applied sciences

MDPI

Editorial

Foreword to the Special Issue on Deep Learning and Neuromorphic Chips

Xuemeng Fan [1,2] and Yishu Zhang [1,2,*]

[1] School of Micro-Nano Electronics, Zhejiang University, Hangzhou 311200, China
[2] ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 311200, China
* Correspondence: zhangyishu@zju.edu.cn

With the advent of the Internet of Things and the era of big data, the ability of machine data processing to reach the level of human brain cognition and learning is an important goal in the field of Internet information technology, including cloud computing, data mining, machine learning, and artificial intelligence (AI). The technological change has brought about the exponential generation of data, and the "memory wall" caused by frequent data movement between memory and processor has aggravated the "von Neumann bottleneck". With the end of Moore's Law, it becomes difficult to improve the performance of chips and systems through chip minification. Therefore, it is urgent to innovate a computing paradigm to solve the power-consumption and computing-power problems induced by the rapid development of data-centric applications. Meanwhile, the demand for energy-sensitive electronic devices with superior computing power is ubiquitous in the era of the Internet of Things, especially for edge applications. Devices are the basis of the computing network of the Internet of Things. Millions of devices not only support the communication between machines but also constantly generate data to exchange information by controlling the complex network of intelligent infrastructure. With the improvement in the computing power of devices, the scale of the network grows larger and larger, and the learning ability grows stronger and stronger. However, the ability to integrate low-cost and high-density devices and the array size are also crucial for evaluating the feasibility of chip integration.

The neuromorphic system simulates the storage, computing, and information expression architecture of the biological brain, which processes massive amounts of data with extremely low power consumption and high speed. It is the core technology supporting the development of emerging applications such as artificial intelligence. Shi et al. proposed to use the 28 nm process to manufacture the neuromorphic computing chip "Tianjic". A single chip supports 40,000 neurons and 10 million synapses, which can simultaneously support computer-science-oriented and neuroscience-oriented computing modes, but it is still difficult to achieve the scale and power consumption of the human brain [1]. The rapid development of information processing depends on the progress of underlying electronic devices. In the research and development of neuromorphic devices, many materials show resistive-switching characteristics under electric fields or current stimulation. The performance of memristors based on traditional bulk materials is highly dependent on atomic scale defect structures and has limited adjustability to neuromorphic functions. However, they have the advantages of good CMOS compatibility, multilevel states, and simple chemical composition, such as HfO_2, ZnO, NiO, TiO_2, WO_3, and TaOx [2]. Due to the advantages of dangling bond-free lattices and easily available single-layer (less than 1 nm in thickness) structures, emerging two-dimensional materials (such as WSe2, h-BN, BP, and MoS2) have become popular research topics in the field of hyper-scaling electronic devices, such as matrix computing performance enhancement, continuous scaling of transistor size, logic gates, and highly integrated chips, to achieve more efficient memory and computing integration. Yang et al. integrated a single-layer MoS_2 transistor with a metal oxide resistor

check for updates

Citation: Fan, X.; Zhang, Y. Foreword to the Special Issue on Deep Learning and Neuromorphic Chips. *Appl. Sci.* **2022**, *12*, 11189. https://doi.org/10.3390/app122111189

Received: 2 November 2022
Accepted: 3 November 2022
Published: 4 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

random access memory (RRAM), created two transistor double resistor (2T2R) transition metal hydride TCAM (TMD-TCAM) cells, made full use of the high-performance and low leakage performance of two-dimensional materials, obtained a high resistance ratio, realized the parallel search for a large number of data bits, and was compatible with the back-end (BEOL) metal interconnection process [3]. Nevertheless, there is still a lack of chip-level experimental verification for the application of 2D materials, and it is a great challenge to produce 2D materials with large-area and high-quality growth.

To achieve large-scale parallel and efficient brain-like computing in devices, strict requirements are put forward for the uniformity, reliability, and stability of device performance. Memory integration and uniformity are the primary conditions, while endurance and retention are the key factors. Optimizing the application characteristics of memristor devices (such as switching energy and speed, retention, and endurance) helps to maximize the computing area, energy efficiency, and memory capacity of memristor-based neural networks. Yang et al. reported a random dynamic artificial neuron based on a diffusion-type memristor device, which can realize unsupervised synaptic weight update and pattern classification of integral memory convolution neural network using artificial neurons and break through the limitation of silicon-based CMOS neural morphology computing chips [4]. Yet, most related research on neuromorphic computing focuses on a single device or a small array. Still, the problems of sneak path current and wire resistance also challenge the realization of a large-scale and high-density crossbar array. The introduction of third-generation neural networks–spike neural networks has promoted deep-learning networks to develop a series of new logical structures and algorithms and filled the gap between neural science and machine learning, which provides the possibility for the construction of large-scale neural networks. Although deep-learning algorithms show excellent performance in many fields, they need to consume a lot of storage and computing resources when performing data calculation and analysis and must have competitive advantages in training ability, speed, and algorithm accuracy [5]. This also reflects the urgent need for non-Von Neumann-architecture-based chips. Leblebici et al. improved the tolerance of algorithms, training speed, and power of machine learning on large-scale artificial neural network chips based on nonvolatile synaptic devices [6]. However, this research does not improve the data-transmission ability between neural network modules. Burr et al. proposed a design of synaptic unit cells to simulate DNN training. By combining a hardware array with circuit simulation through software and hardware mixed experiments, it was estimated that its efficiency and throughput were 280 times and 100 times higher than those of the latest GPU, respectively, providing a reference for training on neuromorphic chips [7].

Although network-level research on applications such as neuromorphic devices, crossbar arrays, and neuromorphic systems is booming, there is still an urgent need to obtain knowledge from neuroscience research. The purpose of this Special Issue on "Deep Learning and Neuromorphic Chips" is to provide corresponding research opportunities to solve the "memory wall" and "power wall" problems in the Von Neumann architecture in terms of new materials, new device characteristics, new algorithms, and integration technologies to achieve brain level computing capabilities and ultra-low power consumption, as well as support the rapid development of artificial intelligence and other fields. I believe that the above research will inspire the development of deep learning and the application of neuromorphic chips.

Funding: This research was funded by the Young Scientists Fund of the National Natural Science Foundation of China [grant number: 62204219].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pei, J.; Deng, L.; Song, S.; Zhao, M.; Zhang, Y.; Wu, S.; Wang, G.; Zou, Z.; Wu, Z.; He, W.; et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **2019**, *572*, 106–111. [CrossRef] [PubMed]
2. Molas, G.; Nowak, E. Advances in Emerging Memory Technologies: From Data Storage to Artificial Intelligence. *Appl. Sci.* **2021**, *11*, 11254. [CrossRef]
3. Yang, R.; Li, H.; Smithe, K.K.; Kim, T.R.; Okabe, K.; Pop, E.; Fan, J.A.; Wong, H.-S.P. Ternary content-addressable memory with MoS2 transistors for massively parallel data search. *Nat. Electron.* **2019**, *2*, 108–114. [CrossRef]
4. Wang, Z.; Joshi, S.; Savel'ev, S.; Song, W.; Midya, R.; Li, Y.; Rao, M.; Yan, P.; Asapu, S.; Zhuo, Y.; et al. Fully memristive neural networks for pattern classification with unsupervised learning. *Nat. Electron.* **2018**, *1*, 137–145. [CrossRef]
5. Davies, M.; Srinivasa, N.; Lin, T.H.; Chinya, G.; Cao, Y.; Choday, S.H.; Dimou, G.; Joshi, S.; Imam, N.; Jain, S.; et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* **2018**, *38*, 82–99. [CrossRef]
6. Burr, G.W.; Narayanan, P.; Shelby, R.M.; Sidler, S.; Boybat, I.; Di Nolfo, C.; Leblebici, Y. Large-scale neural networks implemented with non-volatile memory as the synaptic weight element: Comparative performance analysis (accuracy, speed, and power). In Proceedings of the 2015 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 7–9 December 2015.
7. Ambrogio, S.; Narayanan, P.; Tsai, H.; Shelby, R.M.; Boybat, I.; di Nolfo, C.; Sidler, S.; Giordano, M.; Bodini, M.; Farinha, N.C.P.; et al. Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **2018**, *558*, 60–67. [CrossRef] [PubMed]