Contribution ID: **108** Type: **WP2**

Implementation of low latency, fast inference neural networks on FPGA for trigger-like systems

In many research and industrial settings, achieving fast, low-latency algorithmic responses is essential. To meet the demands of the upgraded LHC and future High Energy Physics (HEP) detectors, quick and powerful triggers are necessary. In recent years, machine learning algorithms have been widely applied to such tasks, and more recently, solutions based on Field Programmable Gate Arrays (FPGAs) have proven to be effective, offering reduced latency and power consumption compared to GPUs not only in HEP field.
This work presents an overview of the development of fast neural networks on FPGAs for High-level and Level-0 trigger systems in the ATLAS experiment. Additionally, to investigate the performance and scalability of the algorithm on multi-FPGA systems, an AMD Alveo cluster is currently being constructed at the INFN-Naples site.
This work is within the use case "Ultra-fast algorithms running on FPGA"within the WP2 framework.

Giorno preferito

Primary author: Dr LOFFREDO, Salvatore (Istituto Nazionale di Fisica Nucleare)

Presenter: Dr LOFFREDO, Salvatore (Istituto Nazionale di Fisica Nucleare)