Mini-workshop on Real time tracking : triggering events with tracks

Report of Contributions

Contribution ID: **1** Type: **not specified**

Directions in Realtime Tracking : Everything, Everywhere, All at Once

Friday 13 October 2023 09:00 (30 minutes)

Presenter: HAHN, Kristian (Northwestern University (US))

Contribution ID: **2** Type: **not specified**

Level-1 Tracking at CMS for the HL-LHC

Friday 13 October 2023 09:30 (22 minutes)

The success of the CMS physics program at the HL-LHC requires maintaining sufficiently low trigger thresholds to select processes at the electroweak scale. With an average expected 200 pileup interactions, critical to achieve this goal while maintaining manageable trigger rates is in the inclusion of tracking in the L1 trigger. A 40 MHz silicon-tracker based track trigger on the scale of the CMS detector has never before been built; it is a novel handle, which in addition to maintaining trigger rates can enable entirely new physics studies.

The main challenges of reconstructing tracks in the L1 trigger are the large data throughput at 40 MHz and the need for a trigger decision within 12.5 μs out of which 4 μs is for track finding. The CMS outer tracker for HL-LHC uses modules with closely-spaced silicon sensors to read out only hits compatible with charged particles with pT above 2 GeV (”stubs”). These are used in the backend L1 track finding system, based on commercially available FPGA technology. The ever-increasing capability of modern FPGAs combined with their programming flexibility is ideal for a fast track finding algorithm. The L1 tracking algorithm forms track seeds (”tracklets”) from pairs of stubs in adjacent layers of the outer tracker. These seeds provide roads where consistent stubs are included to form track candidates. Track candidates sharing multiple stubs are combined prior to being fitted. A Kalman Filter track fitting algorithm is employed to identify the final track candidates and determine the track parameters. The system is divided into nine sectors in the r-phi plane, where the processing for each sector is performed by a dedicated track finding board.

This presentation will discuss the latest status of the CMS L1 track finding and its implementation, present simulation studies of the estimated performance, and discuss the developments of hardware demonstrators.

Co-author: FIORENDI, Sara (University of Tennessee (US))

Presenter: FIORENDI, Sara (University of Tennessee (US))

Contribution ID: **3** Type: **not specified**

A real-time demonstrator of track reconstruction with FPGAs at LHCb

Friday 13 October 2023 11:10 (22 minutes)

The upgraded LHCb detector has started its Run 3 of data taking in 2022, with a completely overhauled DAQ system, reading out and processing the full detector data at every LHC bunch crossing (30 MHz average rate). At the same, an intense R&D activity is taking place, with the aim of further improving the real-time data processing performance of LHCb, in view of a further luminosity upgrade of the experiment ("Upgrade II").
In this work, we describe the experience gained with a prototype device for a 30 MHz real-time tracking in the LHCb VELO detector, implemented in state-of-art PCIe-hosted FPGA cards interconnected by fast optical links.
The system is capable of processing live LHCb data opportunistically during physics data taking, thanks to a dedicated testbed facility fed by the experiment monitoring system. We describe, amongst other things, the system used to organize and optimize the high-speed distribution of data to the components, and the synchronization with the most updated alignment constants to be used in track reconstruction.

Primary authors: CONTU, Andrea (INFN); LAZZARI, Federico (Universita di Pisa & INFN Pisa (IT)); TERZUOLI, Francesco (Università di Siena & INFN Pisa (IT)); BASSI, Giovanni (SNS & INFN Pisa (IT)); PUNZI, Giovanni (Universita & INFN Pisa (IT)); TUCI, Giulia (Heidelberg University (DE)); MORELLO, Michael J. (SNS and INFN-Pisa (IT)); FANTECHI, Riccardo (INFN - Sezione di Pisa); KOTRIAKHOVA, Sofia (Universita e INFN, Ferrara (IT)); BALDINI, Wander (Universita e INFN, Ferrara (IT))

Presenter: TERZUOLI, Francesco (Università di Siena & INFN Pisa (IT))

Contribution ID: **4** Type: **not specified**

Real time event-level track pattern recognition with Graph Neural Networks in CMS

Detecting the signals of very low-pT muons with traditional track reconstruction algorithms, such as Kalman filters, is very challenging. In case of the decay of a tau lepton decaying into three muons, the signature includes three very low pT muons in the forward region of the CMS detector. Some or all of these muons might not carry enough pT to reach all stations of the CMS muon system. Even for muons reaching the CMS muon system, individual low momentum muon track reconstruction at hardware trigger level is prohibited with the traditional reconstruction algorithms, due to multiple scattering, the nonuniform magnetic field and large combinatorics especially in high pile up environment at the LHC/High-Luminosity LHC. An alternative approach is presented where a Graph Neural Network is trained to make use of the correlation between hits in the muon detectors to detect the presence of the τ -> 3 μ signature. The muon hits form the nodes of the graph and are connected by edges encoding their relative position. Based on this architecture a classifier is developed for use in the upgraded L1 trigger of the CMS detector for the HL-LHC. With this approach, a significant improvement of a factor of 5 to 10 in acceptance for τ -> 3 μ events is achieved compared to previous studies in CMS phase 2 muon system and Level-1 TDR. Support for GNNs designed in pytorch geometric is implemented into the hls4ml toolkit, which enables us to generate a FPGA implementation of the model for use in the L1 trigger.

Primary authors: SIMON, Benjamin (Purdue University (US)); SCHULTE, Jan-Frederik (Purdue University (US)); LIU, Miaoyuan (Purdue University (US))

Co-authors: GUERRERO, Daniel (Fermi National Accelerator Lab. (US)); YUN, Hyeon-Seo (Purdue University (US)); KONIGSBERG, Jacobo (University of Florida (US)); LI, Pan; MIAO, Siqi (Purdue University)

Presenter: LIU, Miaoyuan (Purdue University (US))

Contribution ID: **5** Type: **not specified**

Track reconstruction for the ATLAS Phase-II High-Level Trigger using Graph Neural Networks on FPGAs

Friday 13 October 2023 11:54 (22 minutes)

The High-Luminosity LHC (HL-LHC) will provide an order of magnitude increase in integrated luminosity and enhance the discovery reach for new phenomena. The increased pile-up foreseen during the HL-LHC necessitates major upgrades to the ATLAS detector and trigger. The Phase-II trigger will consist of two levels, a hardware-based Level-0 trigger and an Event Filter (EF) with tracking capabilities. Within the Trigger and Data Acquisition group, 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 modeling 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 and considerable speedup over CPU-based execution is possible on FPGAs.

The focus of this talk is a study of track reconstruction for the Phase-II EF system using GNNs on FPGAs. We explore each of the steps in a GNN-based EF tracking pipeline: graph construction, edge classification using an interaction network (IN), and track reconstruction. Several methods and hardware platforms are under evaluation, studying optimizations of the GNN approach aimed to minimize FPGA resources utilization and maximize throughput while retaining high track reconstruction efficiency and low fake rates required for the ATLAS Phase-II EF tracking system. These studies include IN model hyperparameter tuning, model pruning and quantization-aware training, and sequential processing of sub-graphs over the detector.

Primary author: GUPTA, Sachin (Ruprecht Karls Universitaet Heidelberg (DE))

Presenter: GUPTA, Sachin (Ruprecht Karls Universitaet Heidelberg (DE))

Contribution ID: **6** Type: **not specified**

Standalone track reconstruction and matching algorithms for the GPU-based High Level Trigger at LHCb

Friday 13 October 2023 09:52 (22 minutes)

The LHCb Upgrade in Run 3 has changed its trigger scheme for a full software selection in two steps. The first step, HLT1, will be entirely implemented on GPUs and run a fast selection aiming at reducing the visible collision rate from 30 MHz to 1 MHz.
This selection relies on a partial reconstruction of the event. A version of this reconstruction starts with two monolithic tracking algorithms, the VELO-pixel tracking and the HybridSeeding on Scintillating-Fiber tracker, which reconstructs track segments in standalone subdetectors. Those segments are then matched through a matching algorithm in order to produce 'long'tracks, which form the base of the HLT1 reconstruction. We discuss the principle of these algorithms as well as the details of their implementation which allows them to run at a high-throughput configuration. An emphasis is put on the optimizations of the algorithms themselves in order to take advantage of the GPU architecture. Finally, results are presented in the context of the LHCb performance requirements for Run 3.

Co-authors: AGAPOPOULOU, Christina (CERN); HENRY, Louis (EPFL - Ecole Polytechnique Federale Lausanne (CH))

Presenter: HENRY, Louis (EPFL - Ecole Polytechnique Federale Lausanne (CH))

Contribution ID: **7** Type: **not specified**

Trigger Level Tracking With Neural Networks on Heterogeneous Computing Systems

Friday 13 October 2023 11:32 (22 minutes)

The high luminosity upgrade of the LHC aims to better probe the higgs potential and self coupling. The Event Filter task force has been charged with exploring novel approaches to charged particle tracking to be employed in the upgraded ATLAS trigger system, capable of analyzing high luminosity events in real time. We present a neural network (NN) based approach to predicting and identifying hits left by particles at trigger level. In this bottom-up approach, the complexity of and input to the NN are kept minimal to allow the NN to be implemented in a heterogeneous computing system, such as with FPGA or GPU. This hardware based approach allows for increased data throughput, shorter latency and, crucially, flexibility to improve the algorithm in the future.

Primary author: GEKOW, Alex (Ohio State University (US))

Presenter: GEKOW, Alex (Ohio State University (US))

Contribution ID: **8** Type: **not specified**

FPGA-based architecture for a real-time track reconstruction in the LHCb Scintillating Fibre Tracker beyond Run 3

Friday 13 October 2023 10:15 (22 minutes)

Finding track segments downstream of the magnet is some of the most important and computationally expensive task of the first stage of the new GPU-based software trigger of the LHCb Upgrade I, that has started operation in Run 3. These segments are essential to form all good physics tracks with a very high precision momentum measurement, when combined with those reconstructed in the vertex track detector, and to reconstruct long-lived particles, such as Kshort and strange baryons, decaying after the vertex track detector, largely boosting the physics reach of the experiment. In this talk, we discuss the collaboration plans to install a real-time tracking device based on distributed system of FPGAs, dedicated to the reconstruction of particles trajectories in the forward Scintillating Fibre tracker detector, with the aim to preserve the full physics potential of the experiment in Run 4, and in view of Run 5 (Upgrade II) at higher instantaneous luminosity. This system will enhance the DAQ system of the experiment, and will run in real time during physics data taking, reconstructing tracks on-the-fly at the LHC collision rate, before the software trigger processing begins. The design and the expected performance of this device, with the capability of processing events at the full LHC collision rate of 30 MHz is discussed.

Primary authors: CONTU, Andrea (INFN); XU, Ao (Universita & INFN Pisa (IT)); DE OYANGUREN CAMPOS, Arantza (Univ. of Valencia and CSIC (ES)); Mr JASHAL, Brij Kishor (IFIC, Univ. of Valencia, CSIC (ES) and Tata Institute of Fundamental Research (TIFR)); MENDOZA, Diego (Instituto de Física Corpuscular - CERN); LAZZARI, Federico (Universita di Pisa & INFN Pisa (IT)); TERZUOLI, Francesco (Università di Siena & INFN Pisa (IT)); BASSI, Giovanni (SNS & INFN Pisa (IT)); PUNZI, Giovanni (Universita & INFN Pisa (IT)); TUCI, Giulia (Heidelberg University (DE)); ZHUO, Jiahui (Univ. of Valencia and CSIC (ES)); HE, Jibo (University of Chinese Academy of Sciences (CN)); PICA, Lorenzo (SNS & INFN Pisa (IT)); MARTINELLI, Maurizio (Universita & INFN, Milano-Bicocca (IT)); MORELLO, Michael J. (SNS and INFN-Pisa (IT)); SHI, Qi (University of Chinese Academy of Sciences (CN)); FANTECHI, Riccardo (Universita & INFN Pisa (IT)); KOTRIAKHOVA, Sofia (Universita e INFN, Ferrara (IT)); BALDINI, Wander (Universita e INFN, Ferrara (IT))

Presenter: MORELLO, Michael J. (SNS and INFN-Pisa (IT))

Contribution ID: **9** Type: **not specified**

Studies on track finding algorithms based on machine learning with GPU and FPGA

Friday 13 October 2023 12:16 (22 minutes)

Track finding in high-density environments is a key challenge for experiments at modern accelerators. In this presentation we describe the performance obtained running machine learning models studied for the ATLAS Muon High Level Trigger. These models are designed for hit position reconstruction and track pattern recognition with a tracking detector, on a commercially available Xilinx FPGA: Alveo U50, Alveo U250, and Versal VCK5000. We compare the inference times obtained on a CPU, on a GPU and on the FPGA cards. These tests are done using TensorFlow libraries as well as the TensorRT framework, and software frameworks for AI-based applications acceleration. The inference times obtained are compared to the needs of present and future experiments at LHC.

Presenter: DI BELLO, Francesco Armando (INFN e Universita Genova (IT))

Contribution ID: **10** Type: **not specified**

Accelerated Large Graph Creation with CUDA: Empowering Graph Neural Networks in Particle Physics

Graph neural networks have emerged as a powerful tool in various physics studies, particularly in the analysis of sparse and heterogeneous data. However, as the field of particle physics advances towards utilizing graphs in high-luminosity scenarios, a new challenge has emerged: efficient graph creation. While GNN inference is highly optimized, graph creation has not received the same level of attention. This becomes a significant bottleneck when working with graphs that have a large number of nodes such as those arising from HL-LHC collisions.
In this talk, we will delve into the specifics of this issue and explore the impact of graph creation on performance. We will highlight that even for graphs with a specific structure in mind, the process of creating them can be several orders of magnitude slower on a CPU compared to running inference with a network. To overcome this limitation, we will investigate how hardware-based acceleration, specifically utilizing CUDA on GPUs, can address these challenges.
By leveraging CUDA acceleration, we can unlock the full potential of graph neural networks without being hindered by inefficient graph creation. This talk aims to highlight the importance of fast and efficient graph creation in the context of large-scale graphs in particle physics and how leveraging hardware acceleration can propel research in this exciting field.

Co-authors: SHIELDS, Edward; KAKATI, Nilotpal (Weizmann Institute of Science (IL))

Presenter: KAKATI, Nilotpal (Weizmann Institute of Science (IL))