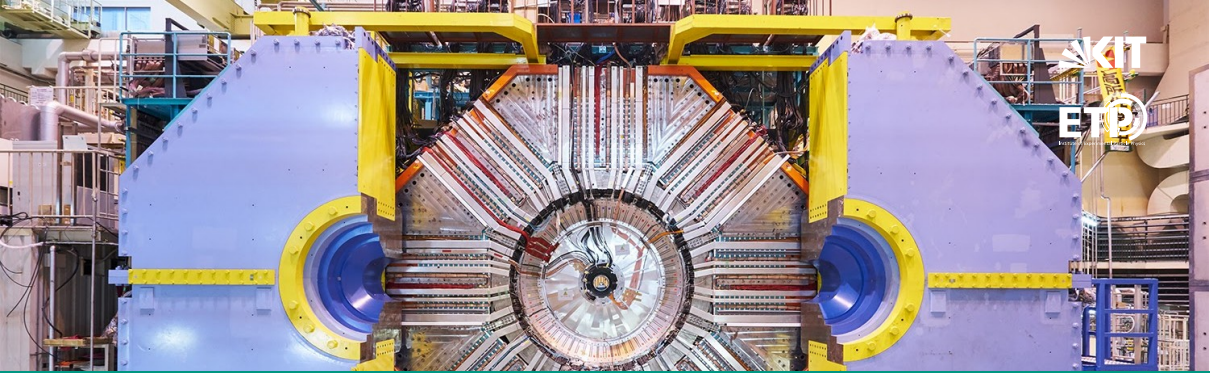


A Real-Time GNN-based Clustering Algorithm for the Level 1 Calorimeter Trigger at Belle II

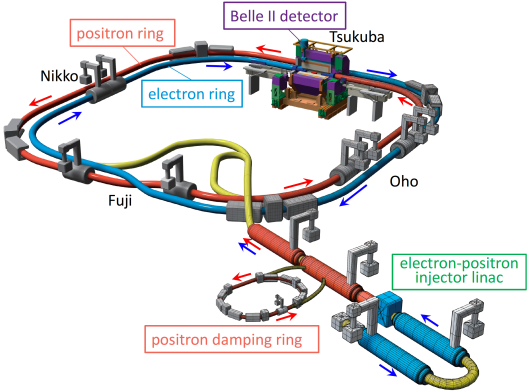
Fast Machine Learning for Science Conference 2025, 5. September 2025

Isabel Haide* (isabel.haide@kit.edu), Frank Baptist*, Jürgen Becker+, Torben Ferber*, Thomas Lobmaier*, Marc Neu+

*Institute of Experimental Particle Physics (ETP), +Institut für Technik der Informationsverarbeitung (ITIV)

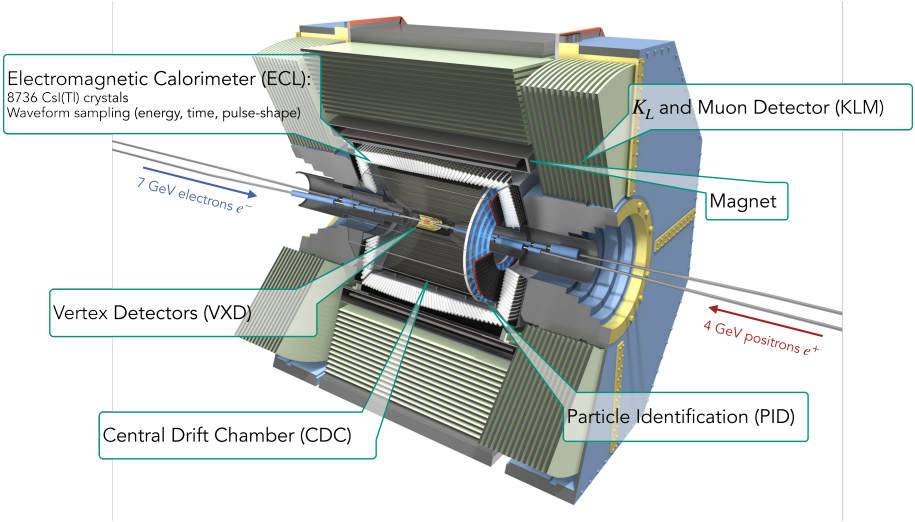
The Belle II Level 1 Trigger

- Belle II is a high precision particle physics experiment located at the SuperKEKB electron-positron collider
- SuperKEKB operates at a bunch crossing frequency of **250 MHz**
- Level 1 (L1) trigger has to select interesting events with a maximum rate of **30 kHz**
- Detector readout frequency is **8 MHz** due to decay times of electromagnetic calorimeter crystals $\rightarrow$ throughput requirement for the L1 trigger
- Detector buffer size allows a total L1 latency of **5 μs** from signal readout to trigger decision

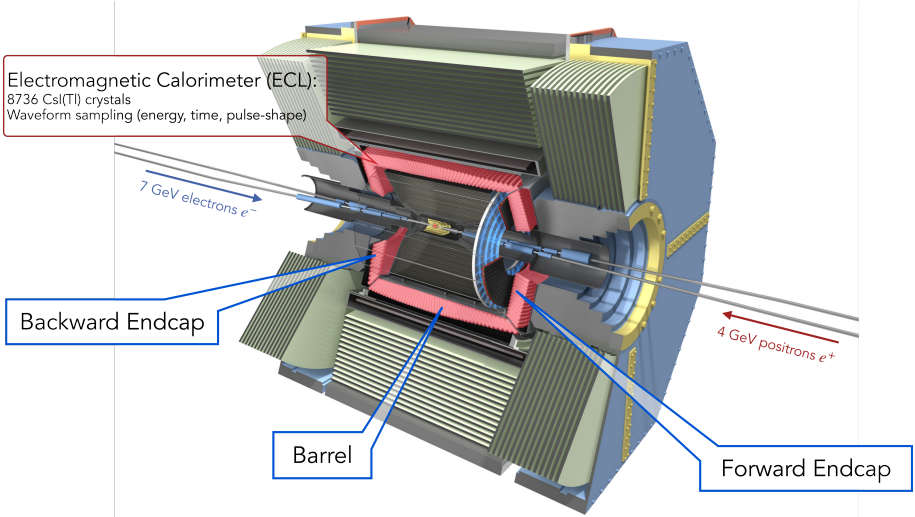


$\Rightarrow$ **Strict throughput and latency requirements, high filter rate**

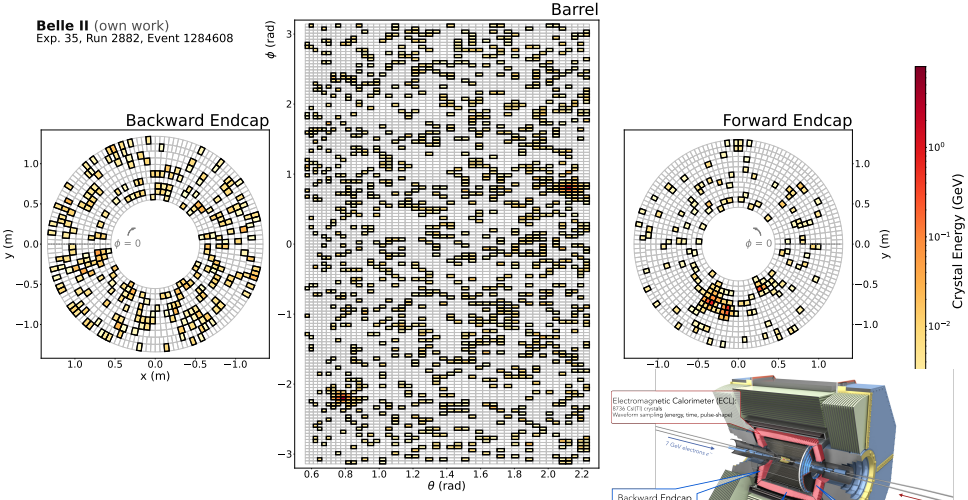
The Belle II Detector



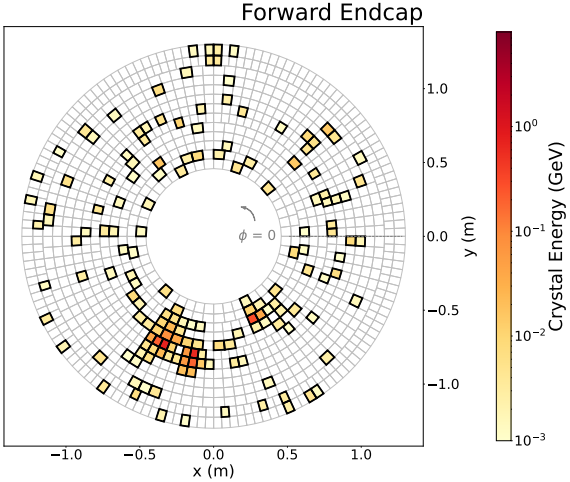
The Belle II Detector - ECL



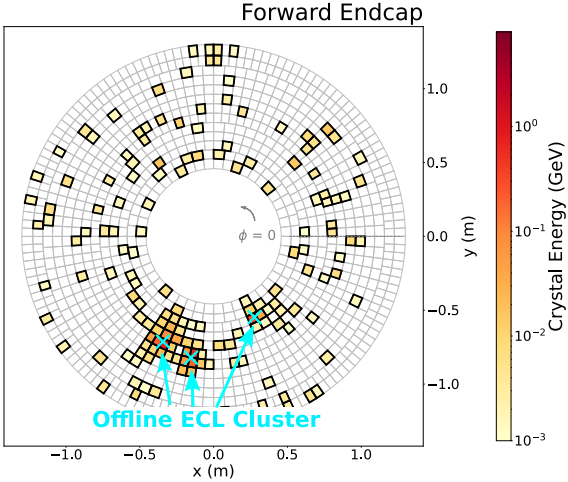
The Belle II Calorimeter - Crystal Readout



The Belle II Calorimeter - Crystal Readout

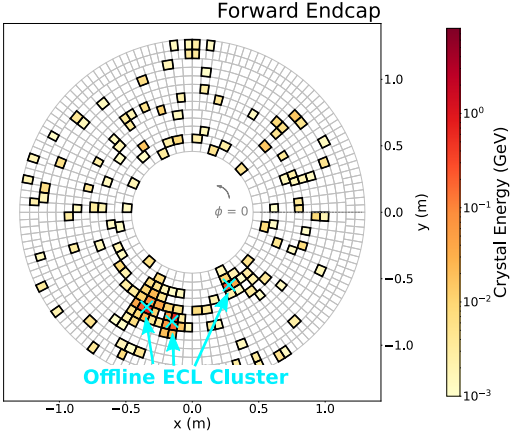


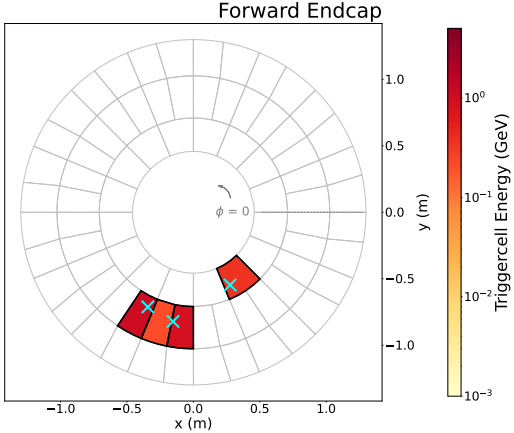
The Belle II Calorimeter - Offline ECL Cluster



Offline = reconstructed using the full Belle II software framework

The Belle II Calorimeter - Crystals to Triggercells

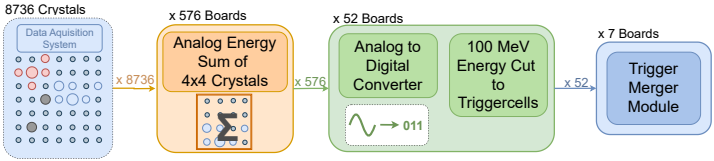


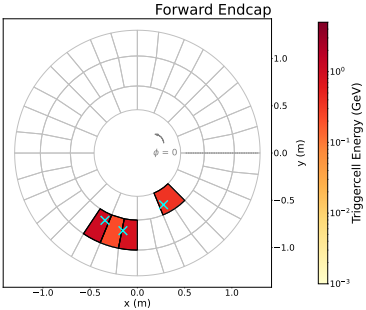


1. Sum energy of 16 crystals into 1 triggercell

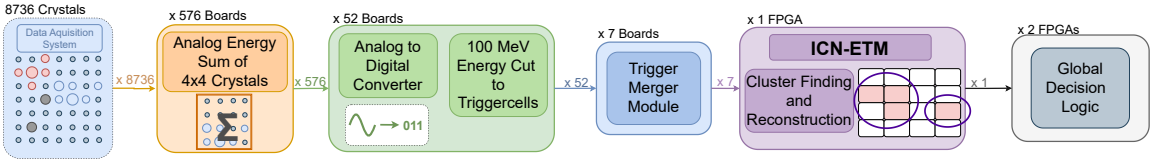
2. Apply 100 MeV energy threshold to triggercells

The Current ECL Trigger: Inputs



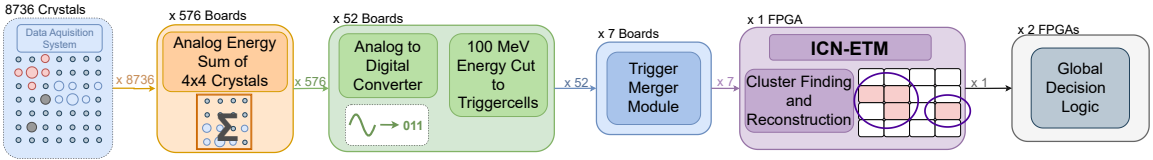


The Current ECL Trigger: ICN-ETM

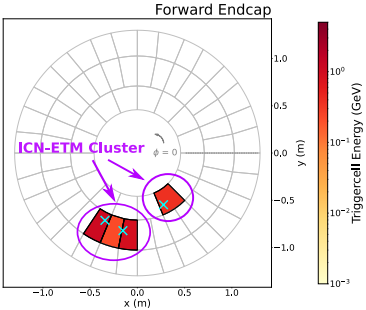


ICN: Number of Isolated Clusters
ETM: ECL Trigger Module

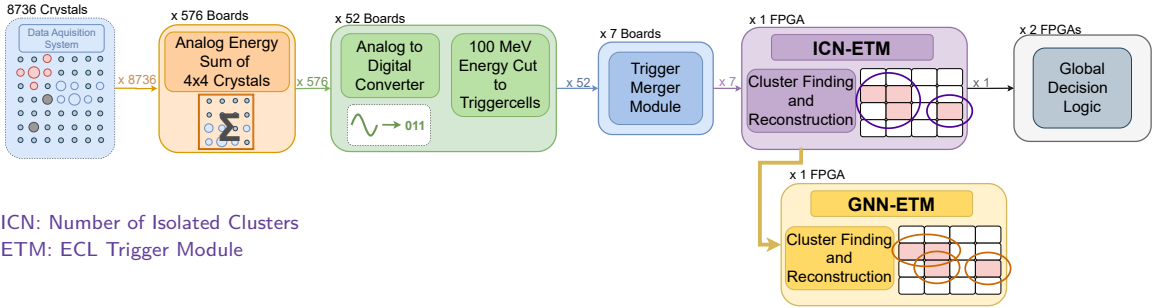
The Current ECL Trigger: ICN-ETM



ICN: Number of Isolated Clusters
ETM: ECL Trigger Module

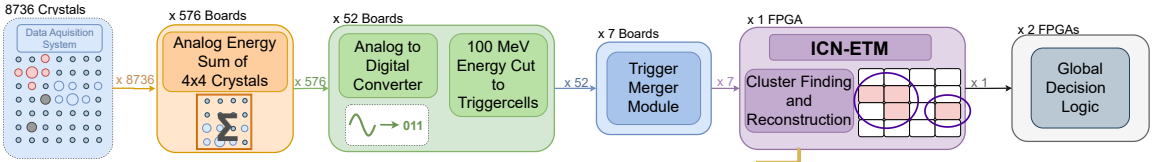


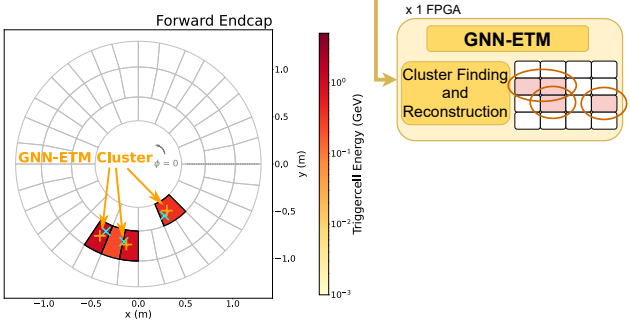
The New ECL Trigger: GNN-ETM



ICN: Number of Isolated Clusters
ETM: ECL Trigger Module

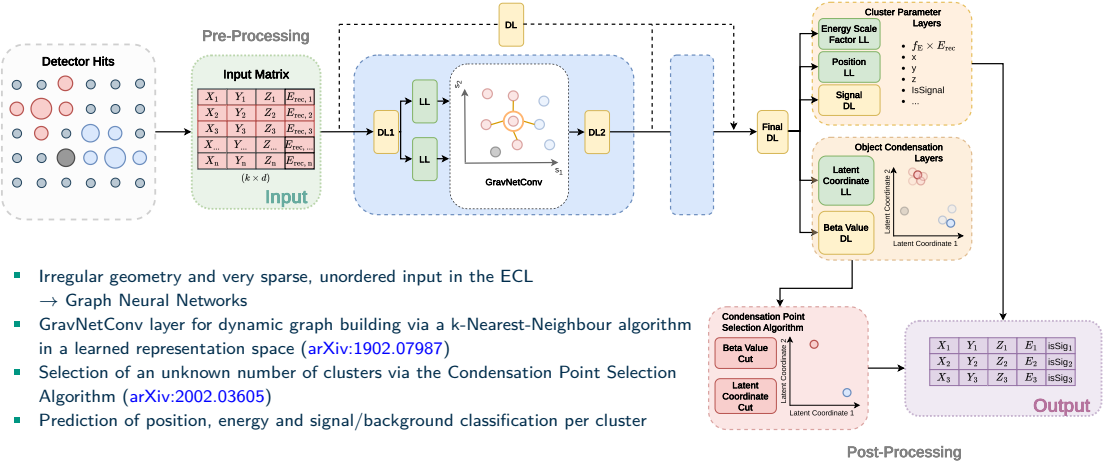
The New ECL Trigger: GNN-ETM





ICN: Number of Isolated Clusters
ETM: ECL Trigger Module

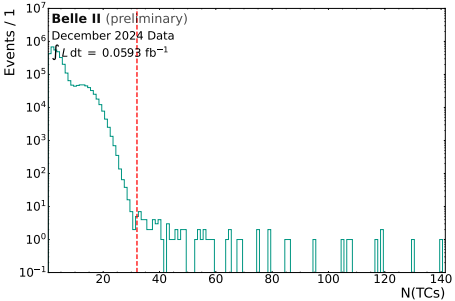
Network Design of GNN ECL Trigger Algorithm



- Irregular geometry and very sparse, unordered input in the ECL
  $\rightarrow$ Graph Neural Networks
- GravNetConv layer for dynamic graph building via a k-Nearest-Neighbour algorithm in a learned representation space (arXiv:1902.07987)
- Selection of an unknown number of clusters via the Condensation Point Selection Algorithm (arXiv:2002.03605)
- Prediction of position, energy and signal/background classification per cluster

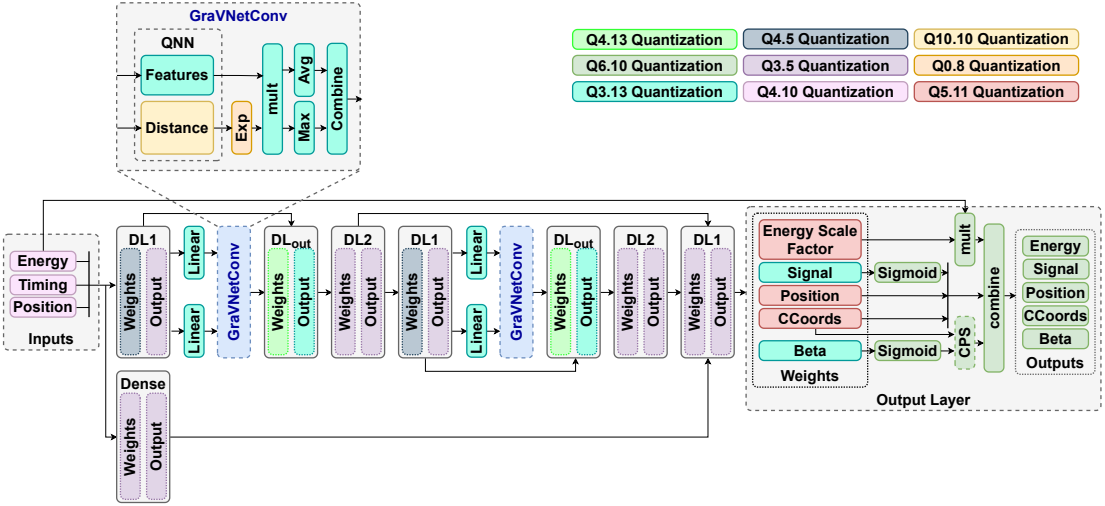
Optimizations towards a Deployment on FPGA

- 32 maximum input TCs to reach required throughput
- Quantization-aware training using `QKeras` with mixed quantization
- Replacing resource-intensive Euclidean distance with Manhattan distance
- Low magnitude-pruning with 40% pruning
- Post-training quantizations for exponential-based calculations in distance weighting and Sigmoid function



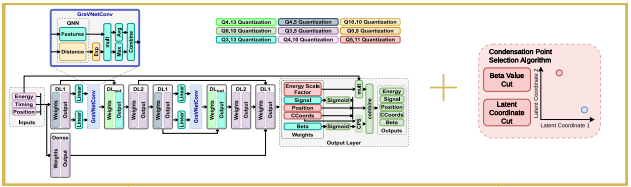
$\Rightarrow$ **Final model has 15 layers and 4800 parameters**

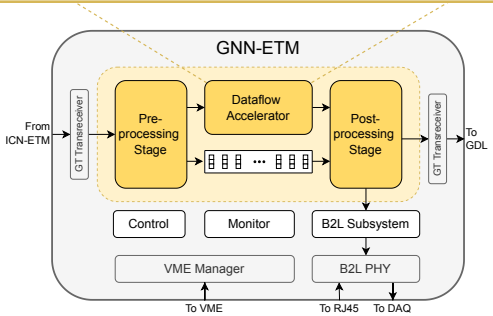
CaloClusterNet Design



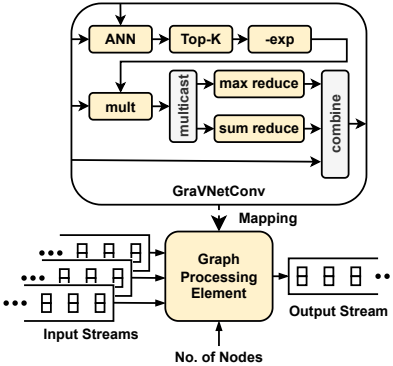
The GNN-ETM

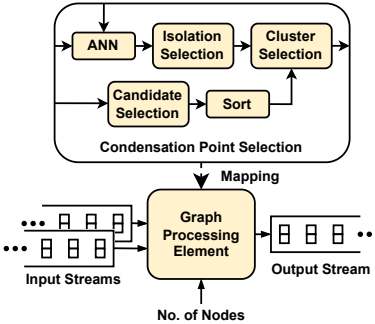
- Templated-based deployment of CaloClusterNet as dataflow accelerator
- Dynamic Graph Operators
  - GraVNetConv
  - Condensation Point Selection
- Preprocessing Stage
  - Sparsity compression for sparse input TCs
  - Event time calibration to calculate time relative to highest-energetic TC per event





Template-Based Design of Custom Architectures

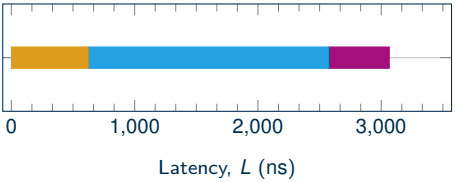


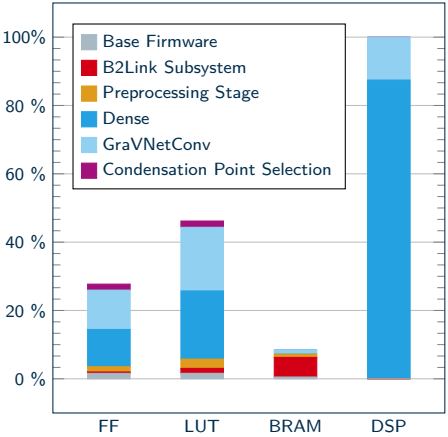


- We follow a template-based deployment approach (see arXiv:2507.05099), similar to FINN or hls4ml
- We design templates for **stall-free deep-pipelined** hardware modules for modules such as the GravNetConv block or the Condensation Point Selection

GNN-ETM Evaluation

- The CaloClusterNet including pre- and postprocessing is deployed on the Universal Trigger Board 4 (AMD Ultrascale XCVU 190).
- The CaloClusterNet is running at f_{sys} = 127 MHz, pre- and postprocessing at 254 MHz
- The design achieves the target throughput of 8 MHz and latency of less than 3.1 μs
- Parasitically included in full Belle II L1 trigger system during collision data taking in December 2024

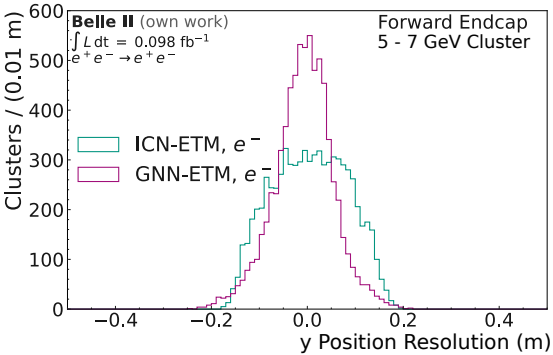




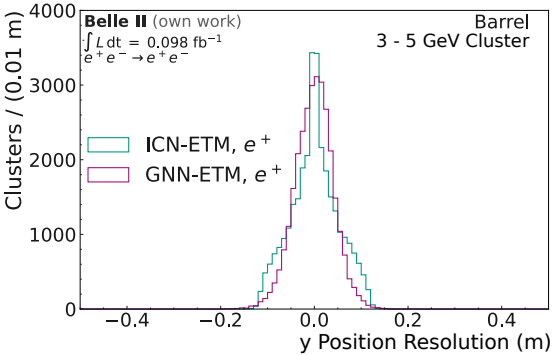
Position Resolution for $e^+e^- \to e^+e^-$

New evaluation of GNN-ETM data with improved weights:

Electron Cluster in Forward Endcap



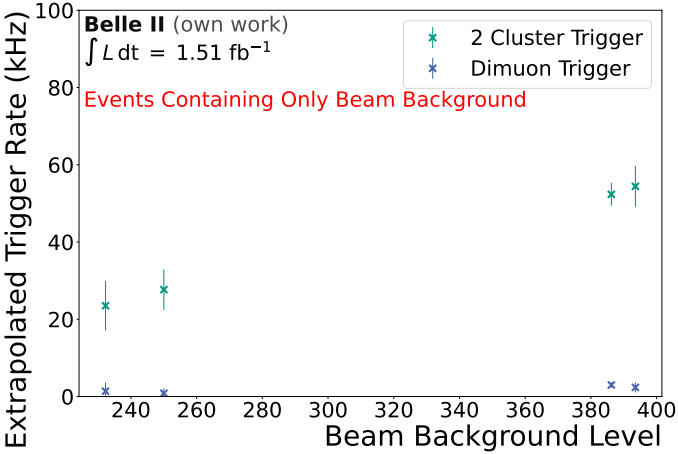
Positron Cluster in Barrel



$\to$ Position resolution for high-energy clusters improved over the full detector

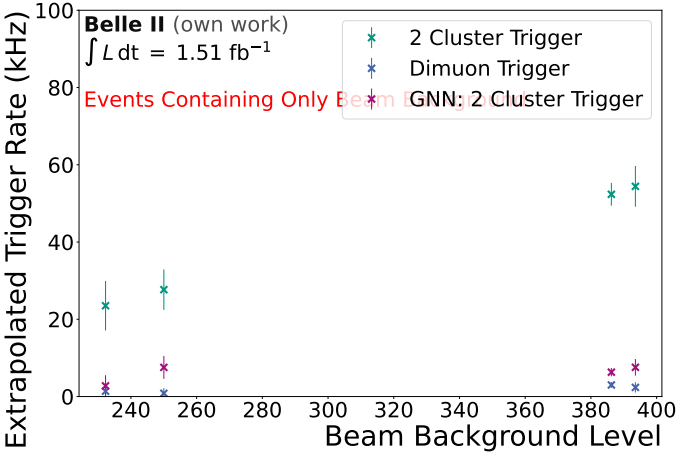
Trigger Rate In High Beam Backgrounds

- Events triggered by random trigger
  → Only beam background in these events
- Low-bias trigger bits, e.g. 2-Cluster bit, have very high rates due to beam background
  → Will need to be downscaled
- Improving beam background trigger rate with GNN predictions and signal/background classifier
  → No need for downscaling



Trigger Rate In High Beam Backgrounds

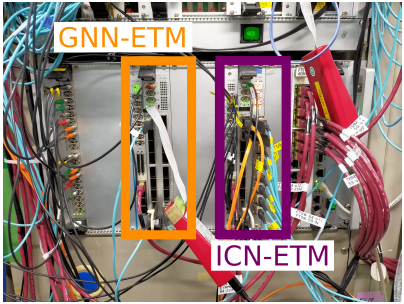
- Events triggered by random trigger
  → Only beam background in these events
- Low-bias trigger bits, e.g. 2-Cluster bit, have very high rates due to beam background
  → Will need to be downscaled
- Improving beam background trigger rate with GNN predictions and signal/background classifier
  → No need for downscaling



Summary

Summary:

- We have designed and implemented a GNN-based clustering algorithm for the Belle II ECL trigger
- This is the **first GNN-based reconstruction algorithm** running in a real-time environment at a particle physics experiment
- The GNN-ETM algorithm improves **position resolution, cluster finding efficiency, and separation of close-by clusters**
- A signal/background classification of trigger clusters reduces the **trigger rate on background events**

Outlook:

- Further reduction of network size and reduced quantization for inclusion into L1 trigger with active trigger decisions ongoing with minimal performance loss
- Full inclusion in the upcoming Belle II data taking in autumn
- Work ongoing towards implementation on boards with Versal AI Engines and different calorimeter granularities in the context of future Belle II upgrades