Ph.D. in Information Technology
Thesis Defenses

October 9th, 2025
At 10:30 a.m.
Room Alpha – Building 24

Davide BRIDAROLLI – XXXVII Cycle

FABRICATION AND CHARACTERIZATION OF PLANAR AND 3D-VERTICAL RESISTIVE RANDOM-ACCESS MEMORY ARRAYS FOR IN-MEMORY COMPUTING
Supervisor: Prof. Daniele Ielmini

Abstract:

The energy consumption of information and communication technology is increasing worldwide, and the rise of Artificial Intelligence (AI) leads to massive amounts of data that need to be processed and stored. Data-centric tasks, such as training and inference in AI models, impose a heavy computational burden on traditional digital computing systems based on the Von Neumann architecture. Motivated by these trends, the investigation of next-generation non-volatile computing and memory devices in which data and logic coexist is vital. The in-memory computing concept (IMC) is a promising paradigm shift where highly specialized tasks are executed within memory arrays, improving the efficiency of the data-intensive workloads, such as deep learning on the edge, by removing the data shuttling between computational and memory elements.

Resistive random access memory (RRAM) is one of the most promising emerging memories, due to a broad range of attractive properties for IMC, such as non-volatile analog storage, low power consumption, and high memory density. 3D-Vertical RRAM (3D-VRRAM) is a promising option to achieve high memory cell capacity with low fabrication cost, toward large-scale integration competitive with state-of-the-art 3D NAND flash.

This Doctoral dissertation presents the fabrication and characterization of planar 2D and 3D-VRRAM crossbar arrays (CBA) based on HfO_x cells and capable of IMC with precise multilevel programming. Proof-of-concept open-loop and closed-loop IMC experiments are carried out on 3D-VRRAM, and to further support the application of IMC on real-life scenarios, the work also reports demonstration of relatively large-size image compression problems adopting 2D-RRAM.

In addition, the dissertation presents results on the heterointegration between RRAM and spin-Hall nano-oscillators (SHNO), in the context of achieving a hybrid neural network. The RRAMs provide the non-volatile storage of synaptic weights, while SHNO act as fast oscillating neurons.

Matteo PORZANI – XXXVII Cycle

ELECTROCHEMICAL RANDOM ACCESS MEMORY (ECRAM) FOR ANALOG IN-MEMORY COMPUTING

Supervisor: Prof. Daniele Ielmini

The enormous popularization of artificial intelligence, and the exponential growth in complexity of deep neural network (DNN) architectures is also followed by a growing need for computational power, requiring energy-efficient hardware to handle the enormous amount of operations needed for their training. Moreover, such training algorithms require an enormous amount of data to be continuously exchanged between the CPU and the memory in a traditional von Neumann architecture, resulting in a performance bottleneck known as the "memory wall". This causes latency and great power dissipation, making it unsustainable for the next-generation DNN models. In-memory computing (IMC) has recently emerged as a new computing paradigm, where memory units also function as computational kernels to directly process data in the devices where they are stored, minimizing data movement. IMC promises low latency operation, thanks to its massively parallelized architecture, and low power consumption, thanks to the adoption of emerging resistive memory devices as key elements of memory arrays, paving the way to alleviate the von Neumann bottleneck. Among many interesting emerging technologies that are being actively researched as a next-generation resistive memory device, the Electrochemical Random Access Memory (ECRAM) is gaining interest due to its low-power consumption and good switching properties. The ECRAM is a three-terminal resistive memory with a transistor-like structure, with a gate insulation composed of a solid-state electrolyte and an ion reservoir layer. The ECRAM exploits the voltage-driven migration of ions through the electrolyte to modulate the conductance of its channel, achieving resistance switching. In particular, ECRAMs based on transition metal oxides, such as tungsten oxide, are of great interest because of their compatibility with the CMOS process and their ability to be seamlessly fabricated in the "back end of the line" (BEOL) fabrication phase, allowing monolithical integration with state-of-the-art CMOS logic. The ECRAM device promises high linearity and symmetry, as well as a high number of states and fast programming times, making it optimal for IMC applications. These devices seem to provide better linearity properties and less stochastic problems than more established resistive memory technologies, like two-terminal RRAM, paying the price of a more complex three-terminal elementary cell. Despite these interesting properties, ECRAM is a newly proposed device that still needs research effort to best characterize its resistive switching mechanism, optimizing scaling, yield, and feasibility in establishing it as a valid alternative for implementing beyond-CMOS applications in hardware neural network accelerators. This Ph.D. thesis deals with the fabrication of ECRAM devices, their electrical characterization, the physical modeling of the observed experimental results, and finally some demonstrations of parallel algebraic operations realized in ECRAM arrays, paving the way for more complex implementations of ECRAM in future IMC architectures. This dissertation is organized as follows: • Chapter 1 introduces the end of Moore's law of electronics and the challenges faced by modern electronics. A brief overview of established memory technologies is also given before delving into the von Neumann bottleneck in modern computer architectures. Emerging memories will be presented as the key technology, making it possible to implement in-memory computing as a solution to overcome the bottleneck. Finally, an in-depth look is given to ECRAM memories, the topic of thesis, with a brief review of the state-of-the-art. • Chapter 2 deals with the microfabrication of ECRAM devices carried out in cleanroom. First,

an overview of the main semiconductor microfabrication technologies is given. Then, a fabrication study of one of the most important materials for ECRAM, WO3, is shown, highlighting the role of oxygen concentration in its electrical characteristics. An overview of the structure of the metal-oxide-based ECRAM devices fabricated for this work will be shown, and finally, the complete fabrication process flow that was developed for the realization of ECRAM crossbar arrays will be presented and commented upon. • Chapter 3 presents the experimental characterization of ECRAM devices. Quasi-static experiments are performed on the devices to observe the large-signal conductance modulation. Afterwards, pulsed experiments will be performed, studying the ECRAM response as a function of the programming parameters. Two-terminal programming will also be explored, identifying an analytical compact model to describe device conductance modulation in the small signal regime. • Chapter 4 is devoted to the physics-based modeling of ECRAM devices. The resistive switching mechanism will be described in terms of nonlinear migration of oxygen vacancies through the device. An analytical model will be derived to describe the conductance update induced by the application of a single voltage pulse, and will be also extended to account for two-terminal operations. The model will be solved numerically, correctly describing both large and small signal behavior of ECRAM devices. Lastly, a compact analytical formula will be derived from the physical picture, paving the way for a physics-based approach to the description of ECRAM in IMC applications. • Chapter 5 is dedicated to in-memory computing applications with ECRAM arrays. Matrix-vector multiplication and transposed matrix-vector multiplication will be demonstrated, along with a closed-loop programming algorithm for ECRAM arrays. Scalar product, outer product, and stochastic multiplication will be demonstrated on 2x2 arrays, and will be formalized in terms of equations derived from the physical model. Finally, large DNN training simulations will be performed with realistic device models as unit cells in IMC hardware accelerators.

PhD Committee

Prof. Luca Bertulessi, **Politecnico di Milano**

Prof. Yoeri Van de Burgt, **Eindhoven University of Technology**

Prof. Andrea Padovani, **Università degli Studi di Modena e Reggio Emilia**