficiency; high mobility, which is important for increasing switch speed of the devices; and high sensitivity/selectivity to certain stimulations, which might benefit near- or in-sensor computing. Dielectric materials for memristor-type neuromorphic devices should also be subject

to significant attention, as the mechanisms of a memristive neuromorphic device, including redox, phase change, magnetic tunneling, and ferroelectric polarization, are mostly determined by the dielectric materials. Therefore, we collected several sections under this topic, and we promise that with the exploring of emerging materials and refining of the conventional materials, these will offer new opportunities for the development of neuromorphic electronics.

Applications and tools for neuromorphic devices and systems. Another important research field with respect to neuromorphic devices is to make them practicable. Emerging applications based on neuromorphic devices that could advance the conventional digital system in certain aspects, like energy efficiency, error tolerance, plasticity, and so on, will be highly pursued. For example, the realization of artificial perception based on neuromorphic devices has potential implications for future robotics that might have humanlike sophisticated sensorimotor capabilities within a very narrow energy budget. Besides, to boost efficiency from design to manufacturing, it is very necessary to develop electronic design automation (EDA) tools, and it would be possible to extend the application and generality of neuromorphic computing and engineering by developing automated synthesis and mapping techniques.

Perspectives for neuromorphic computing. The ultimate aim of neuromorphic devices is to build a neuromorphic system that operates similarly to our neural networks (NNs). A short-term goal would be focused on the implementation of nonvolatile memory (NVM) technique for improving the energy efficiency of machine learning algorithms. The NVM array has shown excellent energy efficiency in dealing with matrix–matrix multiplication, which is a computing- and memory-intensive workload and is not efficient in conventional CPU-like von Neumann architectures. In this trend, system-on-chip (SoCs) accelerating NN computing is widely reported for different specific applications, such as image recognition, keyword spotting, and natural language processing. However, several tough challenges should be addressed before taking full advantage of the neuromorphic device-based SoCs. Development in this trend will also bring with it unprecedented opportunities that could come from many aspects, like algorithm–circuit–architecture codesign, introduction of event-driven process, novel NN structures, and so on.

1. From memristors to neuromorphic computing

Xiaohe Huang and Peng Zhou*

Fudan University, People's Republic of China

E-mail: pengzhou@fudan.edu.cn

1.1. Status

Resistive memory originates from the reversible resistive effect induced by electrical pulses reported in the 1960s [1]. In 1971 Chua proposed a theoretical model of a memristor, predicting its resistive state would change with the history of applied voltage [2]. By 2008 experimentally demonstrated TiO_2 -based resistive random-access-memory (RRAM) devices first linked RRAM with memristors [3]. With a simple 'sandwich' structure, memristors have been developed to a very small size with low power consumption, and they can perform the function of storing and even processing information excellently [4]. When it is unknown whether memristors can shake the position of mature flash memory as the mainstream NVM, it is expected to serve as the most desirable basic device unit for implementing synaptic and neuronal functions in the emergence of neuromorphic computing. The data transfer and speed gap between conventional computing and memory make the data processing waste a huge amount of energy and time. The so-called memory wall problem generated by the von Neumann bottleneck becomes more and more urgent. Neuromorphic computing, which mimics the nervous system in the biological brain, uses analog signals to process information with high parallelism, low power consumption, and convergence of storage and computing [5]. So neuromorphic computing has a natural advantage over digital systems in solving computational tasks faster and more efficiently in a smaller area.

Since the first memristor device simulation to implement synaptic functions in 2010 [6], memristor-based neuromorphic devices are now able to mimic the basic functions of biological neurons and synapses. Among them, memristors have been the most studied as simple artificial synapses, whereby integrated arrays can directly rely on physics laws to perform efficient parallel in-memory computations [7], such as accelerating vector-matrix multiplication in artificial neural network (ANN) training and inference. Classical machine learning tasks, such as information encoding [8], data classification [9], and reinforcement learning [10], can already be implemented in them. Neuromorphic computing is more efficient than the current traditional computing paradigm in the demanded environment of internet of thing (IoT), which requires available ultralow-power intelligent edge computing. Memristors are the most promising underlying devices since artificial neurons that simulate biological functions based on physical principles enable a simpler and more efficient circuitry than CMOS [11]. Although further development of artificial neural devices is a great challenge, it is a promising way to implement complex human brain systems in hardware (figure 1).

1.2. Current and future challenges

Although neuromorphic devices have made some progress from the device level to the array level, they are still a long way from real application, and many technical bottlenecks remain to be solved [12]. Device yields and uniformity for memristors are still not sufficient. The microscopic process of resistive material change is inherently stochastic [13]. The random movement of ions or thermal activation of defects may lead to fluctuation of device parameters, thus reducing uniformity and yield. Even though the yield and uniformity of NVM for binary values can be easily achieved, memristors for multilevel states face higher requirements [14]. Nonlinearity and asymmetry constrain the application of the devices, and the limited number of tunable weight states limits the computing accuracy [15]. Other metrics to evaluate memristors as neuromorphic devices include retention, endurance, and on/off ratio. Devices that meet the requirements in all performance metrics are still lacking to facilitate the design of general-purpose systems. How to design and select materials and mechanisms for neuromorphic devices to manufacture high-precision devices is still a central issue for researchers to consider.

At the array level, the sneak-path currents generated by the applied voltage on unselected devices hinder the read/write operation of the memristors, limiting the large-scale integration of crossbar arrays [16]. On the other hand, if the size scales down and the array scales up, the wire resistance inside the arrays will inevitably increase, which leads to a voltage drop on the wire, significantly affecting the signal transmission, and limits the accuracy of the calculation while also increasing power consumption [17]. At the system application level, algorithms with their practical applications that fully manifest the advantages of neuromorphic computing are important factors to drive the industry forward. Most of the current research is modifying device properties to better fit on available algorithms, which limits the inherent properties of materials and devices, so algorithms that are more closely tied to the device are needed. For example, there is a lack of proven algorithms especially for spiking neural networks (SNNs) [5].

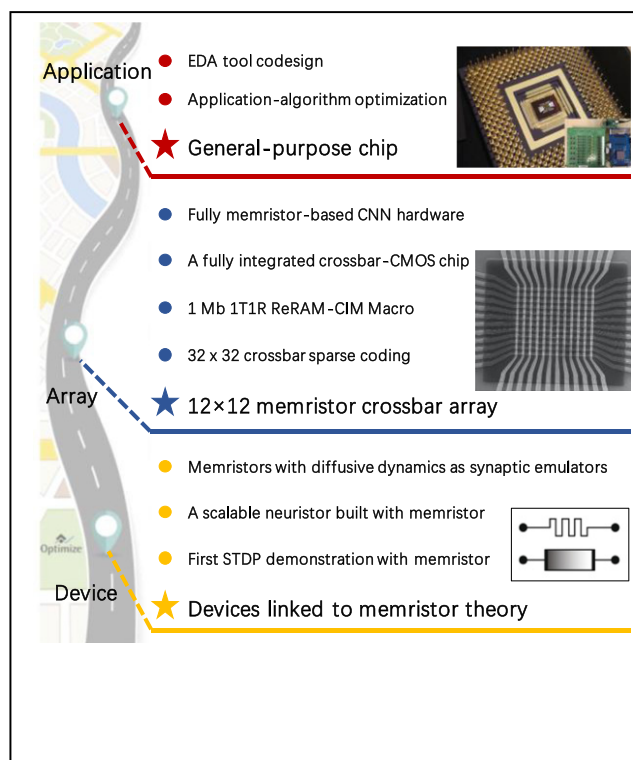
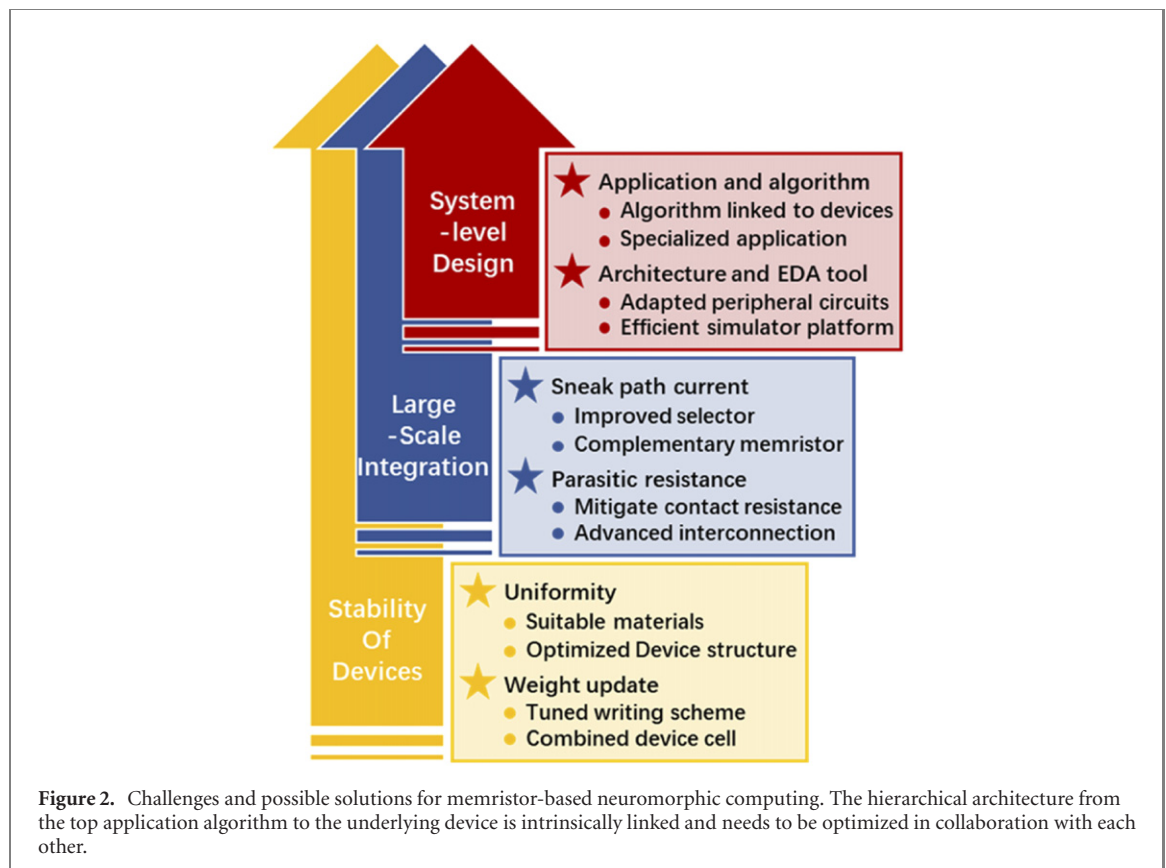


Figure 1. Roadmap for the development of memristor-based neuromorphic computing. Memristors have made milestones on the road to neuromorphic computing, from devices to arrays, but demonstrations of their systems and applications are still in the initial stages. Applicable systems and algorithms are necessary for the future realization of the goal of general-purpose neuromorphic computing in the long term. Panel inset ‘devices linked to memristor theory’ is reproduced from reference [3]. Panel inset ‘12 × 12 memristor crossbar array’ is reproduced from reference [7]. Panel inset ‘fully memristor-based CNN hardware’ is reproduced from reference [11].

1.3. Advances in science and technology to meet challenges

The development and maturation of a new computing paradigm will inevitably confront many obstacles. With the combined efforts and wisdom of the research community and the advancement of science and technology, it is feasible to overcome the challenges encountered at all levels from device to array to the system. For the underlying device stability, the uniformity of the device can be improved by limiting the position of the conductive filament (CF) (for example, by introducing dislocation defects [18] and localized doping [19]) or by increasing the device $R_{\text{off}}/R_{\text{on}}$ ratio (by reducing device area [20] to increase R_{off} for the filamentary type) to compensate for resistance fluctuation. The nonlinear weights can also be alleviated with optimized programming methods [15] using programming pulses that are adapted to the device characteristics for operation. The problem of asymmetry can be overcome by combining multiple devices as a unit structure [21]. Just as silicon is to CMOS technology, finding the materials and structures that best fit the function of the memristors may be the fundamental solution to these device problems.

To allow the array size to be scaled up, it is possible to minimize the sneak-path current by using better selectors [22], and another attempt is to build complementary memristors [23]. An ideal selector is as critical as the memristor in determining the ceiling of the array size. Parasitic resistance between interconnects is also a major challenge affecting array performance. Finding a better conductive material for the wire or using more advanced interconnects to reduce parasitic resistance is a goal that researchers have been working on. A further compromise to mitigate the need for large arrays is combining multiple small arrays to implement different layers of computing. The design of the top-level system influences the collaboration between each level from top to bottom. The algorithms developed according to the properties of the underlying devices can mitigate the stringent requirements for the devices and arrays [17]. A successful algorithm can also lead to customized, specialized applications, unlocking the potential of brain-inspired computing. A generic, proven architecture is the bedrock for building a system, where the design of peripheral circuitry incurs a substantial overhead. Compatible high-performance conversion circuits and controllers enable more efficient data flow within the system. What can be learned from the mature CMOS technology is that EDA tools can effectively collaborate on the design of the entire system at a hierarchical level. Creating an efficient simulation platform will facilitate integrated development from the bottom to the top (figure 2).



1.4. Concluding remarks

The traditional computing paradigm has become unsustainable under the demands of today's big data and IoTs. Neuromorphic computing with memristors is a shock to the ecosystem that has long been dominated by silicon-based CMOS technology and von Neumann architecture. Although neuromorphic computing has the theoretical potential to be very disruptive, the progress made so far has not yet demonstrated its advantages for immediate application. The challenges it faces are not intangible; to overcome them requires joint research efforts across disciplines. It is expected that breakthroughs in materials science or even at the physical mechanism level will greatly improve the performance of neuromorphic devices. Theoretical advances in neuroscience are also contributing to a better understanding of biological computational behavior, which could potentially inspire more efficient algorithms. The mutual development of software, hardware, and theory will lay the foundation for the realization of ultimate brain-inspired intelligence.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61925402, 61851402, 62004040, and 62090032), Science and Technology Commission of Shanghai Municipality (19JC1416600), Shanghai Education Development Foundation, and Shanghai Municipal Education Commission Shuguang Program (18SG01).

2. Two-dimensional two-terminal memristors

Lin Chen* and Tian-Yu Wang

Fudan University, People's Republic of China

E-mail: linchen@fudan.edu.cn and tywang@fudan.edu.cn

2.1. Status

The application of AI technology has promoted the progress of the information age and put forward higher requirements for computing functions for the electronic devices. Traditional computers with von Neumann architecture are facing bottlenecks, such as bandwidth mismatch and huge energy consumption, inspiring the development of novel computing paradigm with in-memory computing (IMC) capability and low power consumption. Emerging devices and materials are needed for the development of novel computing architecture. Neuromorphic computing memristors are promising physical devices to construct IMC system, which show advantages of scalability, high density integration, high-speed operation, and low power consumption [24–26].

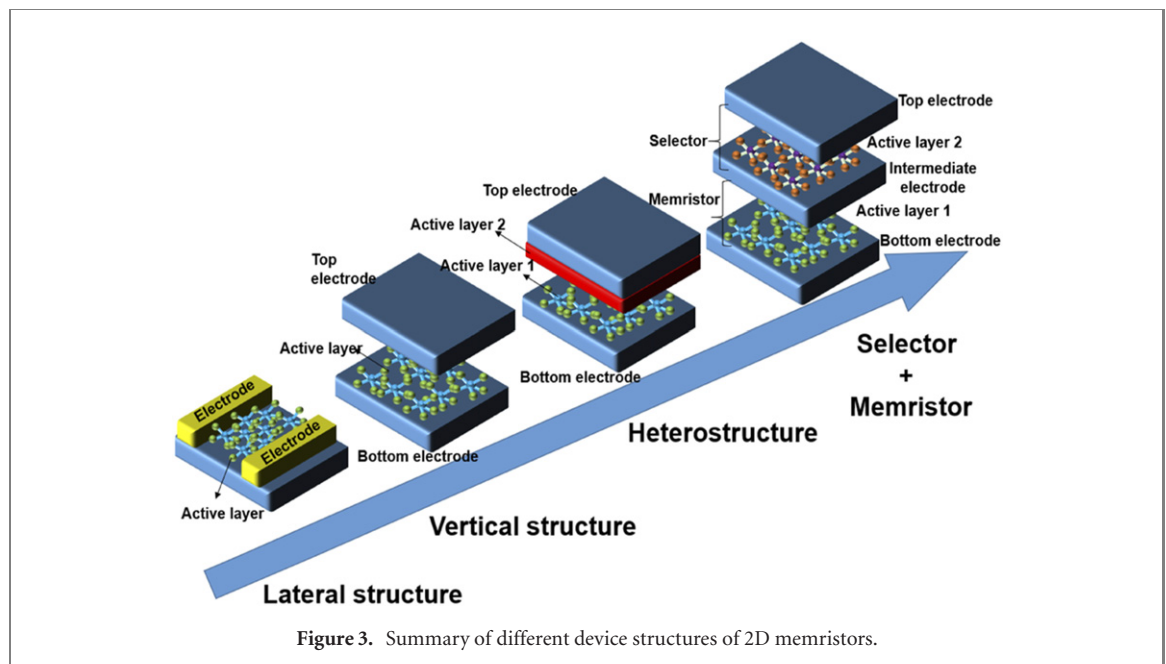
As the feature size of integrated circuits (ICs) gradually moves toward physical limits, it is hard to follow Moore's law under the feature size below 1 nm for traditional silicon-based semiconductor devices. Two-dimensional (2D) layered materials as a promising candidate could work with a thickness of sub-1 nm, showing great potential in next-generation semiconductor materials due to high mobility, capability of overcoming short channel effects, and great size reduction, such as graphene, hexagonal boron nitride (h-BN), molybdenum disulfide (MoS₂), tungsten disulfide, and indium selenide [27–33]. Efforts have been made to prepare 2D memristors for neuromorphic computing applications. Various device structures have been proposed to improve the performance of 2D memristors, including lateral structure, vertical structure, heterostructure, and 'one-selector one-memristor' (1S1M) structure [4, 34–36] (figure 3). The lateral structure-based memristor consists of two horizontal electrodes, and it is beneficial to introducing ions or other materials for modification research. The vertical structure-based memristor consists of a top electrode, a bottom electrode, and a middle active layer, which is the mainstream device structure. There is one more functional layer in heterostructure-based memristors than the vertical structure. For the circuit-level application of neuromorphic computing, the 1S1M structure that combines a selector and a memristor was proposed. Vacancy defect accumulation and CF migration are the main mechanisms in most reported 2D memristors. The 2D memristors may provide more possibilities for a new era of IMC.

2.2. Current and future challenges

Over the past 10 years the 2D memristors have been studied to act as the core components of neuromorphic computing system. However, the CMOS process compatibility, integration capability, reliability, stability, and unclear mechanisms of 2D neuromorphic computing memristors still limit their industrialization and promotion [37]. The reports of 2D memristors are focused on the demonstration of a single device, and the challenge is mainly in the growth of wafer-scale 2D materials for memristor array. Although 2D films could be prepared by chemical vapor deposition, the high temperature is not CMOS compatibility, which limited the promotion of this process. Due to the high temperature and long process time, it is hard to obtain a large-area 2D film for heterogeneous integration with silicon-based ICs. Second, the challenge for 2D neuromorphic computing memristors is reliability and stability. Due to the random defect density in different samples, the electrical properties of 2D memristors are hard to control. More studies focused on the optimization of crystallinity by developing the thin-film growth process to obtain uniform device performances. Third, the sneak-path current of crossbar array limited the integration capability of 2D neuromorphic computing devices. Although 2D materials could easily scale down to sub-nanometer level and integrated to 3D structure, the sneak-path current problem should be solved by integrating different 2D functional units. For future 2D neuromorphic computing electronics, the unclear mechanisms of devices and biological NN need to be investigated for better simulating biological behaviors by 2D memristors and achieving high-efficiency computing.

2.3. Advances in science and technology to meet challenges

From the perspective of material growth, it is in urgent need to develop a suitable process compatible with CMOS technology to fabricate 2D memristors. Atomic layer deposition and advanced transfer technology provide more possibility for fabrication of large-area 2D memristors. From the perspective of device design, further optimization of device structure and lattice defects with the help of advanced electron beam lithography process and rapid thermal annealing process would be helpful to improve performance of device. From the perspective of array integration, the development of a two-terminal selector, a three-terminal transistor, and self-rectifying device would be helpful to solve the problem of leakage and sneak-path current in neuromorphic computing crossbar array based on 2D memristors [38, 39]. From the perspective of mechanism,



advanced characterization methods in biological and IC disciplines, such as *in situ* electron beam microscope and conductive atomic force microscopy, would be helpful to understand the work mechanism.

2.4. Concluding remarks

The 2D memristors have shown great potential in low power consumption, small size, and bandgap adjustable and flexible neuromorphic computing electronics. Although the CMOS process compatibility, integration capability, reliability, stability, and work mechanisms of 2D memristors are still challenges, the advanced fabrication process, transfer technology, characterization methods, and novel device structure could help to break the development bottlenecks and push the promising 2D memristor-based neuromorphic computing system.

Acknowledgements

L Chen and T Wang contributed equally to this work. This work was supported by the National Key Research and Development Program of China (2021YFA1202600), National Natural Science Foundation of China (92064009, 61904033, and 62004044), Program of Shanghai Subject Chief Scientist (18XD1402800), National Postdoctoral Program for Innovative Talents (BX2021070), Zhejiang Lab's International Talent Fund for Young Professionals, China Postdoctoral Science Foundation (2021M700026), the Young Scientist Project of the Ministry of Education Innovation Platform, and Support Plans for the Youth Top-Notch Talents of China.

3. Metal-oxide memristors for IMC

Yi Li^{1,2}, Kan-Hao Xue^{1,2}, Yu-Hui He^{1,2} and Xiang-Shui Miao^{1,2}

¹ Huazhong University of Science and Technology, People's Republic of China

² Hubei Yangtze Memory Laboratories, People's Republic of China

Email: liyi@hust.edu.cn, xkh@hust.edu.cn, heyuhui@hust.edu.cn and miaoxs@hust.edu.cn

3.1. Status

The memristor concept was theorized in 1971 by Professor Chua, and in 2008 researchers at HP Labs linked the concept to the experimentally studied oxide resistive devices [40, 41]. Owing to the advantages of high speed (down to 100 ps), low power consumption (down to femtojoule), high endurance ($>10^{12}$), high scalability (down to 2 nm size), three-dimensional (3D) stacking capability ($4F^2/n$; n is the number of 3D stacked layers), and the back end of line (BEOL) process compatibility, not only can the memristors meet the requirements of high-performance NVM applications (in this case, RRAMs), but they also show the ability to perform basic computation operators, such as Boolean logic and dot product operations. The idea of memristors as a cornerstone of memory and IMC in the post-Moore era has gained widespread acceptance (figure 4) [4, 11, 42].

Dating back to the 1960s, Hickmott reported the low-frequency negative resistance phenomenon in metal oxide devices, and after decades of exploration a large group of oxides with metal/insulator/metal device configurations have been reported to show frequency-dependent pinched hysteresis loops, typical fingerprints predicted by Chua [40, 41]. Compared with complex oxides, like SrTiO_3 , BiFeO_3 , and $\text{Pr}_{0.7}\text{Ca}_{0.3}\text{MnO}_3$, simple binary amorphous metal oxides (especially HfO_x , TaO_x , WO_x , and AlO_x) are in the spotlight of academia and industry because of their SOTA performance and low process cost. A number of oxide-based memristor array prototypes (up to Mb-class capacity) have already been developed (table 1).

In addition to the next-generation universal memory receiving the most attention, the binary switching behavior of memristors has been proposed to perform stateful logic to complement conventional charge-based logic, and high reconfigurability has been achieved with these memristors. Generic memristor arithmetic logic unit and some application-specific hardware for logic-intensive computing scenarios (data query, Hamming distance calculation, exclusive-OR [XOR] encryption, etc) have been demonstrated to provide efficient alternatives.

Besides, in recent years, the multilevel or analog properties of memristors have been fully exploited to emulate the weight updating and storage behaviors of biological synapses, and thus applied to the development of highly intriguing brain-inspired neuromorphic devices and systems. In addition, based on the analog matrix computing capability of crossbar arrays, accelerators for various machine learning algorithms (such as clustering, classification, regression, deep networks, and combinatorial optimization) have been developed with astounding improvements in computing power and energy efficiency (>100 tera-operations per second [TOPS] W^{-1}) over conventional processors, like CPU and graphic processing unit (GPU).

3.2. Current and future challenges

To outstand from various device candidates for IMC, memristors must overcome challenges, including elucidating the physical origin, device optimization, large-scale integration, and industrial applications.

Although it has been intensively exploited by first-principle calculation, *in situ* characterization, and device modeling [43, 44], an in-depth understanding of resistive switching behaviors in the vast range of memristor materials is still to be achieved. For a typical binary oxide HfO_x that works under the binary switching mode, the CF has recently been revealed to be crystalline metal phases [44], whose formation and rupture account for the large resistive window. However, for doped oxides and the memristive materials designed for gradual resistance modulation, the exact physical origin, such as metal filaments, interfacial switching, or defective leakage paths, still needs to be clarified for various materials and/or doping schemes, which is also the basis for further optimization of device performance. Such understanding is particularly critical for analog computing applications, where a stable, high-precision programming is highly desirable and meanwhile nonlinearity, variability, and other nonideal factors should be effectively suppressed.

At the array level, the successfully demonstrated one-transistor one-resistor macros are limited by the large transistor area [45–48]. One-selector (Ovonic threshold switches, insulator–metal transition switches, or metallic filament-based threshold switches) one-resistor architecture is a viable compact alternative yet far from mature [49, 50]. Here the biggest challenge is to improve the overall performance, including drive capability, selectivity, endurance, and uniformity. In favor of low leakage and large switching window, the metallic CF-type oxide selectors are superior to Ovonic or metal–insulator transition types. Nonetheless, the advantages of CF selectors are offset by the random nature of ion migration, slow relaxation process, and atomic

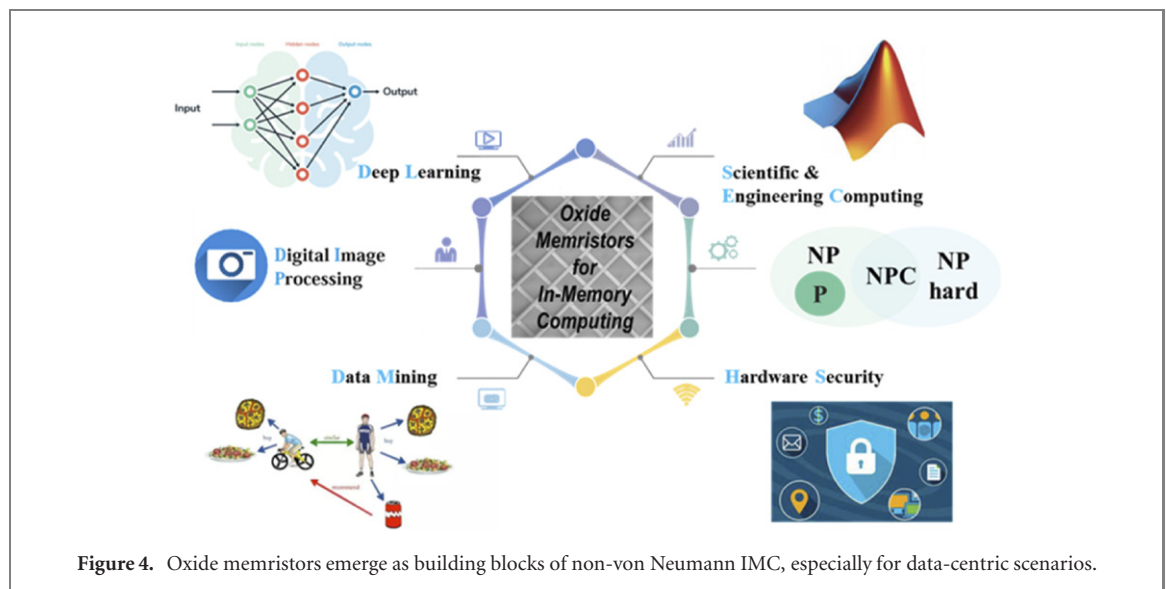


Figure 4. Oxide memristors emerge as building blocks of non-von Neumann IMC, especially for data-centric scenarios.

Table 1. Summary of SOTA oxide memristor arrays.

Type	Passive array	1T1R array ^{23,23} 1T1R: 1 transistor and 1 resistor			4T2R	1S1R array		Self-rectifying array	Special 3D stacking
Year	2019 [11]	2020 [45]	2021 [46]	2021 [47]	2020 [48]	2013 [49]	2021 [50]	2021 [51]	2020 [25]
Device	Au/ WO _x / Au	TiN/ TaO _x / HfO _x / TiN	W/ TiN/ TiO ₂	O _x ReRAM	Ta/ TaO _x / Pt	W/AlO _x / TaO _y / Ta ₂ O _{5-x} / Pt/TiN/ AsTeGeSiN/ TiN	V/VO _x / HfWO _x /Pt	Hf _{0.8} Si _{0.2} O ₂ / Al ₂ O ₃ /Hf _{0.5} Si _{0.5} O ₂	Ti/ HfO ₂ / Ta
Device size (μm ²)	0.25	0.25	0.025	0.022	0.16–0.29	0.5	100	900	0.6
Transistor node (nm)	NA	180	22	14, FinFET	16	NA	NA	NA	NA
Array size	5 kb	2 kb	4 Mb	1 Mb	~1 kb	64 bit	1 kb	100 kb	4 kb
Application	Analog IMC (machine learning)	Analog IMC (CNN)	Binary IMC (CNN)	Memory	Binary IMC (TCAM)	Memory	Analog IMC (SNN)	Analog IMC (matrix multiplication)	Binary IMC (CNN)

accumulation that leads to the drawbacks of large variation, long latency, and limited lifetime. Other nonlinear devices, like self-rectifying memristors, are candidates with 3D integration potential [25, 51, 52]. However, a dense oxide tunneling or barrier layer introduced to provide reverse rectification usually leads to a significant increase in the operating voltage (exceeding 5 V or even 10 V), and thus unfavorable power consumption.

At the application level, a new functionality proposes new demands on device performance [42]. First, *in situ* learning requires frequent updating of device conductance, implying that endurance improvement needs to be focused on. Breakthroughs in computing energy efficiency also mean that the power consumption (both write and read) must be consistently reduced. Second, quantization and variation-aware algorithms to compensate for the limited accuracy and non-idealities also have to be developed. In addition, efficient programming methods in arrays, peripheral circuits to support parallel computing, and analog-to-digital and digital-to-analog conversion modules remain technical challenges. Besides, how to realize high-precision computing using low-precision memristors appears to be a substantial challenge with great scientific value [53].

3.3. Advances in science and technology to meet challenges

To address the challenges mentioned previously, a collaborative strategy of hardware and software is required, with the pull of specific applications. For the NN accelerators or neuromorphic computing systems, recent developments with vacancy-deficient/vacancy-rich structures (such as TaO_{2-x}/Ta₂O_{5-x} and Al₂O₃/AlO_x) are particularly promising since high performance and reliability have been demonstrated while the materials and

processes are compatible with CMOS [49]. The asymmetric structure can create a vacancy profile to control CF formation, which in turn precisely modulates the device conductivity. Besides, this multilayer structure helps to enhance the overall resistance, and hence reduces the read power consumption. In this way, the energy efficiency of the inference process may be significantly improved. For hardware/software collaboration, write-with-verify or closed-loop methods are widely employed to realize high precision (6–7 bits) programming. Other designs, such as quantization or binary NNs, have also been proposed to mitigate the harsh requirements of device analog properties. Efficient training algorithms, along with improvement in device endurance, together offer the possibility of *in situ* training implementation.

In contrast to the IMC with compelling analog devices, binary IMC does not require high-precision programming and has the potential for commercialization in the short to medium term using the currently available device technology, and therefore deserves better attention from researchers. For instance, binary memristor-based ternary content-addressable memories (TCAMs) have been experimentally demonstrated for pattern matching, showing superior performance over the traditional static random-access-memory (SRAM)-based TCAMs [48, 54]. Particular attention has been paid to strategies of leveraging digital IMC to speed up hyperdimensional computing, one-shot learning, and database query, which involve a large number of comparison and logic operations on binary sequences, with significantly higher energy efficiency over CMOS processors [54, 55]. Besides, the use of the bit slicing concept in binary crossbar, redundancy programming strategy, and mixed-precision architecture has been proved to be efficient approaches to improve the precision of IMC [42, 53, 56].

3.4. Concluding remarks

In summary, the promising incorporation of high-performance memory and computing capabilities allows metal-oxide memristors to become one of the building blocks for beyond-Moore non-von Neumann computing architecture. The International Roadmap for Devices and Systems and the Institute of Electrical and Electronics Engineers Rebooting Computing Committee believe that at the crossroads where Moore's law is difficult to sustain, the IMC paradigm is the way to go, with the collaborative development of new technologies from the device, circuit, architecture, and algorithm levels. To promote the commercial application of memristors, efforts must be paid to mechanism characterization, device design and optimization, BEOL integration, and killer applications research. Interdisciplinary and cross-level innovation is needed to understand and overcome the challenges. Although IMC is still in its fancy and needs further extensive investigation in the next 5–10 years, especially on the generality and energy efficiency limits of this technology, the maturity of oxide memristor technology will lay a solid device foundation for high-performance computing (HPC) systems. We also believe that in the ubiquitous cloud and edge computing, memristor-based IMC will play a pivotal role.

Acknowledgements

The authors gratefully acknowledge support from the National Key R&D Program of China (Grant No. 2019YFB2205100), National Natural Science Foundation of China (Grant No. 61874164 and 92064012), Hubei Key Laboratory of Advanced Memories, Hubei Engineering Research Center on Microelectronics, and Chua Memristor Institute.

4. Neuromorphic devices and applications based on phase-change random-access memory

Li Xi, Xie Chenchen, Chen Houpeng and Song Zhitang

Chinese Academy of Sciences, People's Republic of China

E-mail: itluck@mail.sim.ac.cn, xcc@mail.sim.ac.cn, chp6468@mail.sim.ac.cn and ztsong@mail.sim.ac.cn

4.1. Status

The advent of big data and AI era has brought unprecedented challenges to the conventional hardware platforms based on von Neumann architecture, such as ‘storage wall’ and ‘power wall’. To address these issues, a novel conception of neuromorphic computation was first proposed by Mead in 1990 [57]. It has attracted extensive attention with high parallelism, low power consumption, and integration of storage and computation. In the past decade, great progress has been made in constructing neuro-inspired computing systems by imitating the information processing mechanism of biological brain nervous system through mainstream CMOS technology, including Neurogrid [58], TrueNorth [59], Loihi [60], and Tianjic [61]. However, the Moore’s law constraint, complex peripheral circuits, volatility, and incapability of online learning hinder the development of more advanced transistor-based neuro-inspired systems.

On the other hand, depending on unique mechanism, more and more novel structure devices start to directly mimic the behavior of synapses and neurons from the physical level, so as to construct the neuromorphic computing systems. Thus, these devices are also called neuromorphic devices. Most neuromorphic devices are expanded from emerging NVMs with their superior speed and power consumption than flash and excellent analog conductance regulation, like RRAM [46], ferroelectric random-access memory (FeRAM) [62], and magnetic random-access memory (MRAM) [63]. Phase-change random-access memory (PCRAM), as a competitive candidate, utilizes the large resistance contrast between amorphous and crystalline states of chalcogenide materials to store information and shows tremendous potentials in the field of neuromorphic computations [64]. The research of IBM shows that by optimizing the material composition and device structure of PCRAM, more than 1000 resistance states of memory cell can be realized [65], which lays a foundation for realizing multi-precision analog computation. Taiwan Semiconductor Manufacturing Company (TSMC) compared the image recognition performance of convolutional neural networks (CNNs) based on PCRAM, RRAM, and MRAM, respectively, and the result proved that PCRAM can achieve higher inference accuracy due to its larger resistance ratio [66]. In 2020, Peking University and Shanghai Institute of Microsystem and Information Technology realized the local high-precision training of CNNs [67], eligibility traces for energy-efficient reinforcement learning [68], and uncertainty quantification applications [69] with the multi-resistance-level of PCRAM chip. In 2021, a 14 nm CMOS and phase-change-memory (PCM)-based IMC core with 10.5 TOPS W^{-1} energy efficiency and 1.59 TOPS mm^{-2} performance density is presented by IBM. Therefore, the application of PCRAM as a neuromorphic device is the most promising approach to break through the bottleneck of energy efficiency, integration, memorability, and online learning in the field of neuro-inspired computing systems.

4.2. Current and future challenges

Despite the privileged advantages of PCRAM as a neuromorphic device to construct the neuro-inspired computing system due to the increasing process of materials science and the technology accumulated in mass production, there are still some inevitable obstacles impeding its further development in higher-level intelligent applications. First, higher reliability, lower power consumption, and window value of conductance are the most fundamental requirements of PCRAM as kernel device in a high-performance and energy-sensitive neuromorphic computing hardware. The reliability is embodied in the program/read speed, thermal stability, retention, and endurance, which are directly related to the performance of hardware and also urgent issues to be solved as novel memory. The power consumption of PCRAM mainly originates from the programming pulses driven by instantaneous current, which acts as the update of synaptic weight in neuro-inspired computation. The window value of device conductance is associated with the bit width of synaptic weight and will eventually affect the accuracy of computation. Second, the key point when imitating the behavior of NNs is to mimic the synaptic plasticity that refers to a characteristic or phenomenon of adjustable connection strength between neurons. Whether long-term synaptic plasticity (LTSP), short-term synaptic plasticity, or spike-timing-dependent plasticity (STDP), all these kinds of synaptic plasticity are embodied as electric-induced controllable conductance in neuromorphic device. However, owing to the fast phase transition and unique asymmetric switching mechanism of PCRAM, limited linearity and symmetry of conductance regulation can be achieved currently. In addition, since the biological neural system is generally composed of billions of neurons connected by trillions of synapses, in the application of PCRAM devices imitating NNs on a large-scale, the device variability has to be a hard nut to crack in mass production. Last but not least, as a

unique phenomenon of PCRAM device, resistance drift, mainly caused by structural relaxation [70], is also an ineluctable challenge in practical application scenarios.

4.3. Advances in science and technology to meet challenges

To address the mentioned challenges, it is essential to form cross-level collaborative innovation and optimization strategies, including materials, devices, manufacturing process, circuits, and algorithms. Research on materials science, like optimizing composition and doping engineering, cannot only reduce the melting point of PCRAM device to cut down the energy consumption [71], but it also retards the crystallization to obtain the linear and progressive conductance regulation [72]. Furthermore, high-thermal-stability doping can enhance the reliability so as to avoid the excessive complexity of hardware training and the loss of computing accuracy. Employing new materials, like carbon nanotubes or graphene as electrode and advanced device shield, can obviously improve the heating efficiency during write program and restrict inter-device variability, respectively. In terms of device design and structure, the specific thin conducting surfactant layers [73, 74] and the phase-change heterostructure [75] have been proved to be further effective in restricting resistance drift. As for circuit and array design, differential mTnR cell structure and reference cell-based resistance tracking provide a set of feasible solution for symmetrical conductance regulation and device uniformity. Finally, associated with the algorithms, PCRAM-based neuromorphic computing hardware platforms with high accuracy and energy efficiency will lead to a promising future in practical applications.

4.4. Concluding remarks

Neuro-inspired computing is a systematic research involving multiple disciplines, and various emerging memory devices have been tested recently for their suitability. The PCRAM technology, as the most mature and promising neuromorphic device, is expected to become the core technology to replace existing von Neumann architecture and realize the in-memory-computing, which can achieve great breakthrough in computing and energy efficiency.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (91964204 and 92164302), Strategic Priority Research Program of Chinese Academy of Sciences (XDB44010200), and Science and Technology Council of Shanghai (19JC1416801), and in part by the Shanghai Research and Innovation Functional Program (17DZ2260900).

5. Physically transient memristors for neuromorphic computing

Hong Wang* and Yue Hao

Xidian University, People's Republic of China

E-mail: hongwang@xidian.edu.cn

5.1. Status

Physically transient electronics is a new class of technology that the devices and/or systems are designed to physically disappear after completing the specific functions [76]. This concept is proposed by John Rogers in 2012, and electronic devices based on materials that can dissolve in water or bio-fluids have been reported. Physically transient devices based on biodegradable materials demonstrate excellent potential for hardware security, bio-integrated electronics, and implantable medical system applications [77, 78]. On the other hand, brain-inspired neuromorphic computing systems using the co-location of logic and memory concepts, offering a promising solution for the lower power consumption and highly efficient computing that can break through the limits of the von Neumann bottleneck [4]. Combining the advantages of physically transient electronics and neuromorphic computing to develop physically transient neuromorphic computing systems would have a unique advantage and excellent potential for brain-machine communication, such as prosthetics, *in situ* bio-signal recording and feedback, and implantable and/or bio-integrated neural medicine like adaptive biohybrid interfaces, and repair communication between neurons and security computing system applications.

Various emerging devices, such as PCM, FeRAM, and RRAM, have been proposed to perform as artificial neuromorphic devices [79]. Among these, memristor is a promising candidate due to its simple structure, flexible material selection, low power, and high efficiency in computing. So physically transient memristor-based neuromorphic computing has great advantages in brain-computer interface, security electronics, and bio-integrated systems. Physically transient devices, including transistor, diode, and sensors, using biocompatible and biodegradable materials were firstly demonstrated in 2012 [76]. The physically transient memristor based on natural biopolymer with dissolvable properties was proposed in 2016 [80]. Transient memristor to mimic biological synaptic functions had already been demonstrated in 2018 [81]. Thereafter, material optimization and device design of memristor for physically transient neuromorphic devices have demonstrated their feasibility to simulate biological synapses and neurons [82–84]. Materials, including Mg, W, and Mo for electrodes and MgO, ZnO, and MoO_x for switching layers, have shown great advantages for transient memristor application. Although the electrical performance of physically transient memristors still cannot compare with conventional memristors based on transition metal oxides, and integrated arrays are still lacking, the development of physically transient memristors is at a very early stage and faces great challenges, but it has a promising future for physically transient neuromorphic systems.

5.2. Current and future challenges

Physically transient memristors have made rapid progress from materials to the larger-area array fabrication technology, but there remain several fundamental and technical challenges to be resolved. Stability and degradability of transient memristors always show a competitive relationship. However, they should meet the requirements of maintaining stable electrical characteristics with negligible variation with time and physically disappearing immediately after completing their functions or encountering a trigger signal. The currently available transient memristors are mostly based on biodegradable and dissolvable materials, which suffer from degradation in ambient environment. The mechanism and methods to balance the stability and degradability are still a challenge. The device performance of transient memristors, especially endurance and device-to-device uniformity characteristics, is far behind the demand of neuromorphic computing. How to design materials and device constructions to meet the requirements for long-term stability and quick degradability as well as the remarkable electrical performance for neuromorphic device application are basic and central issues.

For integrated array, compatibility with a larger area and with the high-density nano-fabrication technology is the key to commercialization and is still a challenge. On the other hand, feasibility of device size scaling down needs to be verified since the currently available physically transient memristors are still in micron scale. Therefore, wafer-scale fabrication technology for nanodevices and integrated arrays is required. For bio-integrated systems, it is desirable that the operation voltage of transient memristor neuromorphic devices is compatible with the action potential of bio-systems to meet the requirements of brain-machine communication [85]. But the operation voltages of transient memristors far exceed the low bio-voltage. In addition, how to connect artificial devices with living biological systems to realize intelligence brain-machine interfaces needs to be explored. For general applications, triggering methods for transient electronic devices disappearing are very important. Careful and accurate designing of transient neuromorphic systems is needed to control and program their degradation rates and disappearance times.

5.3. Advances in science and technology to meet challenges

The development of physically transient neuromorphic systems from concept to application will meet serious challenges. With the advances in science and technology through efforts from multiple areas of the research community, overcoming challenges from materials to devices to integrated systems is desirable. Because of the contradiction between stability and degradability, materials with different chemical properties are expected to form unique multilayer stacked structures to maintain stable electrical performance in air and rapid degradation in biofluids and/or water. In addition, designing of special chemical bonds in the material system is also possible to achieve such demands. Protection or encapsulation layers are an effective choice for transient memristors to ensure stability of their electrical performances before they disappear. Based on developed materials, the endurance and uniformity of the devices can be enhanced by referring to the proposed methods for conventional memristors, such as controlling the size and shape of CFs.

For large-scale integrated arrays, selecting of wafer-scale and nano-fabrication technology-compatible materials is the most effective option for material and device-level feasibility. Furthermore, a technology called transfer printing, which can transfer devices and circuits on top of wafers to various hetero-substrates such as degradable substrates, is an ideal fabrication technology for integrated transient neuromorphic systems [86]. With advances in materials and transfer printing, it is no doubt that the size of devices can scale down. Fully implemented transient memristor arrays also depend on the progress of the whole transient electronic area. Transient transistors that can fit with transient memristors would greatly boost the realization of physically transient neuromorphic systems. For bio-integrated systems, deep understanding of ionic transport properties both from materials and biological view is helpful in controlling ionic transport to design bio-voltage transient memristors for brain-machine communication. The degradation of transient devices that results in their disappearance can be controlled by applying external forces, such as thermal force, light, and fluids, to these devices [87]. With combined effects from the areas of materials, physics, engineering, etc, and with advances both in software and hardware, physically transient memristors will show promising applications (figure 5).

5.4. Concluding remarks

Physically transient electronic devices that can disappear on demand would lead to transformative new applications. Physically transient memristors, which integrate the advantages of transient electronics and neuromorphic computing, may yield various innovative applications, such as biological interactive intelligence and security computing. The development of transient memristors is just beginning, and great advantages for application have not yet been demonstrated. The main challenges range from materials design to fabrication technology to system application, including the contradiction between stability and degradability, wafer-scale and nano-fabrication technology, and degradation trigger methods. To overcome these challenges needs effort and cooperation from different disciplines. We believe that breakthroughs in chemistry and materials science will resolve the contradictions between stability and degradability for transient memristors. Research progress of materials and electrical engineering will achieve larger-scale integration technology of physically transient neuromorphic devices. Joint research of materials, devices, integration technology, and even biology both from theoretical and experimental views will pave the way for biological integrated intelligence.

Acknowledgements

This work was supported by the National Key Research and Development Program under Grant 2018YFB2202900, National Natural Science Foundation of China (61574107), Opening Project of Key Laboratory of Microelectronic Devices and Integrated Technology, and Institute of Microelectronics of the Chinese Academy of Sciences.

6. Recent advances in transistor-based organic synapses in China: from device to system

Junyao Zhang and Jia Huang

Tongji University, People's Republic of China

E-mail: 1910586@tongji.edu.cn and huangjia@tongji.edu.cn

6.1. Status

Driven by the development of next-generation artificial intelligent systems, the requirement for neuromorphic electronics that can emulate fundamental functions of the human nervous system is significantly increasing [88]. The human nervous system, which consists of $\sim 10^{11}$ neurons and $\sim 10^{15}$ synapses, can be divided into the peripheral nervous system (PNS) and the central nervous system (CNS) [89]. The CNS is capable of calculating and storing massive information simultaneously obtained from the PNS, which has merits of lower energy consumption and faster information processing compared with von Neumann computers. Synaptic plasticity and its event-driven update functions through repetitive spike-form stimulations can be considered as primary elements of energy-efficient computing. Thus, the development of artificial intelligent computing systems at a hardware level demands the construction of neuromorphic electronics with the ability to simulate key factors for processing and memory performance of the CNS. Furthermore, the PNS is composed of sensory and motor nerves and organs. Human beings detect numerous stimuli from the complicated environment and make proper instructions through the cooperation of the CNS and PNS. External stimuli are perceived through the sensory part of the PNS and processed through the CNS, and then corresponding behavioral instructions are given by the motor part of the PNS. Hence, artificial intelligent systems can be developed with the cooperation of neuromorphic electronics, which are capable of accurate perception, efficient data processing, and motor coordination.

Organic synapses are representative components of neuromorphic electronics, which enable to mimic synaptic plasticity and correlative functions [90]. Organic materials have many merits of easy processability for solution-processing techniques, low cost, and simple regulation by molecular properties [91]. Moreover, compared with inorganic materials, organic materials with outstanding ductility and biocompatible mechanical performance are desired for flexible devices, which have promising application prospects in future wearable and bioinspired electronics and robotics. Organic synapses have attracted wide interests of researchers all over the world, and then different kinds of device configurations are proposed. Compared with two-terminal synaptic devices, three-terminal synaptic transistors can execute signal transmission and self-learning concurrently [91]. Also, they can have multiple gates to receive signals from diverse sources at the same time, so dendrite integration and spatiotemporal effect can be realized. Through suitable structure design and material selection, synaptic transistors can transform external stimuli into electrical signals, which makes them possible to realize neuromorphic electronics that directly respond to the surrounding environment. Chinese researchers have made many remarkable contributions in the field of transistor-based organic synapses.

6.2. Current and future challenges

One of the biggest challenges in transistor-based organic synapses is to imitate various biological synaptic characteristics. According to Hebb's theory, short-term plasticity, long-term plasticity, STDP, and spike-rating-dependent plasticity are basic forms of synaptic plasticity [92]. Consequently, it is believed that synaptic devices should simulate these neuroplastic behaviors. Also, it is expected that full synaptic characteristics can be emulated in the next 10 years. Next, energy efficiency is regarded to be a significant factor in constructing an artificial intelligent system based on synaptic devices. A biological synapse consumes ~ 10 fJ per synaptic event; it is essential to find effective ways to decrease the power consumption [91]. Most of the present transistor-based organic synapses have a large gap with the biological counterparts in terms of energy consumption and device size [88]. Besides, since flexibility and stretchability are the extremely important advantages of organic transistors, it is rational to apply transistor-based organic synapses in flexible, conformable, and stretchable neuromorphic electronics [89].

With the successive advancement of transistor-based organic synapses, hardware achievements of artificial intelligent systems have been in the spotlight. The major research objective of neuromorphic electronics is to simulate the neuromorphic computing functions of the human brain in hardware. Besides basic synaptic functions, remarkable progressive neuromorphic computing functions should also be emulated. Transistor-based organic synapses must be integrated in the form of arrays to implement system-level neuromorphic computing. Most studies have still utilized algorithms to establish ANNs for the simulation of system-level neuromorphic computing based on the characteristics of one or several devices. The hardware implementation of ANNs based on organic synaptic transistors requires further efforts, which can be one goal for the next

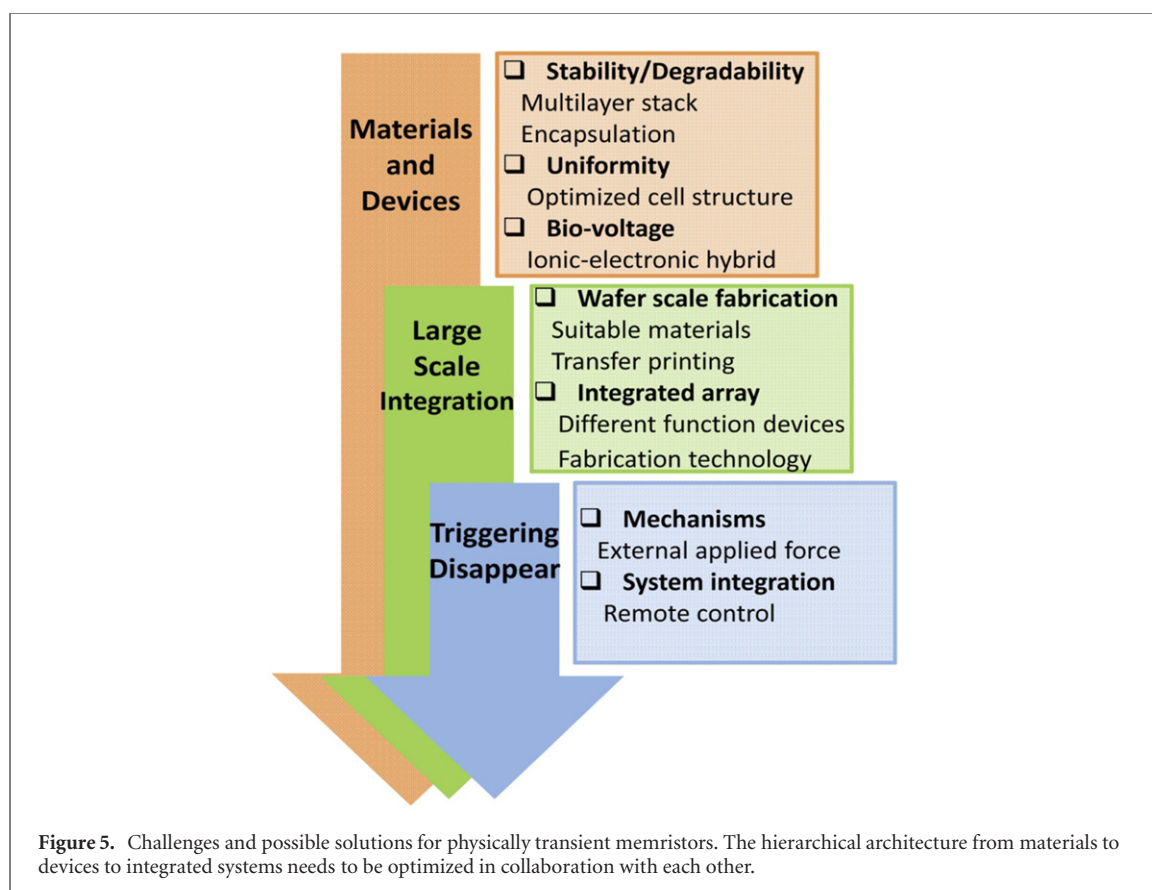


Table 2. Summary of transistor-based organic synapses in China.

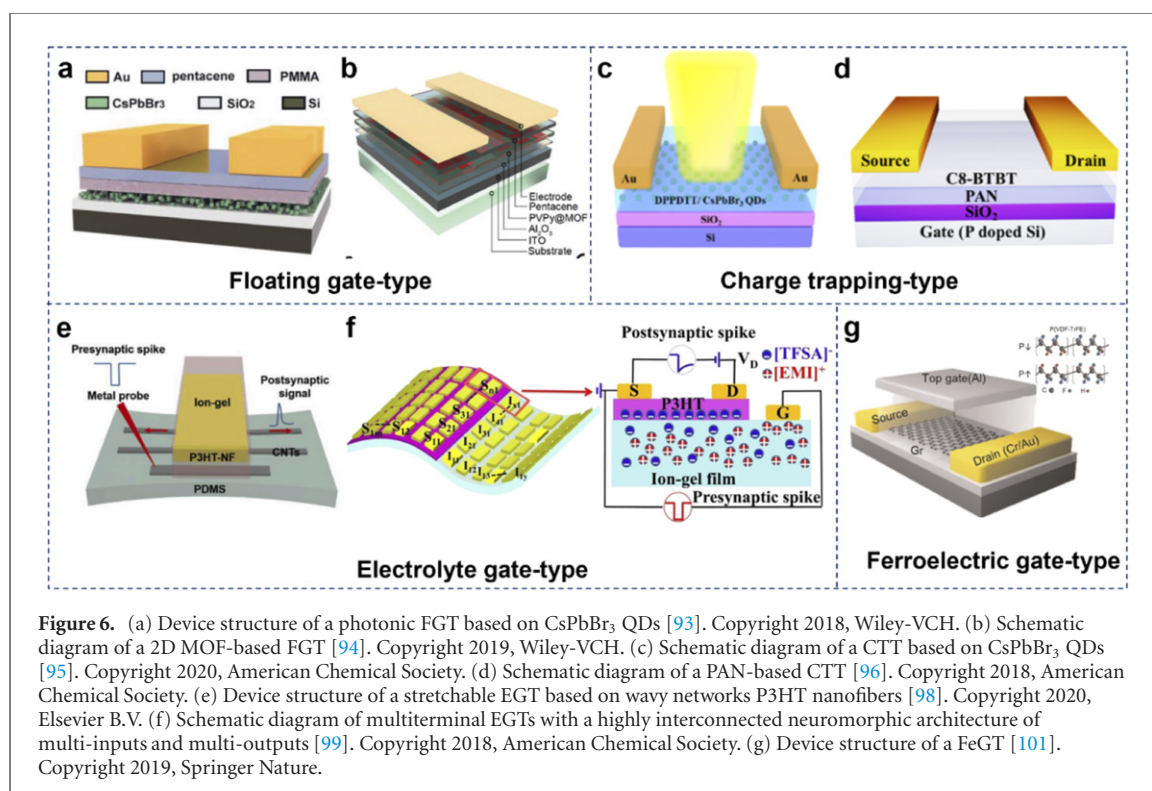
Type	Materials
FGT	Floating gate: C ₆₀ , perovskites, reduced graphene oxide, metal–organic frameworks, topological insulators, carbon dots, 2D imine polymers, black phosphorus–ZnO hybrid nanoparticles, and upconverting nanoparticles Tunneling layer: PMMA, PS, and silk Channel: pentacene, IDTBT, and PDPP4T Active additives: perovskites, MoS ₂ , porphyrin, PLA, and chlorophyll Dielectric: PAN and SiO ₂
CTT	Channel: carbon nanotubes, C8-BTBT, DNTT, PDPP4T, DPPDTT, TIPS-pentacene, PQT-12, and PTCDA Electrolyte insulator: ionic liquid, ion gels, and solid-state electrolyte
EGT	Channel: PQT-12, C8-BTBT, P3HT, IDT-BT, PEDOT: PSS, DNTT, and rubrene Ferroelectric insulator: P(VDF-TrFE)
FeGT	Channel: graphene and PIID-BT

10 years. The hardware implementation of ANNs may also encounter next-level issues, such as uniformity, endurance, reliability, interconnection, and complicated processing problems.

In addition to neuromorphic computing applications, transistor-based organic synapses should also be explored to build artificial perception and motor systems. Through proper suitable structure design, material selection, and device integration, artificial perception systems should be able to convert external stimuli (light, pressure, smell, sound, etc) from the external environment into electrical signals. Then, the corresponding behavioral instructions can be manipulated through artificial motor systems. These bionic processes are similar to sensing and motoring functions of biological sensory systems. It is of great significance to develop a comprehensive artificial intelligent system that can realize multiple perception and motor functions in the next 5 years.

6.3. Advances in science and technology to meet challenges

Through desirable device configuration and material selection, different kinds of transistor-based organic synapses for the hardware implementation of next-generation artificial intelligent systems have been widely studied. So far, four kinds of transistor-based organic synapses have been presented, including floating gate transistor (FGT), charge trapping transistor (CTT), electrolyte gate transistor (EGT), and ferroelectric gate transistor (FeGT) (table 2).

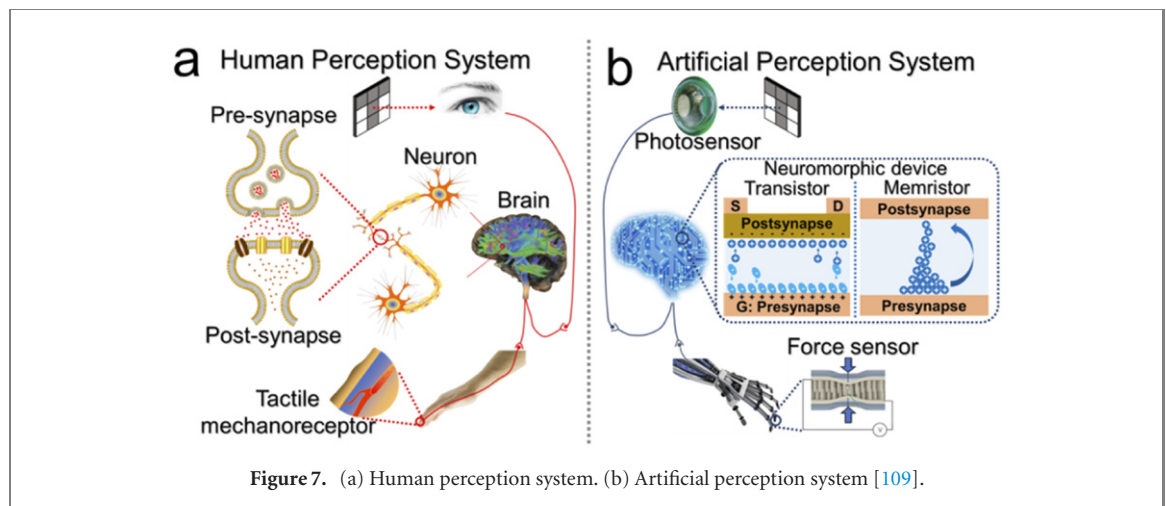


FGTs possess a large on/off ratio and a controllable channel conductance. Thus, they are suitable to achieve long-term synaptic potentiation and depression functions. Both 0D materials [93] and 2D materials [94] (figures 6(a) and (b)) have been extensively explored for the construction of FGTs. For CTTs, charge carriers can be trapped at active additives and active additives/organic semiconductor (OSC) interfaces [95] (figure 6(c)) or dielectric/OSC interfaces [96] (figure 6(d)). Trapped charges offer an extra electric field, leading to a regulatable channel conductance, which is a platform for emulating synaptic characteristics. EGTs are capable of using ions in electrolyte dielectric to adjust the channel conductance. Conformal EGTs with the combination of flexible material components have been reported, which presented remarkable conformability on 3D curved surfaces [97]. Besides, to satisfy compatibility with dynamic surfaces for practical applications, stretchable EGTs were explored [98] (figure 6(e)). Biologically, a neuron is capable of receiving multiple input signals, and each received signal can be transmitted to numerous neurons. Similarly, multiterminal EGTs with a highly interconnected neuromorphic architecture of multi-inputs and multi-outputs were demonstrated [99] (figure 6(f)). More importantly, EGTs with low operating voltage are appropriate for low-energy-consumption synaptic devices. A low power consumption of 0.29 fJ per synaptic event was implemented in organic single-crystalline nanoribbon-based EGTs [100]. Ferroelectric insulator with spontaneous polarization states is the critical component in ferroelectric gate transistors. The channel conductance of FeGTs can be gradually controlled through the modulation of polarization states in the ferroelectric insulator. FeGTs utilizing P(VDF-TrFE) ferroelectric polymer as the gate dielectric were demonstrated [101] (figure 6(g)). Through the gate adjusting, both hole-dominated and electron-dominated transport could be realized, leading to tunable bipolar characteristics.

For neuromorphic computing systems, significant progressive functions, including logic operation [95], filtering [102], and association learning [98], have been successfully achieved. In addition, different kinds of artificial perception systems have been explored, such as visual- [103], tactile- [104], olfactory- [105], and auditory-perception systems [106]. Artificial motor systems have also been built with the combination of an organic synaptic transistor and an artificial motor element [107]. These bionic processes are similar to sensing and motoring functions of biological nervous systems.

6.4. Concluding remarks

Many efforts have been made by Chinese researchers in both material and structure to design novel transistor-based organic synapses that can emulate biological synapses with similar functions and energy consumption level. Although neuromorphic systems based on organic synaptic transistors have been implemented in a wide range of applications, the development of artificial intelligent systems is still in an early stage, and various challenging issues remain to be addressed for practical applications. For example, only a small number of synaptic functions have been emulated. Full biological synaptic characteristics and complete human perception systems should be imitated. Up to now, artificial perception systems have the capability of sensing only one



type of sensing signals, which limit the efficient resolution for the coupling issue of sensing information in the future. By integrating diversified fields, such as materials science, microelectronics, medicine, and computer science, we speculate that the large-scale deployment of artificial intelligent systems will no longer be a dream.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (62074111), Science & Technology Foundation of Shanghai (19JC1412402 and 20JC1415600), Shanghai Municipal Science and Technology Major Project (2021SHZDZX0100), and Shanghai Municipal Commission of Science and Technology Project (19511132101).

7. Brief discussion of oxide neuromorphic transistors in China

Zheng Yu Ren^{1,2,3}, Li Qiang Zhu^{1,2} and Qing Wan⁴

¹ Ningbo University, People's Republic of China

² Chinese Academy of Sciences, People's Republic of China

³ ShanghaiTech University, People's Republic of China

⁴ Nanjing University, People's Republic of China

7.1. Status

Von Neumann architecture is highly efficient in dealing with structural data. However, it is less energy efficient in dealing with intelligent tasks. Especially with the developments of IoT technology, it is necessary to process massive amounts of data in an energy-efficient way. Thus, new computation architectures are highly desirable in dealing with such tasks. In 1950s, John McCarthy *et al* have proposed AI, aiming to mimic human brain [108]. Especially after the success of the AI robot 'AlphaGo', studies on AI have aroused worldwide attention. With deep learning technology, AI has entered into a new era in the fields of speech, image processing, and natural language. Conventionally, AI is built based on von Neumann architecture and complex computer programming codes. However, high energy consumption is always needed because of the limited parallel computation that originated from physical separation of computation modules and memory units in von Neumann architecture (i.e. the von Neumann bottleneck).

As comparison, our brain is a highly parallel biological computation system. It has $\sim 10^{11}$ neurons and $\sim 10^{15}$ synaptic connections. Neurons are the computation engines, while synapses are the basic units of signal processing. Due to the huge number of parallel synaptic computations and the unique synaptic plasticity activities, brain computation is highly reliable with strong fault tolerance. Thus, our brain can handle unstructural real-time data and complex tasks with a low energy consumption of ~ 20 W, including perception, learning, thinking, memory, and decision making. Therefore, it is interesting to mimic biological synapses and neurons on hardware devices. Recently, studies on brain-inspired neuromorphic devices are getting an important branch of AI.

Additionally, our body is a multifunctional perception learning system. External stimuli are received on sensory organs and are transmitted to the CNS through afferent nerves. The CNS processes information, resulting in the formation of perception activities. Thus, designing bionic NNs at hardware level based on neuromorphic devices is becoming a research hot spot, which would provide a possible solution to make energy-efficient AI, as schematically shown in figure 7 [109]. Presently, investigation of artificial perception learning system based on neuromorphic devices is also getting a new research branch. It would provide new opportunities for extending human intelligence.

7.2. Current and future challenges

Imitation of synaptic computation and neural information processing mode in our brain on hardware devices is an important step to realize the brain-inspired hardware-based neuromorphic system, which will greatly promote the development of information technologies. Two-terminal synaptic devices have been proposed to mimic synaptic functions, including memristors, PCMs, atomic switches, etc. They have priorities, including simple structure, low operation power consumption, small physical size, and easy 3D integration. In contrast, transistors have the advantages of controlling electrical performances much easier. Additionally, multigates can be integrated in a single device. Thus, transistors have also been proposed for neuromorphic device applications. In neuromorphic transistors, the gates and the channel are generally regarded as pre- and post-synapses, respectively. Channel conductance is always deemed as synaptic weight. Because of high carrier mobility, excellent optical transparency, and large area preparation, oxide thin-film transistors have been widely investigated for next-generation displays. Recently, oxide transistors have also been proposed for neuromorphic electronic applications. Several synaptic functions have been imitated. In addition, some complex synaptic activities and NN algorithms have been imitated, including spatiotemporal information integration [110], pattern recognition [111], and dendritic algorithms [112].

An artificial perception learning system usually adopts sensors and neuromorphic devices. Sensors are responsible for perceiving and transforming external stimuli into electrical signals. External stimuli perceived by sensors, including photosensor, force sensor, etc, can be delivered to neuromorphic devices. Neuromorphic devices can process population coding in a way similar to that in brain NN. These processes require synaptic transmission and synergistic algorithm to complete complex neural perception activities. Presently, artificial perception learning systems have also attracted increasing attention. To date, some artificial perception systems based on oxide neuromorphic transistors have been reported, such as tactile perception systems [113], visual perception systems [110], and artificial sensory neuron (ASN) with visual-haptic fusion [114]. Basically, a single neuromorphic device generally cannot achieve functions of perception and sensing at the same time. It

Table 3. Summary of oxide neuromorphic transistors in China.

Type	Multi-gate	Gate dielectric	Channel
EGT ^a	Single gate	Ionic liquid or ion gels	WO ₃ , IZO, In ₂ O ₃
		Solid-state electrolyte	InO _x , ITO, In ₂ O ₃ , ZnO NWs
	Dual gate	Ionic liquid; ion gels	SrCoO _{2.5}
		Solid-state electrolyte	IZO, ITO
		Ionic liquid or ion gels	IGZO
FeFET	Multi-gate	Solid-state electrolyte	IZO, ITO
Conventional transistor	Single gate	PZT	IGZO
	Single gate	SiO _x , SiN _x , AlO _x , HfO _x , SiO ₂	IGZO

^aAcronyms used in table: EGT (electrolyte gated transistor), PZT (PbZr_{0.2}Ti_{0.8}O₃), SiO_x (Silicon oxide), SiN_x (Silicon nitride), AlO_x (Aluminum oxide), HfO_x (Hafnia oxide), SiO₂ (Silicon dioxide), WO₃ (Tungsten trioxide), IZO (Indium zinc oxide), In₂O₃ (Indium oxide), InO_x (Indium oxide), ITO (Indium tin oxide), ZnO (Zinc oxide), NWs (Nanowires), SrCoO_{2.5} (Strontium chromium oxide) and IGZO (Indium gallium zinc oxide)

is necessary to develop multifunctional electronic devices with perception and memory to imitate the sensory activities and reflex activities of our body.

7.3. Advances in science and technology to meet challenges

There are several reports on oxide transistor-based neuromorphic devices, including EGTs, ferroelectric field-effect transistor (FeFET), and conventional transistor. Here, we briefly discuss the recent progresses in China, as shown in table 3.

In terms of EGT-based neuromorphic transistors, different electrolytes have been proposed, including ionic liquids and ionic gels [115] and solid-state ionic liquid electrolytes [116]. Additionally, inorganic solid-state electrolytes have several priorities, including good chemical stabilities, CMOS process compatibilities, etc. Thus, inorganic solid-state electrolytes have also been adopted in oxide neuromorphic transistor [117]. Furthermore, laterally coupled oxide neuromorphic transistors have also been fabricated. Spatiotemporally correlated signal processing has been mimicked [118]. Due to the lateral coupling activities, multi-gates could be integrated into a single neuromorphic device to receive multi-inputs, resembling biological neurons well. Thus, laterally coupled neuromorphic transistors show great potentials for building ANNs. Advanced NN algorithms can be demonstrated [112, 119], including a proof-of-principle visual system for emulating lobula giant motion detector neuron, dendrite integration, orientation tuning functions, and neuronal gain control (arithmetic) in the rate coding scheme. With the nimble selection of multiple inputs in a multi-gate oxide neuromorphic transistor, complex hybrid functions could be realized.

Ferroelectric dielectrics have spontaneous polarization states, which can be altered by an external electric field. Thus, modulation could be precisely controlled by accurately providing the gate voltage. In other words, multiple-channel conduction states can be obtained. Thus, FeFETs also have broad application prospects in neuromorphic engineering [120]. Oxide transistors gated with conventional gate dielectrics also possess the priorities of large-scale integration and compatibility with CMOS technology. Recently, such transistors have been proposed for neuromorphic electronic applications [111, 121]. However, it is difficult to mimic LTSP for oxide transistors gated with conventional dielectrics. Fortunately, light-induced persistent photoconductivity can overcome this problem well. Therefore, conventional dielectric gated transistors also have potentials in optoelectronic neuromorphic devices.

Additionally, oxide neuromorphic transistors have also been proposed to construct artificial perception systems, which can mimic the information sensing and processing in a biological system. The main strategy is to connect neuromorphic transistors with different sensing units. To date, some artificial perceptual systems based on oxide neuromorphic transistors have been proposed in China, including tactile perceptual system [113, 122, 123], visual perceptual system [124–126], and auditory perceptual system [127], as shown in table 4. Recently, mimicking the biological perception activities is getting an important branch for the cognitive platforms.

7.4. Concluding remarks

Oxide neuromorphic transistors have been proposed to simulate various synaptic activities, such as synaptic plasticity (e.g. excitatory postsynaptic current [EPSC], paired-pulse facilitation [PPF], post-tetanic potentiation, and long-term potentiation [LTP]), synaptic learning rules (e.g. STDP and spike-rate-dependent plasticity [SRDP]), and advanced learning activities (e.g. associative learning, pattern memory, pattern recognition, and dendritic integration). In addition, artificial perceptual systems based on oxide neuromorphic transistors have also been proposed, such as tactile perceptual system, visual perceptual system, and auditory perceptual

Table 4. Summary of artificial perceptual systems based on oxide transistors in China.

Type	Structure	Stimuli	Channel	References
Tactile perception system	Pressure sensor + transistor	Pressure	ITO, IGZO	[113, 122, 123]
Visual perceptual system	Phototransistor	Light + electricity	IGZO, IGCO, ^a ITO, In ₂ O ₃	[124–126]
Auditory perception system	In-plane synaptic transistors	Electricity	IGZO	[127]

^aIGCO: Indium gallium cadmium oxide

system. However, it should be noted here that more efforts are still needed. Only limited perceptual functions have been mimicked. As comparison, our body perceptual system is a multisensory hybrid perceptual system, which makes our body to respond to outer surroundings in an energy-efficient way. Therefore, it is highly desirable to realize multi-perceptual functions on a single perceptual platform in a synergic mode. Fortunately, multi-gate oxide neuromorphic transistors would act as fundamental building blocks to receive different stimuli from multi-input units. It should be noted here that the number of gates depends on the adoption of materials and the device processing. However, it is possible that tens or hundreds of gates would be integrated to a single-oxide neuromorphic transistor. Thus, the device processing should be updated. Furthermore, it is also a great challenge to realize the interconnection and large-scale integration of oxide neuromorphic transistors in a single chip.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (51972316), Ningbo Key Scientific and Technological Project (2021Z116), and Zhejiang Provincial Natural Science Foundation of China (LR18F040002).

8. Neuromorphic devices based on functional oxides

Jianyu Du and Chen Ge*

Chinese Academy of Sciences, People's Republic of China

E-mail: 1910586@tongji.edu.cn, dujianyu@email.tjut.edu.cn and gechen@iphy.ac.cn, huanjia@tongji.edu.cn

8.1. Status

Building an 'intelligent machine', which could process like a human brain, has always been the dream of scientists. The famous work, 'Intelligent machinery', was written by Turing in 1948 [128]. He proposed a new unorganized machine that can mimic 'cognitive' functions of humans (i.e. 'learning' and 'problem solving'). Turing also investigated the feasibility of fabricating an intelligent machine. However, due to the limitation set by the condition at that time, the assumption of the intelligent machine cannot be realized.

Over the past few decades, due to the rapid development of neuroscience, the functions of neurons and synapses have been uncovered and documented. The neural behaviors and synaptic plasticity are the main research content of the neural systems. In the meantime, materials sciences were also growing rapidly, which paved the way for mimicking the biological behaviors by manipulating electronic materials.

The concept of 'neuromorphic computing' was first proposed by Mead in the late 1980s, and the first neural-inspired chip was fabricated by Mead [129]. Based on the early development, remarkable progresses have been achieved during the past decade. For example, the CMOS-based neuromorphic chips TrueNorth (IBM) and Loihi (Intel) were launched, which provided an impressive industrial debut [130, 131]. With the rapid growth of the big data application market, traditional computers are limited to the 'von Neumann bottleneck', which cannot afford cost-effective and sustainable scaling. The high-power consumption and bandwidth problem, brought on by von Neumann architecture, has not been fundamentally solved. To resolve these issues, an increasing number of new devices, called neuromorphic devices, were proposed as basic neuromorphic units, which can provide high speed, low power consumption, and high bandwidth. Over the past decade, various sorts of neuromorphic devices have emerged successively [132], which mainly fall into three categories via their operation principles: resistive random access memory (ReRAM) [133, 134], PCM [135, 136], FeFET [137], and others. All these emerging devices were used to simulate the function of biological synapses and neurons, and all have been widely applied to neuromorphic circuits to accelerate matrix–vector multiplications to realize machine learning tasks. Of course, like most new technologies, these emerging devices also face various challenges.

8.2. Current and future challenges

Although neuromorphic devices have greatly developed in recent years, there are many remaining challenges that need to be addressed. For fully realizing the goal of neuromorphic computation, the artificial neuromorphic device should have good characteristics, such as high switching ratio, ultralow power consumption, good endurance behavior, great potential to scale down, super running speed, etc. Therefore, the challenge that has been presented to us is to develop new materials and fabricate new devices that can simultaneously meet all the requirements. Among the emerging materials, functional oxides seem likely to have the potential to achieve this goal. In the last few years, functional oxides are widely used in the field of neuromorphic device, which was supported by various operation principles, such as drift, diffusive, phase change, ferroelectric, and magnetic. These devices, based on different mechanisms, may perform in one aspect, while still suffer in other aspects. The benchmark of the SOTA performance of devices with different mechanisms is shown in figure 8. Take the ReRAM for instance. The intrinsic randomness is associated with the growth and rupture of conducting filaments, which can be used to simulate both neuronal and synaptic dynamics. Nevertheless, the ReRAM materials, such as TaO_x, HfO_x, and TiO_x, often face reliability issues [134, 138, 139]. Similarly, although FeFET exhibits some promising characteristics as an electronic synapse, such as fast programming operations, symmetric potentiation and depression curves, and large switch ratio, it suffers from scaling limitations.

For the PCM devices, the resistivity change originates from phase change. There are two kinds of phase transitions. One is that PCM materials transform between crystalline phase and amorphous phase, and the other one is that PCM materials transform between two different crystalline phases. For the first one, temperature plays a very important role in the transformation, so thermal management is a challenge, particularly for scaled devices. For the other one, the structure of the functional oxides changes without breaking the lattice framework, which is also called topotactic phase transformation [140]. In this situation, oxygen stoichiometry has a great influence on the physical and electrochemical functionalities. In the topotactic phase transformation, many novel physical properties can emerge due to the coupling between lattice, charge, and spin degrees of freedom. These novel properties can provide new opportunities for the development of neuromorphic device, and many recent studies have been focused on this field. Ge *et al* realized reversible topotactic

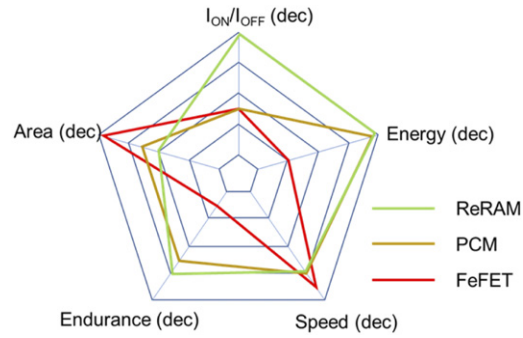


Figure 8. Benchmark of the SOTA performance of devices with different mechanisms.

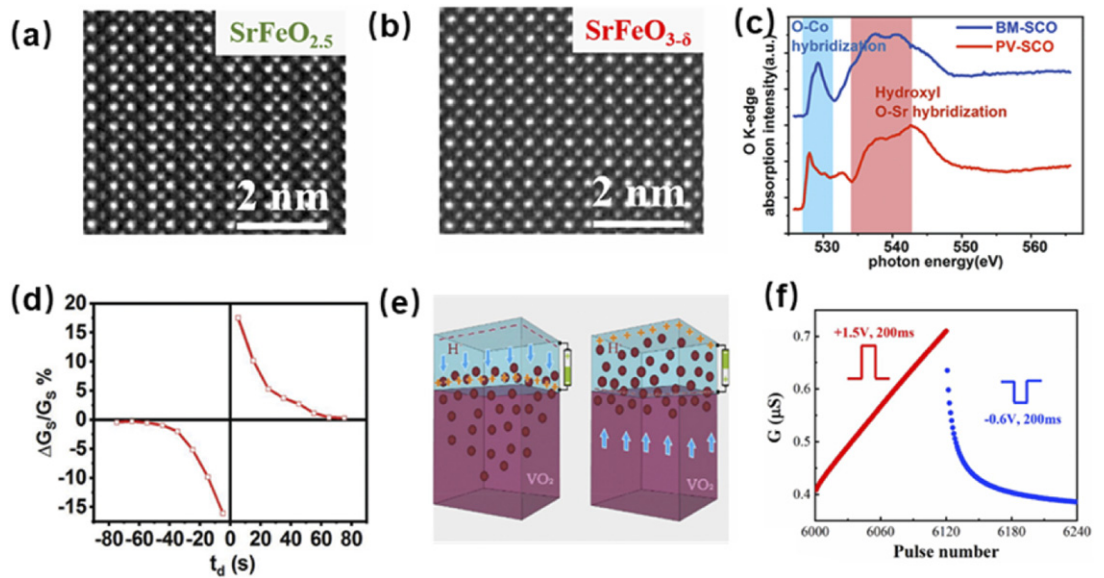
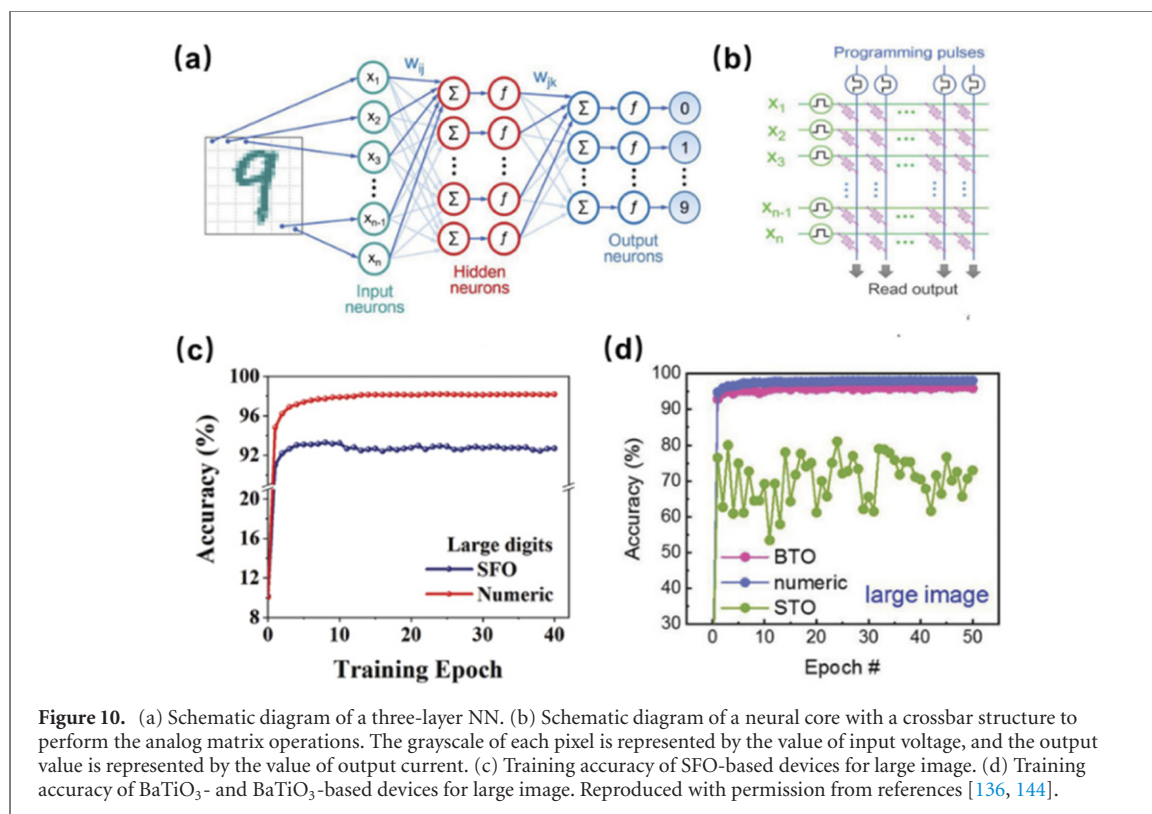


Figure 9. (a) and (b) Cross-sectional HAADF-STEM image of SFO. (c) X-ray absorption spectra of the O-K edges. (d) Asymmetric STDP implemented in the SCO-based synapses. (e) Schematic diagram of hydrogen ion movement in the VO_2 -based devices. (f) Cycle for gating-induced long-term potentiation and long-term depression processes of the VO_2 -based devices. Reproduced with permission from references [135, 136, 141].

phase transformation in strontium ferrite, SrFeO_x (SFO), which can be applied to neuromorphic computing [136] (figures 9(a) and (b)). The SFO-based synaptic transistor exhibited high performance and offered new options for neuromorphic device. A similar phenomenon has been found in SrCoO_x (SCO) and was applied to the simulation of synaptic functions [141] (figures 9(c) and (d)). It should be mentioned that the typical Mott material VO_2 may be the promising candidate for the neuromorphic device due to its stable, reversible, and ultrafast metal–insulator transition [135, 142] (as shown in figures 9(e) and (f)). In other words, the coupling of multi-degree of freedom in the functional oxides provides multiple possibilities for the simulation of synaptic behavior.

From the application aspect, the neuromorphic device should be integrated into deep neural networks (DNNs) for simulating the functionality of the CNS. Pattern classification is one of the most important applications for the DNN (figure 10(a)). The crossbar array made up of neuromorphic devices was proposed for hardware implementation (figure 10(b)). The synaptic weights were mapped into the neuromorphic devices in the crossbar array, and the core operation of vector-by-matrix multiplication is accelerated by the matrix of conductance multiplication at the physical level. During the operation of the DNN, the synaptic weights are adjusted. Regardless of the performance of device units, the number of data states, linearity, and symmetry have always been the top priorities to improve the accuracy of the pattern classification [143]. The synaptic devices based on functional oxides are potential candidates to benefit from the high performance of gradual resistance change and high on/off ratio. The three-terminal synaptic transistor based on SFO exhibits good linearity and symmetry behavior. The accuracy of the SFO-based synaptic transistors came up to 92.7%, which is much higher than the obtained phase-change memory devices [136] (figure 10(c)). Li *et al* demonstrated



an artificial synapse based on ferroelectric tunnel junction. The artificial synaptic device showed excellent linearity, and as many as 200 conductance states were obtained. The simulated ANN exhibited a very high recognition accuracy of 96.4% for the handwritten data set [144] (figure 10(d)). Despite the fact that these new types of synaptic devices exhibited excellent linearity and enough conductance states, more work must be done to further improve performance and adaptability, for example, by combining the excellent characteristics of emerging devices with the conventional CMOS platform [145].

In summary, despite a scene of prosperity of the network-level research is spreading out, we still need to obtain inspiration from neuroscience studies and apply novel physical properties to the simulation of complex neural behaviors. Under such conditions, the modern characterization methods are essential to find new physical mechanisms, which can speed up discovery in materials science to meet the requirement. For commercial availability, further attempts should be made to develop the fabrication procedures, algorithms, and the way of integration for neuromorphic computing.

8.3. Concluding remarks

During the past years, neuromorphic devices based on functional oxides have been a research hot spot with the rapid development of AI. Some prototypes of neuromorphic chips have been demonstrated and have exhibited their superiority in some applications. The future development of neuromorphic devices and systems will be focused on the discovery of new physical mechanisms and materials, new fabrication procedures, efficient algorithms, and a new way of integration. We firmly believe that neuromorphic computation is one potential pathway to break through the ‘von Neumann bottleneck’, which can promote computer power growth to a great extent.

Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2017YFA0303604 and 2019YFA0308500), Youth Innovation Promotion Association of Chinese Academy of Sciences (No. 2018008), and National Natural Science Foundation of China (Nos. 12074416 and 12222414).

9. Development of 2D-based neuromorphic transistor in China

Yang Liu¹, Guanglong Ding², and Ye Zhou² and Su-Ting Han^{1,*}

¹ Shenzhen University, People's Republic of China

² Shenzhen University, People's Republic of China

E-mail: sutinghan@szu.edu.cn

9.1. Status

Building the human brain-like neuromorphic computing system on a hardware level is considered an effective way to overcome problems faced by traditional computers in this data explosion era (e.g. low parallelism and high power consumption) [146, 147]. To achieve this goal, different neuromorphic devices have been developed by using various functional materials to mimic the basic behaviors of synapses, which is the basis of biological NN. A synapse can be roughly divided into three parts: presynaptic membrane, synaptic cleft, and postsynaptic membrane (figure 11) [148]. Under stimulation, a neurotransmitter can be released from the presynaptic membrane into the synaptic cleft and then act on the postsynaptic membrane. In this process, the signal transmission efficiency between two neurons, that is, synaptic weight, can be adjusted. Up to now, synapses have been successfully implemented by different types of electronic devices. Among them, neuromorphic transistors have attracted a lot of attention due to their multiple terminals, excellent stability, clear operation mechanism, and relatively controllable device performances [149, 150].

Due to the atomic thickness, high surface-to-volume ratio, and extreme sensitivity to charge transfer or electrostatic modulations at the interface, 2D materials are considered as one of the most promising potential choices for developing diverse neuromorphic transistors with powerful features [151]. Moreover, van der Waals heterostructures can be composed of a combination of 2D materials, which opened a window for developing various types of electronic/optoelectronic synaptic and heterosynaptic devices [152, 153].

9.2. Current and future challenges

Because the development of 2D material-based neuromorphic transistors is still in the early stage, there are some problems and challenges that need to be solved. First, the physical size of the device array is relatively large, and the fabrication processing technology of uniform 2D material channels is not compatible with the conventional CMOS technology, which discourages large-scale integration [154]. Second, the biological synapse connections are arranged in 3D space, resulting in a great challenge in the interconnection of neuromorphic devices [155]. Third, only a small fraction of biological synaptic behaviors have been imitated [91].

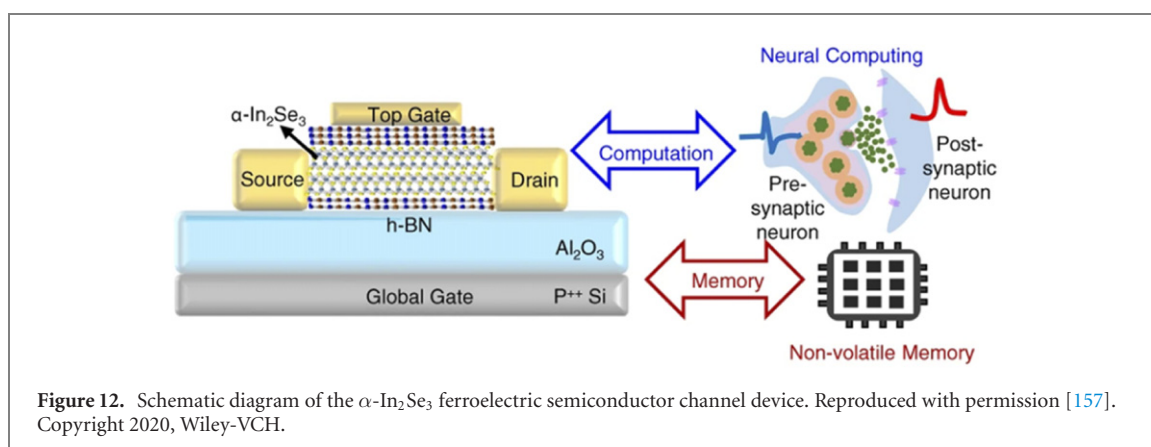
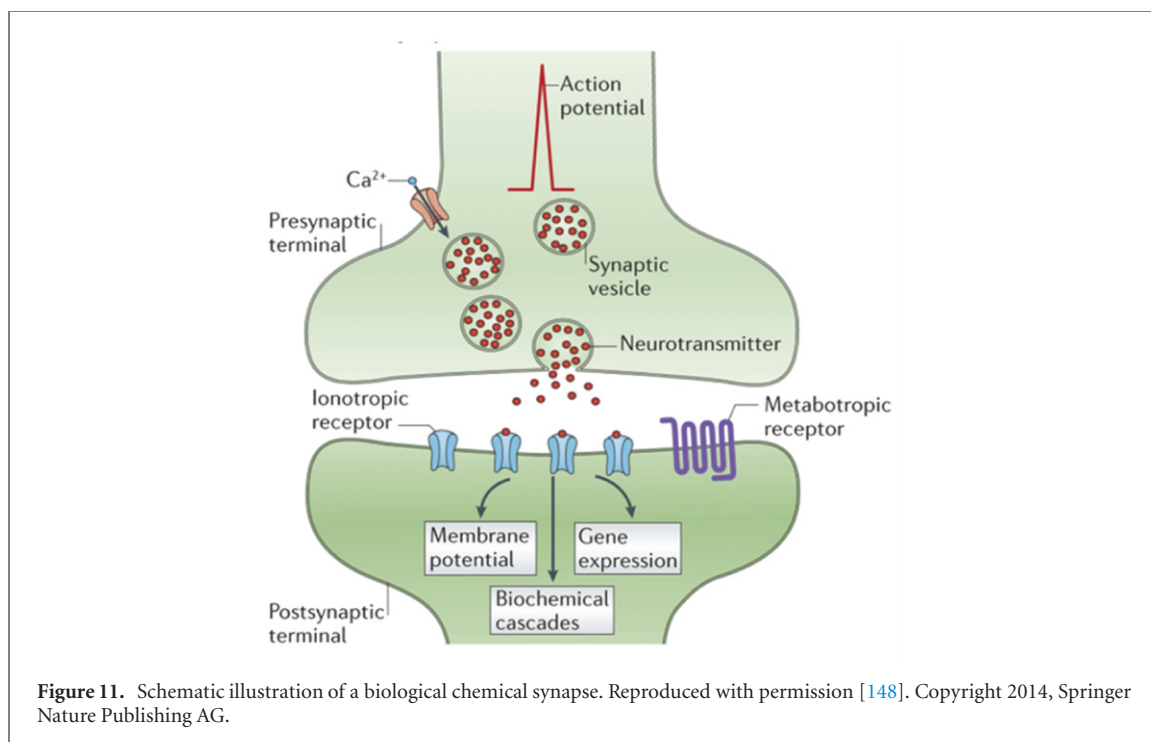
9.3. Advances in science and technology to meet challenges

At present, domestic research still focuses on the realization of high-performance synaptic devices, and gate electrode, device channel, drain electrode, and channel conductance are regarded as the presynaptic membrane, synaptic cleft, postsynaptic membrane, and synaptic weight, respectively. Inspired by the ion transport mode in biological synapses, 2D material-based electrolyte-gated synaptic transistors were developed [151, 156, 157]. In 2018, Zhu *et al* demonstrated an ionic gating-modulated neuromorphic transistor by using WSe₂ [151]. This device has remarkable linearity, symmetry, reproducibility, and ultralow energy consumption of ~30 fJ per spike. In addition, the transistor successfully mimicked various synaptic behaviors, including EPSC, PPF, SRDP, and dynamic filtering.

Distinct from the electrolyte-gated transistors, charges tunneling between the channel and the floating gate can achieve gate-tunable nonvolatile channel conductance, which provides an alternative approach to emulate synaptic plasticity [158, 159]. Jin *et al* incorporated the HfO_x/HfS₂ heterostructure into a WSe₂ flash memory device, achieving a high on/off current ratio of ~10⁵, a large memory window over 60 V, good endurance over 250 cycles, and a long retention time over 10³ s [160]. By applying gate voltages with different polarities, the synaptic depression and potentiation can be further mimicked in the devices.

FeFETs have also attracted considerable attention as a promising platform for mimicking biological synapses, because the carrier concentration of FeFETs can be precisely and gradually modulated by changing the polarization state of ferroelectric materials using the gate voltage [161, 162]. Recently, Wang *et al* demonstrated the α -In₂Se₃ ferroelectric transistors with ultrafast write speed (40 ns), improved endurance, flexible adjustment of neural plasticity, ultralow energy consumption of 234/40 fJ per event for excitation/inhibition, and thermally modulated 94.74% high-precision iris recognition classification simulation (figure 12) [162].

Using light to control the synaptic weight may implement synaptic devices with large bandwidth, low interconnection energy loss, and ultrafast signal transmission, and can help construct novel ANN architectures [163, 164]. Cheng *et al* experimentally demonstrated a novel photoelectrically modulated neuromorphic device based on a vertical 0D–CsPbBr₃–quantum–dots/2D–MoS₂ hybrid-dimensional van der Waals heterojunction [163]. Photoelectrically modulated spiking Boolean logics, dendritic integrations in both temporal



and spatial modes, and Hebbian learning rules can be successfully mimicked in the devices using specific intriguing optical and electrical synergy approach.

9.4. Concluding remarks

In summary, based on the contents discussed previously, we consider that future studies should combine the conventional semiconductor micro–nano manufacturing technology with 2D materials, achieving large-scale application of 2D neuromorphic transistors. Moreover, persistent efforts should be made by researchers to design and fabricate more complicated neuromorphic networks based on 2D synaptic transistors. Last but not least, high-performance and multifunctional devices should be proposed to achieve more biological synaptic behaviors.

Acknowledgements

Y Liu and G Ding contributed equally to this work. The authors acknowledge grants from the National Natural Science Foundation of China (Grant Nos. 62074104, 61974093, and 51902205), Guangdong Province Special Support Plan for High-Level Talents (Grant No. 2017TQ04X082), Guangdong Provincial Department of Science and Technology (Grant No. 2018B030306028), Science and Technology Innovation Commission of Shenzhen (Grant Nos. JCYJ20180507182042530, RCYX20200714114524157, and JCYJ20180507182000722), National Taipei University of Technology–Shenzhen University Joint Research Program, and Natural Science Foundation of Shenzhen University.

10. Neuromorphic transistor for bionic perception

Changjin Wan and Qing Wan

Nanjing University, People's Republic of China

E-mail: cjwan@nju.edu.cn and wanqing@nju.edu.cn

10.1. Status

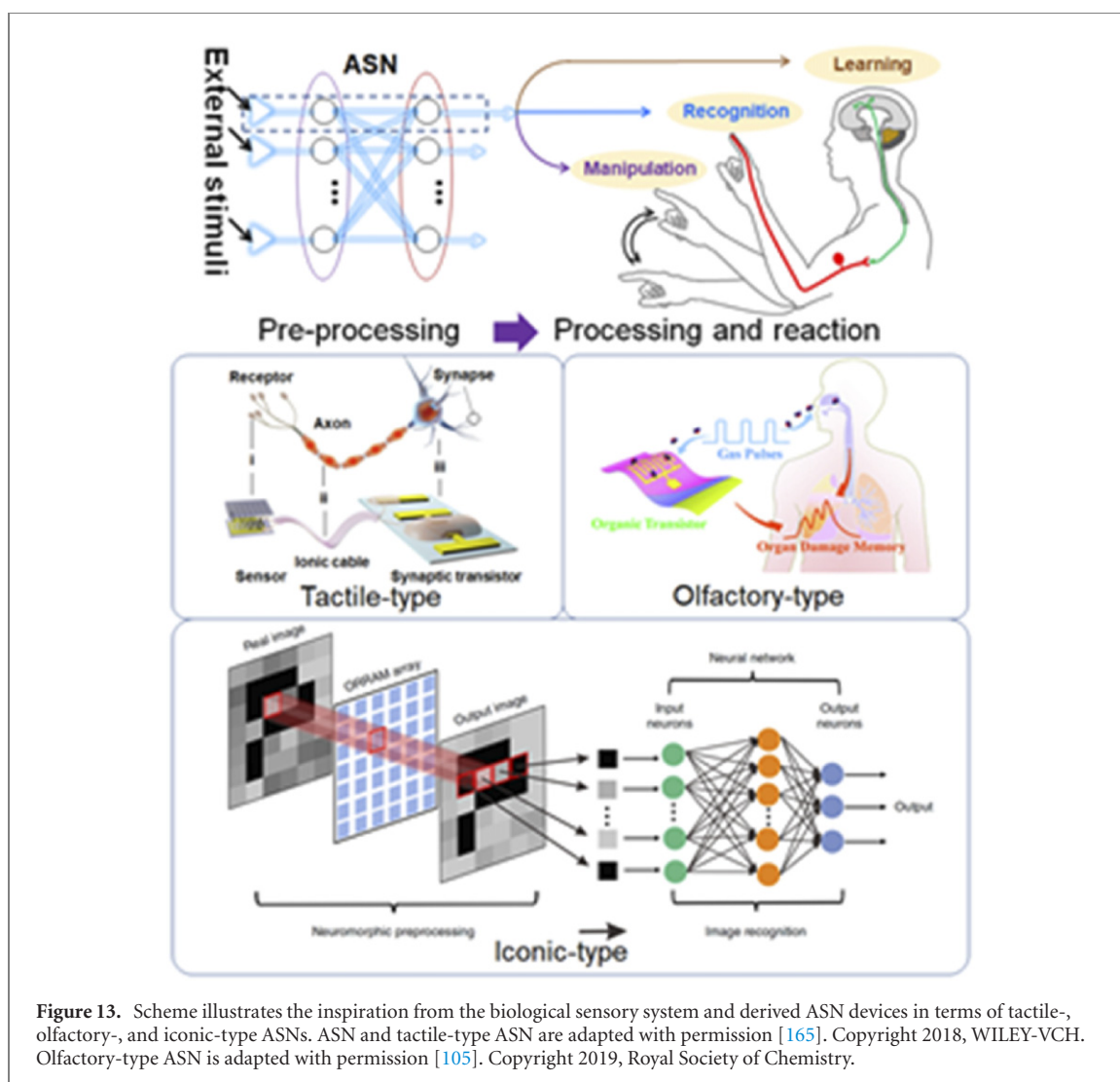
In 2018, a novel type of bionic device called ASN came into sight, which was aimed to imitate the biological perceptual capabilities of electronic devices [165]. After that, an increased number of ASN was proposed, pursuing the emulations of sensory neuron with various modalities, such as touch, vision, and olfactory sense. In 2018, Xu *et al* from Nankai University developed the tactile-type ASN for extracting spatiotemporal features of tactile pattern to enhance recognition accuracy and for motion control of actuators to construct the artificial reflex arc [166]. In 2019, Song *et al* from Northeast Normal University developed a damage memory system based on organic transistors, which mimics our olfactory system to protect the subject from hazardous gas [105]. In the same year, Zhou *et al* from Hong Kong Polytechnic University developed an artificial visual neuron based on optoelectronic memristor for neuromorphic visual pre-processing functions [167]. Among them, neuromorphic transistor is the most used as the information processing component in an ASN, due to its unique advantages, such as spatiotemporal integration property and scalability with other functional components. Transistor-based ASN is able to integrate inputs from spatial isolated terminals, like the dendrite in a neuron without additional wiring [118], and it can be used for implementing multisensory fusion by integrating with multiple sensors [114].

The basic form of the ASN is the integration of neuromorphic and sensing components as well as some additional parts, like signal generation and/or conduction media. The sensing components acted as the receptor of a sensory neuron, which is responsible for transducing the external stimuli into spike train or analog signal via the aforementioned additional parts. By selecting or designing proper sensors, the devices could be built for mimicking different types of sensory neurons. The neuromorphic component, which acted as the synapse, is responsible for the preprocessing of sensing information and feeds the followed perceptual learning processes. The development of this area has shifted from pursuing the biological fidelity of a sensory neuron to exploit the advantages of biology-like perception for empowering the bioinspired machineries. More importantly, with the gradual deepening of understanding on in-sensor computing paradigm that has been long evolved in biological sensory system, achieving of bionic perception from the bottom up has gained increasing attention. It is very promising for propelling autonomous AI and solving the von Neumann bottleneck problem in conventional computing systems, which are hash-rate and power-hungry machines (figure 13).

10.2. Current and future challenges

Currently, the main challenge comes from the poor understanding of sensory neurons and the sensory system, which is vitally important for the conceptual design of an ASN. Our sensory system can filter large-scale external stimuli by the sensory neurons and interpret them into simplified presentations through the NN. This process greatly reduces the scale of the sensory data, facilitates the learning and recognition process, and enables ultralow-energy operation. However, the detailed mechanisms are still unclear, which increase the difficulty of the translation into electronic implementation. The secondary challenge might be from the device level. What kind of functional components should be integrated with the neuromorphic transistors and how to integrate them to realize the essential functions of a sensory neuron are the main issues belonging to the device-level challenge. Previously, the ASN incorporates a synaptic transistor with one or more sensors with similar properties of a certain receptor, which have been demonstrated with apparent sensory processing power. However, the biological sensory neurons are far more complex. They communicate using spike trains, form a vast number of interconnections in the gray matter, and are more adaptive and energy efficient. What is more, conventional devices do not have the intriguing mechanical properties as biological counterparts, such as stretchability and self-healing. Another challenge might come from the material level. Although a vast number of emerging materials were developed for expanding the functionalities of ASNs, such as 2D materials, perovskite, and MXene, optimizing and innovating are still highly required for empowering the current ASN fundamentally.

Furthermore, important issues for system-level integration have not been considered yet, such as the connection between device to device and the device density. In a biological system, the receptor density is highly varied from region to region, and they are connected to the neural system through the nerve fibers. This is absolutely different from the IC technologies. A revolutionary or at least a comprised solution should be addressed for the replication of the biological sensory system.



10.3. Advances in science and technology to meet challenges

Confronted with these current and future challenges, there are some possible solutions. For example, with the mapping of a simple NN of a nematode by serial electron microscopy reconstructions [168], understanding of the connection, distribution, and mechanism has increased significantly. So, how a neuron and the NN operate and how functions and mechanisms relied on the complex network structures could be revealed for not too long. In terms of devices, some stimulation-sensitive materials (e.g. triboelectric materials [169], hydrogels [170], and conducting polymers [171]) could also serve as the gate dielectric or channel materials, which can simplify the integration of the sensing and processing components of an ASN and may empower the in-sensor-memory process. With the development of some efficient and bioinspired oscillators, like carbon nanotube-based ring oscillators [172] and diffusive memristor-based oscillators [173], the ASN could encode the information as the way of their biological counterparts with lower energy consumption and higher error tolerance. The biological mechanical properties could benefit from the introduction of intrinsic stretchable active materials, such as the conjugated polymer/elastomer phase separation-induced elastic semiconductor [166] and rubber-based semiconductor [174], nanomaterials (including carbon nanotubes and silver nanowire), and novel structure designs (like the serpentine, micro-cracked, and waved [175]).

Transistors with a multigate configuration could be served as a good example for constructing the complex network connections, as they can form interconnections on demand and without precise and defined wiring [118]. This would greatly simplify the reproduction of the highly complex connections in the NN. Furthermore, by integration of several kinds of neuromorphic devices, the network could be more sophisticated as the biological one. An example is the formation and degeneration by integrating with memristors. Meanwhile, we are still waiting for breakthroughs in both scientific and technological aspects to build a highly autonomous artificial perceptual system.

10.4. Concluding remarks

The ultimate goal is to build an autonomous artificial intelligent system with human-like perceptual, cognitive, and active capabilities. However, several challenges should be addressed before starting practical applications. The main obstacle is from the neuroscience aspect, in which the mechanisms for how neurons and NNs operate are still not clear. However, efforts paid to develop electronic implementations for imitating their biological counterparts will not be in vain. On the one hand, bioelectronics boomed with the creation of novel electronic paradigms based on neural mechanisms and by endowing existing electronics with biological plausible intelligence. On the other hand, it can serve as the supplementary path for understanding the basic neural mechanisms in parallel to the neuroscience methodology. Despite its infancy, it has become appealing to the scientific community, including materials science, flexible electronics, soft robotics, neuromorphic engineering, and bioelectronics. Among the various candidates, neuromorphic transistors are thought to be one of the most promising information processing components for ASNs. We believe that the development of ASN based on neuromorphic transistors would eventually give birth to the autonomous artificial intelligent system, which has profound significance in improving our lives and manufacturing processes.

Acknowledgements

The authors are grateful for the financial supports from the National Key R & D Program of China (Grant No. 2021YFA1202600) and the National Natural Science Foundation of China (Grant No. 62174082).

11. Neuromorphic computing going efficient

Guosheng Wang¹, Xiao Yu^{1,*} and Bing Chen^{2,*}

¹ Research Center for Intelligent Chips, Zhejiang Lab, Hangzhou 311121, People's Republic of China

² Zhejiang University, People's Republic of China

E-mail: wangg@zhejianglab.com, yuxiao@zhejianglab.com and bingchen@zju.edu.cn

11.1. Status

Research on neuromorphic ICs started as early as the 1980s [12]. However, dilemma in parallelism and power consumption of neuromorphic computing has hindered their development. Recently, rapid growing of data-centric computation rearoused attention on the neuromorphic ICs. The implementation of the neuromorphic ICs cannot only dramatically enhance the efficiency of data processing but also promote the applications of AI in various edge computing scenarios. In addition, it will simulate the operation mode of the human brain to a certain extent, which will further enhance the development of neuroscience and in return will lay the foundation for the development of strong AI. However, although electron devices are superior in general-purpose computations nowadays due to their much lower latency compared with neurons, the huge efficiency gap between ICs and the nervous system is a major stumbling block on neuromorphic computing.

There are two main approaches to realize neuromorphic computing; one is based on the CMOS technique and the other on the NVM technique. Figure 14 summarizes the energy efficiency and power consumption of the recently reported neuromorphic computing system based on CMOS and NVM with ANN and SNN; research from or participated by Chinese scientists are highlighted. Although the SNN shows lower energy efficiency compared with the ANN, the event-driven process of the SNN would have great potential archiving ultralow-power applications. Moreover, the NVM-based architectures are more energy efficient.

For CMOS-based ANN structure [176–178], a chip using SRAM has been proposed, and the recognition of handwriting digits at an energy of 630 pJ has been achieved. To further improve the energy efficiency, one way is to mimic the mechanism of the brain with IC in the SNN architecture [59, 61, 185]. For example, the Darwin chip utilizing SRAM supports 2048 neurons and over four million synapses, consuming 0.84 mW [185].

However, the CMOS devices limit the integration and energy of chips. Therefore, it is necessary to explore a new architecture based on novel devices. Up to now, there have been many NVM-based ANN and SNN structures [45, 179–184]. Compared with the CMOS-based architecture, the new structures are generally more energy efficient.

Due to the development of neuromorphic computing-oriented novel design methodology and NVM devices, great progress has been achieved in neuromorphic ICs in terms of complex tasks with limited power consumption. In the future, continuously improving the efficiency of the devices and ICs will change the environment of neuromorphic computing.

11.2. Current and future challenges

There have been many progresses in neuromorphic computing, but lots of challenges to reduce energy cost need to be addressed. Difficulties arise from different levels. For chips with a large scale of cores, the network-on-chip (NoC) connection strategy has been widely used to communicate between the process elements (PEs). In this part, power cost occurred mostly in routers and communication link. For instance, the enormous inter-core communication will lead to congestion of NoC and occupy amount of storages, which would dramatically reduce performance.

For every core, the analog circuits are used to realize the computing in memory, which requires interfaces to connect the digital and analog circuit components. However, the utilization of analog-to-digital converter (ADC) and the digital-to-analog converter (DAC) results in large power dissipation, especially for the ADC. Meanwhile, in a CMOS-based core, six transistors are needed to realize at least a simple synapse or neuron. To implement more functions with electronic synapse and neuron, more transistors have to be utilized in the circuits. As a result, rapid rise in area and power cost is inevitable. As the number of cores increases, the power inefficiency brought by these difficulties becomes severe.

Recently, increasing works realizing the neuromorphic chip based on the NVM to further simulate neural function and improve energy efficiency have been reported. However, there are also many special difficulties to optimize the low-power brain-inspired chips without considering the device reliability and yields, especially for training process. First, the programming voltages of the NVM are sometimes higher than the VDD of core devices due to the device characteristic, which leads to huge energy cost when tuning the conductance. Moreover, the most common NVM devices have a nonlinear characteristic of modulating the conductance, which is nonideal for conductance tuning. It is necessary to propose some schemes to solve the problem. Meanwhile, most of the approaches are implemented using the lookup table (LUT) or calibration circuits, which increase the energy cost in the training process. In addition, to perform the on-chip training, the STDP rule is also

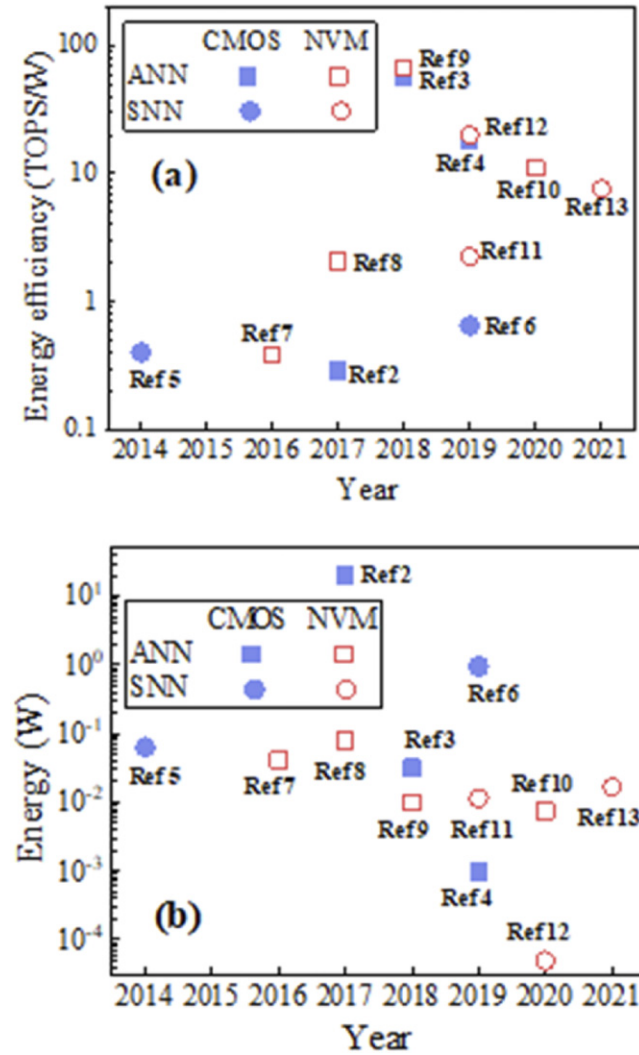


Figure 14. (a) Energy efficiency and (b) power consumption of the recently published neuromorphic computing. Research from or participated by Chinese scientists are highlighted.

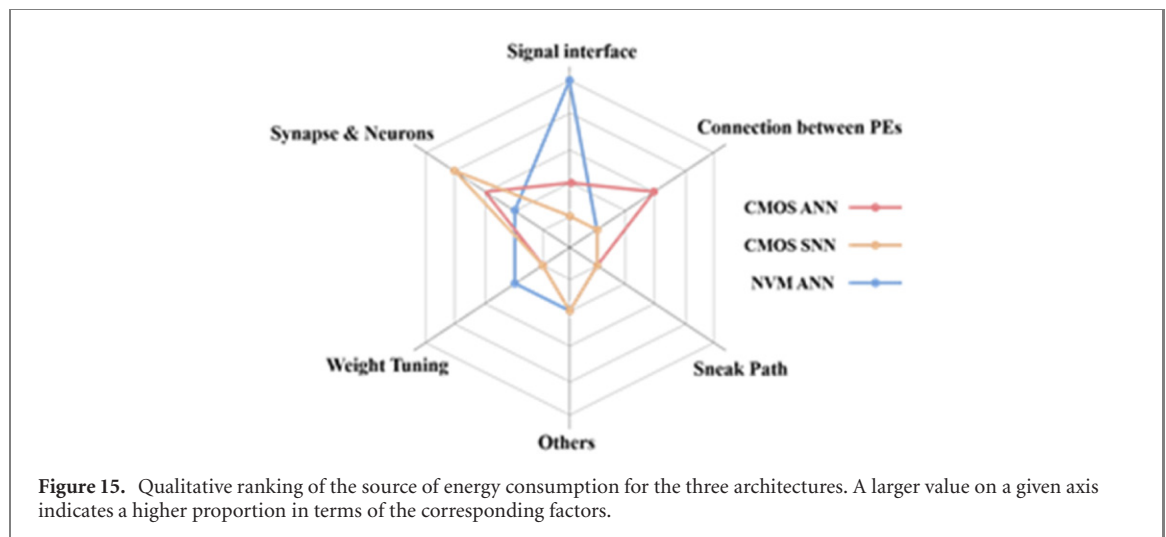
utilized with complex programming signal, which would lead to an enormous waste of power. Furthermore, the sneak-path leakage of the crossbar also results in a certain power consumption. Some typical factors that influence the energy consumption are summarized in figure 15.

11.3. Advances in science and technology to meet challenges

To meet the aforementioned challenges, many new technologies need to be proposed by researchers in China. For the NoC, it is critical to reduce the energy consumption of data transmission and cache. Using sparse and low-bit data is a good approach, because routers can transmit less data and become asleep rapidly. In addition, the strategy to balance the distance and congestion between routers is important. An efficient mapping solution called memetic algorithm-based mapping method, which can reduce the average latency by 63% and the average energy consumption by 69% by maximizing bandwidth utilization, has been reported [186].

Besides, achieving a low-power interface is a critical issue for IMC. Since throughput and resolution are the main sources of power consumption in ADC, reducing these two categories without compromising accuracy would be necessary. Therefore, the reuse of ADC and DAC is beneficial to enable an efficient brain-inspired system. An RRAM-based accelerator named area efficient and power efficient [180] has been proposed, which improved the power efficiency by reducing the number of DACs and balancing the trade-off between the algorithm accuracy of deep CNNs and the resolution of ADCs.

In addition, it is important to establish electronic synapses and neurons with low-power circuits or devices. Meanwhile, designing the programming scheme and training algorithm to realize an efficient on-chip learning is an attractive challenge. Fortunately, many preliminary efforts aiming at these goals have been made. A silicon neuron with the behaviors of regular spiking, intrinsically bursting, and fast spiking has been realized



[187]. In addition, neuro-transistors, by integrating dynamic pseudo-memcapacitors as the gates of transistors, have been proposed, by which a capacitive NN has been built [188]. Excitingly, a leaky integrate and fire neuron, using a capacitor-less leaky FeFET with ultralow hardware cost, has been demonstrated [189]. FeFETs with symmetrical and linear-like weight tuning, which uses a fixed voltage, have also been proposed [190]. Besides, a bilayer transparent memristor has been realized with set and reset voltages of 14 mV and 0.3 V, respectively, and it is composed of indium tin oxide solely [191]. These prototype devices not only reduce the operation voltage of traditional memristors, but they also pave the way for simplifying the device structure and the fabrication process.

11.4. Concluding remarks

Ultralow-power neuromorphic computing ICs have become a center of research recently and have achieved fruitful results. Until now, some CMOS-based architectures have already been proposed. To further reduce energy consumption and improve integration, a lot of explorations and efforts have been made from devices to the chip architecture. Yet, there are still some challenges that need to be addressed, including simulating a synapse or neuron with fewer devices, designing efficiency communication mechanisms, reducing the power consumption of tuning conductance, etc. By far, a unified framework or theory in the field of low-power neuromorphic computing has not been proposed yet. Meanwhile, there is not even a standard for the manufacture of devices and the selection of materials. In summary, it has come to a point for us to think more about where to go and how to develop new devices, chip architectures, and ways of integration for low-power neuromorphic computing.

Acknowledgements

The authors acknowledge support from NSFC (No. 92064001), the Major Scientific Research Project of Zhejiang Lab (Grant No. 2021MD0AC01), and the Zhejiang Province Key R & D programs (Grant No. 2022C01232, and 2021C05004)

12. Automated synthesis and mapping

Zhufei Chu*, Lunyao Wang and Yinshui Xia
Ningbo University, People's Republic of China
E-mail: chuzhufei@nbu.edu.cn

12.1. Status

The new computing paradigm is inseparable from dedicated EDA tools [192]. There are two essential parts of the traditional design flow, namely, front-end and back-end syntheses. The front-end synthesis is concerned with how to translate design abstraction into a technology-specific netlist under the constraints of power, performance, and area, while the netlist to physical layout is generally implemented in the back end. In addition, verification runs through the whole process. Automated synthesis is the core part of the front-end synthesis. Synthesis includes translating the high-level description into a gate-level netlist, technology-independent optimization, and technology mapping [193, 194].

Neuromorphic computing is quite different from standard-cell CMOS technologies and LUT field-programmable gate array components. Memristors sandwiched in crossbar arrays present new logical abstractions and challenges to (technology) mapping. Memristive devices are characterized by their nonvolatile binary storage capability and their ability to store continuous conductance values. Binary storage is used to execute NOR or implication logic operations and other digital logical operations. In contrast, storing a continuum of conductance values facilitates the computation of Ohm's law and Kirchhoff's current summation laws [194]. In previous research, an NN was mapped to a crossbar based on physical laws, which enables matrix–vector multiplication to be more efficient. The weights of the network are stored in memristive devices in this scenario [12, 45]. Neuromorphic architecture, however, requires additional effort when incorporated into a general design, such as digital circuits. Recently, a mixed-signal neuromorphic architecture was proposed to unite analog efficiency and digital programmability [195]. Furthermore, the neuromorphic computing architecture can be realized using a bottom-up approach using the self-assembly method for the fabrication of nanowires and nanodevices. As a result of self-assembly, non-idealities are inevitable when it comes to computing. In addition to improving yield from a hardware perspective, fault-tolerant/defect-tolerant EDA tools are also beneficial. In general, a correct function is obtained by either utilizing existing defective devices or by avoiding certain areas of defective devices. Defect-tolerant logic mapping on a digital logic architecture for 'CMOS/nanowire/molecular' hybrid circuits [196] was widely addressed recently [197, 198].

It would be possible to extend the application and generality of neuromorphic computing and engineering by developing automated synthesis and mapping techniques, the design flow of which is shown in figure 16. AI in advance brings a great deal of profit to EDA (e.g. the simulator for the mapped crossbar array), while EDA techniques could be used to enable AI-based hardware. Therefore, deep integration of neuromorphic computing and EDA is a promising goal.

12.2. Current and future challenges

In recent decades, logic synthesis has been developed mainly for CMOS technology, where logic is abstracted by using traditional AND/OR, NAND/NOR, or XOR. For neuromorphic applications, however, memristive devices can implement stateful implication logic [199] or majority-of-three (MAJ) logic [200]. Memristive devices could have multilevel states, which allow the computation of multi-values instead of binary-value logic. The following is a general view of the challenges of automated synthesis and mapping.

Logic representation. Logic optimization relies on the underlying logic representations. In the early ages (i.e. the 1960s), truth table is used for representation of small-scale functions. Later, as the programmable logic array technology natively implements the sum of product (SOP) form, two-level logic representations were extensively studied at that time. The standard logic representation began shifting from SOPs to directed acyclic graphs (DAGs) with the advent of large-scale integration. The notable DAG-based representation examples are binary decision diagrams, AND-inverter graphs (AIGs), and MAJ-inverter graphs [201], the nodes in which act as 2:1 multiplexer, two-input AND, and three-input MAJ, respectively. In neuromorphic computing, the diversity of memory devices leads to different logic abstractions. Although current SOPs or AIGs can be transformed into representations suitable for memristive devices, the naïve one-to-one mapping could bring lots of redundancy.

Logic mapping. Mapping the optimized logic network to a crossbar array is the main task for logic mapping. The crossbar array supports both serial and parallel computing, thanks to its regular fabric. Hence, the logic mapping process can be further divided into three stages: (1) scheduling the logic computing to partition the designs and generate instruction sequence, (2) defect-tolerant mapping of the designs, and (3) the online/offline mapping of an NN. The challenges mainly come from the non-idealities. An efficient simulator for the crossbar is highly in demand. The simulator should capture the sources of non-idealities of both

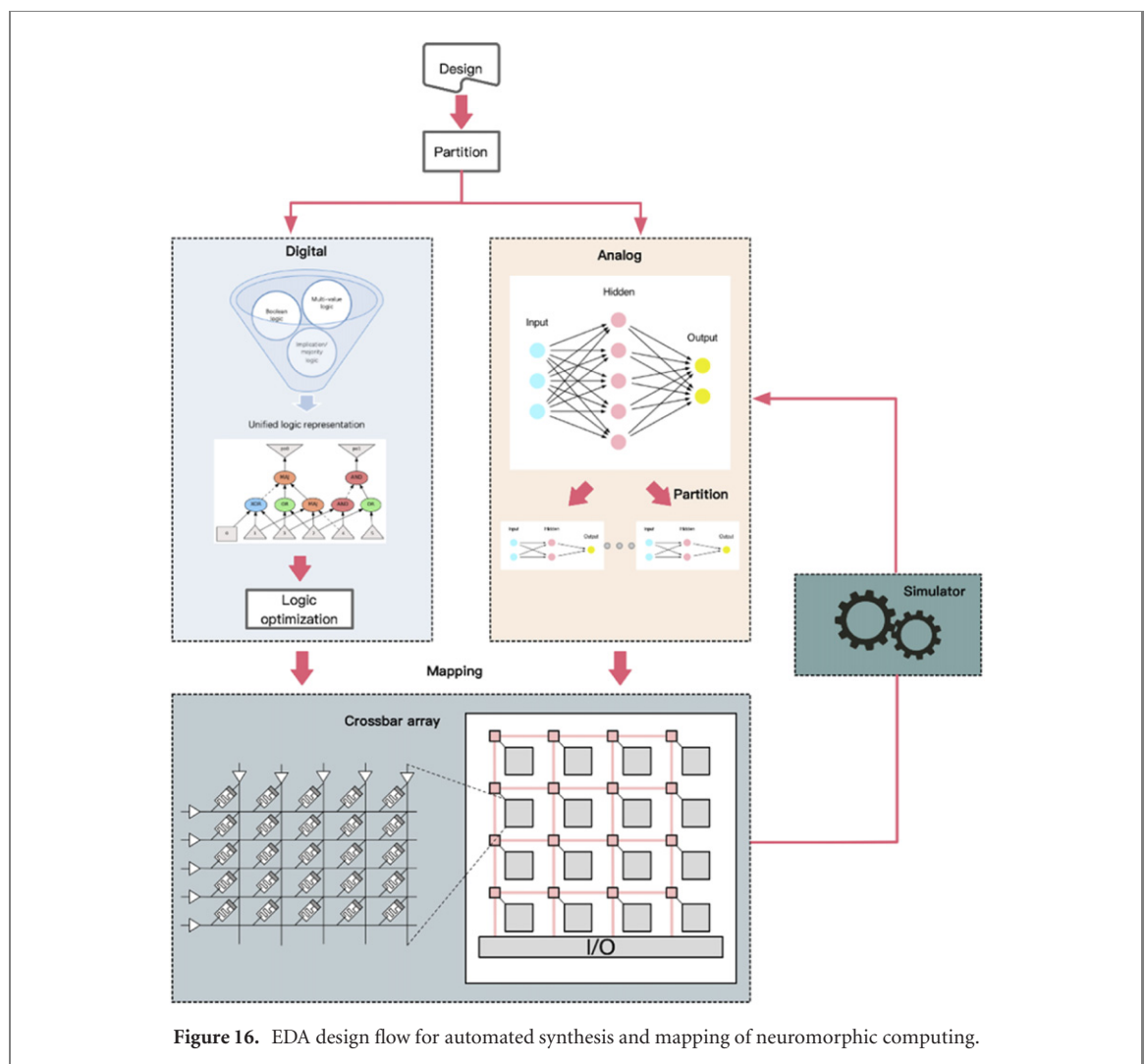


Figure 16. EDA design flow for automated synthesis and mapping of neuromorphic computing.

non-data dependent- (linear) or data-dependent (nonlinear) types. The nonlinearity could be a big challenge. By using data training and inference, NNs can be used to deal with this issue [202].

Multi-value states. The continuous conductance values are usually used to store the weights of the NN. There is less attention paid to multi-value logic (MVL) synthesis, but multi-value representation is considered as a key enabler for next-generation and high-information-density digital electronics [203]. MVL is not a new concept. It is restricted since the CMOS technology works better for binary-value logic. MVL can now be operated with significantly reduced complexity and remarkable new capabilities, thanks to the development of new electronic materials and devices. The exploitation of multi-value states of memristive devices could bring lots of new opportunities.

12.3. Advances in science and technology to meet challenges

Logic representation. To keep the diversity of the memristive devices, the logic representation for neuromorphic computing should be flexible to represent all the candidates, from two-input AND/OR to three-input majority. The majority logic operation over variables x , y , and z is $M(x, y, z) = xy + xz + yz$. It inherently incorporates AND/OR since $M(x, y, 0) = xy$ and $M(x, y, 1) = x + y$. Hence, the majority logic is more expressive and can be used as a unified logic representation. Combined with an inverter, the majority logic is functionally complete for all the functions. XOR is widely used in arithmetic circuits; the majority logic with three-input XOR could be a promising logic representation [204]. For example, a full adder has two outputs that are sum out and carry out, which correspond to a three-input XOR and majority logic, respectively.

Logic mapping. To exploit the parallel computing properties by the crossbar array, it is important to have a design partitioned from both coarse-grain and fine-grain perspectives. Also, the partitioned design should be scheduled to optimize the number of instructions and computing resources. The scheduling process is highly related to the logic representation under the physical constraints of the crossbar array. Because the logic mapping actually performs the placement and routing step compared in the traditional flow, a mapping-aware logic synthesis could also benefit the neuromorphic computing design flow. For the simulator, the AI-based

simulation method could be a promising path to address the nonlinearity fault. Finally, the synthesizer, the simulator, and the mapper make up the essential parts for the EDA tools of neuromorphic computing.

Multi-value states. MVL synthesis was addressed in the 1990s [205]. Logic representations, optimization algorithms, and intermediate file format were discussed. However, a significant amount of effort and research needs to be done. The development of new electronic materials and devices can verify the synthesized results. In turn, the synthesizer enables the development of new devices. Moreover, once the number of multi-value states is large enough, the gap between digital and analog circuits could be narrowed. The multi-value state has more sampling of the signals.

12.4. Concluding remarks

EDA toolchains are indispensable for neuromorphic computing and engineering. The three important parts for design automation of neuromorphic computing are synthesizer, mapper, and simulator. The three parts are not independent existence but have strong coupling with each other. Various device candidates, device nonideal behaviors, and multi-value states pose great challenges to promoting neuromorphic computing technologies for practical application. Consequently, there needs continuous work on unified logic representation, defect/fault tolerance logic mapping, and MVL synthesis. With the advances in EDA and large-scale integration, neuromorphic computing is promising to be a critical computing paradigm.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61871242 and in part by the State Key Laboratory of ASIC & System under Grant 2021KF008.

13. Survey on machine learning SoC: from near-memory computing to IMC

Chen Mu, Feng Lin and Chixiao Chen

Fudan University, People's Republic of China

E-mail: 20112020022@fudan.edu.cn, linfeng1990@live.com and cxchen@fudan.edu.cn

13.1. Status

The recent surge in AI drives the exploration of machine learning hardware, especially chip designs. It turns out that conventional CPU-like von-Neumann architectures are not suitable for machine learning tasks in terms of energy efficiency. Optimized for complex controlling problems, current CPUs focus more on out-of-order dispatch, branch prediction, and cache coherence. Only 9% of the entire processor power dissipation is consumed by data path-related operations. In addition, the on-chip memory of CPUs is often limited. As a result, the CPU requires tremendous external memory access and latency to meet computing- and memory-intensive workloads, such as matrix–matrix multiplication in NNs. Due to these reasons, the commercial CPU performances of HPC have been saturated for decades [206]. In contrast, the computing operation number of SOTA AI algorithms doubles every 3.4 months [207], resulting in a big gap between hardware and software. The GPU was the first viable hardware solution for SOTA machine learning algorithms. It employs single-instruction multiple-thread architectures to perform large-scale parallel computing. However, the power consumption of GPUs normally exceeds hundreds of watt, which is too power hungry to deploy in edge and embedded scenarios.

The aforementioned facts drive domain-specific computer architectures and corresponding silicon prototypes for machine learning. In this trend, SoCs accelerating NN computing are widely reported for different specific applications, such as image recognition, keyword spotting, and natural language processing. Compared with conventional CPU and GPU, these machine learning SoCs rearrange the hardware organization and even instruction set. Tensor processing unit from Google [208], for example, employs systolic arrays to efficiently implement matrix–matrix multiplication efficiently. Only five instructions remain for the HPC coprocessor. Memory hierarchies are also optimized for data movement. Meanwhile, researchers from academia are seeking more efficient implementation, such as Cambricon [209] and Eyeriss [210]. In this review, we categorize these solutions as near-memory computing (NMC), that data processing is no longer externally but close to where data reside—memory.

More recently, a new type of computer architecture called IMC prevails. As the name implies, IMC refers to that data processing takes place inside the memory; therefore, (part of) data movement is eliminated. IMC reuses storage devices, for example, memristor of NVM [45], to perform analog computing. It features not only good energy efficiency but less data movement. In this review, both NMC- and IMC-based machine learning SoCs, illustrated in figure 17, will be investigated.

13.2. Current and future challenges

There are two main challenges inside machine learning SoC designs, normally known as the *power wall* and the *memory wall*.

Power wall. Due to the end of Dennard scaling, the power dissipation density concerning area keeps increasing as the fabrication technology downscales from sub-100 to sub-10 nm. This trend results in a severe problem of chip cooling. To ensure that transistors work properly under a reasonable temperature range, there exists a maximum thermal limit, known as the power wall. SOTA HPC chips have been already approaching this limit when using 7–14 nm FinFET technology [211]. As a result, CPU performance (maximum clock rate) does not increase when better technology is adopted. It is critical to investigate more energy-efficient circuits and architectures to implement HPC. The potential of conventional digital circuits is running low, while analog and mixed signal-based computing circuits are promising [212]. However, analog circuits are sensitive to process variation, noise, and varying voltage/temperature, especially for accurate computing. A trade-off between precision and power consumption is of great concern.

Memory wall. Another issue of machine learning SoC comes from the data transportation between computing logic blocks and memory blocks. For each multiplication and accumulation operation, it requires four times memory access in the conventional computer architectures. In other words, an $N \times N$ matrix multiplication would trigger $4N^3$ times memory access, which consumes much longer latency and power consumption than parallel computing if the memory is off-chip. Note that the actual input and output data sizes are $2N^2$ and N^2 , respectively. Therefore, a good design should target an optimized solution reducing the memory access gap from $O(N^3)$ to $O(N^2)$. Historically, this limit is known as the memory wall or the von Neumann bottleneck. This challenge is not scaling down-friendly either. Although the power and area consumption of the logic decreases as the technology scales, the pad size and the external data bandwidth are almost maintained, if not using a high-speed link interface, across different technology nodes. A straightforward solution to the

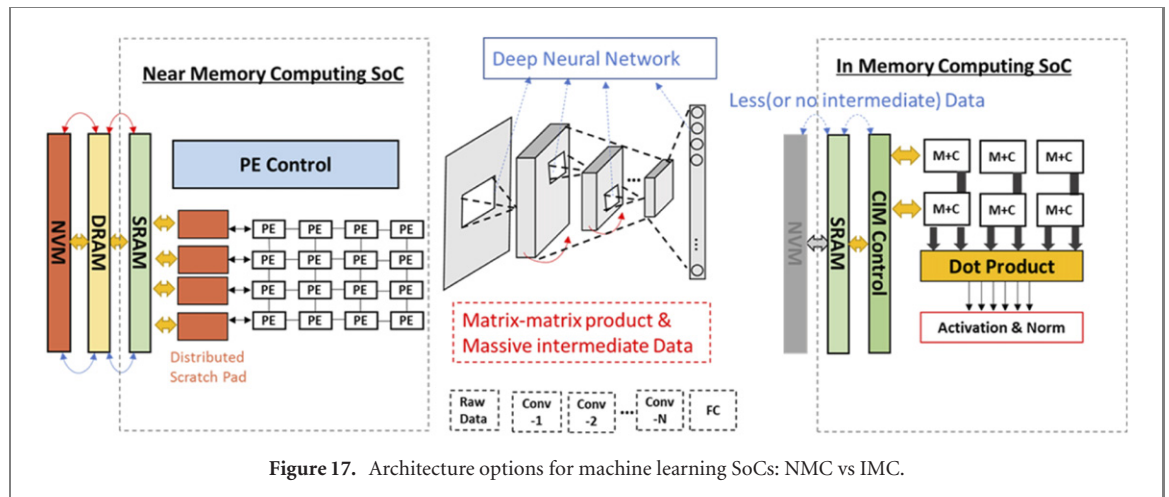


Figure 17. Architecture options for machine learning SoCs: NMC vs IMC.

memory wall is embedding an extremely large (>100 MB) on-chip memory inside SoCs [213]. But it is not a cost-effective way. Detailed understanding of algorithms shows that data reuse inside NNs can be a key knob to solve the memory wall. Systolic arrays, for example, are a common architecture to avoid intermediate results of matrix computing transferring outside the computing logic circuits.

13.3. Advances in science and technology to meet challenges

Many innovations were proposed to address these challenges. We categorize some of them into three major advanced trends, notifying that there is some significant work that is not covered due to the limited length.

13.3.1. Reconfigurable data flow optimization

Reconfigurable NMC architectures feature software-defined hardware fabric connecting bunches of PEs and local memory blocks. In NN processors, different NN data flows can be applied to the architecture, reducing the data movement. For example, Thinker in [214] is a reconfigurable hybrid NN processor, which has two reconfigurable heterogeneous PE arrays supporting on-demand partitioning and reconfiguration for parallel processing different NNs. Each PE supports bit-width adaptive computing to meet variant bit widths of different neural layers. Another reconfigurable deep learning accelerator, iFPNA [215], achieves both energy efficiency and flexibility. This processor has both the programmable data flow engine with custom instruction sets and reconfigurable PE arrays. The method also supports application-specific design, such as a biomedical AI processor (BioAIP) in [216]. The reconfigurable BioAIP with adaptive learning is compatible with an NN and biomedical signal processing engine.

13.3.2. Algorithm–circuit–architecture codesign

Original deep learning algorithms have redundant margins, causing extra power consumption and memory access. Sophisticated NN compression and modification techniques can remove the cost of these redundancies, improving the overall performance. For example, a block-circulant algorithm can unify CNN/fully connected/recurrent neural network (RNN) workloads with transpose-domain acceleration. Based on this algorithm, a unified NN processor has been proposed [217], providing $8 \times -128 \times$ storage reduction. Also, to solve the problem of huge computations and storage of CNN and DNN, a binarized depth-wise separate convolution neural network has been designed [218]. This lightweight NN architecture reduces memory footprint and computations compared with traditional CNN for the task of keyword spotting. In addition, an RNN accelerator, on-chip incremental-learning enhanced artificial neural network (OCEAN) in [219], implements both inferencing and training on the same hardware by utilizing a mixture of analytical and numerical gradient decent. OCEAN achieved on-chip RNN training with a low cost of hardware overhead.

13.3.3. Analog/mixed-signal in-memory computing-based SoC

Exploiting analog computing within memory bit cells, IMC circuits feature both good energy efficiency and less data movement. IMC-based machine learning SoC targets extending macro advances to system performance. For example, an IMC-friendly dynamic sparsity scaling (DSS) architecture was designed in [220]. This accelerator realized activation and weight sparsity-aware acceleration and lower power by DSS. Moreover, to reduce more off-chip data accesses in IMC, an attention-based context-breaking method is presented in [221]. It reduces data movement up to 30.3% by removing weak context connections in RNNs. In addition, time-domain analog computation with less toggle activity provides an alternative in [222]. Figure 18 shows that with

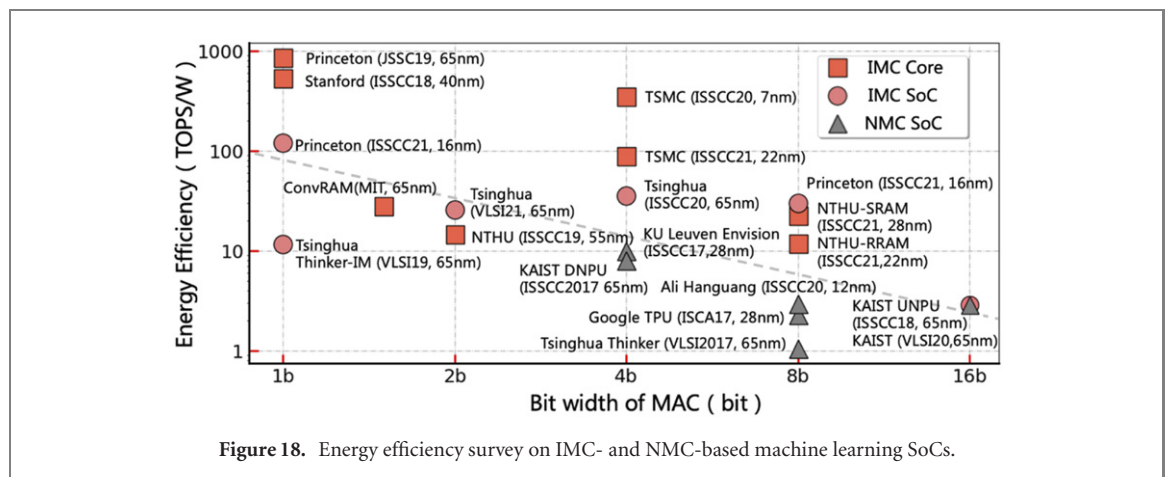


Figure 18. Energy efficiency survey on IMC- and NMC-based machine learning SoCs.

the assistance of analog/mixed-signal computing, SoTA IMC SoCs achieve nearly ten times energy efficiency improvement than NMC SoCs.

13.4. Concluding remarks

As one of the most representative domain-specific architectures and corresponding chips, machine learning SoC aims to break the power wall and memory wall of HPC hardware. A bunch of verified silicons demonstrates that reconfigurable data flow optimization, algorithm–circuit–architecture codesign, and analog/mixed-signal IMC helps to improve overall SoC performances.

Acknowledgements

This work was supported by the National Key Research and Development Program of China under Grant 2019YFB2205000 and Shanghai Rising-Star Program under Grant 20QA1407300.

14. Market-oriented neuromorphic SoC solution

Bojun Cheng¹, Yannan Xing², Weitao Zeng², Hong Chen³, Lei Yu⁴, Giacomo Indiveri^{1,5} and Ning Qiao^{1,2,5,*}

¹ SynSense AG, Switzerland

² Chengdu SynSense Tech. Co. Ltd., People's Republic of China

³ Tsinghua University, People's Republic of China

⁴ Wuhan University, People's Republic of China

⁵ University of Zurich and ETH Zurich, Switzerland

E-mail: ning.qiao@synsense.ai

14.1. Status

The human brain consumes only approximately 20 W of power to perform complex tasks that outperform SOTA supercomputers by several orders of magnitude in energy efficiency and volume. Neuromorphic computing emulates the principles of computation of the human brain, which utilizes asynchronous events as information carriers. Compared to conventional computers with centralized processing architecture, neuromorphic computing is distributed, massively parallel, and adaptive. With this approach, neuromorphic computing can overcome the von Neumann bottleneck.

Since the early 2000s, endeavors of emulating the human brain have promoted the design of large-scale neuromorphic chips. More recently, spiking neuron networks (SNNs) that exchange information via spikes have gained increasing attention in the field. SNNs are naturally compatible with neuromorphic hardware due to their spatiotemporal dynamics, diverse coding schemes, and event-driven characteristics [223]. In addition, SNNs feature extremely low power consumption, just like the brain does [224].

Thanks to the SNN-based neuromorphic chip's low power consumption and low chip area requirements, performing real-time pattern recognition tasks in an edge computing scenario has become possible. For instance, object recognition and classification with SNN chips are advantageous in energy consumption and area compared with CPUs.

Previous research on neuromorphic chips mainly investigated spike-driven computation. Specifically, learning with 'spikes' has been demonstrated by different academic and industrial research groups. Table 5 shows a summary of representative neuromorphic platforms, with focus on many China-based groups (rows 4–7). In 2014, IBM introduced the TrueNorth chip with one million neurons [225]. The proof-of-concept chip possesses the majority mathematical functionality but occupies an area of 430 mm², which is too large and expensive for IoT applications [226]. Most of the successive neuromorphic chips comprised less than 200 000 neurons, as the neuron density is often restricted to ~2000 to 3000 neurons per mm². The current mainstream to design effective neuromorphic hardware exploits digital and asynchronous hardware. In this domain, asynchronous circuits can consume up to ten times less power but can require a larger overhead area than synchronous circuits. To break the bottleneck of neuron density, Intel Loihi 2 was manufactured by the latest Intel 4 technology node, and the neuron density reaches 32 000 mm², which is 15 times larger than its first generation [227].

14.2. Current and future challenges

Nevertheless, all of these research chips are facing the same problem toward IoT commercial applications—the chip area is too large and the cost is too high. Currently, estimation of chip price lies in the range of 10 to 50 US dollars per million neurons. This is because most of the aforementioned chips are designed based on general-purpose architectures. Such architecture is highly reconfigurable and can be adapted to various computational models, which has tremendous advantages for research purposes but does not fit well with application-specific market-oriented products. As a matter of fact, edge applications are very sensitive to the cost and power consumption of solutions. The demand for low-power applications requires a dedicated solution with optimized architecture to save power consumption and manufacture cost with a constraint to the number of neurons and chip size. Moreover, many real-time tasks also require dedicated input/output (I/O) interfaces to form a complete SoC to process various signals and be compatible with common analog and/or digital sensors. A neuromorphic application-specific design approach that can directly support these sensors with adequate I/O interfaces is urgently needed for commercial applications.

14.3. Advances in science and technology to meet challenges

Multiple steps are required to make such a market-oriented chip design approach successful. First, a clear definition of the system is necessary to define the fundamental neuromorphic chip specifications, such as power consumption, cost, computing capacity, extendibility, I/O interface, in/out protocol, etc. Then, a market-oriented definition based on real applications further defines the algorithm development/optimization and

Table 5. Summary of existing neuromorphic computing platforms.

	Power (mW)	Craft (nm)	Area (mm ²)	Number of neurons	Neuron density(mm ²)	Method	Cost/million neurons per chip (US dollar ^a)
IBM TrueNorth [225]	63–300	28	430	1×10^6	2.3×10^3	Syn + Asyn full digital	30.3
Intel Loihi [60]	74	14	60	1.3×10^5	2.2×10^3	Asyn full digital	44
Intel Loihi 2 [227]	/	Intel 4	31	1×10^6	3.2×10^4	Asyn full digital	7.65 ^b
Tsinghua Univ. Tianjic [61]	400–950	28	14.44	4×10^4	2.8×10^3	Syn full digital	23.4
Zhejiang Univ. Darwin-1 [185]	58.8	180	25	2×10^3 – 3.2×10^4	82–1311	Syn full digital	/
Zhejiang Univ. Darwin-2	~1000	55	/	1.47×10^5	/	/	/
SynSense	<1	22	12	1×10^6	8.3×10^4	Asyn full digital	0.86
DYNAP-CNN [228]							

^aBased on 12 inch wafer without considering mask price: 28 nm from Semiconductor Manufacturing International Corporation, 22 nm from GlobalFoundries, and 14 nm from TSMC.

^bIntel 4 refers to TSMC 5 nm.

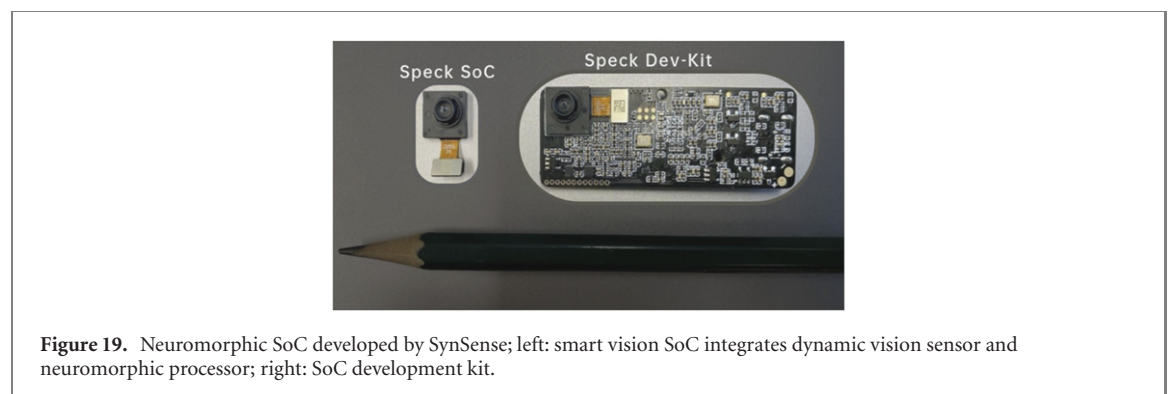


Figure 19. Neuromorphic SoC developed by SynSense; left: smart vision SoC integrates dynamic vision sensor and neuromorphic processor; right: SoC development kit.

modeling. Finally, an algorithm and circuit codesign effort are crucial for efficiently contributing to the production of the most efficient neuromorphic architecture implementation.

Such an approach has been recently implemented by SynSense with the release of the DYNAP-CNN [228] neuromorphic chip. This approach has led to 30–50 times higher neuron density, 30–50 times lower cost, 10–100 times reduction of power consumption, and ten times higher computation efficiency compared to general-purpose neuromorphic research chips, like TrueNorth [225], Loihi [60], and Tianjic [61].

Similarly, the market-oriented SoC solution ‘SPECK’, shown in figure 19, has recently been demonstrated. Driven by market demands, a complete end–end event-based dynamic vision neuromorphic intelligent SoC, with an embedded SNN architecture comprising a total of 32 700 spiking neurons, has been realized. The SoC is fully configurable and has been optimized to run SNNs that can be used for human behavior detection, gesture recognition, motion detection, and various types of intelligent vision-based scenarios.

14.4. Concluding remarks

Neuromorphic computing technologies are ideal for intelligent edge computing tasks that are power and/or size constrained. Market-oriented neuromorphic SoCs need to be driven by both real-world applications and algorithm codesign. For a complete neuromorphic solution, considerations need to be made on cost, power consumption, and inter-chip communication at the system level. This covers a significant number of specialized designs in circuit structure, power management, compatible algorithm development, and optimization.

Acknowledgments

The authors acknowledge that this work was partially supported by the Science and Technology Department of Sichuan Province under the project ‘Neuromorphic Technology based Ultra-low Power Smart Vision SoC for Edge Computing’ with App No. 2021YFG0134.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Qing Wan  <https://orcid.org/0000-0003-4588-7592>
 Changjin Wan  <https://orcid.org/0000-0002-3210-6673>
 Yuchao Yang  <https://orcid.org/0000-0003-4674-4059>
 Peng Zhou  <https://orcid.org/0000-0002-7301-1013>
 Lin Chen  <https://orcid.org/0000-0002-7145-7564>
 Yi Li  <https://orcid.org/0000-0003-1930-8550>
 Kan-Hao Xue  <https://orcid.org/0000-0002-2894-7912>
 Xiang-Shui Miao  <https://orcid.org/0000-0002-0546-7692>
 Junyao Zhang  <https://orcid.org/0000-0002-6300-5875>
 Jia Huang  <https://orcid.org/0000-0002-5823-4817>
 Li Qiang Zhu  <https://orcid.org/0000-0002-8249-3421>
 Chen Ge  <https://orcid.org/0000-0002-8093-940X>
 Ye Zhou  <https://orcid.org/0000-0002-0273-007X>
 Giacomo Indiveri  <https://orcid.org/0000-0002-7109-1689>

References

- [1] Hickmott T W 1962 Low-frequency negative resistance in thin anodic oxide films *J. Appl. Phys.* **33** 2669
- [2] Chua L 1971 Memristor-the missing circuit element *IEEE Trans. Circuit Theory* **18** 507
- [3] Strukov D B, Snider G S, Stewart D R and Williams R S 2008 The missing memristor found *Nature* **453** 80
- [4] Xia Q and Yang J J 2019 Memristive crossbar arrays for brain-inspired computing *Nat. Mater.* **18** 309–23
- [5] Roy K, Jaiswal A and Panda P 2019 Towards spike-based machine intelligence with neuromorphic computing *Nature* **575** 607
- [6] Jo S H, Chang T, Ebong I, Bhadviya B B, Mazumder P and Lu W 2010 Nanoscale memristor device as synapse in neuromorphic systems *Nano Lett.* **10** 1297
- [7] Prezioso M, Merrih-Bayat F, Hoskins B D, Adam G C, Likharev K K and Strukov D B 2015 Training and operation of an integrated neuromorphic network based on metal-oxide memristors *Nature* **521** 61
- [8] Sheridan P M, Cai F, Du C, Ma W, Zhang Z and Lu W D 2017 Sparse coding with memristor networks *Nat. Nanotechnol.* **12** 784
- [9] Li C *et al* 2018 Efficient and self-adaptive *in situ* learning in multilayer memristor neural networks *Nat. Commun.* **9** 2385
- [10] Wang Z *et al* 2019 Reinforcement learning with analogue memristor arrays *Nat. Electron.* **2** 115
- [11] Cai F, Correll J M, Lee S H, Lim Y, Bothra V, Zhang Z, Flynn M P and Lu W D 2019 A fully integrated reprogrammable memristor-CMOS system for efficient multiply-accumulate operations *Nat. Electron.* **2** 290–9
- [12] Zhang W, Gao B, Tang J, Yao P, Yu S, Chang M-F, Yoo H-J, Qian H and Wu H 2020 Neuro-inspired computing chips *Nat. Electron.* **3** 371–82
- [13] Wu H *et al* 2017 Device and circuit optimization of RRAM for neuromorphic computing 2017 *IEEE Int. Electron Devices Meeting (IEDM)* (2–6 December 2017)
- [14] Yu S 2018 Neuro-inspired computing with emerging nonvolatile memories *Proc. IEEE* **106** 260
- [15] Chen P, Lin B, Wang I, Hou T, Ye J, Vruthula S, Seo J, Cao Y and Yu S 2015 Mitigating effects of non-ideal synaptic device characteristics for on-chip learning 2015 *IEEE/ACM Int. Conf. Computer-Aided Design (ICCAD)* (2–6 November 2015)
- [16] Jeong Y, Zidan M A and Lu W D 2018 Parasitic effect analysis in memristor-array-based neuromorphic systems *IEEE Trans. Nanotechnol.* **17** 184
- [17] Li C *et al* 2019 Long short-term memory networks in memristor crossbar arrays *Nat. Mach. Intell.* **1** 49
- [18] Choi S, Tan S H, Li Z, Kim Y, Choi C, Chen P-Y, Yeon H, Yu S and Kim J 2018 SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations *Nat. Mater.* **17** 335
- [19] Wang Z, Kang J, Fang Y, Yu Z, Yang X, Cai Y, Wang Y and Huang R 2016 Localized metal doping effect on switching behaviors of TaO_x-based RRAM device 2016 *16th Non-Volatile Memory Technology Symp. (NVMTS)* (17–19 October 2016)
- [20] Yang J J *et al* 2010 Diffusion of adhesion layer metals controls nanoscale memristive switching *Adv. Mater.* **22** 4034
- [21] Ambrogio S *et al* 2018 Equivalent-accuracy accelerated neural-network training using analogue memory *Nature* **558** 60
- [22] Woo J, Peng X and Yu S 2018 Design considerations of selector device in cross-point RRAM array for neuromorphic computing 2018 *IEEE Int. Symp. Circuits and Systems (ISCAS)* (27–30 May 2018)
- [23] Yang Y, Sheridan P and Lu W 2012 Complementary resistive switching in tantalum oxide-based resistive memory devices *Appl. Phys. Lett.* **100** 203112
- [24] Jung S *et al* 2022 A crossbar array of magnetoresistive memory devices for in-memory computing *Nature* **601** 211–6
- [25] Lin P *et al* 2020 Three-dimensional memristor circuits as complex neural networks *Nat. Electron.* **3** 225–32
- [26] Wang T-Y *et al* 2020 Three-Dimensional nanoscale flexible memristor networks with ultralow power for information transmission and processing application *Nano Lett.* **20** 4111–20
- [27] Sangwan V K and Hersam M C 2020 Neuromorphic nanoelectronic materials *Nat. Nanotechnol.* **15** 517–28
- [28] Meng J-L *et al* 2021 Flexible boron nitride-based memristor for *in situ* digital and analogue neuromorphic computing applications *Mater. Horiz.* **8** 538–46
- [29] Yang F-S *et al* 2020 Oxidation-boosted charge trapping in ultra-sensitive van der Waals materials for artificial synaptic features *Nat. Commun.* **11** 2972

- [30] Wu L *et al* 2021 Atomically sharp interface enabled ultrahigh-speed non-volatile memory devices *Nat. Nanotechnol.* **16** 882–7
- [31] Wang T Y, Meng J L, He Z Y, Chen L, Zhu H, Sun Q Q, Ding S J, Zhou P and Zhang D W 2020 Ultralow power wearable heterosynapse with photoelectric synergistic modulation *Adv. Sci.* **7** 1903480
- [32] Wang C-Y *et al* 2020 Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor *Sci. Adv.* **6** eaba6173
- [33] Meng J, Wang T, Zhu H, Ji L, Bao W, Zhou P, Chen L, Sun Q-Q and Zhang D W 2022 Integrated in-sensor computing optoelectronic device for environment-adaptable artificial retina perception application *Nano Lett.* **22** 81–9
- [34] Yuan J and Lou J 2015 Memristor goes two-dimensional *Nat. Nanotechnol.* **10** 389–90
- [35] Wang M *et al* 2018 Robust memristors based on layered two-dimensional materials *Nat. Electron.* **1** 130–6
- [36] Zhu X, Li D, Liang X and Lu W D 2019 Ionic modulation and ionic coupling effects in MoS₂ devices for neuromorphic computing *Nat. Mater.* **18** 141–8
- [37] Huh W, Lee D and Lee C H 2020 Memristors based on 2D materials as an artificial synapse for neuromorphic electronics *Adv. Mater.* **32** 2002092
- [38] Sun L *et al* 2019 Self-selective van der Waals heterostructures for large scale memory array *Nat. Commun.* **10** 3161
- [39] Sivan M *et al* 2019 All WSe₂ 1T1R resistive RAM cell for future monolithic 3D embedded memory integration *Nat. Commun.* **10** 5201
- [40] 2018 The memristor revisited *Nat. Electron.* **1** 261
- [41] Chua L 2019 Resistance switching memories are memristors *Handbook of Memristor Networks* ed L Chua, G Sirakoulis and A Adamatzky (Berlin: Springer)
- [42] Chen J, Li J, Li Y and Miao X 2021 Multiply accumulate operations in memristor crossbar array for analog computing *J. Semicond.* **42** 013104
- [43] Sun W, Gao B, Chi M, Xia Q, Yang J J, Qian H and Wu H 2019 Understanding memristive switching via *in situ* characterization and device modeling *Nat. Commun.* **10** 3453
- [44] Zhang Y *et al* 2021 Evolution of the conductive filament system in HfO₂-based memristors observed by direct atomic-scale imaging *Nat. Commun.* **12** 7232
- [45] Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Joshua Yang J and Qian H 2020 Fully hardware-implemented memristor convolutional neural network *Nature* **577** 641–6
- [46] Xue C-X *et al* 2021 16.1 A 22 nm 4 Mb 8b-precision ReRAM computing-in-memory macro with 11.91 to 195.7 TOPS/W for tiny AI edge devices 2021 *IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 245–7
- [47] Yang J *et al* 2021 A 14 nm-FinFET 1 Mb embedded 1T1R RRAM with a 0.022 μm^2 cell size using self-adaptive delayed termination and multi-cell reference 2021 *IEEE Int. Solid-State Circuits Conf. (ISSCC)* vol 64 (IEEE)
- [48] Graves C E, Li C, Sheng X, Miller D, Ignowski J, Kiyama L and Strachan J P 2020 In-memory computing with memristor content addressable memories for pattern matching *Adv. Mater.* **32** 2003437
- [49] Lee M-J *et al* 2013 A plasma-treated chalcogenide switch device for stackable scalable 3D nanoscale memory *Nat. Commun.* **4** 2629
- [50] Fu Y, Zhou Y, Huang X, Gao B, Fu Y, He Y, Li Y, Chai Y and Miao X 2021 Forming-free and annealing-free V/VO_x/HfWO_x/Pt device exhibiting reconfigurable threshold and resistive switching with high speed (<30 ns) and high endurance (>10¹²/>10¹⁰) 2021 *IEEE Int. Electron Devices Meeting (IEDM)*
- [51] Jeon K, Kim J, Ryu J J, Yoo S-J, Song C, Yang M K, Jeong D S and Kim G H 2021 Self-rectifying resistive memory in passive crossbar arrays *Nat. Commun.* **12** 2968
- [52] Ren S-G, Ni R, Huang X-D, Li Y, Xue K-H and Miao X-S 2022 Pt/Al₂O₃/TaO_x/Ta self-rectifying memristor with record-low operation current (<2 pA), low power (fJ), and high scalability *IEEE Trans. Electron Devices* **69** 838–42
- [53] Zidan M A, Jeong Y, Lee J, Chen B, Huang S, Kushner M J and Lu W D 2018 A general memristor-based partial differential equation solver *Nat Electron* **1** 411–20
- [54] Li Y *et al* 2021 Monolithic 3D integration of logic, memory and computing-in-memory for one-shot learning 2021 *IEEE Int. Electron Devices Meeting (IEDM)*
- [55] Giannopoulos I, Singh A, Le Gallo M, Jonnalagadda V P, Hamdioui S and Sebastian A 2020 In-memory database query *Adv. Intell. Syst.* **2** 2000141
- [56] Pedretti G, Mannocci P, Li C, Sun Z, Strachan J P and Ielmini D 2021 Redundancy and analog slicing for precise in-memory machine learning-part: I. Programming techniques *IEEE Trans. Electron Devices* **68** 4373–8
- [57] Mead C 1990 Neuromorphic electronic systems *Proc. IEEE* **78** 1629–36
- [58] Benjamin B V *et al* 2014 Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations *Proc. IEEE* **102** 699–716
- [59] Merolla P A *et al* 2014 A million spiking-neuron integrated circuit with a scalable communication network and interface *Science* **345** 668–73
- [60] Davies M *et al* 2018 Loihi: a neuromorphic manycore processor with on-chip learning *IEEE Micro* **38** 82–99
- [61] Pei J *et al* 2019 Towards artificial general intelligence with hybrid Tianjic chip architecture *Nature* **572** 106–11
- [62] Berdan R, Marukame T, Ota K, Yamaguchi M, Saitoh M, Fujii S, Deguchi J and Nishi Y 2020 Low-power linear computation using nonlinear ferroelectric tunnel junction memristors *Nat. Electron.* **3** 259–66
- [63] Srinivasan G, Sengupta A and Roy K 2016 Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network with on-chip STDP learning *Sci. Rep.* **6** 29545
- [64] Tuma T, Pantazi A, Gallo M L, Sebastian A and Eleftheriou E 2016 Stochastic phase-change neurons *Nat. Nanotechnol.* **11** 693–9
- [65] Kim W *et al* 2019 Confined PCM-based analog synaptic devices offering low resistance-drift and 1000 programmable states for deep learning 2019 *Symp. VLSI Technology* (IEEE)
- [66] Wu J Y, Chen Y S, Khwa W S, Yu S M, Wang T Y, Tseng J C, Chih Y D and Diaz C H *et al* 2018 A 40 nm low-power logic compatible phase change memory technology 2018 *IEEE Int. Electron Devices Meeting (IEDM)* pp 1–4
- [67] Lu Y, Lu X, Lu Y, Yan L, Zhang T, Yang Y, Song Z and Huang R 2020 Accelerated local training of CNNs by optimized direct feedback alignment based on stochasticity of 4 Mb C-doped Ge₂Sb₂Te₅ PCM chip in 40 nm node 2020 *IEEE Int. Electron Devices Meeting (IEDM)* pp 1–4
- [68] Lu Y, Li X, Yan B, Yan L, Zhang T, Song Z, Huang R and Yang Y 2021 In-memory realization of eligibility traces based on conductance drift of phase change memory for energy-efficient reinforcement learning *Adv. Mater.* **34** 2107811

- [69] Yan L, Li X, Zhu Y, Yan B, Lu Y, Zhang T, Yang Y, Song Z and Huang R 2021 Uncertainty quantification based on multilevel conductance and stochasticity of heater size dependent C-doped Ge₂Sb₂Te₃ PCM chip 2021 *IEEE Int. Electron Devices Meeting (IEDM)* pp 605–8
- [70] Ielmini D, Lavizzari S, Sharma D and Lacaite A L 2007 Physical interpretation, modelling and impact on phase change memory (PCM) reliability of resistance drift due to chalcogenide structural relaxation 2007 *IEEE Int. Electron Devices Meeting* pp 939–42
- [71] Song S, Song Z, Lu Y, Liu B, Wu L and Feng S 2010 Sb₂Te₃–Ta₂O₅ nano-composite films for low-power phase-change memory application *Mater. Lett.* **64** 2728–30
- [72] Zhou X, Xia M, Rao F, Wu L, Li X, Song Z, Feng S and Sun H 2014 Understanding phase-change behaviors of carbon-doped Ge₂Sb₂Te₃ for phase-change memory application *ACS Appl. Mater. Interfaces* **6** 14207–14
- [73] Koelmans W W, Sebastian A, Jonnalagadda V P, Krebs D, Dellmann L and Eleftheriou E 2015 Projected phase-change memory devices *Nat. Commun.* **6** 8181
- [74] Bruce R L *et al* 2021 Mushroom-type phase change memory with projection liner: an array-level demonstration of conductance drift and noise mitigation 2021 *IEEE Int. Reliability Physics Symp. (IRPS)* pp 1–6
- [75] Ding K *et al* 2019 Phase-change heterostructure enables ultralow noise and drift for memory operation *Science* **366** 210–5
- [76] Hwang S-W *et al* 2012 A physically transient form of silicon electronics *Science* **337** 1640
- [77] Kang S-K *et al* 2016 Bioresorbable silicon electronic sensors for the brain *Nature* **530** 71
- [78] Lee J *et al* 2017 Ultra-wideband multi-dye-sensitized upconverting nanoparticles for information security application *Adv. Mater.* **29** 1603169
- [79] Zhu J, Zhang T, Yang Y and Huang R 2020 A comprehensive review on emerging artificial neuromorphic devices *Appl. Phys. Rev.* **7** 011312
- [80] Wang H, Zhu B, Ma X, Hao Y and Chen X 2016 Physically transient resistive switching memory based on silk protein *Small* **12** 2715
- [81] Dang B, Wu Q, Song F, Sun J, Yang M, Ma X, Wang H and Hao Y 2018 A bio-inspired physically transient/biodegradable synapse for security neuromorphic computing based on memristors *Nanoscale* **10** 20089
- [82] Dang B *et al* 2019 Physically transient memristor synapse based on embedding magnesium nanolayer in oxide for security neuromorphic electronics *IEEE Electron Device Lett.* **40** 1265
- [83] Sun J *et al* 2019 Physically transient memristive synapse with short-term plasticity based on magnesium oxide *IEEE Electron Device Lett.* **40** 706
- [84] Wang S, Dang B, Sun J, Zhao M, Yang M, Ma X, Wang H and Hao Y 2020 Physically transient W/ZnO/MgO/W Schottky diode for rectifying and artificial synapse *IEEE Electron Device Lett.* **41** 844
- [85] Fu T *et al* 2020 Bioinspired bio-voltage memristors *Nat. Commun.* **11** 1861
- [86] Park T *et al* 2020 Thermally assisted nanotransfer printing with sub-20 nm resolution and 8 inch wafer scalability *Sci. Adv.* **6** eabb6462
- [87] Lee G, Choi Y S, Yoon H-J and Rogers J A 2020 Advances in physicochemically stimuli-responsive materials for on-demand transient electronic systems *Matter* **3** 1031
- [88] van de Burgt Y, Melianas A, Keene S T, Malliaras G and Salleo A 2018 Organic electronics for neuromorphic computing *Nat. Electron.* **1** 386–97
- [89] Lee Y and Lee T-W 2019 Organic synapses for neuromorphic electronics: from brain-inspired computing to sensorimotor nervetronics *Acc. Chem. Res.* **52** 964–74
- [90] Sun J, Fu Y and Wan Q 2018 Organic synaptic devices for neuromorphic systems *J. Phys. D: Appl. Phys.* **51** 314004
- [91] Dai S, Zhao Y, Wang Y, Zhang J, Fang L, Jin S, Shao Y and Huang J 2019 Recent advances in transistor-based artificial synapses *Adv. Funct. Mater.* **29** 1903700
- [92] Abbott L F and Regehr W G 2004 Synaptic computation *Nature* **431** 796–803
- [93] Wang Y, Lv Z, Chen J, Wang Z, Zhou Y, Zhou L, Chen X and Han S T 2018 Photonic synapses based on inorganic perovskite quantum dots for neuromorphic computing *Adv. Mater.* **30** e1802883
- [94] Ding G *et al* 2019 Synaptic plasticity and filtering emulated in metal–organic frameworks nanosheets based transistors *Adv. Electron. Mater.* **6** 1900978
- [95] Hao D, Zhang J, Dai S, Zhang J and Huang J 2020 Perovskite/organic semiconductor-based photonic synaptic transistor for artificial visual system *ACS Appl. Mater. Interfaces* **12** 39487–95
- [96] Dai S, Wu X, Liu D, Chu Y, Wang K, Yang B and Huang J 2018 Light-stimulated synaptic devices utilizing interfacial effect of organic field-effect transistors *ACS Appl. Mater. Interfaces* **10** 21472–80
- [97] Wang H, Yang M, Tang Q, Zhao X, Tong Y and Liu Y 2019 Flexible, conformal organic synaptic transistors on elastomer for biomedical applications *Adv. Funct. Mater.* **29** 1901107
- [98] Wang X, Yan Y, Li E, Liu Y, Lai D, Lin Z, Liu Y, Chen H and Guo T 2020 Stretchable synaptic transistors with tunable synaptic behavior *Nano Energy* **75** 104952
- [99] Fu Y, Kong L-a, Chen Y, Wang J, Qian C, Yuan Y, Sun J, Gao Y and Wan Q 2018 Flexible neuromorphic architectures based on self-supported multiterminal organic transistors *ACS Appl. Mater. Interfaces* **10** 26443–50
- [100] Zhang C, Wang S, Zhao X, Yang Y, Tong Y, Zhang M, Tang Q and Liu Y 2020 Sub-femtojoule-energy-consumption conformable synaptic transistors based on organic single-crystalline nanoribbons *Adv. Funct. Mater.* **31** 2007894
- [101] Chen Y, Zhou Y, Zhuge F, Tian B, Yan M, Li Y, He Y and Miao X S 2019 Graphene-ferroelectric transistors as complementary synapses for supervised learning in spiking neural network *npj 2D Mater. Appl.* **3** 31
- [102] Dai S, Wang Y, Zhang J, Zhao Y, Xiao F, Liu D, Wang T and Huang J 2018 Wood-derived nanopaper dielectrics for organic synaptic transistors *ACS Appl. Mater. Interfaces* **10** 39983–91
- [103] Deng W, Zhang X, Jia R, Huang L, Zhang X and Jie J 2019 Organic molecular crystal-based photosynaptic devices for an artificial visual-perception system *NPG Asia Mater.* **11** 77
- [104] Zang Y, Shen H, Huang D, Di C-A and Zhu D 2017 A dual-organic-transistor-based tactile-perception system with signal-processing functionality *Adv. Mater.* **29** 1606088
- [105] Song Z, Tong Y, Zhao X, Ren H, Tang Q and Liu Y 2019 A flexible conformable artificial organ-damage memory system towards hazardous gas leakage based on a single organic transistor *Mater. Horiz.* **6** 717–26
- [106] Liu Y, Li E, Wang X, Chen Q, Zhou Y, Hu Y, Chen G, Chen H and Guo T 2020 Self-powered artificial auditory pathway for intelligent neuromorphic computing and sound detection *Nano Energy* **78** 105403
- [107] Zhang S, Guo K, Sun L, Ni Y, Liu L, Xu W, Yang L and Xu W 2021 Selective release of different neurotransmitters emulated by a p–i–n junction synaptic transistor for environment-responsive action control *Adv. Mater.* **33** e2007350

- [108] Russell S J and Norvig P 2009 *Artificial Intelligence: A Modern Approach* 3rd edn (Englewood Cliffs, NJ: Prentice-Hall)
- [109] Yu F and Zhu L Q 2019 Ionotronic neuromorphic devices for bionic neural network applications *Phys. Status Solidi* **13** 1800674
- [110] Guo Y B, Zhu L Q, Long T Y, Wan D Y and Ren Z Y 2020 Bio-polysaccharide electrolyte gated photoelectric synergic coupled oxide neuromorphic transistor with Pavlovian activities *J. Mater. Chem. C* **8** 2780–9
- [111] Duan N *et al* 2019 An electro-photo-sensitive synaptic transistor for edge neuromorphic visual systems *Nanoscale* **11** 17590–9
- [112] Wan C J, Zhu L Q, Liu Y H, Feng P, Liu Z P, Cao H L, Xiao P, Shi Y and Wan Q 2016 Proton-conducting graphene oxide-coupled neuron transistors for brain-inspired cognitive systems *Adv. Mater.* **28** 3557–63
- [113] Yu F *et al* 2020 Artificial tactile perceptual neuron with nociceptive and pressure decoding abilities *ACS Appl. Mater. Interfaces* **12** 26258–66
- [114] Wan C *et al* 2020 An artificial sensory neuron with visual-haptic fusion *Nat. Commun.* **11** 4602
- [115] Yang J-T, Ge C, Du J-Y, Huang H-Y, He M, Wang C, Lu H-B, Yang G-Z and Jin K-J 2018 Artificial synapses emulated by an electrolyte-gated tungsten-oxide transistor *Adv. Mater.* **30** 1801548
- [116] Zhu Y, Liu G, Xin Z, Fu C, Wan Q and Shan F 2020 Solution-processed, electrolyte-gated In₂O₃ flexible synaptic transistors for brain-inspired neuromorphic applications *ACS Appl. Mater. Interfaces* **12** 1061–8
- [117] Zhou Y, Li J, Yang Y, Chen Q and Zhang J 2020 Artificial synapse emulated through fully aqueous solution-processed low-voltage In₂O₃ thin-film transistor with Gd₂O₃ solid electrolyte *ACS Appl. Mater. Interfaces* **12** 980–8
- [118] Zhu L Q, Wan C J, Guo L Q, Shi Y and Wan Q 2014 Artificial synapse network on inorganic proton conductor for neuromorphic systems *Nat. Commun.* **5** 3158
- [119] Wan C J, Liu Y H, Feng P, Wang W, Zhu L Q, Liu Z P, Shi Y and Wan Q 2016 Flexible metal oxide/graphene oxide hybrid neuromorphic transistors on flexible conducting graphene substrates *Adv. Mater.* **28** 5878–85
- [120] Zhong G *et al* 2020 Flexible electronic synapse enabled by ferroelectric field effect transistor for robust neuromorphic computing *Appl. Phys. Lett.* **117** 092903
- [121] Wang W, Li X-X, Wang T, Huang W, Ji Z-G, Zhang D W and Lu H-L 2020 Investigation of light-stimulated α -IGZO based photoelectric transistors for neuromorphic applications *IEEE Trans. Electron Devices* **67** 3141–5
- [122] Li Y *et al* 2020 Oxide-based electrolyte-gated transistors for spatiotemporal information processing *Adv. Mater.* **32** 2003018
- [123] Zhang C, Li S, He Y, Chen C, Jiang S, Yang X, Wang X, Pan L and Wan Q 2020 Oxide synaptic transistors coupled with triboelectric nanogenerators for bio-inspired tactile sensing application *IEEE Electron Device Lett.* **41** 617–20
- [124] Zhu L, He Y, Chen C, Zhu Y, Shi Y and Wan Q 2021 Synergistic modulation of synaptic plasticity in IGZO-based photoelectric neuromorphic TFTs *IEEE Trans. Electron Devices* **68** 1659–63
- [125] Yu J J *et al* 2019 Optoelectronic neuromorphic thin-film transistors capable of selective attention and with ultra-low power dissipation *Nano Energy* **62** 772–80
- [126] Jiang D, Li J, Fu W, Chen Q, Yang Y, Zhou Y and Zhang J 2020 Light-stimulated artificial synapse with memory and learning functions by utilizing an aqueous solution-processed In₂O₃/AlLiO thin-film transistor *ACS Appl. Electron. Mater.* **2** 2772–9
- [127] He Y, Nie S, Liu R, Jiang S, Shi Y and Wan Q 2019 Patiotemporal information processing emulated by multiterminal neuro-transistor networks *Adv. Mater.* **31** 1900903
- [128] Turing A 1969 *The Essential Turing* (Oxford: Clarendon Press) p 395
- [129] Mead C 1980 *IEE Proc. I* **128** 18
- [130] Akopyan F *et al* 2015 *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **34** 1537–57
- [131] Davies M *et al* 2018 *IEEE Micro* **38** 82–99
- [132] Zhong H *et al* 2020 *Chin. Phys. B* **29** 040703
- [133] Yang J J, Pickett M D, Li X, Ohlberg D A A, Stewart D R and Williams R S 2008 *Nat. Nanotechnol.* **3** 429–33
- [134] Wang Z *et al* 2017 *Nat. Mater.* **16** 101–8
- [135] Ge C, Li G, Zhou Q-l, Du J-y, Guo E-j, He M, Wang C, Yang G-z and Jin K-j 2020 *Nano Energy* **67** 104268
- [136] Ge C *et al* 2019 *Adv. Mater.* **31** 1900379
- [137] Luo Z-D, Xia X, Yang M-M, Wilson N R, Gruverman A and Alexe M 2020 *ACS Nano* **14** 746
- [138] Torrezan A C, Strachan J P, Medeiros-Ribeiro G and Williams R S 2011 *Nanotechnology* **22** 485203
- [139] Kumar S *et al* 2016 *ACS Nano* **10** 11205–10
- [140] Tian J *et al* 2019 *Adv. Mater.* **31** 1903679
- [141] Huang H Y *et al* 2019 *Adv. Funct. Mater.* **29** 1902702
- [142] Deng X, Wang S Q, Liu Y X, Zhong N, He Y H, Peng H, Xiang P H and Duan C G 2021 *Adv. Funct. Mater.* **31** 2101099
- [143] Jang J-W, Park S, Burr G W, Hwang H and Jeong Y-H 2015 *IEEE Electron Device Lett.* **36** 457–9
- [144] Li J, Ge C, Du J, Wang C, Yang G and Jin K 2020 *Adv. Mater.* **32** 1905764
- [145] Ambrogio S *et al* 2018 *Nature* **558** 60–7
- [146] Upadhyay N K, Jiang H, Wang Z, Asapu S, Xia Q and Yang J J 2019 Emerging memory devices for neuromorphic computing *Adv. Mater. Technol.* **4** 1800589
- [147] Wan Q, Sharbati M T, Erickson J R, Du Y and Xiong F 2019 Emerging artificial synaptic devices for neuromorphic computing *Adv. Mater. Technol.* **4** 1900037
- [148] Pereda A E 2014 Electrical synapses and their functional interactions with chemical synapses *Nat. Rev. Neurosci.* **15** 250–63
- [149] Gao G M *et al* 2019 Photoinduced hysteresis of graphene field-effect transistors due to hydrogen-complexed defects in silicon dioxide *ACS Appl. Mater. Interfaces* **11** 12170–8
- [150] Gao G, Liu X, Liu W, Liu X and Wang X 2017 Chemical environment dominated Fermi level pinning of a graphene gas sensor *Carbon* **124** 57–63
- [151] Zhu J *et al* 2018 Ion gated synaptic transistors based on 2D van der Waals crystals with tunable diffusive dynamics *Adv. Mater.* **30** 1800195
- [152] Tian H *et al* 2017 Emulating bilingual synaptic response using a junction-based artificial synaptic device *ACS Nano* **11** 7156–63
- [153] Xiong X, Kang J, Hu Q, Gu C, Gao T, Li X and Wu Y 2020 Reconfigurable logic-in-memory and multilingual artificial synapses based on 2D heterostructures *Adv. Funct. Mater.* **30** 1909645
- [154] Cao G, Meng P, Chen J, Liu H, Bian R, Zhu C, Liu F and Liu Z 2021 2D material based synaptic devices for neuromorphic computing *Adv. Funct. Mater.* **31** 2005443
- [155] Lee G, Baek J H, Ren F, Pearton S J, Lee G H and Kim J 2021 Artificial neuron and synapse devices based on 2D materials *Small* **17** 2100640
- [156] Hu W, Jiang J, Xie D, Liu B, Yang J and He J 2019 Proton-electron-coupled MoS₂ synaptic transistors with a natural renewable biopolymer neurotransmitter for brain-inspired neuromorphic learning *J. Mater. Chem. C* **7** 682–91

- [157] Sharbati M T, Du Y, Torres J, Ardolino N D, Yun M and Xiong F 2018 Low-power, electrochemically tunable graphene synapses for neuromorphic computing *Adv. Mater.* **30** 1802353
- [158] Kim S, Choi B, Lim M, Yoon J, Lee J, Kim H-D and Choi S-J 2017 Pattern recognition using carbon nanotube synaptic transistors with an adjustable weight update protocol *ACS Nano* **11** 2814–22
- [159] Ren Y, Yang J Q, Zhou L, Mao J Y, Zhang S R, Zhou Y and Han S T 2018 Gate-Tunable synaptic plasticity through controlled polarity of charge trapping in fullerene composites *Adv. Funct. Mater.* **28** 1805599
- [160] Jin T, Zheng Y, Gao J, Wang Y, Li E, Chen H, Pan X, Lin M and Chen W 2021 Controlling native oxidation of HfS_2 for 2D materials based flash memory and artificial synapse *ACS Appl. Mater. Interfaces* **13** 10639–49
- [161] Huang W *et al* 2020 Gate-coupling-enabled robust hysteresis for nonvolatile memory and programmable rectifier in van der Waals ferroelectric heterojunctions *Adv. Mater.* **32** 1908040
- [162] Wang S, Liu L, Gan L, Chen H, Hou X, Ding Y, Ma S, Zhang D W and Zhou P 2021 Two-dimensional ferroelectric channel transistors integrating ultra-fast memory and neural computing *Nat. Commun.* **12** 53
- [163] Cheng Y, Li H, Liu B, Jiang L, Liu M, Huang H, Yang J, He J and Jiang J 2020 Vertical 0D-perovskite/2D-MoS₂ van der Waals heterojunction phototransistor for emulating photoelectric-synergistically classical pavlovian conditioning and neural coding dynamics *Small* **16** e2005217
- [164] Cheng Y, Shan K, Xu Y, Yang J, He J and Jiang J 2020 Hardware implementation of photoelectrically modulated dendritic arithmetic and spike-timing-dependent plasticity enabled by an ion-coupling gate-tunable vertical 0D-perovskite/2D-MoS₂ hybrid-dimensional van der Waals heterostructure *Nanoscale* **12** 21798–811
- [165] Wan C, Cai P, Wang M, Qian Y, Huang W and Chen X 2019 Artificial sensory memory *Adv. Mater.* **32** 1902434
- [166] Kim Y *et al* 2018 A bioinspired flexible organic artificial afferent nerve *Science* **360** 998–1003
- [167] Zhou F *et al* 2019 Optoelectronic resistive random access memory for neuromorphic vision sensors *Nat. Nanotechnol.* **14** 776–82
- [168] Cook S J *et al* 2019 Whole-animal connectomes of both *Caenorhabditis elegans* sexes *Nature* **571** 63–71
- [169] Hua Q *et al* 2020 Atomic threshold-switching enabled MoS₂ transistors towards ultralow-power electronics *Nat. Commun.* **11** 6207
- [170] Yang C and Suo Z 2018 Hydrogel iontronics *Nat. Rev. Mater.* **3** 125–42
- [171] Lenz J, del Giudice F, Geisenhof F R, Winterer F and Weitz R T 2019 Vertical, electrolyte-gated organic transistors show continuous operation in the MA cm^{-2} regime and artificial synaptic behaviour *Nat. Nanotechnol.* **14** 579–85
- [172] Tee B C-K *et al* 2015 A skin-inspired organic digital mechanoreceptor *Science* **350** 313–6
- [173] Midya R *et al* 2019 Artificial neural network (ANN) to spiking neural network (SNN) converters based on diffusive memristors *Adv. Electron. Mater.* **5** 1900060
- [174] Kim H-J, Sim K, Thukral A and Yu C 2017 Rubbery electronics and sensors from intrinsically stretchable elastomeric composites of semiconductors and conductors *Sci. Adv.* **3** e1701114
- [175] Jung Y H, Park B, Kim J U and Kim T-i 2019 Bioinspired electronics for artificial sensory systems *Adv. Mater.* **31** 1803637
- [176] Luo T, Liu S, Li L, Wang Y, Zhang S, Chen T, Xu Z, Temam O and Chen Y 2017 DaDianNao: a neural network supercomputer *IEEE Trans. Comput.* **66** 73–88
- [177] Khwa W-S *et al* 2018 A 65 nm 4 Kb algorithm-dependent computing-in-memory SRAM unit-macro with 2.3 ns and 55.8 TOPS/W fully parallel product-sum operation for binary DNN edge processors *IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 496–8
- [178] Si X *et al* 2019 24.5 A twin-8 T SRAM computation-in-memory macro for multiple-bit CNN-based machine learning *IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 396–8
- [179] Shafiee A, Nag A, Muralimanohar N, Balasubramonian R, Strachan J P, Hu M, Williams R S and Srikumar V 2016 ISAAC: a convolutional neural network accelerator with *in situ* analog arithmetic in crossbars vol 44 *ACM/IEEE 43rd Annual Int. Symp. Computer Architecture (ISCA)* pp 14–26
- [180] Tang S *et al* 2017 AEPE: an area and power efficient RRAM crossbar-based accelerator for deep CNNs *IEEE 6th Non-Volatile Memory Systems and Applications Symp. (NVMSA)* pp 1–6
- [181] Mochida R *et al* 2018 A 4 M synapses integrated analog ReRAM based 66.5 TOPS/W neural-network processor with cell current controlled writing and flexible network architecture *IEEE Symp. VLSI Technology* pp 175–6
- [182] Kulkarni S R, Kadetotad D V, Yin S, Seo J and Rajendran B 2019 Neuromorphic hardware accelerator for SNN inference based on STT-RAM crossbar arrays *26th IEEE Int. Conf. Electronics, Circuits and Systems (ICECS)* pp 438–41
- [183] Zhang X *et al* 2020 Fully memristive SNNs with temporal coding for fast and low-power edge computing *IEEE Int. Electron Devices Meeting (IEDM)* pp 1–4
- [184] Li J *et al* 2020 *In situ* learning in hardware compatible multi-layer memristive spiking neural network *IEEE Trans. Cogn. Develop. Syst.* **14** 448–61
- [185] Ma D, Shen J, Gu Z, Zhang M, Zhu X, Xu X, Xu Q, Shen Y and Pan G 2017 Darwin: a neuromorphic hardware co-processor based on spiking neural networks *J. Syst. Archit.* **77** 43–51
- [186] Zhang L *et al* 2020 MAMAP: congestion relieved memetic algorithm based mapping method for mapping large-scale SNNs onto NoC-based neuromorphic hardware *IEEE 22nd Int. Conf. High Performance Computing and Communications; IEEE 18th Int. Conf. Smart City; IEEE 6th Int. Conf. Data Science and Systems (HPCC/SmartCity/DSS)* pp 640–7
- [187] Zhang J, Wang Y, Zhang X and Huang R 2017 Compact digital-controlled neuromorphic circuit with low power consumption *IEEE Int. Symp. Circuits and Systems (ISCAS)* pp 1–4
- [188] Wang Z *et al* 2018 Capacitive neural network with neuro-transistors *Nat. Commun.* **9** 3208
- [189] Chen C, Yang M, Liu S, Liu T, Zhu K, Zhao Y, Wang H, Huang Q and Huang R 2019 Bio-inspired neurons based on novel leaky-FeFET with ultra-low hardware cost and advanced functionality for all-ferroelectric neural network *Symp. VLSI Technology* pp T136–7
- [190] Peng Y *et al* 2020 Non-volatile field-effect transistors enabled by oxygen vacancy-related dipoles for memory and synapse applications *IEEE Trans. Electron Devices* **67** 3632–6
- [191] Wang Y, Hu L, Wei X and Zhuge F 2020 Ultralow operation voltages of a transparent memristor based on bilayer ITO *Appl. Phys. Lett.* **116** 221602
- [192] Testa E, Soeken M, Amar I G and De Micheli G 2018 Logic synthesis for established and emerging computing *Proc. IEEE* **107** 165–84
- [193] De Micheli G 1994 *Synthesis and Optimization of Digital Circuits* (New York: McGraw-Hill)
- [194] Sebastian A, Le Gallo M, Khaddam-Aljameh R and Eleftheriou E 2020 Memory devices and applications for in-memory computing *Nat. Nanotechnol.* **15** 529–44

- [195] Neckar A, Fok S, Benjamin B V, Stewart T C, Oza N N, Voelker A R, Eliasmith C, Manohar R and Boahen K 2018 Braindrop: a mixed-signal neuromorphic architecture with a dynamical systems-based programming model *Proc. IEEE* **107** 144–64
- [196] Strukov D B and Likharev K K 2005 CMOL FPGA: a reconfigurable architecture for hybrid digital circuits with two-terminal nanodevices *Nanotechnology* **16** 888
- [197] Xia Y, Chu Z, Hung W N, Wang L and Song X 2011 An integrated optimization approach for nanohybrid circuit cell mapping *IEEE Trans. Nanotechnol.* **10** 1275–84
- [198] Zha X and Xia Y 2020 Defective cell reuse based defect-tolerance method for CMOL cell mapping optimization *Microelectron. J.* **105** 104863
- [199] Borghetti J, Snider G S, Kuekes P J, Yang J J, Stewart D R and Williams R S 2010 ‘Memristive’ switches enable ‘stateful’ logic operations via material implication *Nature* **464** 873–6
- [200] Soeken M, Gaillardon P-E, Shirinzadeh S, Drechsler R and De Micheli G 2017 A PLiM computer for the internet of things *Computer* **50** 35–40
- [201] Amaru L, Gaillardon P-E and De Micheli G 2015 Majority-inverter graph: a new paradigm for logic optimization *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **35** 806–19
- [202] Chakraborty I, Ali M F, Kim D E, Ankit A and Roy K 2020 GENIEx: a generalized approach to emulating non-ideality in memristive xbars using neural networks *2020 57th ACM/IEEE Design Automation Conf. (DAC)* (IEEE) pp 1–6
- [203] Yoo H and Kim C-H 2021 Multi-valued logic system: new opportunities from emerging materials and devices *J. Mater. Chem. C* **9** 4092–104
- [204] Chu Z, Li Z, Xia Y, Wang L and Liu W 2020 BCD adder designs based on three-input XOR and majority gates *IEEE Trans. Circuits Syst. II* **68** 1942–6
- [205] Brayton R K and Khatir S P 1999 Multi-valued logic synthesis *Proceedings 12th Int. Conf. VLSI Design (Cat. No. PR00013)* (IEEE) pp 196–205
- [206] Dean J, Patterson D and Young C 2018 A new golden age in computer architecture: empowering the machine-learning revolution *IEEE Micro* **38** 21–9
- [207] Amodei D and Hernandez D 2018 AI and compute *OpenAI*
- [208] Jouppi N P *et al* 2017 In-datacenter performance analysis of a tensor processing unit *Proc. IEEE Int. Symp. Computer Architecture (ISCA)* pp 1–12
- [209] Liu S, Du Z, Tao J, Han D, Luo T, Xie Y, Chen Y and Chen T 2016 Cambricon: an instruction set architecture for neural networks *Proc. IEEE Int. Symp. Computer Architecture (ISCA)* pp 393–405
- [210] Chen Y, Krishna T, Emer J and Sze V 2016 14.5 Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks *2016 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 262–3
- [211] Ouyang J, Du X, Ma Y and Liu J 2021 3.3 Kunlun: a 14 nm high-performance AI processor for diversified workloads *2021 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 50–1
- [212] Lin Z, Zhan H, Chen Z, Peng C, Wu X, Lu W, Zhao Q, Li X and Chen J 2021 Cascade current mirror to improve linearity and consistency in SRAM in-memory computing *IEEE J. Solid-State Circuits* **56** 2550–62
- [213] Jiao Y *et al* 2020 7.2 A 12 nm programmable convolution-efficient neural-processing-unit chip achieving 825 TOPS *2020 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 136–40
- [214] Yin S *et al* 2018 A high energy efficient reconfigurable hybrid neural network processor for deep learning applications *IEEE J. Solid-State Circuits* **53** 968–82
- [215] Chen C, Liu X, Peng H, Ding H and Shi C-R 2018 iFPNA: a flexible and efficient deep neural network accelerator with a programmable data flow engine in 28 nm CMOS *IEEE 44th European Solid State Circuits Conf. (ESSCIRC)* pp 170–3
- [216] Liu J *et al* 2021 BioAIP: a reconfigurable biomedical AI processor with adaptive learning for versatile intelligent health monitoring *2021 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 62–4
- [217] Yue J *et al* 2019 A 65 nm 0.39 to 140.3 TOPS/W 1 to 12b unified neural network processor using block-circulant-enabled transpose-domain acceleration with $8.1\times$ higher TOPS/mm² and 6 T HBST-TRAM-based 2D data-reuse architecture *2019 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 138–40
- [218] Shan W *et al* 2020 A 510 nW 0.41 V low-memory low-computation keyword-spotting chip using serial FFT-based MFCC and binarized depthwise separable convolutional neural network in 28 nm CMOS *2020 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 230–2
- [219] Chen C *et al* 2017 OCEAN: an on-chip incremental-learning enhanced processor with gated recurrent neural network accelerators *43rd IEEE European Solid State Circuits Conf.* pp 259–62
- [220] Yue J *et al* 2020 A 65 nm computing-in-memory-based CNN processor with 2.9 to 35.8 TOPS/W system energy efficiency using dynamic-sparsity performance-scaling architecture and energy-efficient inter/intra-macro data reuse *2020 IEEE Int. Solid-State Circuits Conf. (ISSCC)* pp 234–6
- [221] Guo R *et al* 2021 A 6.54 to 26.03 TOPS/W computing-in-memory RNN processor using input similarity optimization and attention-based context-breaking with output speculation *2021 Symp. VLSI Circuits* pp 1–2
- [222] Yang J *et al* 2021 TIMAQ: a time-domain computing-in-memory-based processor using predictable decomposed convolution for arbitrary quantized DNNs *IEEE J. Solid-State Circuits* **56** 3021–38
- [223] Tavanaei A, Ghodrati M, Kheradpisheh S R, Masquelier T and Maida A 2019 Deep learning in spiking neural networks *Neural Netw.* **111** 47–63
- [224] Werner T, Vianello E, Bichler O, Garbin D, Cattaert D, Yvert B, De Salvo B and Perniola L 2016 Spiking neural networks based on OxRAM synapses for real-time unsupervised spike sorting *Front. Neurosci.* **10** 474
- [225] Akopyan F *et al* 2015 TrueNorth: design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **34** 1537–57
- [226] Alioto M and Shahghasemi M 2018 The internet of things on its edge: trends toward its tipping point *IEEE Consum. Electron. Mag.* **7** 77–87
- [227] Intel Neuromorphic Computing Lab 2021 *Taking neuromorphic computing to the next level with Loihi 2* <https://download.intel.com/newsroom/2021/new-technologies/neuromorphic-computing-loihi-2-brief.pdf>
- [228] Liu Q, Richter O, Nielsen C, Sheik S, Indiveri G and Qiao N 2019 Live demonstration: face recognition on an ultra-low power event-driven convolutional neural network ASIC *IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops*