2D Materials for Next-Generation Computing Technologies

Peng Zhou

State Key Lab of ASIC and Systems，Fudan University
220#, Handan, Yangpu, Shanghai 200433, China
Phone: +86-021-6564-2198 E-mail: pengzhou@fudan.edu.cn

Abstract

The rapid development of digital technology has led to a large increase in computing tasks, and has put forward strict energy efficiency and area efficiency requirements for next-generation computing. Considering the efficiency of data-driven unique computing tasks, in-memory computing and transistor-based computing have become effective technologies for realizing matrix and logic computing to solve perception and reasoning tasks. However, in order to meet future computing requirements, new materials are urgently needed to supplement the existing Si CMOS technology and develop new technologies to further diversify electronics and its applications. The abundance and rich electronic properties of two-dimensional (2D) materials give them the potential to enhance computational energy efficiency, while enabling devices to continue to shrink to the atomic level. This paper discusses the requirements, opportunities and challenges of integrating 2D materials with in-memory computing and transistor-based computing technologies, from the perspective of matrix and logic computing for perception and reasoning tasks.

1. Introduction

According to data processing methods, computing tasks can be divided into perception tasks and inference tasks. For solving perceptual tasks such as image recognition and language processing, highly parallel matrix calculation methods perform better, while for solving inference tasks, serial logic calculation methods have more advantages. At present, the computer architecture is based on the von Neumann architecture of physically separated computing and memory modules. However, researchers are trying to solve both perception and reasoning tasks through this sole architecture. The difference between the essential requirements of perception and reasoning tasks poses challenges to the von Neumann architecture. First, the data shuttle between the processing and the memory module is limited by the von Neumann structure, which results in higher energy and time consumption when processing perception tasks, and may even be greater than the calculation itself [1]. Although the industry has proposed near-memory technologies to improve energy efficiency [2], the limited connection density of through-silicon via (TSV) interconnects is still not enough to achieve energy-efficient computing [3,4]. In addition, solving inference tasks with high performance requires higher transistor density in logic processing modules. However, the device scaling rule proposed by Dennard et al. [5], that is, when the thickness of silicon is less than 3 nm, it faces an inevitable performance degradation [6,7]. Recently, due to the different requirements of calculation methods for perception and reasoning tasks, it is a more effective way to design a computing architecture based on the computing method. In-memory computing assembles storage devices into arrays to perform matrix calculations, which allows to eliminate energy and time-consuming data movement [8]. This technology has broad prospects in matrix calculations, but many technical problems still need to be addressed before commercial applications are feasible. High-energy-consuming memory operations will increase computing power consumption, which will offset the advantages of in-situ computing; memory instability will reduce computing accuracy; non-biologically similar dynamic mechanisms will bring challenges to the actual implementation of low-power biological systems. All the challenges discussed above urgently require the introduction of novel material systems, as these problems are largely due to the limitations of bulk materials.

For logic computing, memory computing technology has proven to be inefficient because of the low logic calculation efficiency caused by its step-by-step calculation process [9]. As an alternative, field-effect transistor (FET) technology is still the most suitable choice. In order to achieve higher transistor density, the feature size of transistors has continued to shrink below 5 nm, and the strong scattering caused by bulk materials with dangling bonds is no longer a suitable material system. With the advantage of no dangling bonds, 2D materials can overcome the limitations of silicon-based transistors and realize an area-efficient structure, which is also suitable for the integration of computing and memory.

In this paper, we focus on the requirements, opportunities and challenges of integrating 2D materials with in-memory computing and transistor-based computing technology into the processing of perception and reasoning tasks from the perspective of matrix and logic computing.

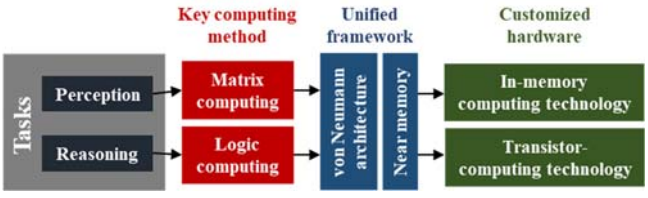

Fig. 1 Conceptual framework of task demand-driven computing method and its hardware implementation technology.

2. 2D materials for matrix computing applications

The storage unit of matrix calculation needs to meet specific requirements in different neural network applications. Specifically, artificial neural networks are expected to have non-volatile analog storage characteristics with multiple conductance modulation, large dynamic range, high linearity and weight update symmetry, while spiking neural networks prefer the ability to realistically simulate biological neurons and synaptic functions. Memory devices that integrate 2D materials show the potential to boost matrix computing in indicators such as power consumption, accuracy, sneak current path problems, and bionic plasticity, which are usually troublesome problems in bulk memory devices.

In matrix computing, low power consumption is very demanding, and memristive devices usually bear high SET currents. For example, phase change memories always require high operating voltages. By introducing 2D materials with atomic thickness, the operating voltage or current can be significantly reduced, thereby reducing power consumption.

The multiple conductance states play a key role in the calculation accuracy of the artificial neural network. In addition, the large dynamic range, high linearity and symmetry of conductance modulation are indispensable. The variability between devices is also a worrying issue, which may result in a significant drop in artificial neural network performance. The sneak current also has an adverse effect on the calculation accuracy of the artificial neural network. Integrating 2D materials and memory devices as network hardware units is a promising strategy to solve these problems and improve calculation accuracy. Moreover, due to the excellent optical properties of 2D materials, light stimulation has become an emerging approach for conductance modulation.

Furthermore, the integration of storage devices and 2D materials exhibits superiority in matrix calculations for SNN, which allows realistic simulation of biomimetic properties. Due to the atomic-scale layered structure and the subsequent special properties, 2D materials have become an optimized platform for ion transistors in multifunctional and high-performance SNN applications. Abundant 2D material libraries and photoelectrically tunable heterojunction interface barriers provide the feasibility for the construction of various 2D synapses and neuronal components through band alignment and structural engineering. And 2D materials without dangling bonds allow the flexible formation of heterostructures and have photoelectric tunability, which provides opportunities for the realization of multiple bioplasticity.

3. 2D materials for logic computing applications

In approaching the limit scale, two physical constraints need to be resolved. First, in order to ensure effective gate control, the thickness of the channel material is close to the quantum limit. In addition, due to the leakage current of the ultra-thin dielectric, the scaling of the power supply voltage has stagnated, so devices with new mechanisms are required to further reduce the voltage.

The lattice without dangling bonds and the easily available monolayer (thickness less than 1 nm) provide the potential for sustainable scaling of 2D materials. For supply voltage scaling, some new device structures have been proposed, such as TFET and NCFET, which usually involve a new mechanism to overcome the subthreshold swing limit (60 mV/dec at room temperature). However, as a compromise, the drive current of the steep sub-threshold swing device is not large enough. By introducing atomic-scale 2D materials, the sub-threshold swing and drive current performance of the device have been significantly improved, simultaneously.

In addition to the performance optimization in single device, thanks to the layered structure, various materials can be transferred to each other regardless of lattice mismatch, which leads to the introduction of 2D materials that can present a more effective logic gate structure and van der Waals integration technology.

4. Conclusions

In summary, we focus the performance improvement and device innovation inspired by the introduction of 2D materials, including but not limited to the power consumption, computing accuracy, and biomimetic properties advantages of in-memory computing technology, the sustainable physical size and voltage scaling and high area-efficient integration potentials of 2D materials in transistor-based computing technology.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61925402, 61851402 and 61734003), Science and Technology Commission of Shanghai Municipality (19JC1416600), National Key Research and Development Program (2017YFB0405600), Shanghai Education Development Foundation and Shanghai Municipal Education Commission Shuguang Program(18SG01).

References

[1] M. Horowitz, *in 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers* (2014) 10-14.
[2] D. U. Lee et al., *in 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers* (2014) 432-433.
[3] D. Liu and S. Park, J. Electron. Packag. **136** (2014) 014001.
[4] M. M. Shulaker et al., Nature **547** (2017) 74-78.
[5] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous and A. R. LeBlanc, IEEE J. Solid-State Circuits **9** (1974) 256-268.
[6] T. Irisawa, T. Numata, T. Tezuka, N. Sugiyama and S. I. Takagi, *in 2006 International Electron Devices Meeting* (2006) 1-4.
[7] K. Uchida et al., *in Technical Digest-International Electron Devices Meeting* (2002) 47-50.
[8] D. Ielmini and H. S. P. Wong, Nat. Electron. **1** (2018) 333-343.
[9] Zhang Z, Wang Z, Shi T,et al. InfoMat. 2020;2:261–290.