****

**Stefano Ambrogio, IBM Almaden Research Center**

**ABSTRACT**

**Novel Technologies for Artificial Intelligence: prospects and challenges**

Artificial Intelligence (AI), exploiting bio-inspired algorithms such as Spike-Timing-Dependent-Plasticity (STDP) and back-propagation schemes as found in Deep Neural Networks (DNNs), is able to perform accurate recognition and classification of large amounts of data. However, to further proceed in the development of AI, novel hardware technologies supporting fast calculation should be developed. Recently, many algorithms have been efficiently mapped into arrays of Non-Volatile Memories, such as Phase-Change Memory (PCM) or Resistive Memory (RRAM).

In this invited talk, we provide a summary of recent progress in hardware acceleration of AI algorithms, such as the training of Fully Connected (FC) DNNs based on large arrays of PCMs and the inference of Long-Short Term Memory (LSTM) networks with more than 2.5 million devices. In such schemes, crossbar arrays of weights encoded as conductances are shown to provide orders of magnitude increases in speed and energy efficiency with respect to current state of the art CPUs and GPUs, while still retaining software-equivalent accuracy.

**BIO**

Stefano Ambrogio obtained his PhD in 2016 in Italy, at Politecnico di Milano, under the supervision of Prof. Daniele Ielmini, studying the reliability of resistive memories and their application on neuromorphic networks. He is now a Research Staff Member at IBM- Research, Almaden, in the Neuromorphic Devices and Architectures Team, working on hardware accelerators based on Non-Volatile Memories for neural networks.