Contribution ID: **59** Type: **Oral Presentation**

Scalable training on scalable infrastructures for programmable hardware (Remote presentation)

Wednesday, 22 March 2023 14:00 (20 minutes)

The increasingly pervasive and dominant role of machine learning (ML) and deep learning (DL) techniques in High Energy Physics is posing challenging requirements to effective computing infrastructures on which AI workflows are executed, as well as demanding requests in terms of training and upskilling new users and/or future developers of such technologies.

In particular, a growth in the request for training opportunities to become proficient in exploiting programmable hardware capable of delivering low latencies and low energy consumption, like FPGAs, is observed. While training opportunities on generic ML/DL concepts is rich and quite wide in the coverage of sub-topics, a gap is observed in the delivery of hands-on tutorials on ML/DL on FPGAs that can scale to a relatively large number of attendants and that can give access to a relatively diverse set of ad-hoc hardware with different hardware specs.

A pilot course on ML/DL on FPGAs - born from the collaboration of INFN-Bologna, the University of Bologna and INFN-CNAF - has been successful in paving the way for the creation of a line of work dedicated to maintaining and expanding an ad-hoc scalable toolkit for similar courses in the future. The practical sessions are based on virtual machines (for code development, no FPGAs), in-house cloud platforms (INFN-cloud infrastructure equipped with AMD/Xilinx Alveo FPGA), Amazon AWS instances for project deployment on FPGAs - all complemented by docker containers with the full environments for the DL frameworks used, as well as Jupyter Notebooks for interactive exercises. The current results and plans of work along the consolidation of such a toolkit will be presented and discussed.

Finally, a software ecosystem called Bond Machine, capable of dynamically generate computer architectures that can be synthesised in FPGA, is being considered as a suitable alternative to teach FPGA programming without entering into the low-level details, thanks to the hardware abstraction it offers which can simplify the interaction with FPGAs.

Primary authors: COSTANTINI, Alessandro (INFN-CNAF); Prof. BONACORSI, Daniele (University of Bologna); SPIGA, Daniele (INFN-PG); SALOMONI, Davide (INFN); MICHELOTTO, Diego (INFN-CNAF); DUMA, Doina Cristina (INFN - CNAF); LORUSSO, Marco (Alma Mater Studiorum - University of Bologna); MARIOTTI, Mirko (Department of Physics and Geology, University of Perugia); VERONESI, Paolo (INFN); Dr TRAVAGLINI, Riccardo (INFN)

Presenter: LORUSSO, Marco (Alma Mater Studiorum - University of Bologna)

Session Classification: Converging Infrastructure Clouds, Virtualisation & HPC

Track Classification: Track 8: Infrastructure Clouds and Virtualizations