RECENT PROGRESSES OF IN-MEMORY COMPUTING: MATERIALS, DEVICES AND ARCHITECTURES

D. Ielmini[1,*], F. Sancandi[1], M. Farronato[1], S. Hashemkhani[1], S. Ricci[1], M. Baldo[1], P. Mannocci[1], N. Lepri[1], L. Cattaneo[1], A. Milozzi[1], A. Glukhov[1]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy
(*) daniele.ielmini@polimi.it

With the end of Moore's law of transistor scaling, there has been a renewed interest in unconventional concepts for computing that can overcome the memory bottleneck typical of von Neumann architectures. Among the novel computing concepts, in-memory computing (IMC) offers the opportunity of bringing computation into the memory, thus virtually suppressing data movement and the associated latency and power consumption. At the same time, IMC is also compatible with the conventional CMOS technology, is scalable and can be operated at room temperature, thus is suitable for edge computing devices and the internet of things (IoT). Figure 1 illustrates the most popular trends of IMC, including (i) accelerators of artificial neural networks, (ii) brain-inspired computing concepts and (iii) hardware accelerators for linear algebra and machine learning. While each of these topics has its own architecture and applications, they all share the fundamental requirements for high energy efficiency, high throughput and good scalability, to compete with the conventional digital CMOS technology.

This talk will review the recent progresses of IMC in terms of materials, devices and architectures. From the materials/device viewpoints, I will review novel concepts of resistive switching random access memory (RRAM) exhibiting volatile switching, thanks to the surface re-diffusion of Ag across the conductive filament. As application cases, a direction-selective IMC circuit [1] and a fully-memristive reservoir-computing circuit based on a nanowire network will be shown [2]. From the architecture viewpoint, I will present a new concept of analogue IMC with closed-loop capable of accelerating linear algebra tasks, such as matrix inverse/pseudoinverse calculation in one step [3]. The main limitations of IMC in terms of accuracy and scaling will finally be discussed.

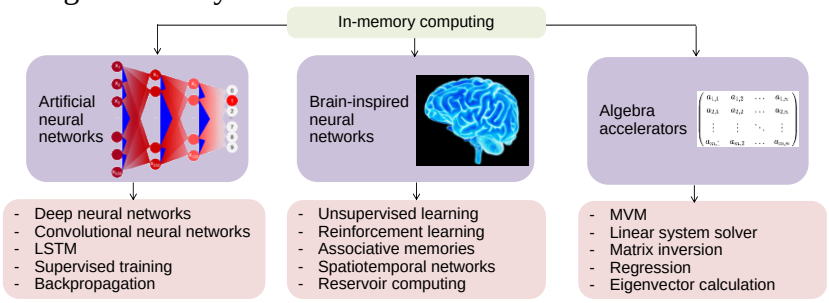

Figure 1. Summary of the recent trends about IMC.

Keywords
In-memory computing, neuromorphic computing, embedded memory, deep learning accelerators.

References

[1] W. Wang et al., "Neuromorphic motion detection and orientation selectivity by volatile resistive switching memories," Advanced Intelligent Systems 3, 2000224 (2020). DOI: 10.1002/aisy.202000224.

[2] G. Milano et al., "In-materia reservoir computing with a fully memristive architecture based on self-organizing nanowire networks", Nature Mater (2021). DOI: 10.1038/s41563-021-01099-9.

[3] P. Mannocci et al., "A fully-analogue, universal primitive for linear algebra operations using resistive memories", IEEE Trans. Circuits and Systems I: Regular Papers (2021).