T 92: Electronics, Trigger, DAQ IV

Time: Friday 9:00–10:15 Location: KH 00.023

T 92.1 Fri 9:00 KH 00.023

Implementation of a two-level AI-enhanced trigger on a single chip with AI cores for live reconstruction — •Patrick Schwäbig for the Lohengrin-Collaboration — Physikalisches Institut, 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 approach of using fast triggers to discard uninteresting and irrelevant data remains 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. In contrast to this stands live track reconstruction for triggering, which was rarely possible due to limited computation power in the past. With the emergence of highly parallelized processors for AI inference, attempts to sufficiently accelerate tracking algorithms become viable. The AMD Versal Adaptive Compute Acceleration Platform (ACAP) is one such technology and combines FPGA and CPU resources with dedicated AI cores.

Our approach is to implement a two-level trigger on a single chip by utilizing the tightly integrated combination of FPGA and AI cores to profit from their individual strengths. In this talk our concept for a two-level trigger setup, implemented on an AMD VC1902, including classical and quantized AI algorithms, Timepix3 readout, as well as first testbeam results, will be shown. A similar setup 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 92.2 Fri 9:15 KH 00.023

Track Finding with Graph Neural Networks at the ATLAS Event Filter — •Giulia Fazzino, Sebastian Dittmeier, and André Schöning — Physikalisches Institut, Universität Heidelberg, Germany

The upcoming High Luminosity upgrade of the Large Hadron Collider will increase the number of simultaneous interactions per bunch-crossing in the ATLAS experiment from $\langle\mu\rangle \approx 56$ to $\langle\mu\rangle \approx 200$.

To cope with the computational demands resulting from the corresponding rise in data rate, the Trigger and Data Acquisition System of the experiment will undergo several upgrades. The trigger will consist of a hardware trigger and a software trigger, the Event Filter. In the latter, charged particle tracks in the Inner Tracker (ITk) will be reconstructed for event selection. To reduce the computational resources required for this task, the possibility of using hardware accelerators and new tracking algorithms has been extensively investigated over the last years.

One promising approach uses Graph Neural Networks (GNNs) for track finding, one of the most computationally expensive steps of track reconstruction. The algorithm first builds a graph from the hits in the ITk, then uses a GNN to score its edges, and lastly applies a segmentation procedure to generate track candidates. The high parallelizability of the method makes it suited for implementation on FPGAs or GPUs.

This talk will present an overview of GNN-based track finding for the ATLAS Event Filter, with a focus on its implementation and optimization for FPGA deployment.

T 92.3 Fri 9:30 KH 00.023

Graph Neural Network based Algorithms for the Belle II Upgrade of the Electromagnetic Calorimeter Trigger on Versal SoCs with integrated AI Engines — •Thomas Lobmaier, Isabel Haide, and Torben Ferber — Institute of Experimental Particle Physics, Karlsruhe Institute of Technology, Karlsruhe, Germany

In order to reach the target design luminosity of Belle II, the instantaneous luminosity has to be increased by adjustments of the SuperKEKB accelerator. Especially with the planned long shutdown II in 2030 a significant gain in luminosity is expected, which directly increases trigger rates.

Planned upgrades for the electromagnetic calorimeter electronics increase the readout granularity, which also allows the subsequent trigger algorithms to access higher granularity inputs. In addition to the associated better energy and position resolutions, this enables the utilization of shower shape information to reconstruct the event more accurately and improves background suppression.

We show a first implementation of a GNN on Versal SoCs with integrated AI Engines, which enables the processing of up to 256 inputs per event. We present the performance on datasets with different input reduction strategies for the Belle II Long Shutdown 2 upgrade.

T 92.4 Fri 9:45 KH 00.023

Low-latency GNN-based hit filtering for the Belle II Level-1 track trigger — •Greta Heine[1], Fabio Mayer[2], Marc Neu[2], Giacomo De Pietro[1], Jürgen Becker[2], and Torben Ferber[1] — [1]Institute of Experimental Particle Physics, Karlsruhe Institute of Technology, Karlsruhe, Germany — [2]Institute for Information Processing Technology, Karlsruhe Institute of Technology, Karlsruhe, Germany

The Belle II experiment encounters increasing beam-induced background with rising instantaneous luminosity, which places tighter requirements on online and offline tracking algorithms. The Level-1 trigger must maintain a high efficiency for physics signals while operating under strict latency and bandwidth constraints. To keep trigger rates within data acquisition limits, effective background suppression, in particular in the Central Drift Chamber (CDC), is essential.

This talk presents a Graph Neural Network (GNN)-based hit filter for the Belle II Level-1 CDC hardware trigger, with a focus on the hardware-aware design workflow for FPGA deployment, including model compression via quantization-aware training and network pruning. Performance results are presented in terms of the trade-off between filtering performance and resource utilization: hit-level background rejection, impact on subsequent track reconstruction and trigger rates, as well as resource usage and sub-microsecond latency for an AMD Ultrascale FPGA implementation that meets timing after place-and-route.

T 92.5 Fri 10:00 KH 00.023

AI-enabled FPGA trigger for autonomous radio detection of cosmic rays — Alperen Aksoy[3], Ilja Bekman[3], Markus Cristinziani[1], Eric-Teunis de Boone[1], •Vesselin Dimitrov[1], Qader Dorosti[1], Chimezie Eguzo[3], Stefan Heidbrink[2], Waldemar Stroh[2], Jens Winter[2], and Michael Ziolkowski[2] — [1]Experimentelle Astroteilchenphysik, Center for Particle Physics Siegen, Universität Siegen — [2]Elektronikentwicklungslabor des Departments Physik, Universität Siegen — [3]Peter Grünberg Institute - Integrated Computing Architectures, Forschungszentrum Jülich

Radio detection of extensive air showers induced by ultra-high-energy cosmic rays provides crucial information on their origin, composition and energy. Radio arrays detect these events, but cosmic-ray signals are exceedingly rare compared to the overwhelming radio noise and RFI. Since storing all data is not feasible, a trigger system must decide in real time which data to record. FPGAs are a fitting option for this requirement because they provide deterministic low latency and low energy consumption compared to CPUs or GPUs. In this collaborative work between the University of Siegen and the Forschungszentrum Jülich, we investigate a novel approach where a machine-learning-based trigger is implemented on an FPGA. The goal is to achieve reliable discrimination between cosmic-ray signals and background, by employing a quantized neural network with minimal latency and power consumption. We aim to validate the quantized model on an FPGA and assess resource usage, latency, power consumption and trigger efficiency, while strongly reducing the false-trigger rate for cosmic-ray events.