

## **Ph.D. in Information Technology: Thesis Defenses**

**February 4th, 2021**

**online by Teams – at 15.00**

### **Stefano BIANCHI – XXXIII Cycle**

**“HARDWARE DESIGN AND IMPLEMENTATION OF MEMRISTIVE-BASED LEARNING SYSTEMS FOR EFFICIENT NEUROCOMPUTING”**

Advisor: Prof. **Daniele Ielmini**

#### **Abstract:**

Brain-inspired computing is probably one of the top research challenges for enhancing the next era of the artificial computation, aiming to reproduce in hardware the learning, adaptation, and elaboration abilities of the brain. At the beginning of the research in computing algorithms and hardware development, the investigation on the artificial intelligence showed a deep interchange with neuroscience and psychology, with highly productive collaborations. In the last years, the interaction has become much less commonplace, since both the research fields have deepened the analysis and the complexity in their research lines, respectively. The principal aim of this doctoral dissertation is to cover this gap, offering novelties to the field of artificial intelligence by looking at the hardware realization of hybrid bio-inspired/artificial neural networks using memristive synaptic devices.

The first part of the thesis deals with the modelling of novel non-volatile memories and bio-inspired spiking neural networks. The principal goal of this study is related to the theorization of a complete set of analytical tools for the investigation of artificial intelligent networks. Such insights are also propaedeutic for the study, implementation, and hardware realization of neuromorphic systems able to overcome historical limitations of the artificial intelligence, such as the “stability-plasticity dilemma”. The research line proposed in this doctoral dissertation also focuses on some specific applications using bio-inspired neurons able to reproduce bio-plausible cognitive behaviours, such as synaptic scaling and reinforcement learning. In particular, a self-referential hardware is described in detail, thus offering an experimental insight over some relevant applications such as decision-

making and optimized navigation in harsh environments. A high-level analysis related to the integrated design of neural accelerators by exploiting the benefits of PCM-based “in-memory computing” is also proposed in order to discuss the most relevant technological contributions. The novelties here introduced, as well as the description of the experimental setups, are key elements to introduce new paradigms in the standard elaboration procedures, thus highlighting the fundamental steps for boosting the next technological era of computation.

### **Irene MUÑOZ MARTIN – XXXIII Cycle**

**“A MIXED-SIGNAL INTEGRATED CIRCUIT BASED ON PHASE CHANGE MEMORY SYNAPSES FOR DEEP NEURAL ACCELERATORS”**

Advisor: Prof. **Daniele Ielmini**

#### **Abstract:**

In-memory computing is an emerging technology that allows to perform data calculations exactly where the data are stored, that is, on the memory. It rises as a very effective method for overcoming the limitations of typical von Neumann architectures since it massively parallelizes the operations and avoids the so called “von Neumann bottleneck”.

In particular, in-memory computing requires the use of memory elements capable of storing data and performing calculations at the same time. New emerging non-volatile memories (NVM), such as phase change memory (PCM) or resistive switching RAM (RRAM) give solution to these requirements, as they have small size and show fast switching, multilevel capability and low-voltage operation. In addition, they show time-dependent dynamics and can be arranged in array architectures. Typically, the application of in-memory computing with NVM exploits the Ohm's and Kirchhoff's laws for performing multiply-and-accumulate (MAC) operations.

Deep learning is a machine learning method based on deep neural networks (DNNs) that trust on backpropagation algorithm for performing intelligent tasks. However, GPU and CPU's based training and testing operations of DNNs are power and time consuming. As the main operations in DNNs are related to dense matrix-matrix multiplications, in-memory computing rises as a suitable solution for implementing DNNs as the power efficiency and speed could be highly improved.

This doctoral dissertation proposes a mixed-signal integrated circuit based on PCM synapses for the development of deep neural accelerators. Following the in-memory computing hardware approach, DNN weights are mapped in NVM arrays. The circuit performs intelligent tasks of recognition of handwritten digits (MNIST dataset) at high speed (256 mega operations per second, MOPS), and relies on a significant robustness with respect to the non-idealities of PCM devices, since the results are resilient both to drift and resistance variability, achieving almost the same software classification accuracy of the MNIST dataset (~85 %).

This work highlights the main features, problems, and requirements for implementing a hardware integrated DNN using PCM cells and proposes several circuitual solutions for obtaining an efficient design.

#### **PhD Committee**

Prof. **Christian Monzio Compagnoni**, DEIB

Prof. **Benjamin Ryad Benosman**, University of Pittsburgh

Prof. **Antonio Gnudi**, Universita' di Bologna