

WP2 FPGA MS9 Status Report

Simone Gennai & Bernardino Spisso

MS9 Deliverables

- **MS9 Intermediate Step:** For this flagship project, MS9 was an intermediate step where significant advancements were not expected for most use cases. Some use cases faced delays due to the late delivery of FPGA hardware by the Terabit project. The open calls have concluded, and new personnel have been hired for a relevant use case. However, due to the long time elapsed between the initial document's creation and the open call, the proponent of the use case has changed the topic, with further details provided later in the report.
- **Personnel and Topic Changes:** The open calls have ended, and new personnel have been recruited for a use case of interest to this flagship project. Due to the significant time gap between the initial document's writing and the occurrence of the open call, the proponent of the use case has changed the topic.

- Trigger, DAQ and on-line processing
 - Development of trigger and anomaly detection algorithms based on FPGAs at event level for the Atlas experiment.
 - Current efforts include testing GNN ML-based implementations, focusing on Level-0 triggers using RPC.
 - Development of track reconstruction on FPGA for LHC-b data acquisition:
 - No report expected for MS9
 - Development of digital trigger logic for a “missing energy” experiment with a positron beam at CERN (POKER/NA64):
 - Due to the long time passed between the original proposal and the open call, the use case is changed concentrating on a data reduction algorithm based on autoencoder
 - Scouting and processing of Level-1 trigger data using FPGA to run on-the-fly momentum object calibration with ML:
 - An intermediate report was already given for MS8 and can be found in the FPGA flagship document. No further development has been done in the meanwhile,

- Developing FPGA tools
 - Development of a Customizable Framework for Multi-FPGA Accelerator Generation via architectures:
 - The framework now supports the Alveo U55C FPGA card, which has been tested and integrated. It has been used to develop a new application, `bmqsim`, for simulating Quantum Computing algorithms, with a working prototype tested on the ZedBoard and Alveo U55C. A new project site has been launched at bondmachine.it with updated content. Ongoing work includes firmware for multi-FPGA systems and integrating convolutional neural networks into the framework.
 - Development and testing of RDMA over converged ethernet (ROCE) on FPGA for data transfer from detectors' front-end to computing servers
 - Although a report was not explicitly targeted for this milestone, the developments are noteworthy. The Padova group implemented the RoCE v2 stack firmware on an AMD-Xilinx VCU118 FPGA and tested data transfer to an NVIDIA-Mellanox Connect-X NIC on a receiver server. The direct connection tests showed no packet loss and achieved 100 Gbps bandwidth
- Terabit (FPGA clusters) : The machines are about to be installed

Updated list of papers and contributions

Tabella 1

Titolo	Autori	Link a paper/conference	Accepted for PUB/PROceeding
Fast Neural Network Inference on FPGAs for Triggering on Long-Lived Particles at Colliders	Andrea Coccaro Francesco Armando Di Bello Stefano Giagu Lucrezia Rambelli and Nicola Stocchetti	https://arxiv.org/pdf/2307.05152.pdf	Andrea Coccaro et al 2023 Mach. Learn.: Sci. Technol. 4 045040
Sviluppo di acceleratori per il Machine Learning e sistemi di Inference as a Service su FPGA	Daniele Spiga, Diego Ciangottini , Giacomo Surace, Giulio Bianchini, Lorian Storchi , Mirko Mariotti	Workshop Loano	
KServe inference extension for a FPGA vendor-free ecosystem	Daniele Spiga, Diego Ciangottini , Giacomo Surace, Giulio Bianchini, Lorian Storchi , Mirko Mariotti	CHEP 2023	EPJ Web of Conferences 295, 11012 (2024)
Deep Learning techniques for reconstruction on ASTRI Mini-Array Monte Carlo data	Saverio Lombardi, Francesco Visconti, Michele Mastropietro	https://pos.sissa.it/444/713/pdf	PoS(ICRC2023)713
A novel explainable approach in radiomics pipeline for local recurrence prediction of lung cancer: a feasibility study exploiting high energy physics potential to evaluate the model	Mariagrazia Monteleone, Simone Gennai, Pietro Govoni, Chiara Paganelli	ACM ISBN 979-8-4007-0815-2/23/09. https://doi.org/10.1145/3632047.3632074	ACM ISBN 979-8-4007-0815-2/23/09. https://doi.org/10.1145/3632047.3632074
Triggerless data acquisition pipeline for Machine Learning based statistical anomaly detection	Gaia Grosso, Nicolò Lai, Matteo Migliorini, Jacopo Pazzini, Andrea Triossi, Marco Zanetti, Alberto Zucchetta	CHEP 2023	G. Grosso et al EPJ Web of Conf., 295 (2024) 02033
40MHz Triggerless Readout of the CMS Drift Tube Muon Detector	Matteo Migliorini, Jacopo Pazzini, Andrea Triossi, Marco Zanetti	TWEPP 2023	M. Migliorini et al 2024 JINST 19 C02050
Front-End RDMA Over Converged Ethernet, real-time firmware simulation	Gabriele Bortolato, Antonio Bergnoli, Damiano Bortolato, Daniele Mengoni, Matteo Migliorini, Fabio Montecassiano, Jacopo Pazzini, Sandro Ventura, Andrea Triossi, Marco Zanetti	TWEPP 2023	G. Bortolato et al 2024 JINST 19 C03038
Front-End Rdma Over Converged Ethernet, real-time firmware simulation	Gabriele Bortolato, Antonio Bergnoli, Damiano Bortolato, Daniele Mengoni, Matteo Migliorini, Fabio Montecassiano, Jacopo Pazzini, Sandro Ventura, Andrea Triossi, Marco Zanetti	TIPP 2023	
The CMS Level-1 trigger data scouting for LHC run 3 and the CMS phase-2 upgrade	Sabrina Giorgetti (Matteo Migliorini, Rocco Ardino, Jacopo Pazzini, Andrea Triossi, Marco Zanetti) on behalf of the CMS Collaboration	La Thuile 2024 - YSF	
Hardware implementation of quantum machine learning predictors for ultra-low latency applications	Lorenzo Borella, Alberto Coppi, Jacopo Pazzini, Andrea Stanco, Andrea Triossi, Marco Zanetti	EuCAIFCon 2024	
Quantum machine learning classifiers implemented on FPGA for ultra-low latency applications	Lorenzo Borella, Alberto Coppi, Jacopo Pazzini, Andrea Stanco, Andrea Triossi, Marco Zanetti	ICHEP 2024	

Status KPIs

[Link to the flagship document](#)

KPI ID	Description	Acceptance threshold	Status up to today
KPI2.2.3.1	Development of triggering algorithms, on-line analyses, data acquisition on FPGA	Submission of 1 paper to a peer-reviewed journal	1 paper already accepted
KPI2.2.3.2	Online scouting	Submission of 1 paper to a peer-reviewed journal	Abstract being submitted to ichep 2024 about scouting
KPI2.2.3.3	Development of tools to integrate several FPGAs together	Submission of 1 paper to a peer-reviewed journal	G. Bortolato et al 2024 JINST 19 C03038
KPI2.2.3.4	Organizing courses about FPGA programming on low and high level	At least two courses organized	1 course done at the end of 2023 1 VHDL course done in February 2024

For what regards WP2-WP4 synergies a series of lessons on ML training opportunity on GPU was organized within the “Artificial Intelligent and Modern Physics” PhD school (<https://aiphy.fisica.unimib.it>) organized by, among others, the WP4 leaders. The school aimed to give an advanced series of lessons on ML related problems and developments. Many of these models required heavy use of GPU computing for their training. The GPUs have been made available by CINECA through their Leonardo Booster nodes.



Artificial intelligence and modern physics: a two-way connection

29 September - 5 October 2024, Monopoli (Bari) - Italy