T 33: DAQ NN/ML – GRID I

Time: Tuesday 17:00–18:30 Location: HSZ/0301

T 33.1 Tue 17:00 HSZ/0301

Track reconstruction with Graph Neural Networks on FPGAs for the ATLAS Event Filter at the HL-LHC — Sebastian Dittmeier and •Sachin Gupta — Physikalisches Institut, Universität Heidelberg

The High-Luminosity LHC (HL-LHC) will enhance the potential to discover new physics with the ATLAS experiment beyond its reach at the LHC. To cope with the increased pile-up foreseen during the HL-LHC, major upgrades to the ATLAS detector and trigger system are required. The trigger system will consist of a hardware-based trigger and an online server farm, called the Event Filter (EF), with track reconstruction capabilities. For the EF, a heterogeneous computing farm consisting of CPUs and potentially GPUs and/or FPGAs is under study, together with the use of modern machine learning algorithms such as Graph Neural Networks (GNNs).

GNNs are a powerful class of geometric deep learning methods for modelling spatial dependencies via message passing over graphs. They are well-suited for track reconstruction tasks by learning on an expressive structured graph representation of hit data. A considerable speed-up over CPU-based execution is possible on FPGAs.

In this talk, a study of track reconstruction for the ATLAS EF system at HL-LHC using GNNs on FPGAs is presented. The main focus is set on model size minimization using quantization aware training, as resource utilization is a key aspect in the application of GNNs on FPGAs.

T 33.2 Tue 17:15 HSZ/0301

Convolutional Neural Networks on FPGAs for Processing of ATLAS Liquid Argon Calorimeter Signals — •Johann Christoph Voigt, Anne-Sophie Berthold, Nick Fritzsche, Rainer Hentges, Christian Gutsche, and Arno Straessner — Institut für Kern- und Teilchenphysik, TU Dresden, Germany

The Phase-II upgrade of the ATLAS Liquid Argon Calorimeter allows for the energy reconstruction of all ~180000 readout channels at the LHC bunch crossing frequency of 40 MHz. Further challenges arise from the increased pile-up due to the planned higher number of simultaneous proton-proton collisions.

For the digital energy reconstruction, we propose the use of Convolutional Neural Networks (CNNs) instead of the previous Optimal Filter. The networks need be able to run on an FPGA with limited resources and are therefore limited in complexity to approximately 100 weight parameters.

This talk focuses on the firmware implementation of these networks in VHDL. The implementation is optimized for DSP usage and latency. To be able to process all readout channels on the available FPGAs, time domain multiplexing is used to process multiple channels per CNN instance. This reduces the number of required instances and increases the frequency the design needs to run at. A multiplexing factor of 12 at a frequency of 480 Mhz is demonstrated for a design processing 384 detector cells. The latest FPGA resource usage estimates are presented.

T 33.3 Tue 17:30 HSZ/0301

Implementation of neural networks for live reconstruction using AI processors — Klaus Desch[1], Jochen Kaminski[1], Michael Lupberger[1,2], and •Patrick Schwäbig[1] — [1]Physikalisches Institut, Universität Bonn, Deutschland — [2]Helmholtz-Institut für Strahlen- und Kernphysik, Universität Bonn, Deutschland

For years, data rates generated by modern detectors and the corresponding readout electronics exceeded by far the limits of data storage space and bandwidth available in many experiments. The solution of using fast triggers to discard uninteresting and irrelevant data is a solution used to this day. Using FPGAs, ASICs or directly the readout chip, a fixed set of rules based on low level parameters is applied as a pre-selection. Only a few years ago, live track reconstruction for triggering was rarely possible but with the emergence of fast and highly parallelized processors for AI inference attempts to sufficiently accelerate tracking algorithms become viable. The Xilinx Versal AI Adaptive Compute Acceleration Platform (ACAP) is one such technology and combines FPGA and CPU resources with dedicated AI cores. Our approach is to utilize the unique combination of FPGA and AI cores to leverage neural networks for live triggering which will be relevant for future experiments and upgrades of already existing setups.

In this talk AI algorithms for track reconstruction, especially their quantized and non-quantized implementation on the Xilinx VC1902, will be shown. They will be used in an envisioned mid-size ultra-high rate fixed-target dark matter experiment (Lohengrin) at the ELSA accelerator at the University of Bonn.

T 33.4 Tue 17:45 HSZ/0301

Profiling of GPU-based neural network trainings — •Tim Voigtländer, Manuel Giffels, Artur Gottmann, Günter Quast, Matthias Schnepf, and Roger Wolf — Karlsruhe Institute of Technology, Karlsruhe, Germany

The training of neural networks has become a significant workload of particle physics analyses. To speed up these trainings and reduce their turnaround cycle, one or more accelerators, e.g. GPUs, are typically utilized. While the increase in computational capacity is greatly beneficial, the heterogeneous hardware also adds layers of complexity to an already opaque process. In order to improve the efficiency in the usage of the available hardware, suitable profiling to identify possible bottlenecks. In this talk, solutions to a number of commonly occurring challenges found in single- and multi-GPU neural network trainings are presented, using the DeepTau neural network training as a case-study.

T 33.5 Tue 18:00 HSZ/0301

Open Science in KM3NeT — •Rodrigo Gracia-Ruiz for the ANTARES-KM3NET-ERLANGEN-Collaboration — FAU-ECAP, Erlangen, Germany

The KM3NeT neutrino detectors are currently under construction at two locations in the Mediterranean Sea, with a first data taking of high-energy neutrino interactions already under way. This scientific data is valuable both for the astrophysics and neutrino physics communities as well as for marine biologists. In order to facilitate FAIR data sharing of the research results, the KM3NeT collaboration is actively working towards an open science infrastructure to provide high-level scientific data, software, and analysis pipelines in an interoperable research environment suited both for research and education. This contribution introduces the open science program of KM3NeT and gives an overview of its current architecture and implementation.

T 33.6 Tue 18:15 HSZ/0301

Towards JupyterHub as one point of entry for the PUNCH4NFDI computing infrastructure — •Luka Vomberg, Philip Bechtle, Oliver Freyermuth, and Peter Wienemann for the PUNCH4NFDI Consortium-Collaboration — Physikalisches Institut Bonn

PUNCH (Particles, Universe, NuClei and Hadrons) is a consortium of the NFDI (Nationale ForschungsDaten Infrastruktur) representing about 9000 physicists in Germany at the Ph.D. level. The goals of PUNCH are the setup of infrastructure that enables physicists to easily manage and publish their data and corresponding analyses in accordance to the FAIR principles. These are Findability, Accessibility, Interoperability and Reproducibility, which are desirable properties that publications and their underlying data should adhere to as much as possible. The infrastructure supplied by PUNCH is meant to make this as easy as possible for individual researchers and collaborations alike. This talk will describe the vision of accessing the PUNCH infrastructure with FAIRness in mind through a JupyterHub infrastructure built on the PUNCH AAI (Authorisation and Authentication Infrastructure).