Karlsruhe Institute of Technology

Institute of Experimental Particle Physics

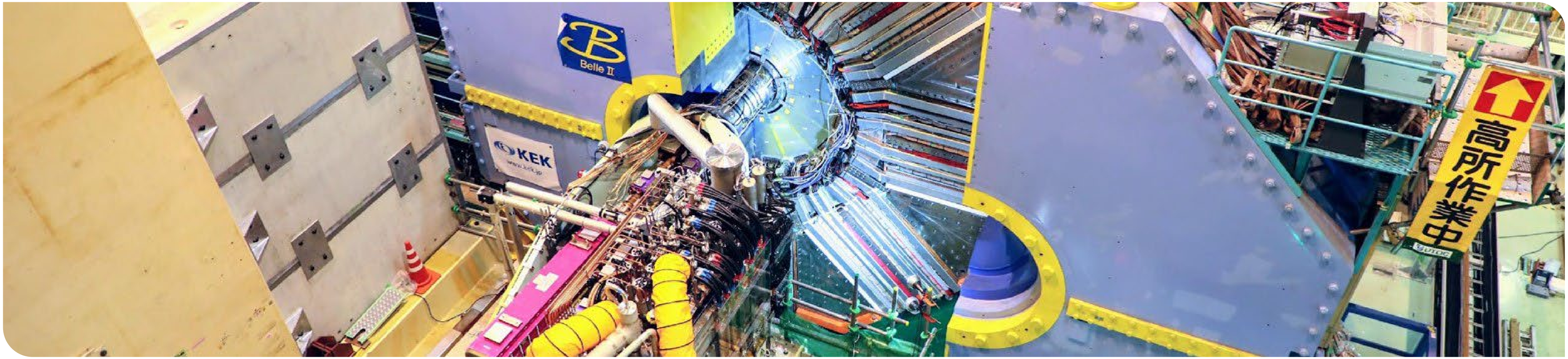
Belle II

Graph-Building and Input Feature Analysis for Edge Classification in the Central Drift Chamber at Belle II

ETP Meeting

Philipp Dorwarth, Torben Ferber, Lea Reuter, Slavomira Stefkova | 05th June 2023

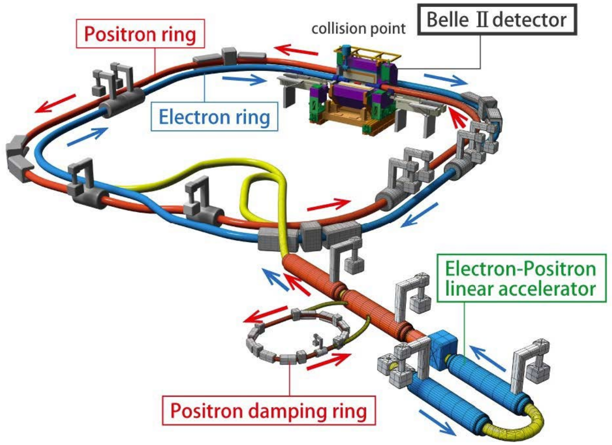
Philipp.Dorwarth@student.kit.edu

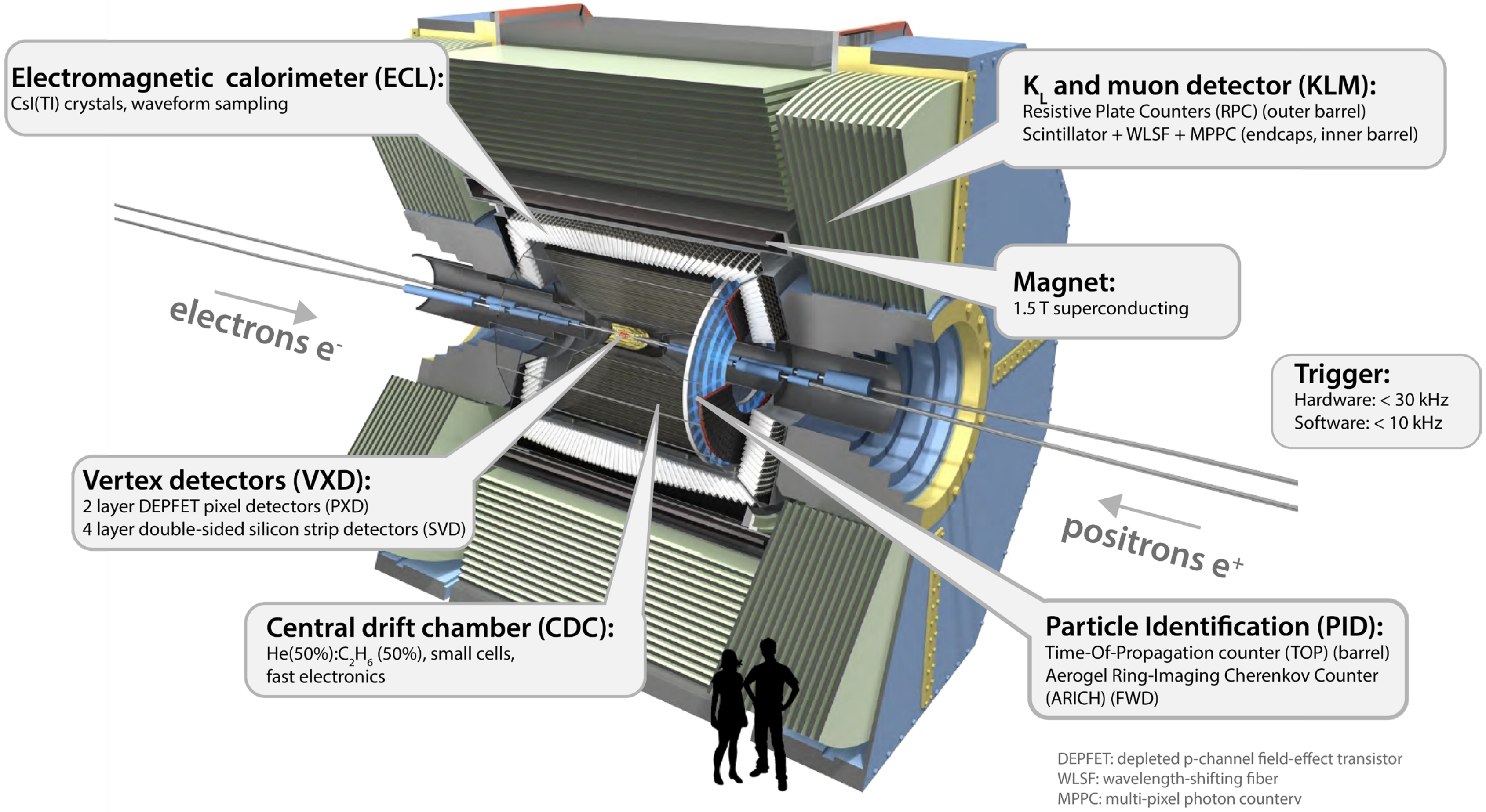


The Belle II Experiment

SuperKEKB

- Asymmetric e^+e^- collider
- $\Upsilon(4S)$ resonance $\rightarrow$ B-factory



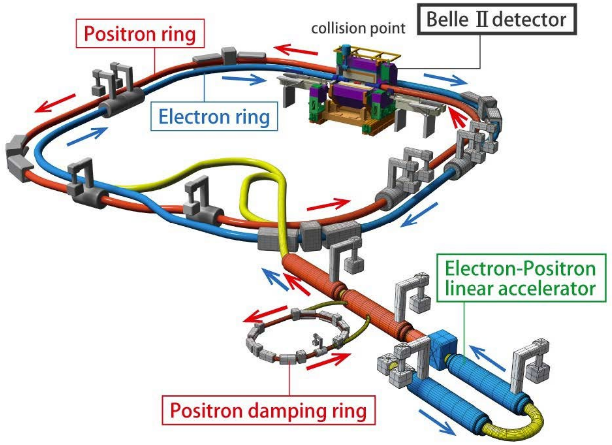


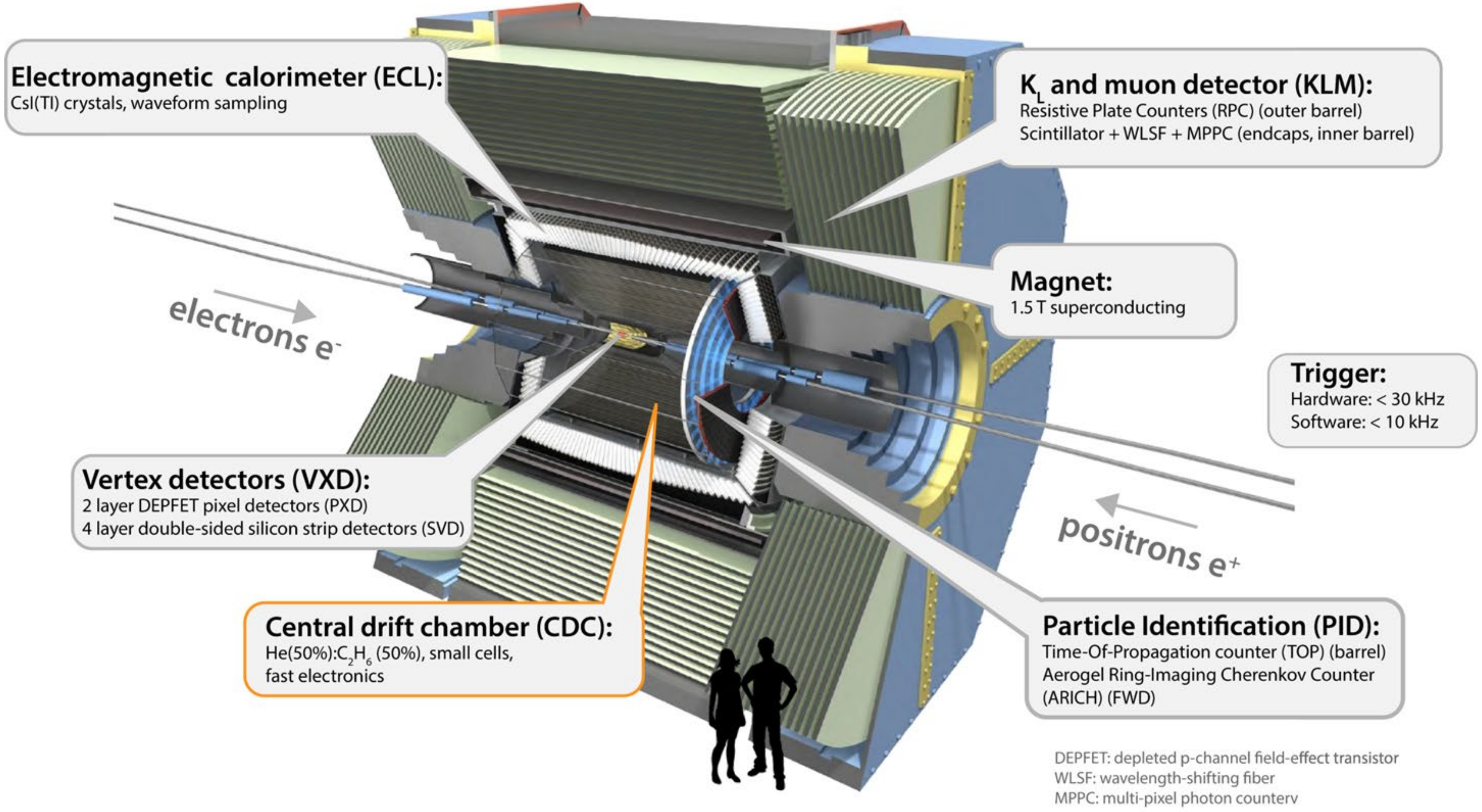
Karlsruhe Institute of Technology

The Belle II Experiment

SuperKEKB

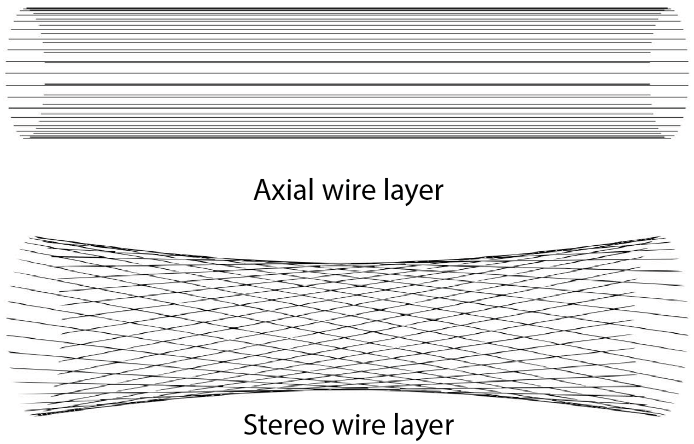
- Asymmetric e^+e^- collider
- $\Upsilon(4S)$ resonance $\rightarrow$ B-factory

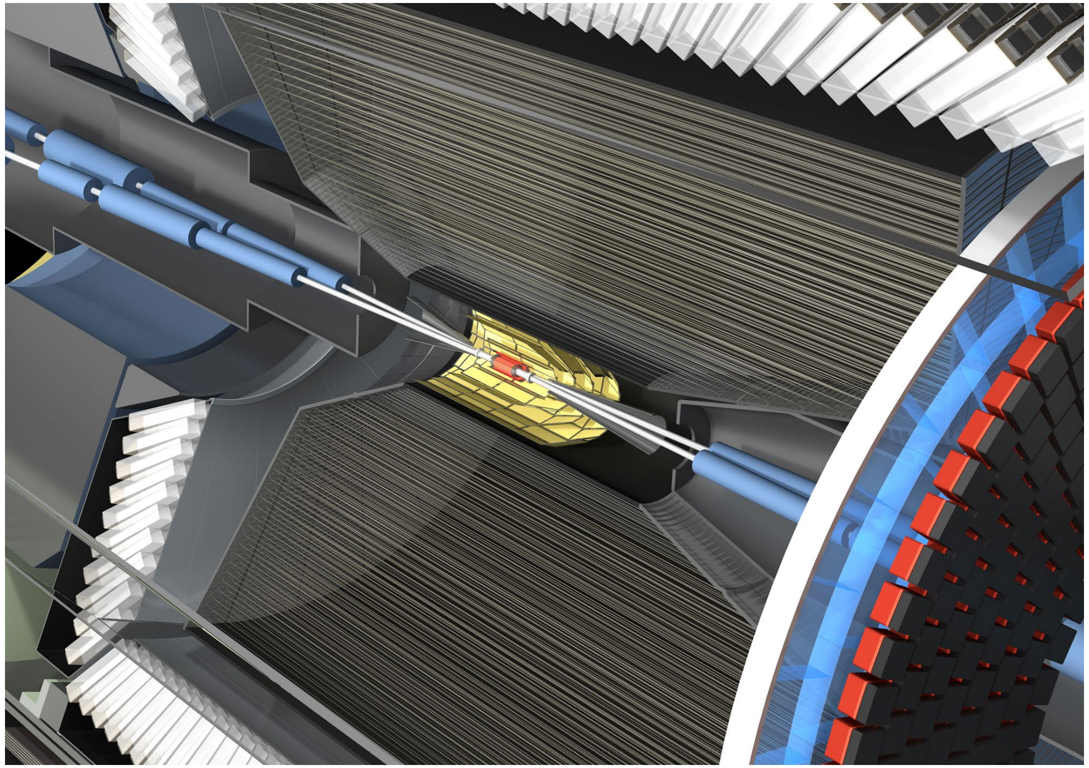




The Central Drift Chamber (CDC) of Belle II

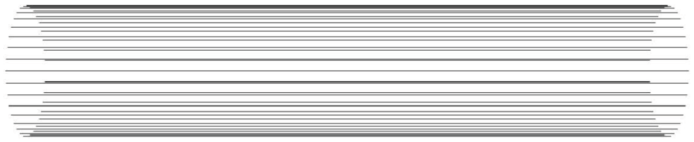
- The L1 trigger system uses the following information from the CDC to identify relevant events.
  - *Hit information*
  - *ADC Count:*
    Pulse height information, referring to the deposited energy
  - *TDC Count*
    Timing information



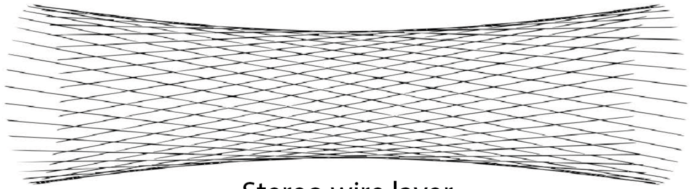


The Central Drift Chamber (CDC) of Belle II

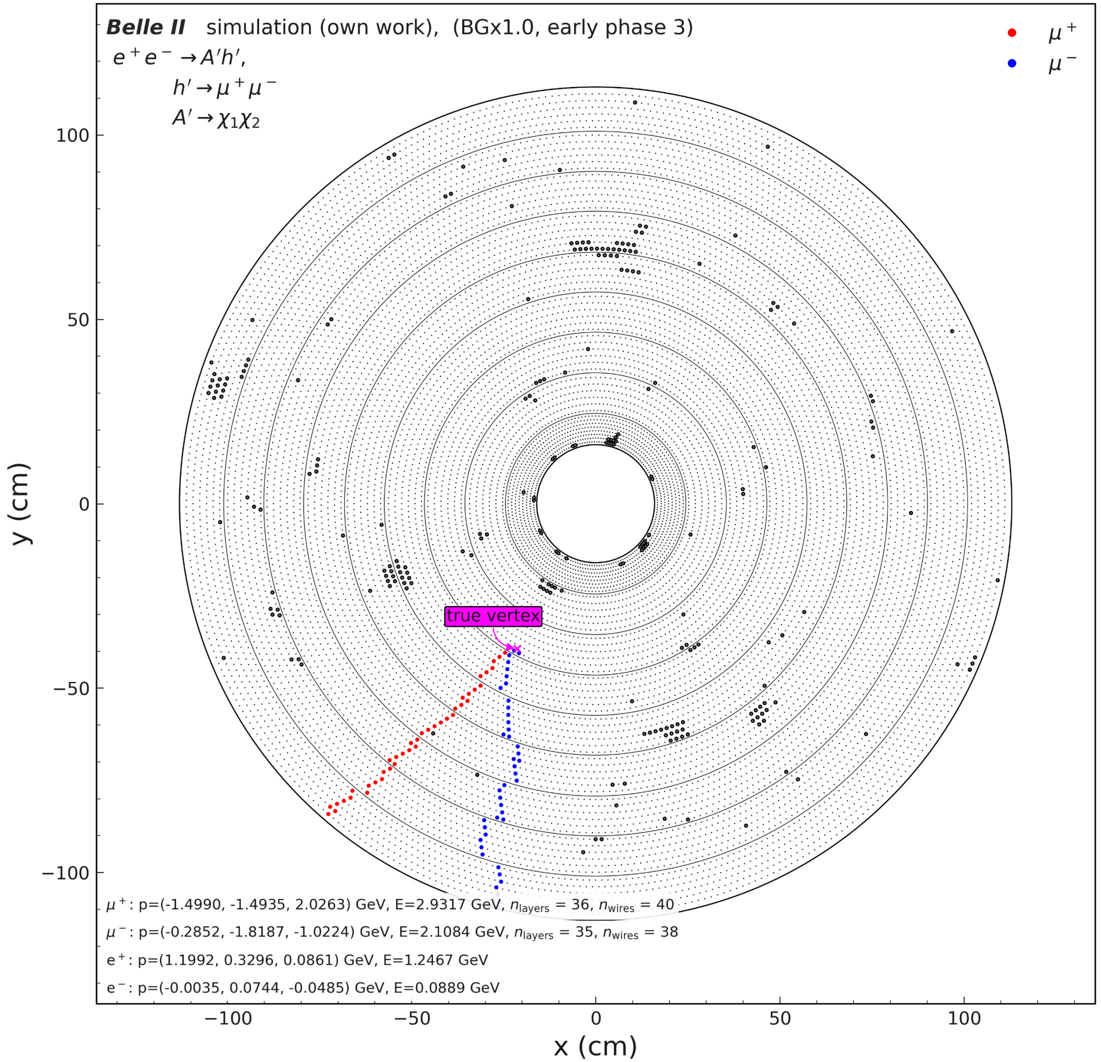
- The L1 trigger system uses the following information from the CDC to identify relevant events.
  - *Hit information*
  - *ADC Count:*
    Pulse height information, referring to the deposited energy
  - *TDC Count*
    Timing information



Axial wire layer

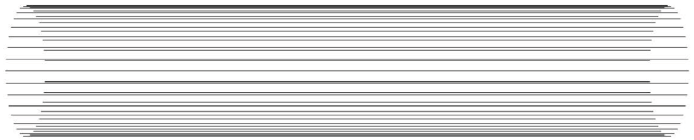


Stereo wire layer

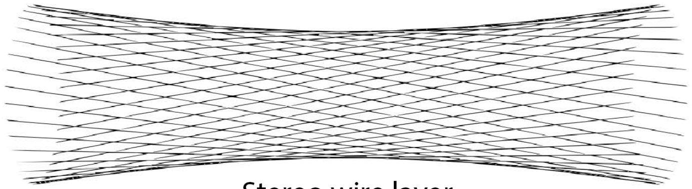


The Central Drift Chamber (CDC) of Belle II

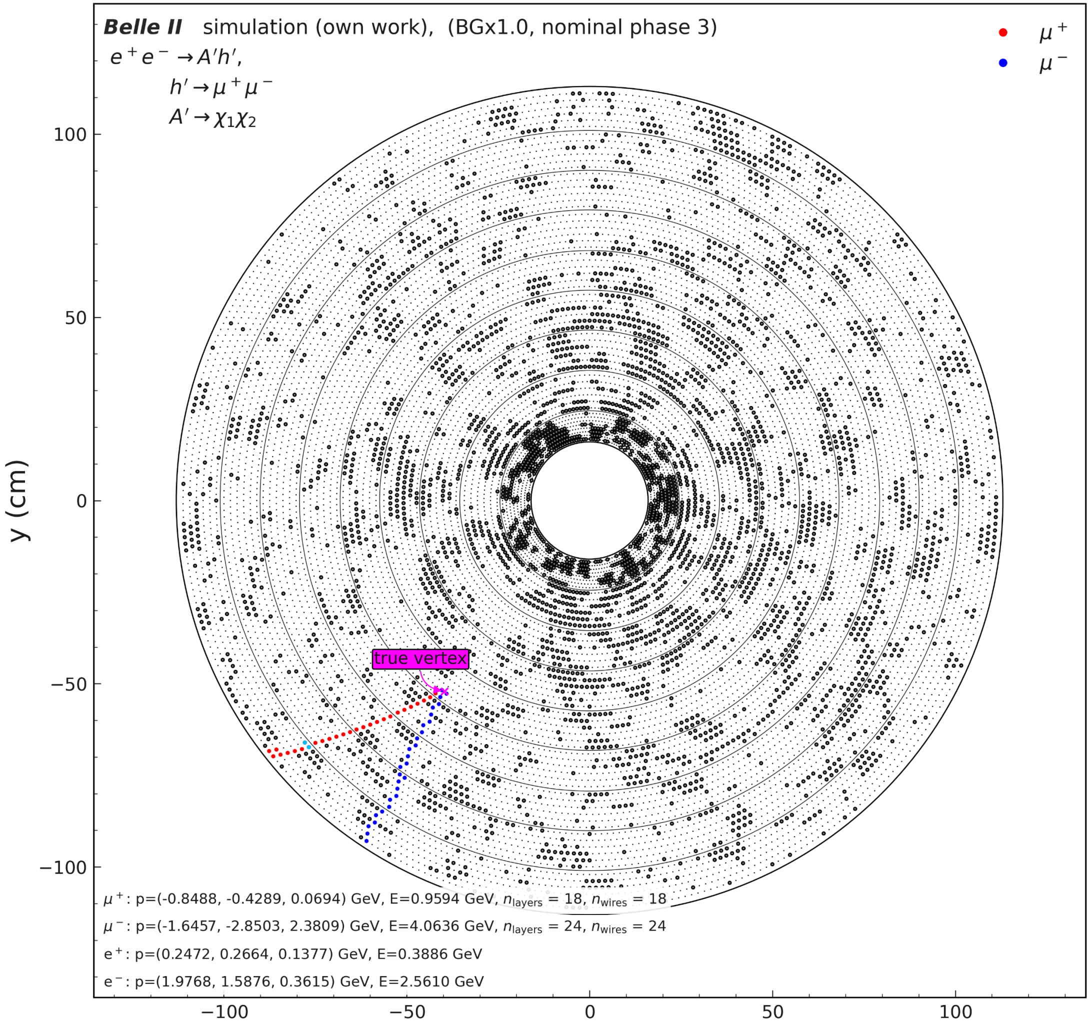
- The L1 trigger system uses the following information from the CDC to identify relevant events.
  - *Hit information*
  - *ADC Count:*
    Pulse height information, referring to the deposited energy
  - *TDC Count*
    Timing information



Axial wire layer



Stereo wire layer

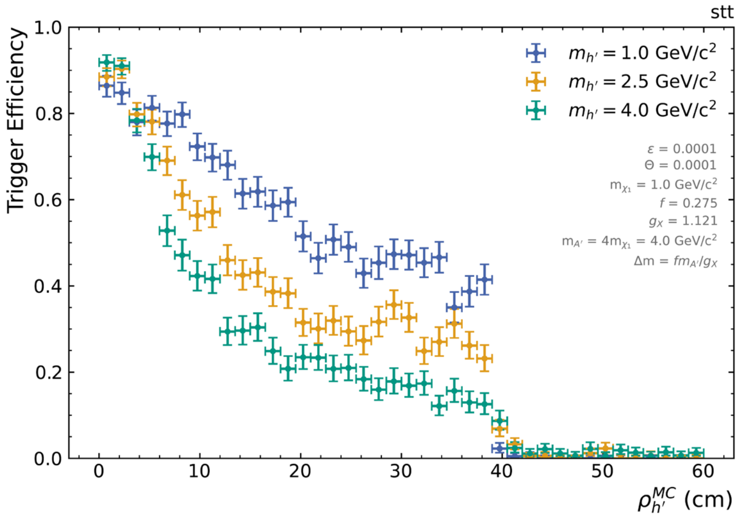


Motivation

Karlsruhe Institute of Technology

Searches for displaced vertices

- Displaced vertices are an important signature in searches for new physics[1]

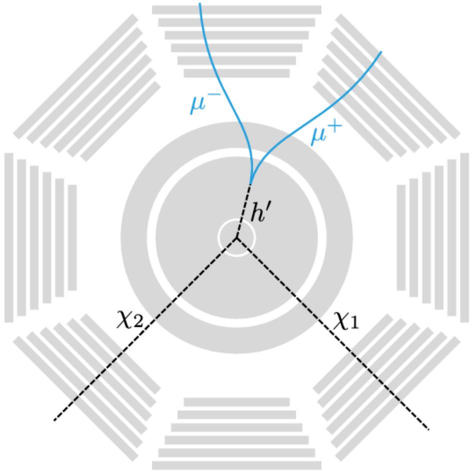


1 Long-lived Dark Higgs and Inelastic Dark Matter at Belle II (arXiv:2012.08595)

Single displaced vertex examples

Signal decay with dark photon A' and dark Higgs h' :

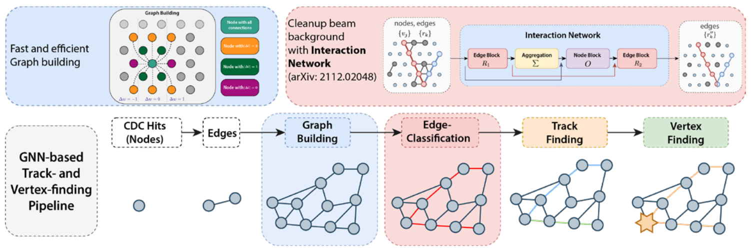
- $e^+e^- \rightarrow A'h'$
- $h' \rightarrow \mu^+\mu^-$,
- $A' \rightarrow \chi_1\chi_2$
- $\chi_2 \rightarrow \chi_1 e^+e^-$ (outside of CDC)



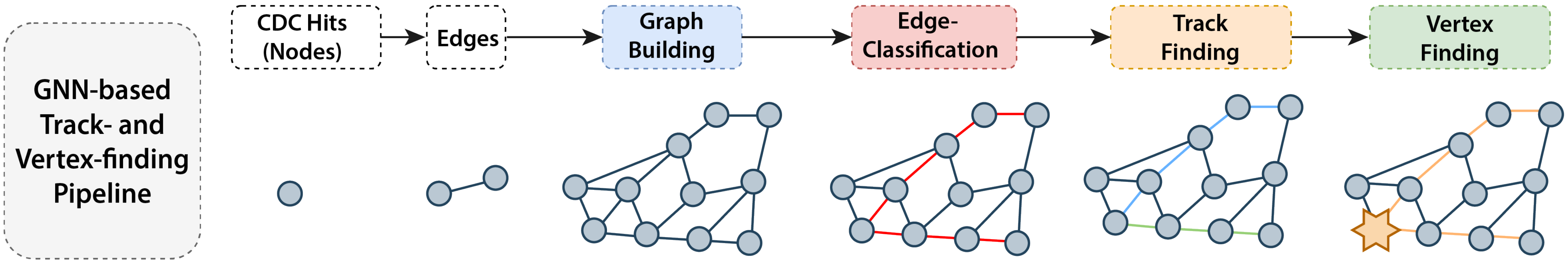
Project Goal:

Improve Track and Vertex Finding using Graph Neural Networks (GNN)

- Online (real-time, $< 2\ \mu s$)
- Offline



GNN-Based Track- and Vertex-Finding Pipeline

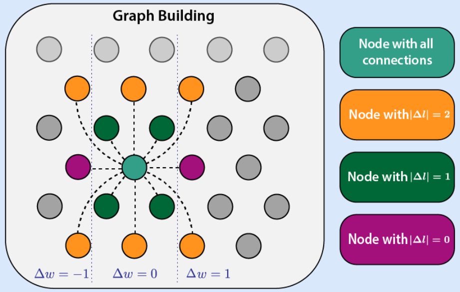


- Incorporate signal hits and signal edges in the graph-building
- Identify signal hits to reduce the input for the track finding
- Limited computing resources for online application demand small neural networks

GNN-Based Track- and Vertex-Finding Pipeline

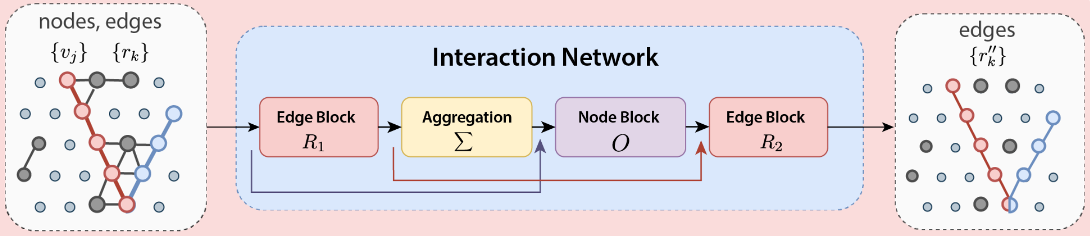
Focus of this presentation

Fast and efficient Graph building

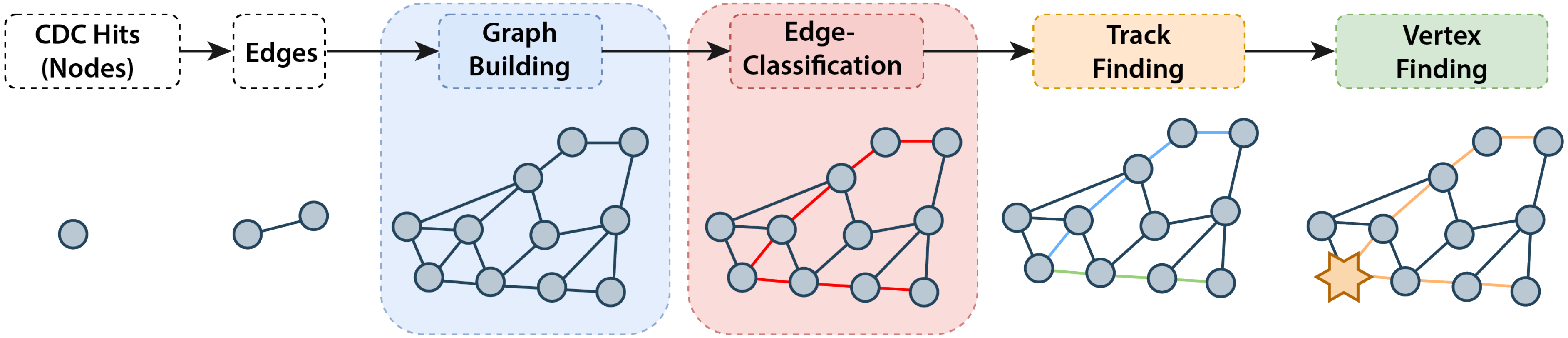


Focus of this presentation

Cleanup beam background with **Interaction Network** (arXiv: 2112.02048)



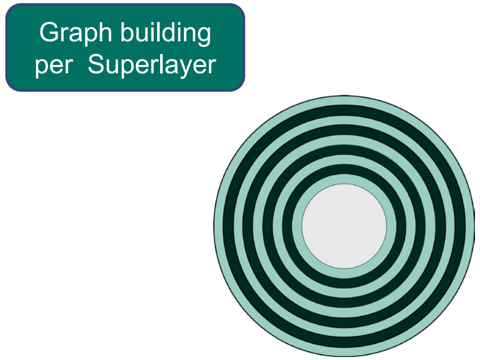
GNN-based Track- and Vertex-finding Pipeline

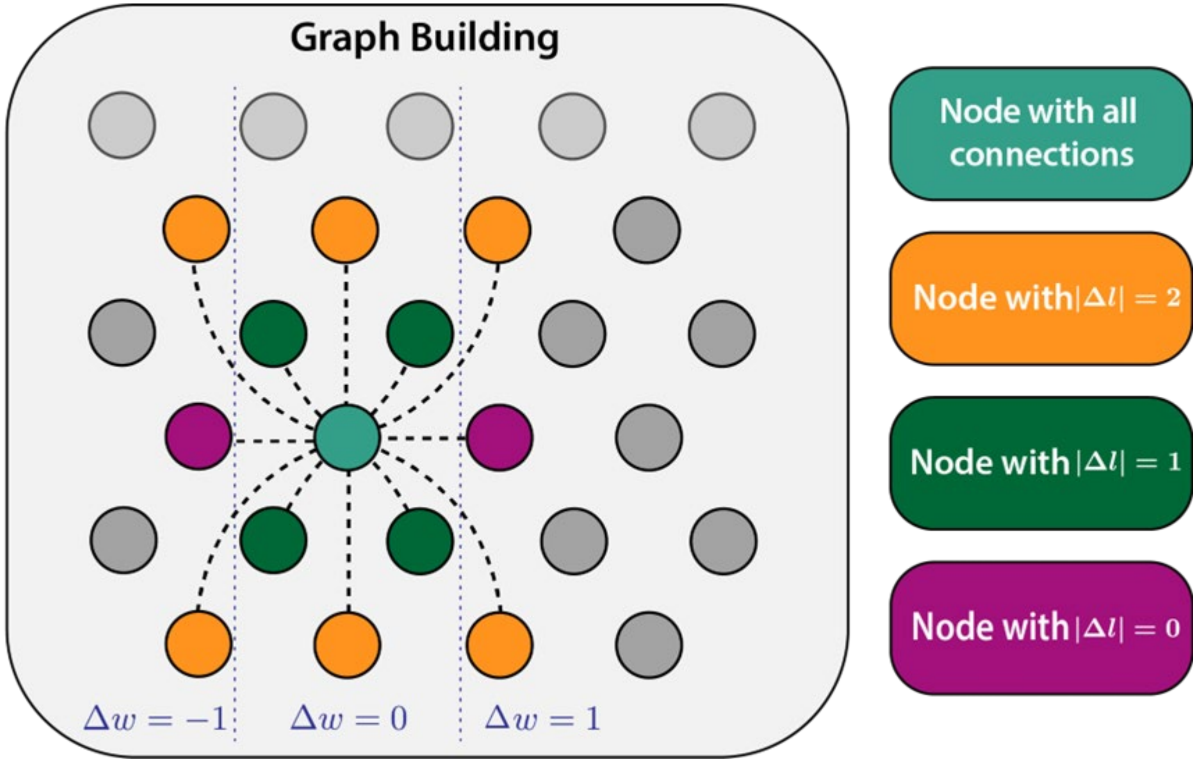


- Incorporate signal hits and signal edges in the graph-building
- Identify signal hits to reduce the input for the track finding
- Limited computing resources for online application demand small neural networks

Graph-Building - Approach

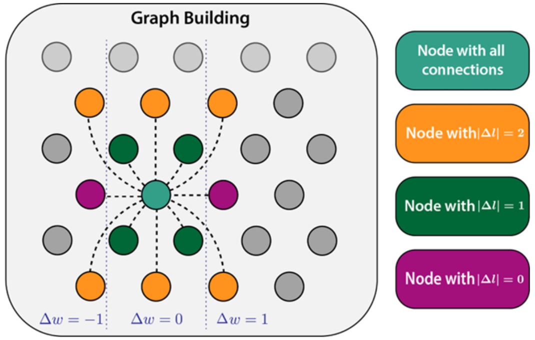
- Graphs are built for each superlayer using layer pairs
- Layer pairs are considered if within a distance of $\Delta l = 0, 1, 2$
- Each node of a layer is compared with all nodes in the corresponding layer
- Edges are built between nodes that meet defined distance criteria



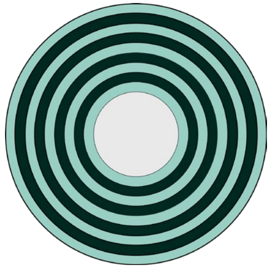


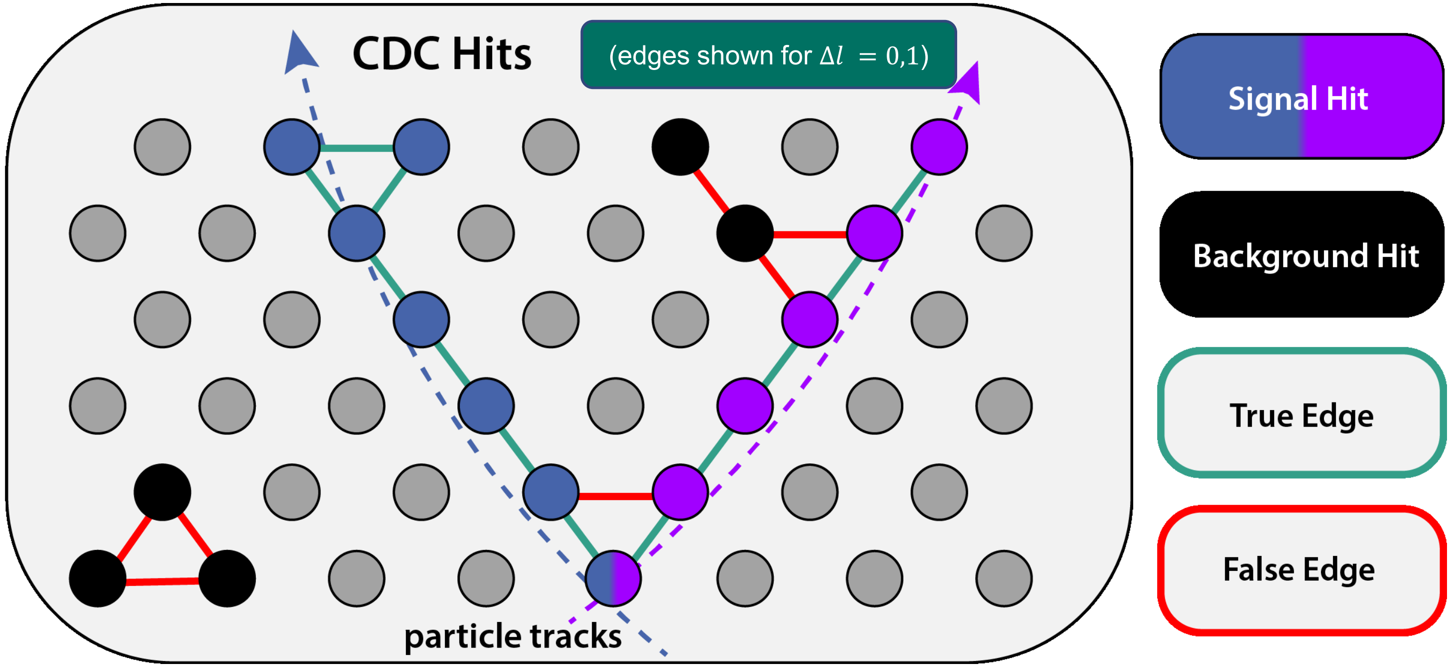
Graph-Building - Approach

- Graphs are built for each superlayer using layer pairs
- Layer pairs are considered if within a distance of
  $\Delta l = 0, 1, 2$
- Each node of a layer is compared with all nodes in the corresponding layer
- Edges are built between nodes that meet defined distance criteria



Graph building per Superlayer

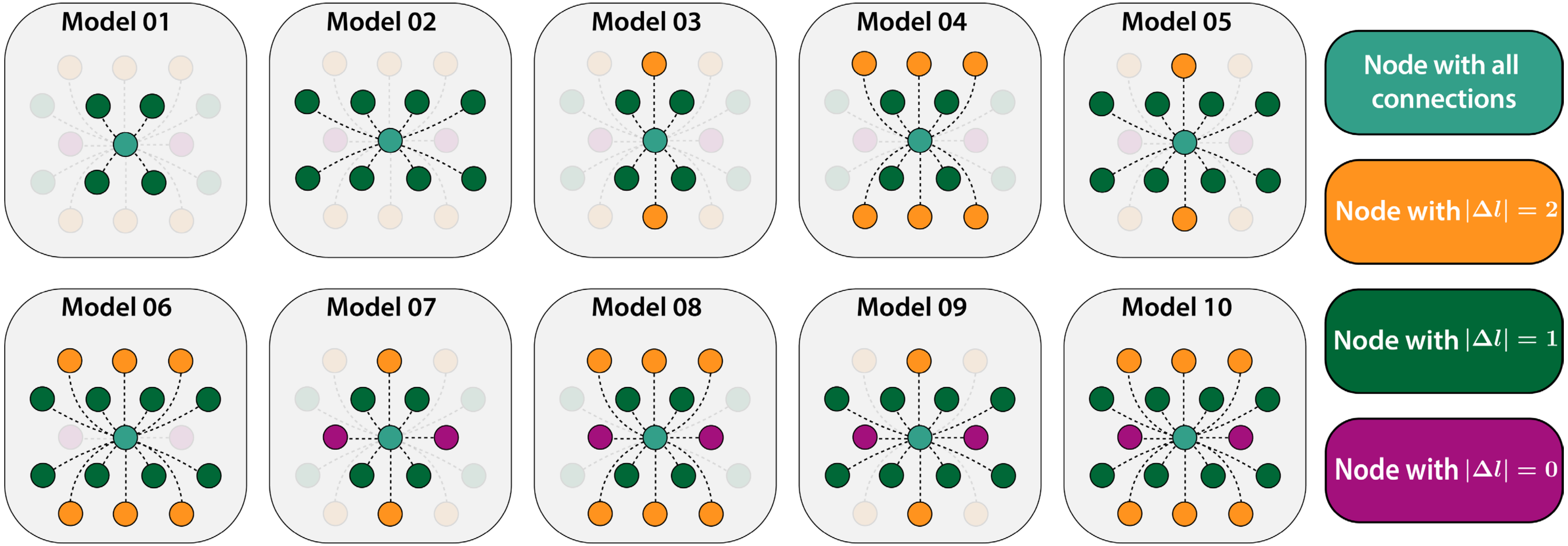




Graph-Building – Define Patterns

Find the best suiting pattern for the *GNN-pipeline.*

Different patterns need to be investigated to understand importance of connections in different wire and layer distances.

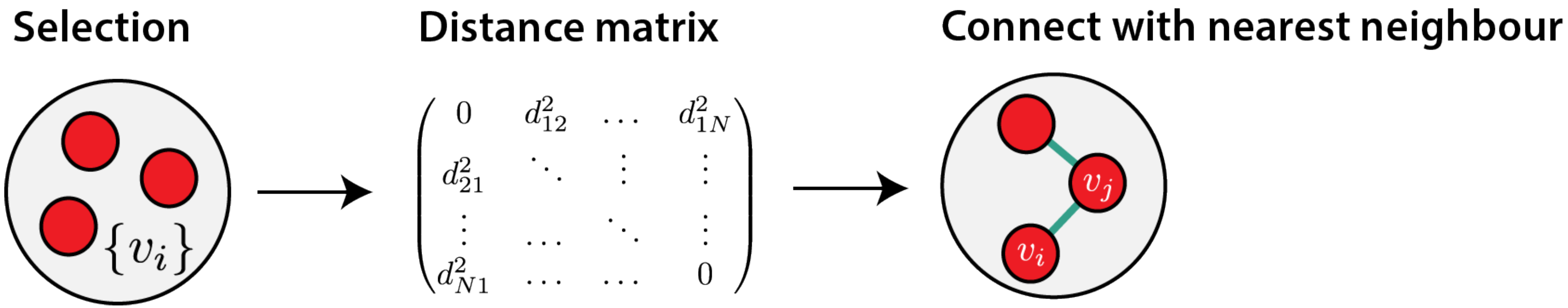


Graph-Building – The True Graph, a Set of True Edges

Karlsruhe Institute of Technology

True graph edges are defined from the sequences of CDC-hits originating from signal particles

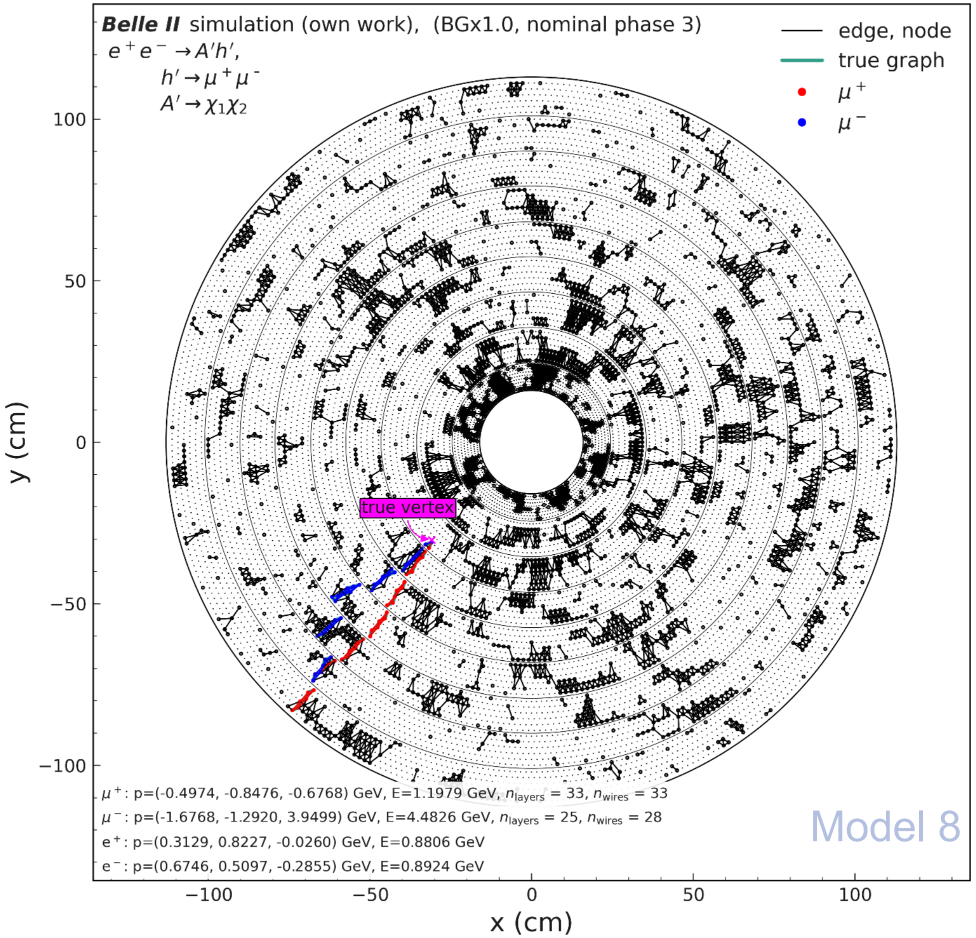
- In this analysis, an approximation for the hit sequence is used
  - CDC-hits are expected to be neighboring
  - Tracks are expected to be mainly outwards going



$\{v_i\}$: Nodes originating from the same signal Monte Carlo particle

Graph-Building – Evaluation

Eventdisplay of the CDC with graph

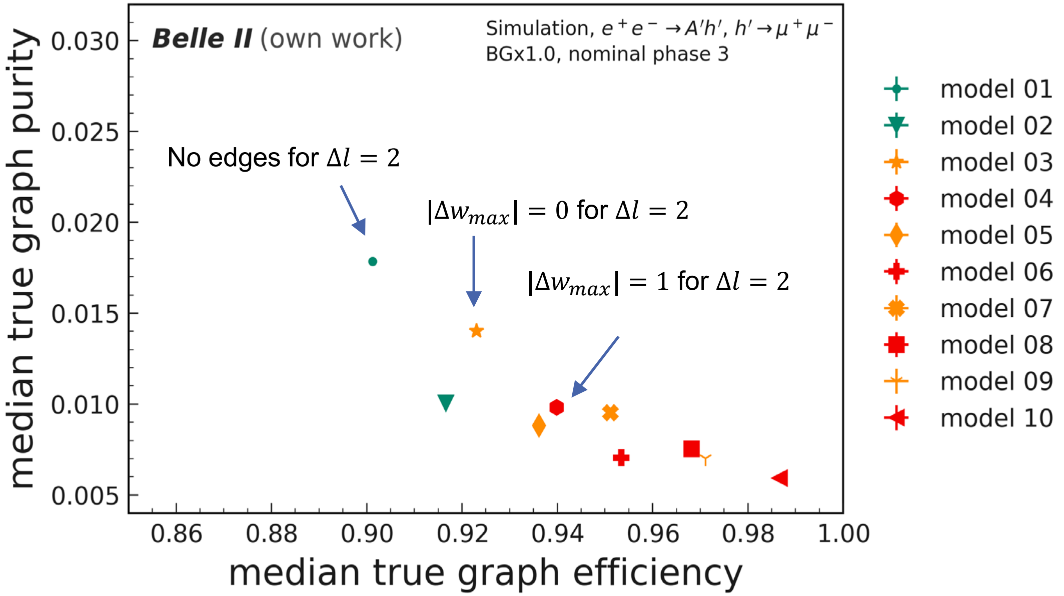


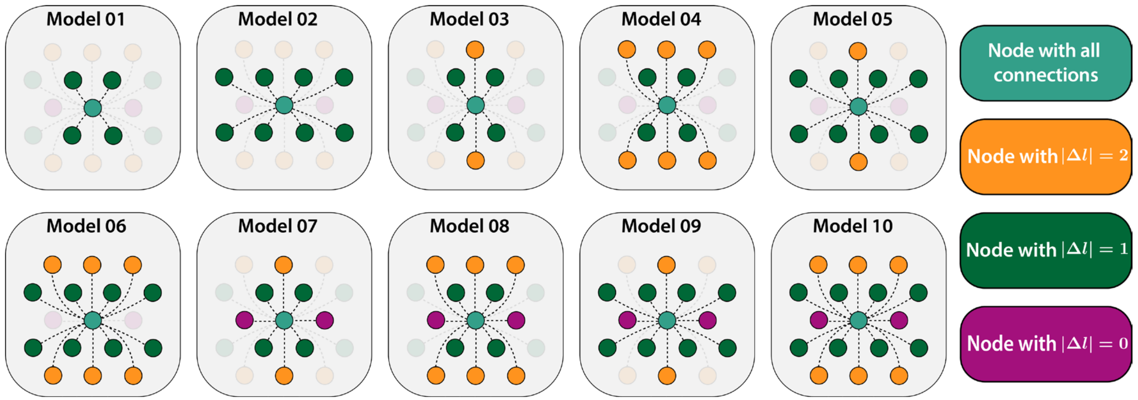
- This analysis uses 30 000 simulated events
- Uniformly distributed vertices, up to $\mathrm{r} = 100\ \mathrm{cm}$
- Full background (nominal phase 3)

$$\text{efficiency} = \frac{n_{\text{incl. true graph edges}}}{n_{\text{true graph edges}}}$$

$$\text{purity} = \frac{n_{\text{incl. true graph edges}}}{n_{\text{total edges}}}$$

Graph-Building – Evaluation



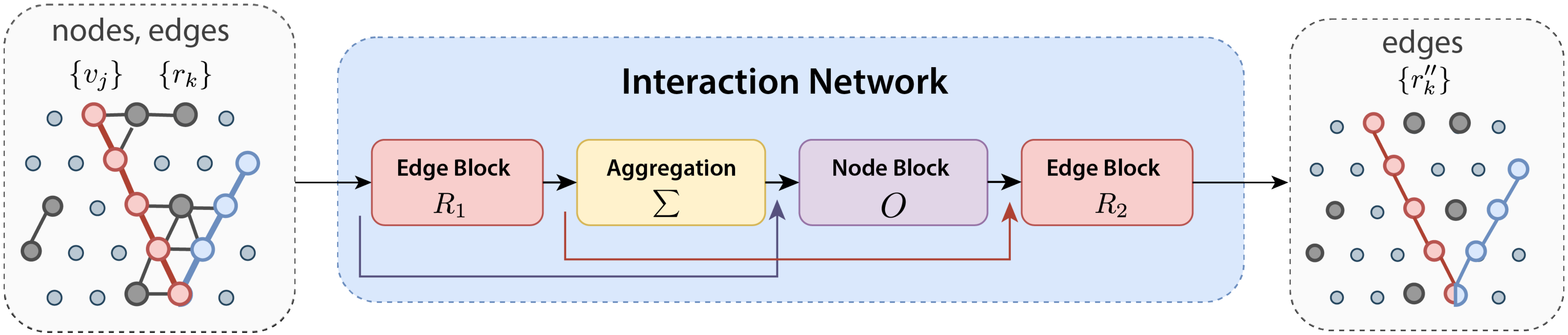


- This analysis uses 30 000 simulated events
- uniformly distributed vertices, up to $\mathrm{r} = 100\ \mathrm{cm}$
- Full background (nominal phase 3)
- Efficiency increases by incorporating more edges with a trade-off in purity

$$\text{efficiency} = \frac{n_{\text{incl. true graph edges}}}{n_{\text{true graph edges}}}$$

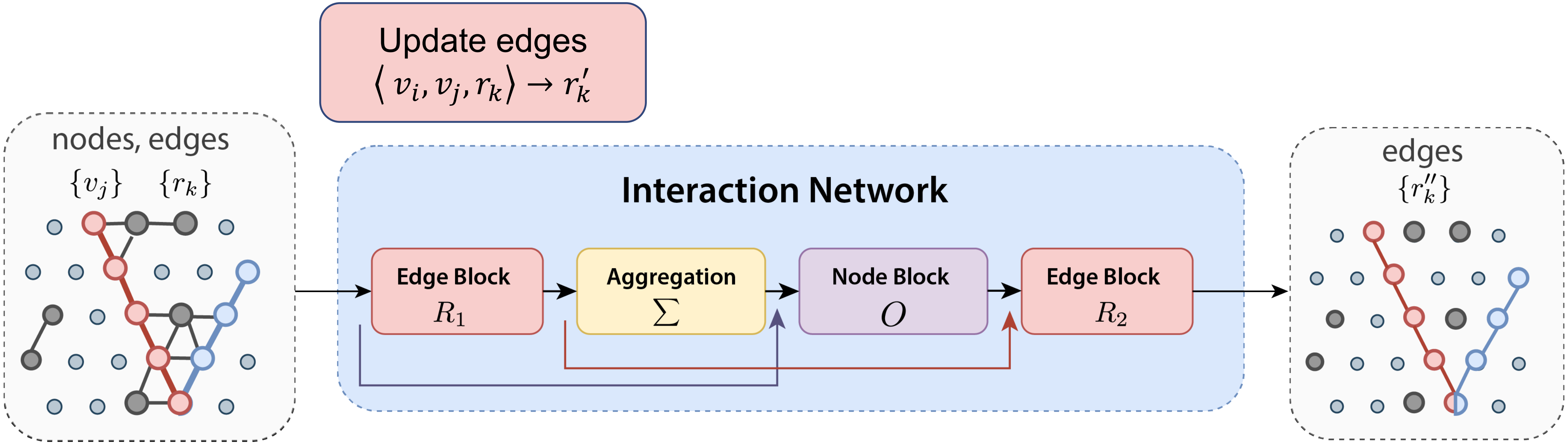
$$\text{purity} = \frac{n_{\text{incl. true graph edges}}}{n_{\text{total edges}}}$$

Edge Classification – The Interaction Network



1 Performance of a geometric deep learning pipeline for HL-LHC particle tracking
2 Interaction Networks for Learning about Objects, Relations and Physics

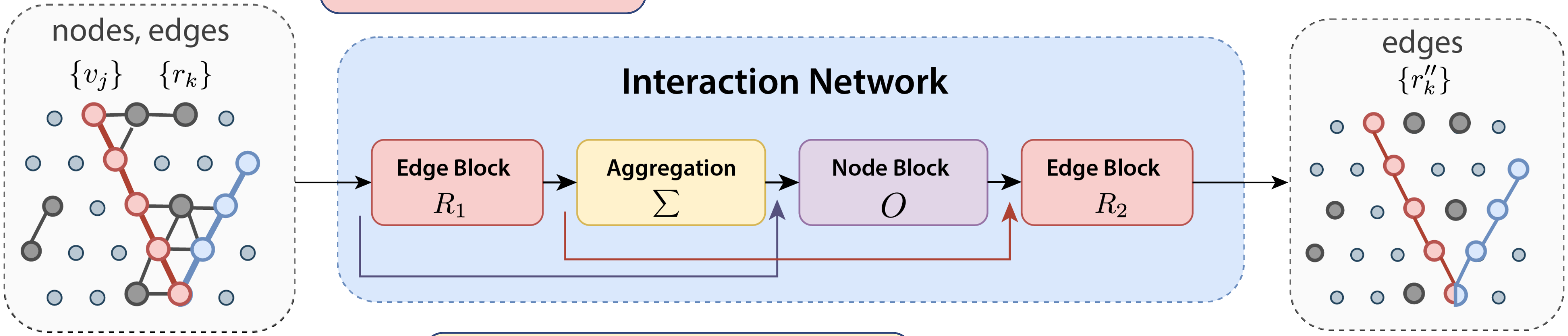
Edge Classification – The Interaction Network



1 Performance of a geometric deep learning pipeline for HL-LHC particle tracking
2 Interaction Networks for Learning about Objects, Relations and Physics

Edge Classification – The Interaction Network

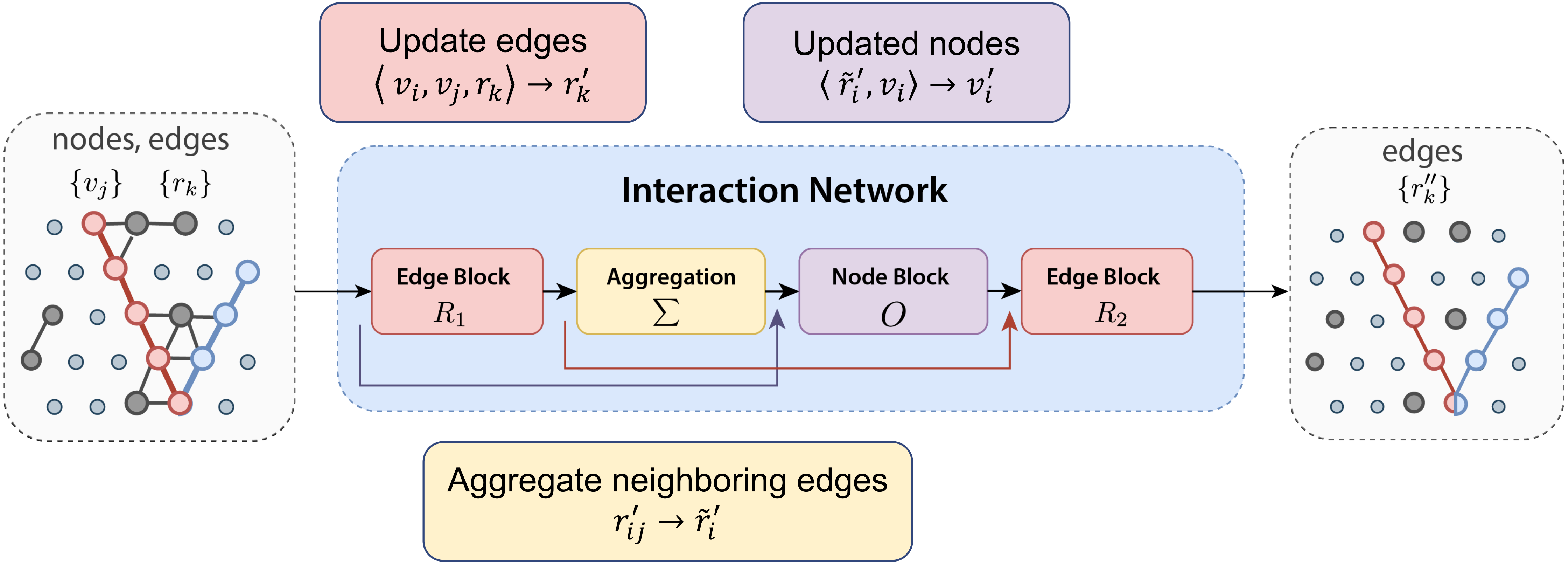
Update edges
$\langle v_i, v_j, r_k \rangle \rightarrow r'_k$



Aggregate neighboring edges
$r'_{ij} \rightarrow \tilde{r}'_i$

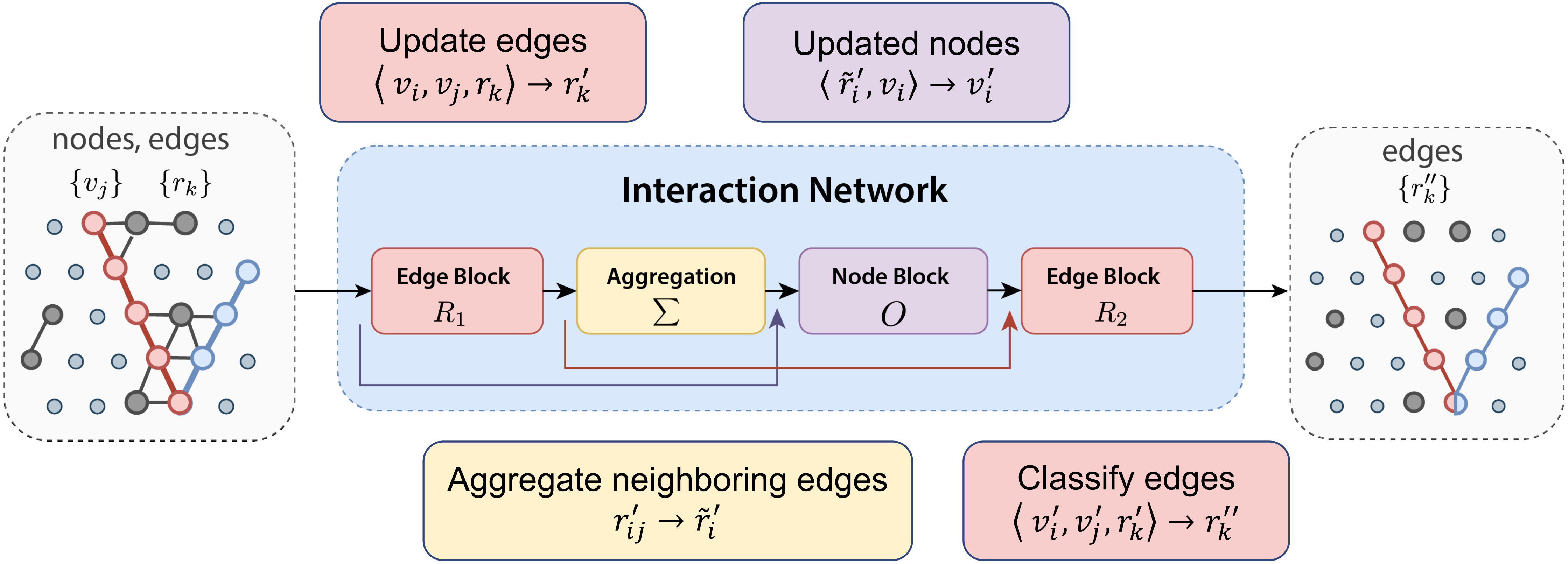
1 [Performance of a geometric deep learning pipeline for HL-LHC particle tracking]
2 [Interaction Networks for Learning about Objects, Relations and Physics]

Edge Classification – The Interaction Network



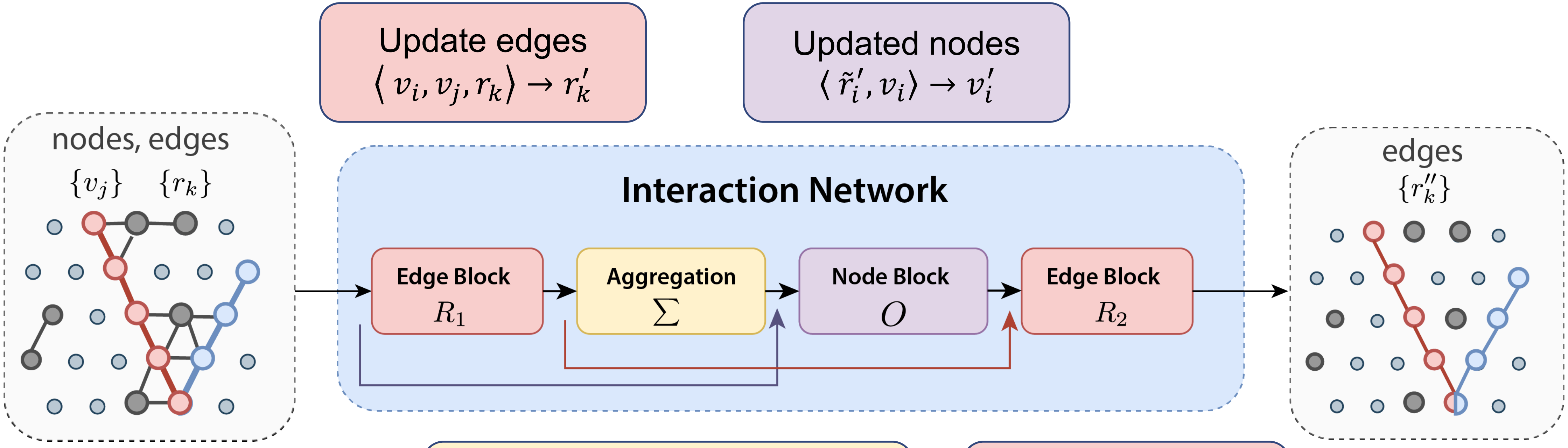
1 Performance of a geometric deep learning pipeline for HL-LHC particle tracking
2 Interaction Networks for Learning about Objects, Relations and Physics

Edge Classification – The Interaction Network



1 Performance of a geometric deep learning pipeline for HL-LHC particle tracking
2 Interaction Networks for Learning about Objects, Relations and Physics

Edge Classification – The Interaction Network



- Input features: x, y, TDC, ADC
- Edge attributes: $\Delta r, \Delta \phi$

1 Performance of a geometric deep learning pipeline for HL-LHC particle tracking
2 Interaction Networks for Learning about Objects, Relations and Physics

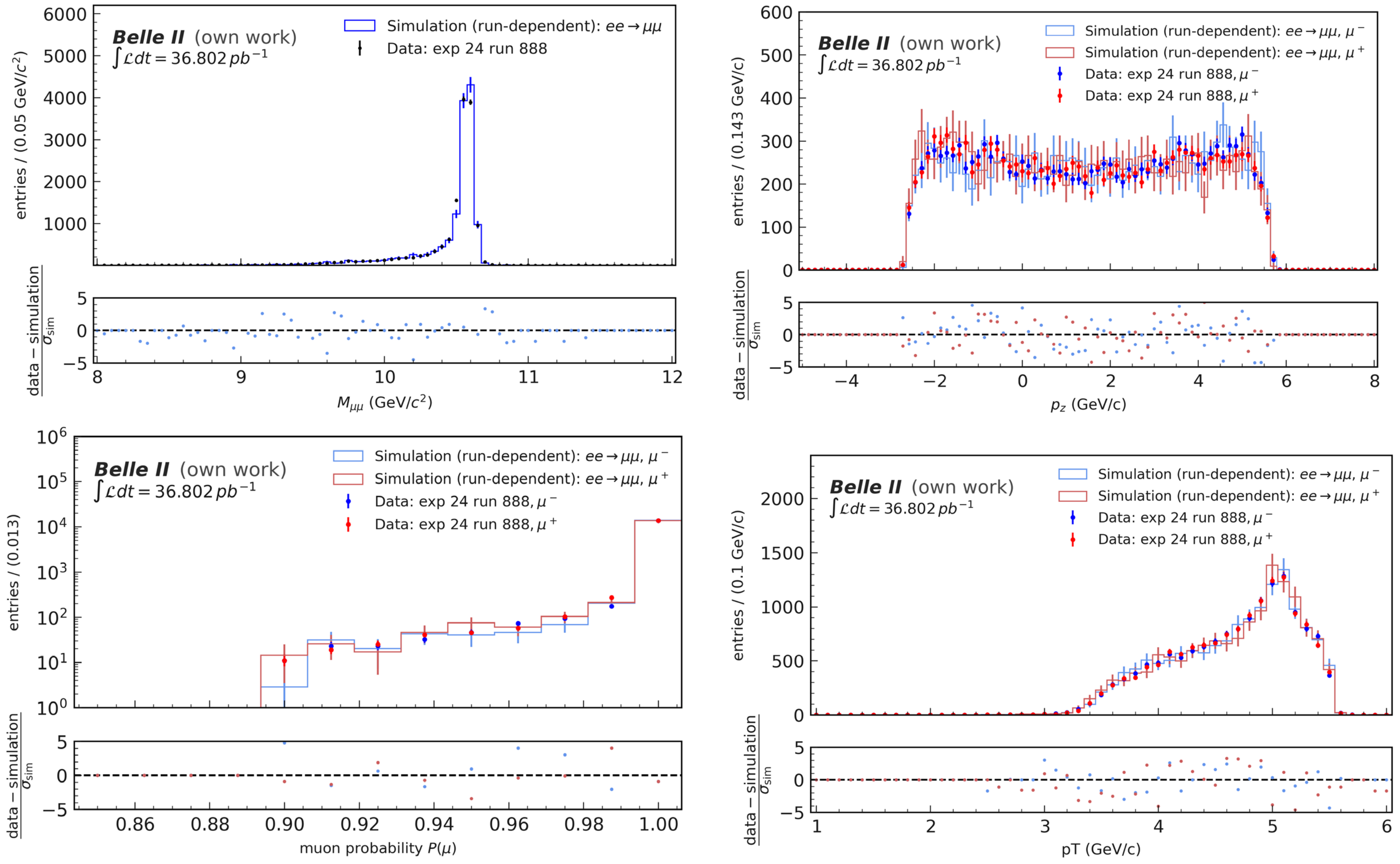
Input Feature Analysis - Methodology

- The aim of this study is to obtain insights into the ADC and TDC distributions and compare the simulation with Data
- Own simulation: $e^+e^- \rightarrow \mu^+\mu^-$

Philipp Dorwarth – Philipp.Dorwarth@student.kit.edu,

Input Feature Analysis - Methodology

- The aim of this study is to obtain insights into the ADC and TDC distributions and compare the simulation with Data
- Own simulation: $e^+e^- \rightarrow \mu^+\mu^-$
- Overall, the shape of the distributions shows good agreement
- Additional application of corrections expected to improve the agreement



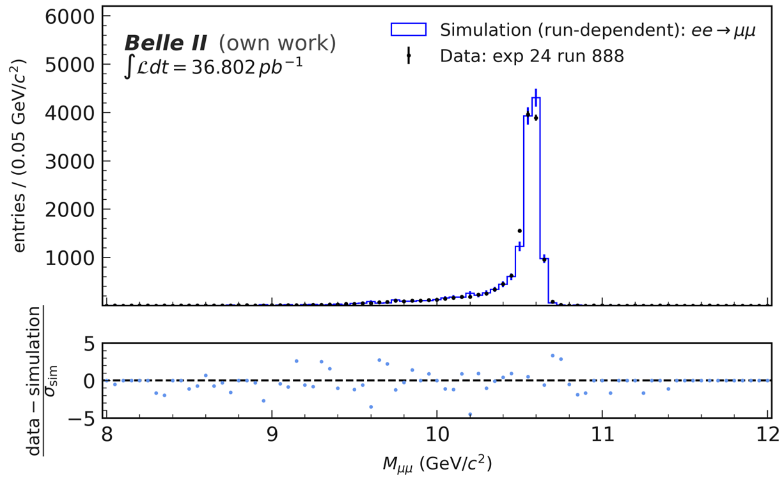
Philipp Dorwarth – Philipp.Dorwarth@student.kit.edu,

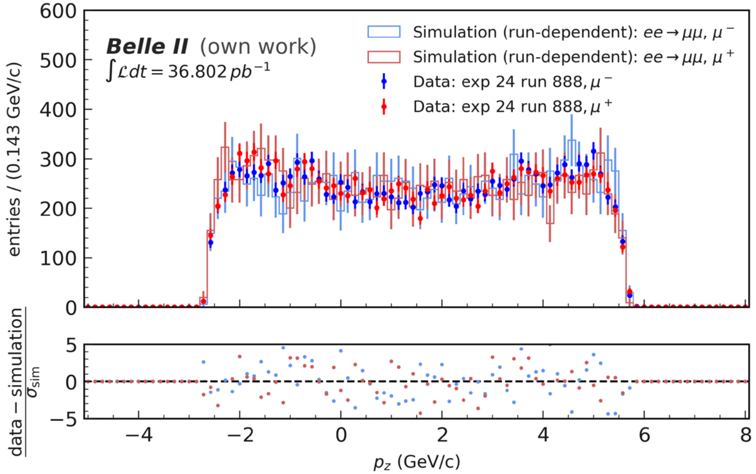
Institute of Experimental Particle Physics (ETP)

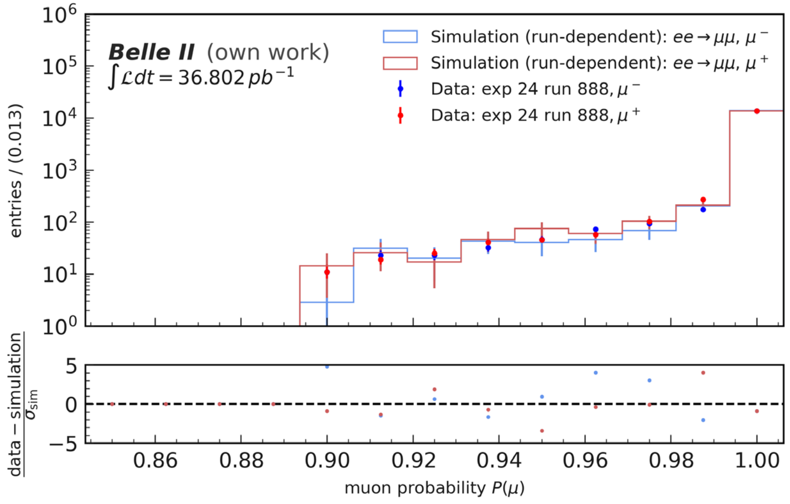
Input Feature Analysis - Methodology

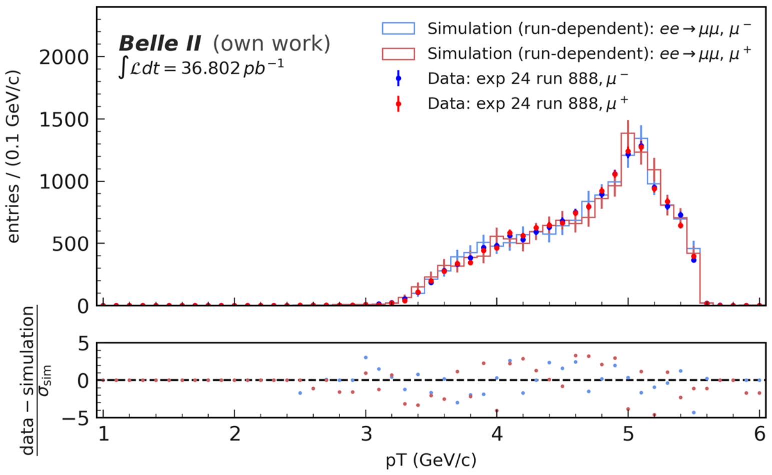
- The aim of this study is to obtain insights into the ADC and TDC distributions and compare the simulation with Data
- Own simulation: $e^+e^- \rightarrow \mu^+\mu^-$
- Overall, the shape of the distributions shows good agreement
- Additional application of corrections expected to improve the agreement

➡ Agreement for the selection was found to be sufficient for the ADC & TDC comparison



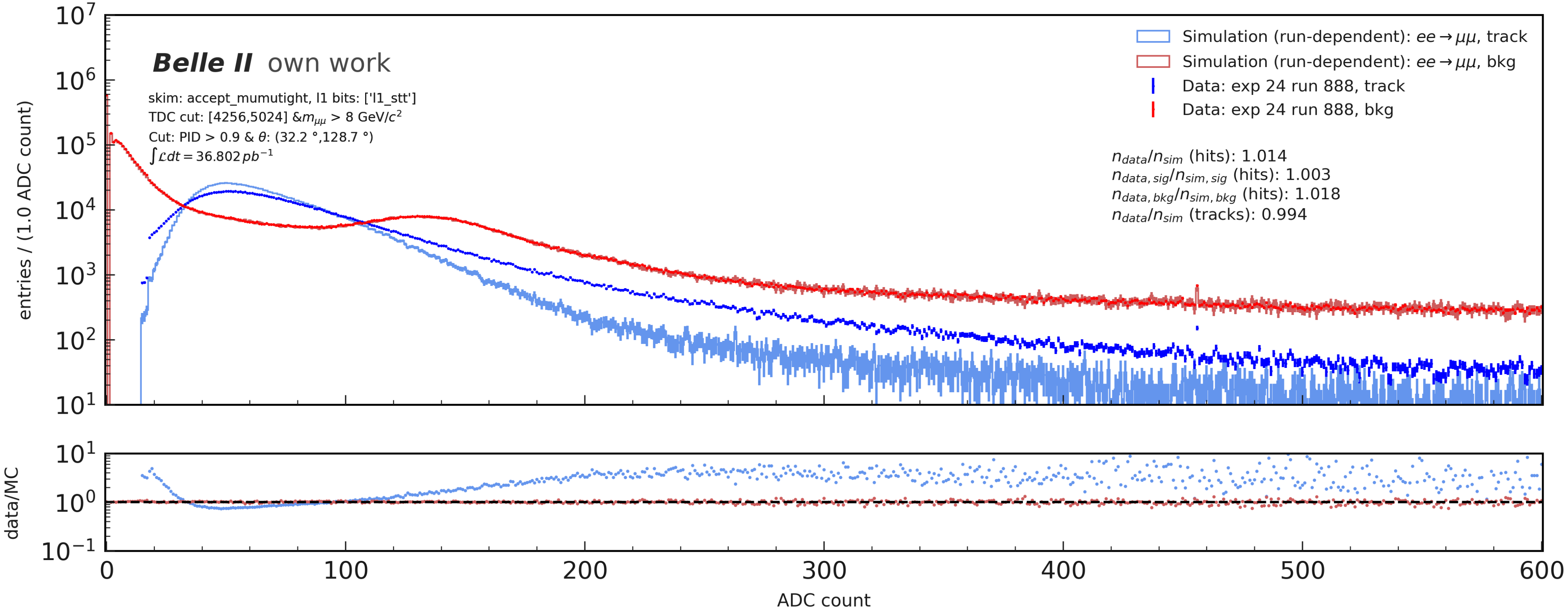






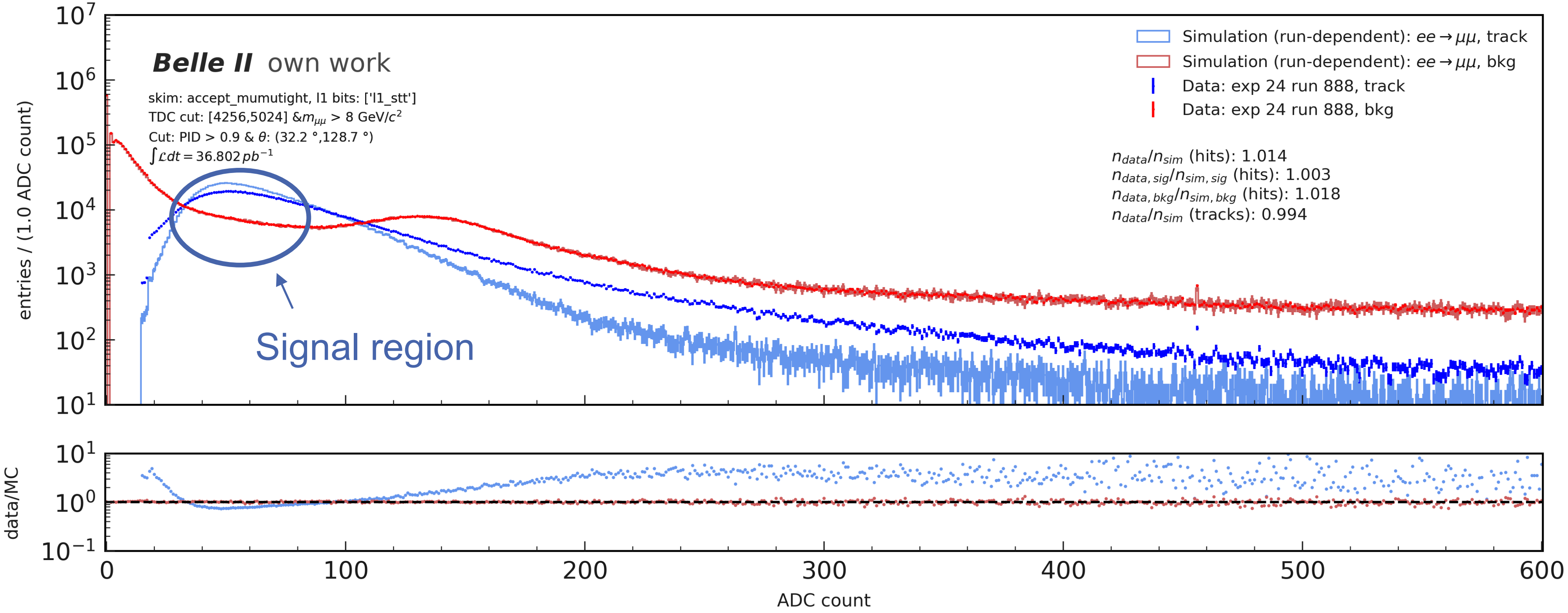
Input Feature Analysis - ADC

- Input feature analysis and MC/Data comparison of dimuon events
- $e^+e^- \rightarrow \mu^+\mu^-$



Input Feature Analysis - ADC

- Input feature analysis and MC/Data comparison of dimuon events
- $e^+e^- \rightarrow \mu^+\mu^-$



Philipp Dorwarth – Philipp.Dorwarth@student.kit.edu,

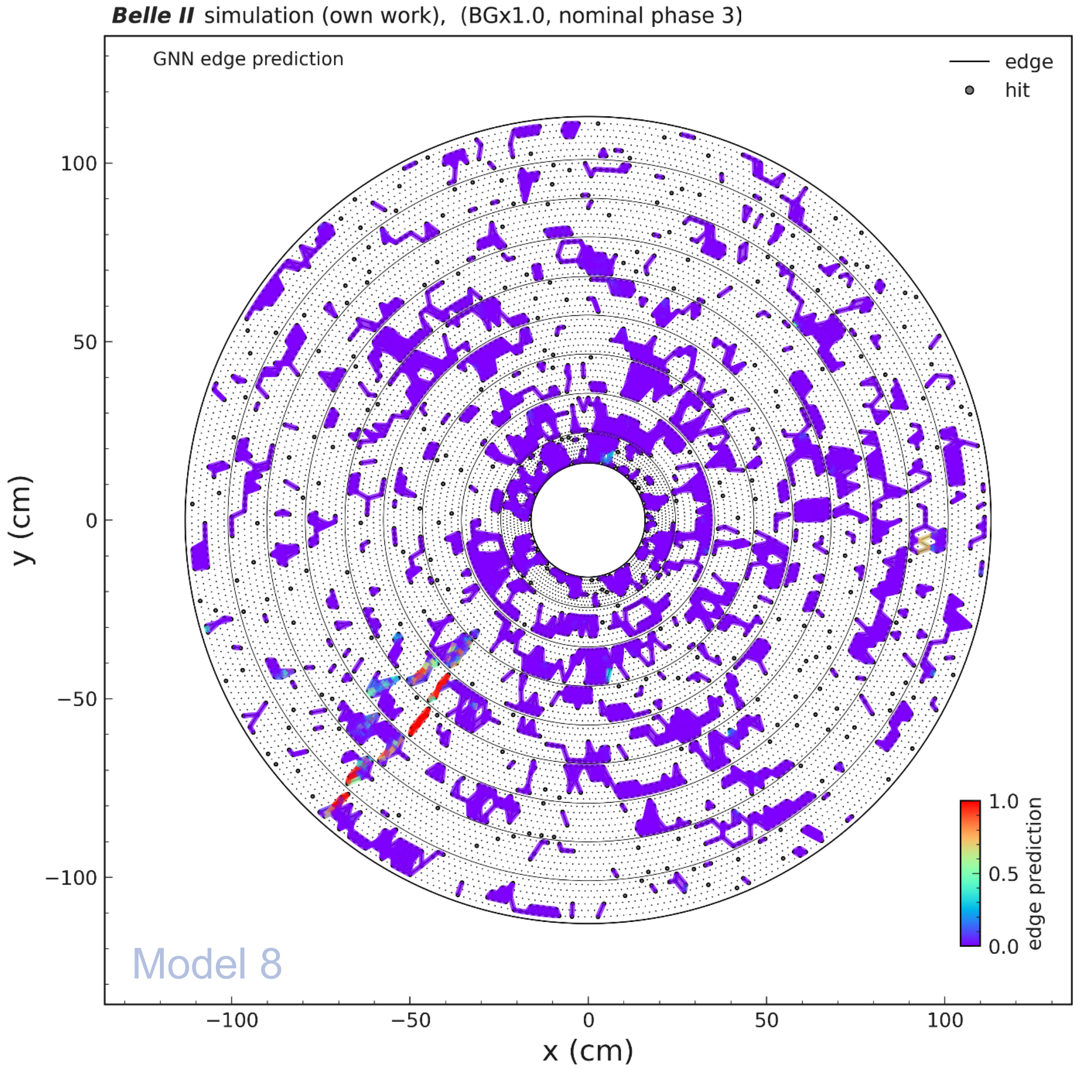
Edge Classification - ADC

- Input feature analysis and MC/Data comparison of dimuon events
- $e^+e^- \rightarrow \mu^+\mu^-$

Data and MC Comparison

- ADC and TDC are discriminatory and provide orthogonal information
- Simulation must be possibly adjusted to align more strongly with the data for training of GNN-Pipeline

Edge Classification – CDC Edge Prediction



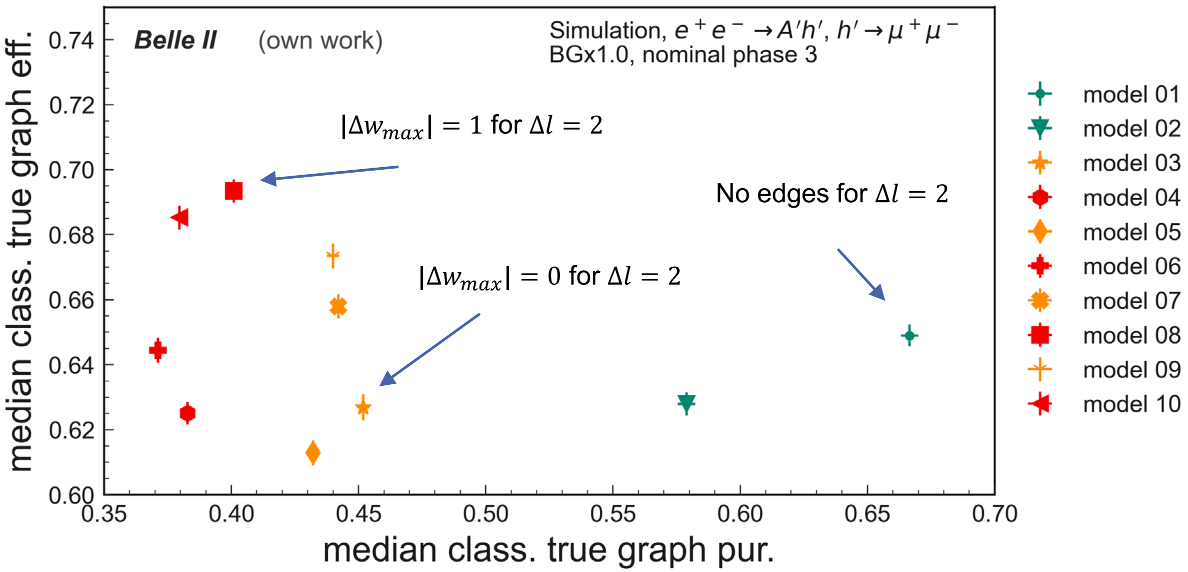
- Training with a train set of 24000 simulated events
- Independent models are trained for each graph-building pattern
- Use of full background level
- Input features: x, y, TDC, ADC
- Edge attributes: $\Delta r, \Delta\phi$
- IN consist of 1987 trainable parameters
- Threshold is obtained by maximizing the F1 score

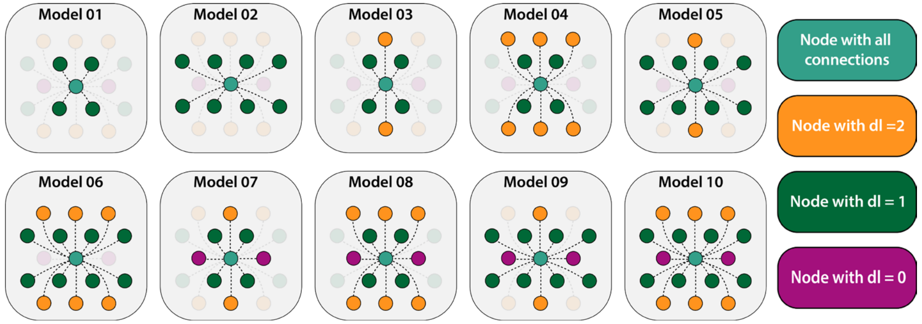
$$\mathrm{F1}_{\mathrm{true\_edge}} = 2\frac{\textit{class. true edge pur.} \times \textit{class. true edge eff.}}{\textit{class. true edge pur.} + \textit{class. true edge eff.}}$$

$$\textit{class. true edge eff.} = \frac{n_{\text{incl. pred. true edges}}}{n_{\text{incl. true edges}}}$$

$$\textit{class. true edge pur.} = \frac{n_{\text{incl. pred. true edges}}}{n_{\text{total predicted edges}}}$$

Edge Classification – True Graph Edges



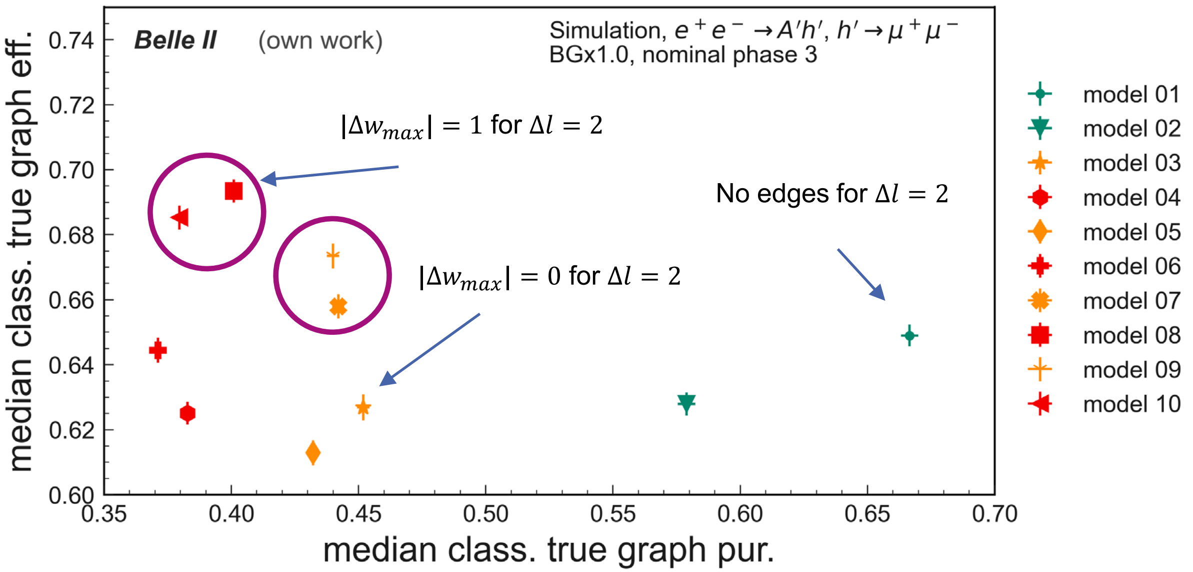


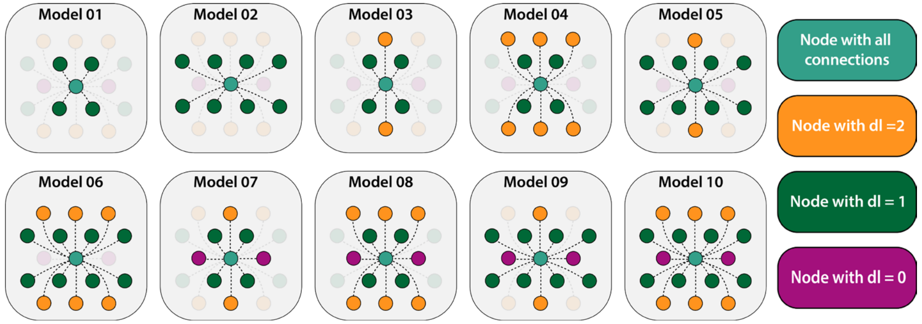
- Efficiency and purity of the estimated true graph after edge classification
- $\Delta l = 2$ connections influence purity

$$\textit{class. true graph eff.} = \frac{n_{\text{incl. pred. true graph edges}}}{n_{\text{incl. true graph edges}}}$$

$$\textit{class. true graph pur.} = \frac{n_{\text{incl. pred. true graph edges}}}{n_{\text{total predicted edges}}}$$

Edge Classification – True Graph Edges



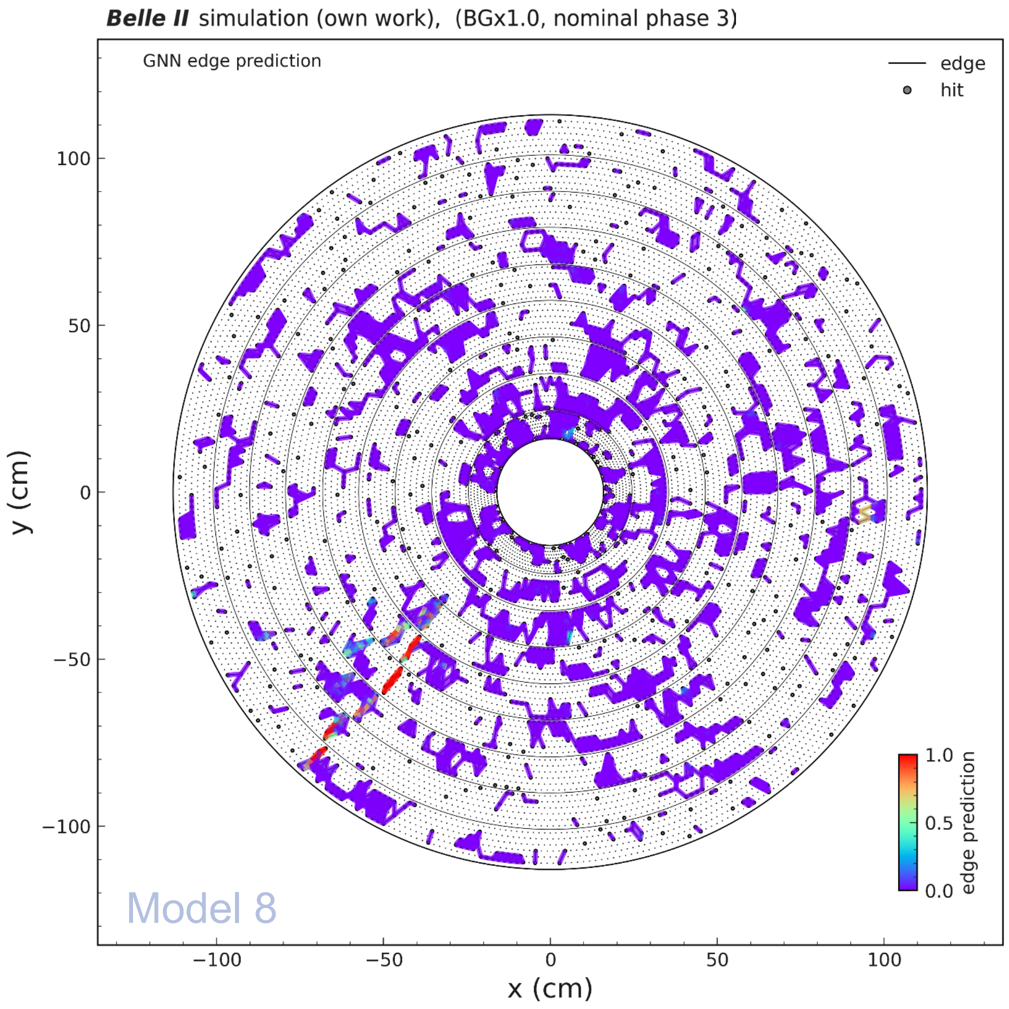


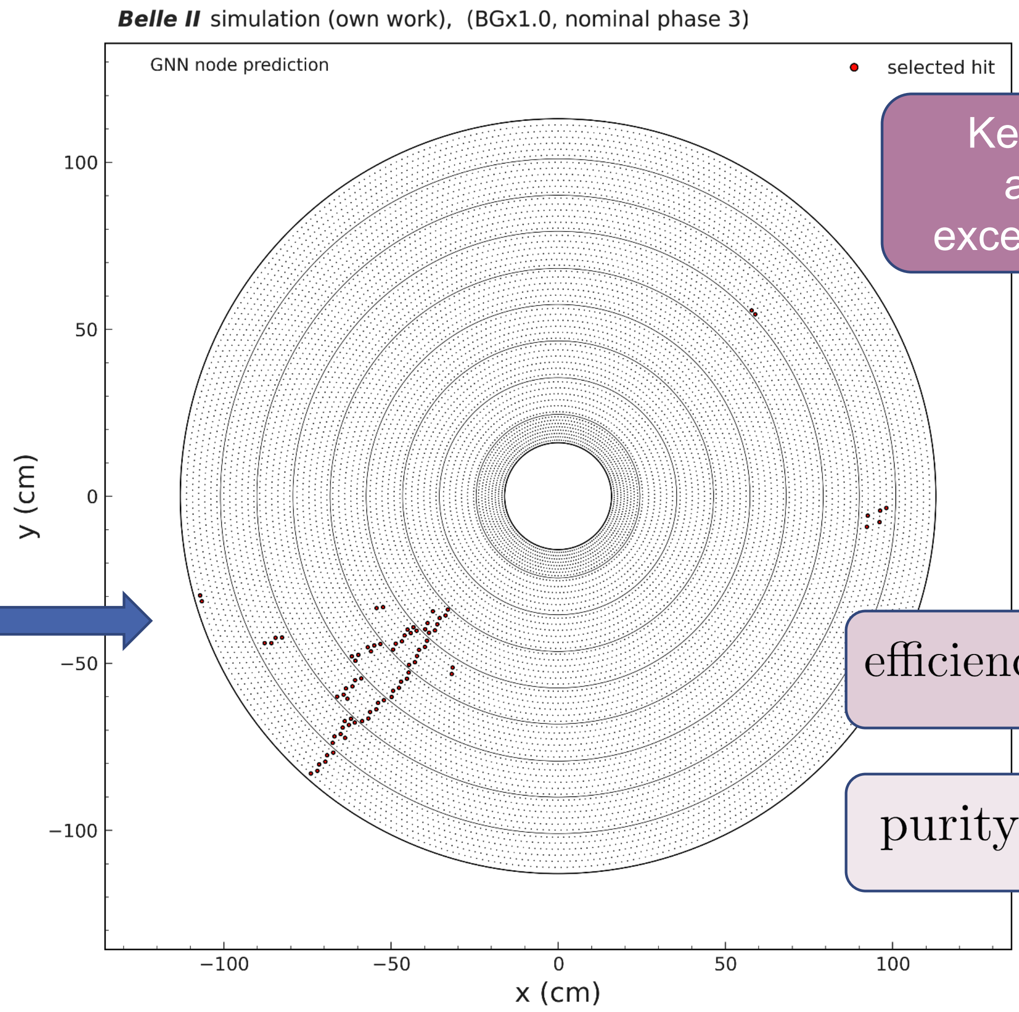
- Efficiency and purity of the estimated true graph after edge classification
- $\Delta l = 2$ connections influence purity
- $\Delta l = 0$ connections influence efficiency

$$class.\ true\ graph\ eff. = \frac{n_{\text{incl. pred. true graph edges}}}{n_{\text{incl. true graph edges}}}$$

$$class.\ true\ graph\ pur. = \frac{n_{\text{incl. pred. true graph edges}}}{n_{\text{total predicted edges}}}$$

Edge Classification – Clean-Up



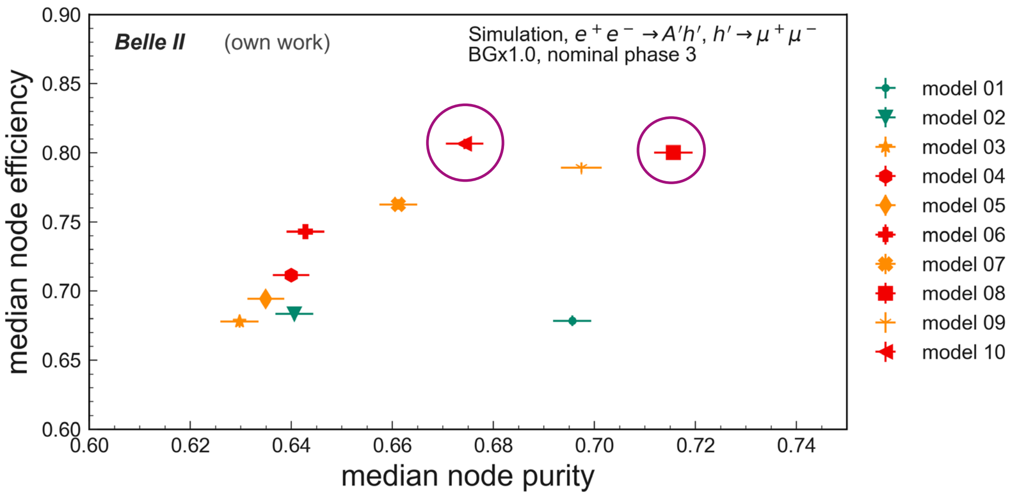


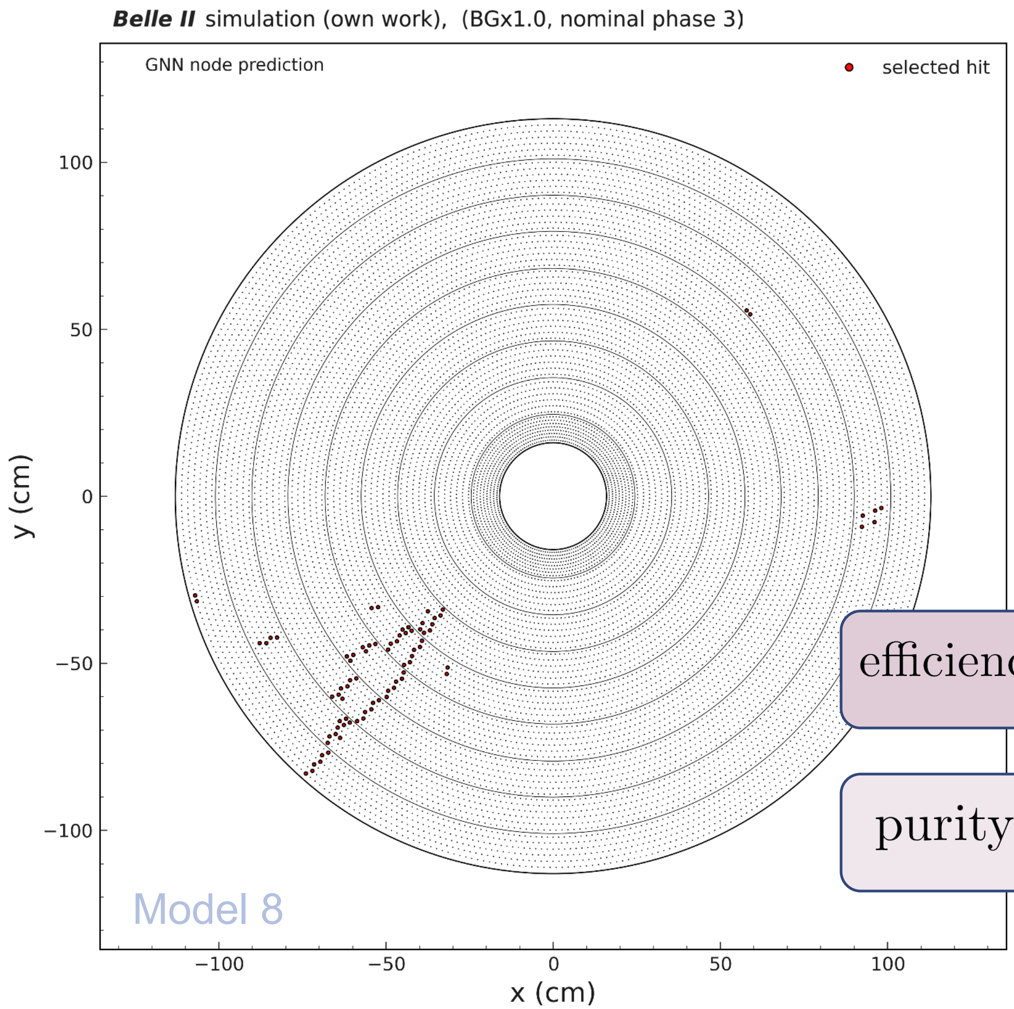
Keep any hit with an associated edge exceeding the threshold

$$\text{efficiency} = \frac{n_{\text{incl. true nodes}}}{n_{\text{true nodes}}}$$

$$\text{purity} = \frac{n_{\text{incl. true nodes}}}{n_{\text{nodes}}}$$

Edge Classification – Clean-Up





$$\text{efficiency} = \frac{n_{\text{incl. true nodes}}}{n_{\text{true nodes}}}$$

$$\text{purity} = \frac{n_{\text{incl. true nodes}}}{n_{\text{nodes}}}$$

- model 10 and model 8 achieve both high efficiency
  - Trade off between purity and efficiency
  - Purity and efficiency can be adjusted with the threshold for each model
- High computational complexity

Online Application – Limitations

Karlsruhe Institute of Technology

Precision:

PyG employs high-precision (32 Bit) operations.

For FPGA, lower precision arithmetic may be needed, potentially affecting network accuracy.

Resource Consideration:

Space multiplexing, used over time multiplexing, increases FPGA resource usage.

The size of the IN and input graphs directly impacts resource usage.

Memory Limitations:

FPGA memory is limited and reliance on external memory can increase latency.

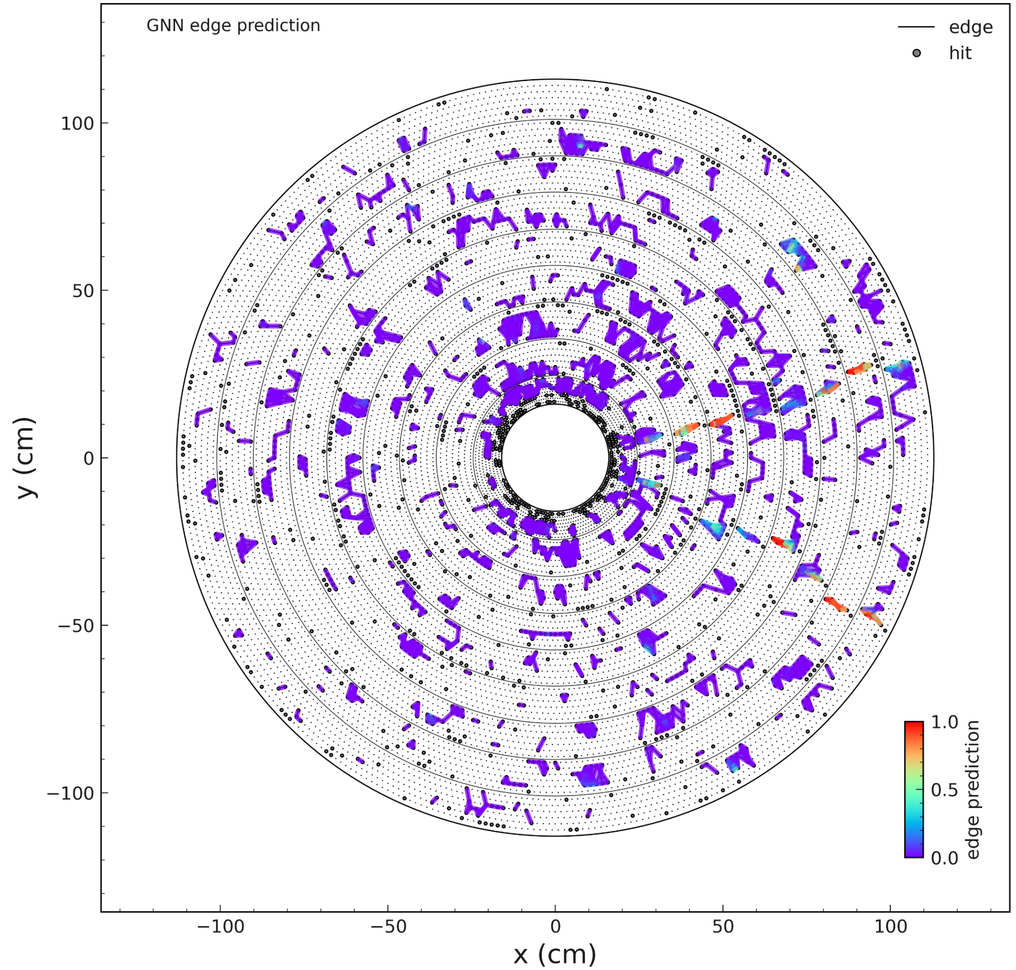
Optimizing input graphs and neural network model sizes is key for low latency.

Input Resolution:

Anticipated reduction in input feature resolution may require modification in data processing techniques and consideration of relative information.

Limitations when comparing PyTorch Geometric to the implementation of FPGAs

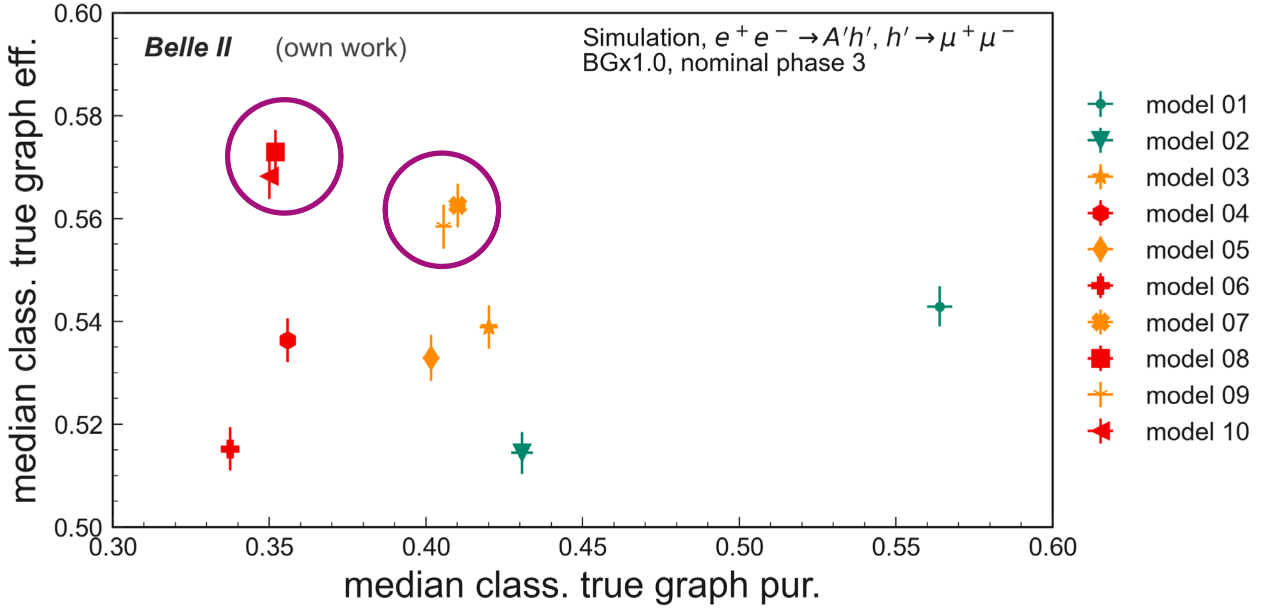
Online Application - Adaption and Mitigation



- Resolution Reductions:
  - Limited CDC sense wire information.
  - TDC resolution drop (1 ns to 2 ns) & absolute value becomes meaningless.
  - Decreased ADC count resolution & sampling rate.
- Adaption Strategies:
  - Using only trigger wires.
  - TDC counts binned in twos.
  - Remove TDC as input feature.
- Mitigation Strategies:
  - *Δ*TDC used as an additional edge attribute.
  - ADC counts are grouped into three ranges.

Online Application – Results With Reduced Resolution

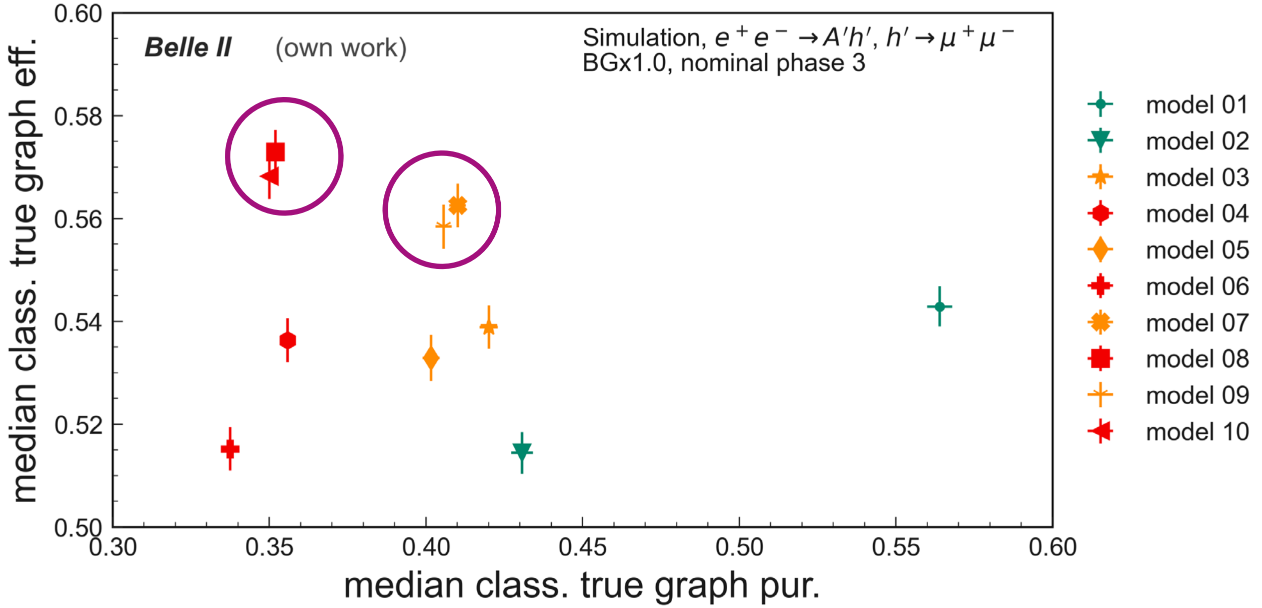
- Efficiency and purity of the estimated true graph after edge classification
- 1.947 trainable parameters
- Input features: x, y, ADC (reduced resolution)
- Edge attributes: $\Delta r, \Delta\phi, \Delta\mathrm{TDC}$ (reduced resolution)



Philipp Dorwarth – Philipp.Dorwarth@student.kit.edu,

Online Application – Results With Reduced Resolution

- Efficiency and purity of the estimated true graph after edge classification
- 1.947 trainable parameters
- Input features: x, y, ADC (reduced resolution)
- Edge attributes: $\Delta r, \Delta\phi, \Delta\mathrm{TDC}$ (reduced resolution)

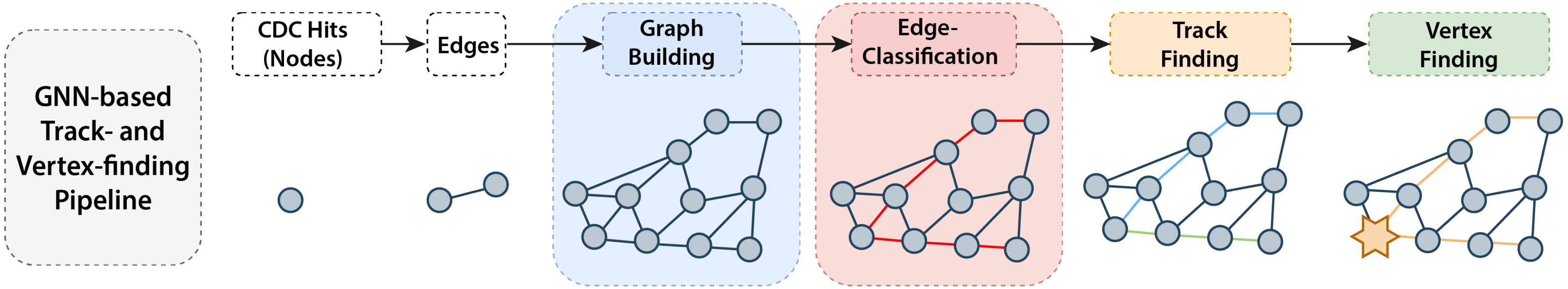


→ A drop of approximately 10 pp.

A trade-off between efficiency and purity

→ Adjust to the needs of track- and vertex finding

Summary & Outlook



Summary

- Developed graph-building models for effective capture of particle track information
- Identified crucial edges for performance measurements
- Achieved encouraging results in edge classification and CDC clean-up offline and online
- This study and related studies have encouraged the Belle II collaboration to make ADC and TDC information available on the trigger level

Outlook

Working together with ITIV (Department of Electrical Engineering and Information Technology at KIT) to implement the GNN-pipeline on FPGA for real-time application

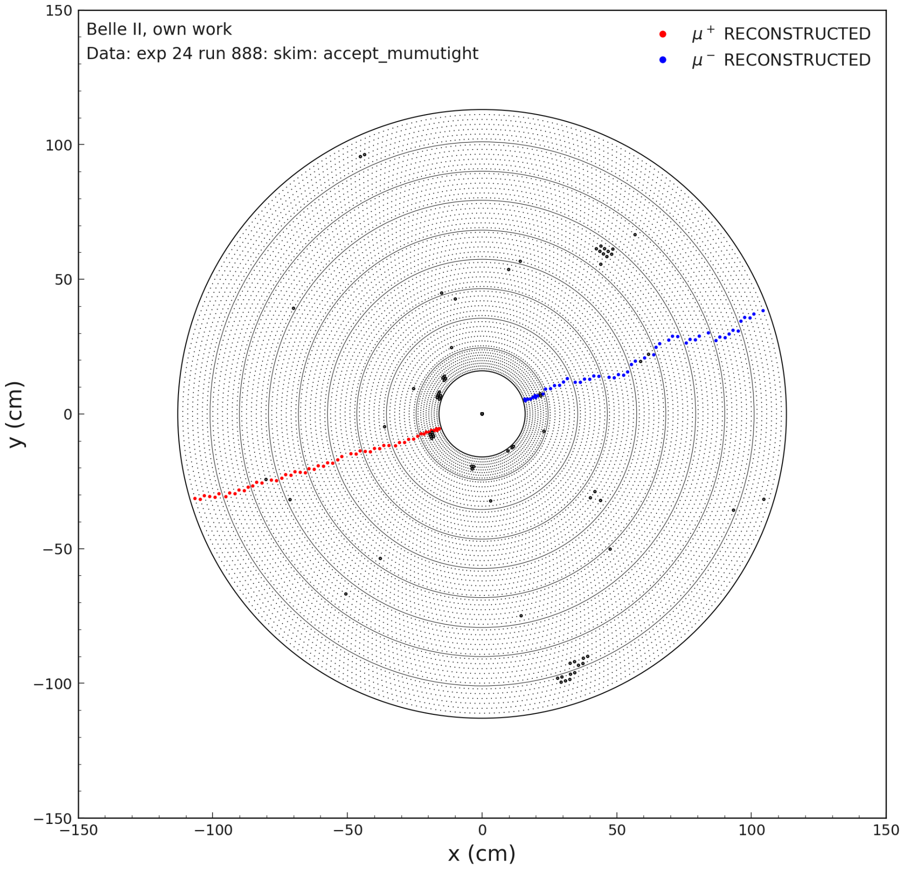
- First implementation of graph building
- Detailed analysis of the subsequent performance of the IN on hardware will provide insights for optimization requirements.

Further Work

- Devised additional metrics for graph-building
- Introduced metric to compare patterns after classification
- Studied computational complexities of various graph-patterns

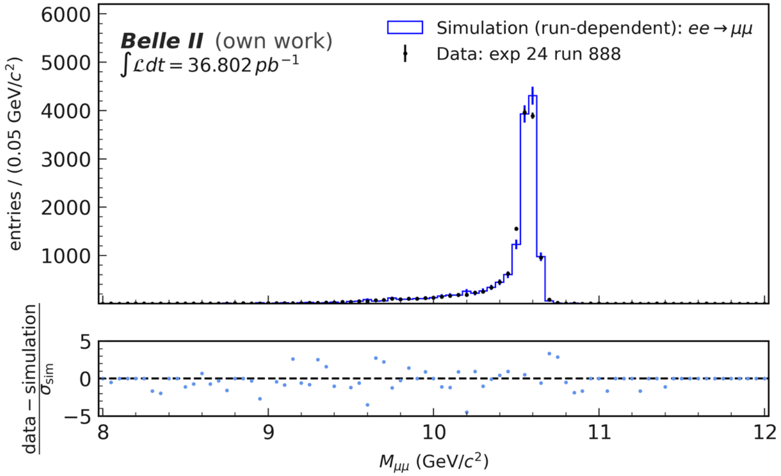
Selection

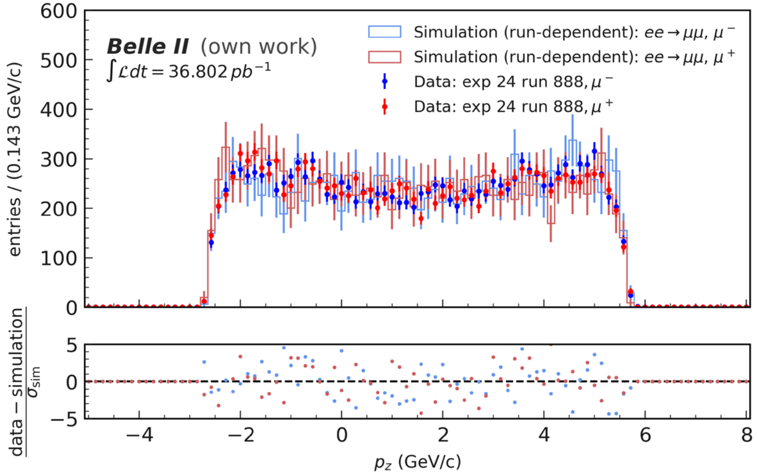
- Skim: accept_mumutight
- L1 Trigger bit: “stt”
- Events with two reconstructed tracks
  - each with a Muon PID > 0.9
  - $\theta = [32.2\,°, 128.7\,°]$ (barrel region)
  - Dimuon mass cut: $(m_{\gamma\gamma} > 8\ GeV/c^2)$
- TDC_{CDCHit} < 5021

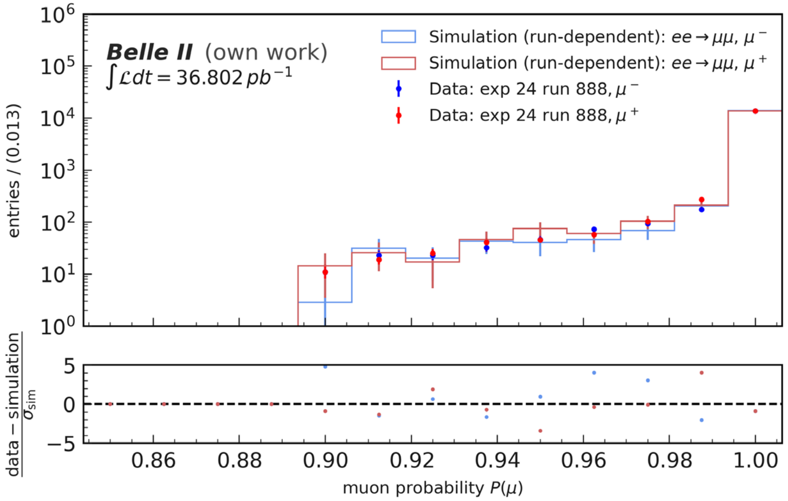


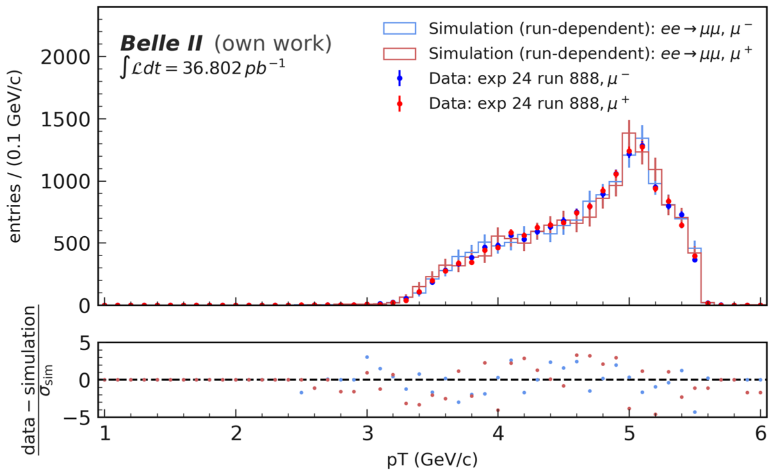
Order No	Selection	Data (exp 24, run 888)	MC (simulation)
1	Theta cut	0.764	0.755
2	L1 selection *stt*	0.949	0.996
3	Dimuon track selection	0.878	0.955
4	Dimuon mass cut	0.947	0.998
5	PID cut	0.944	0.955
	Total efficiency	0.569	0.684

Input Feature Analysis - Track Properties







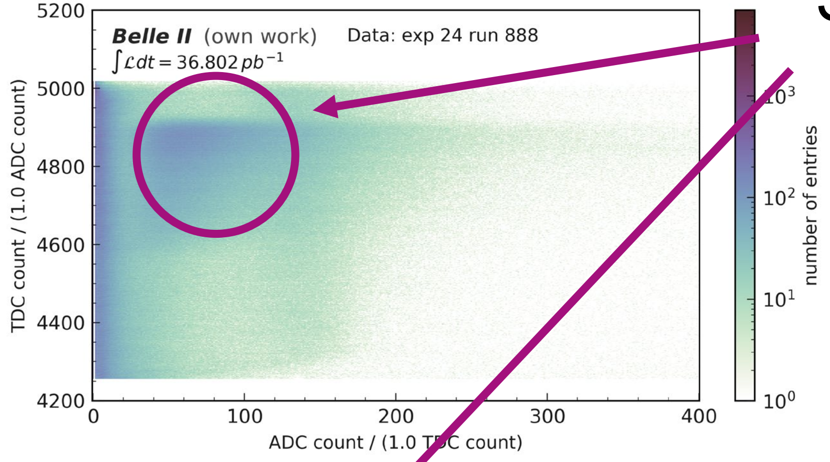


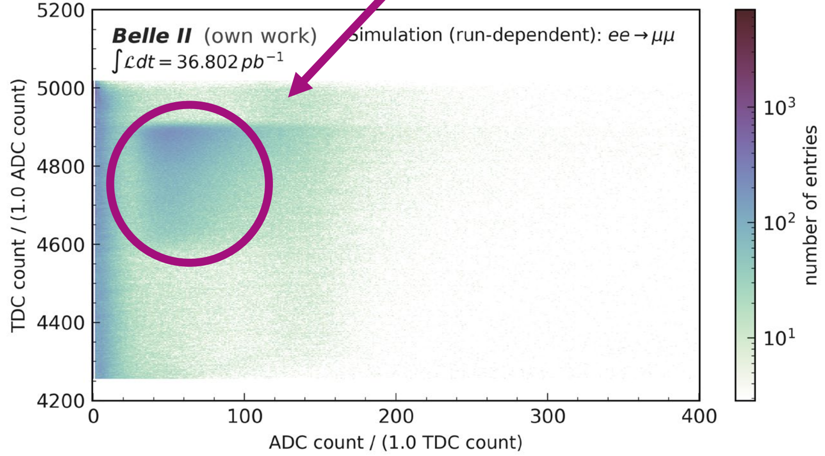
- The aim of this study is to obtain insights into the ADC and TDC distributions and compare the simulation with Data
- Overall, the shape of the distributions shows good agreement
- Additional application of corrections expected to improve the agreement

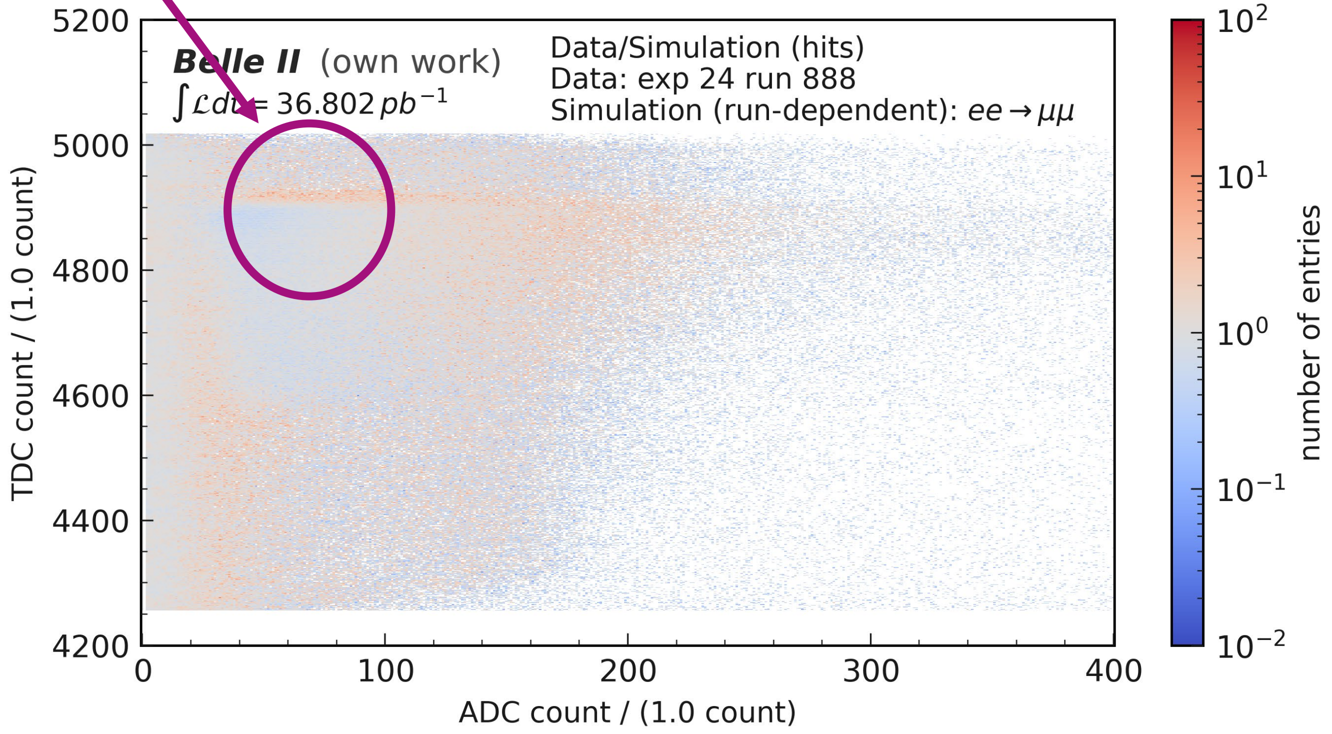
Agreement for the selection was found to be sufficient for the ADC & TDC comparison

ADC and TDC 2D Comparison

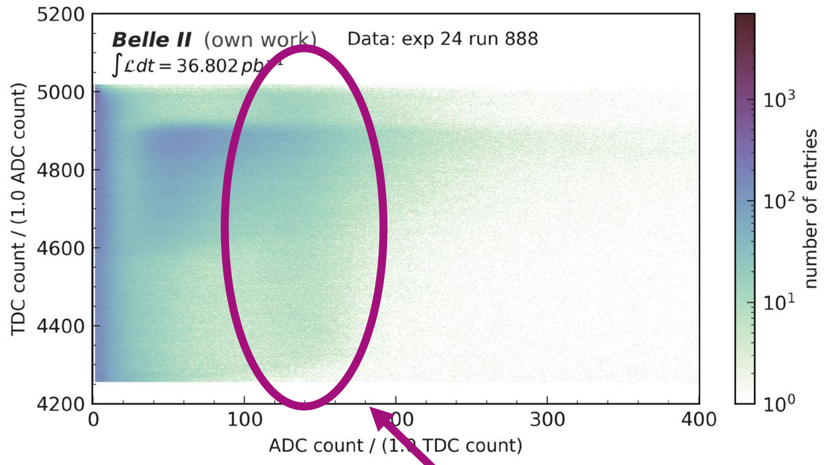
Signal region

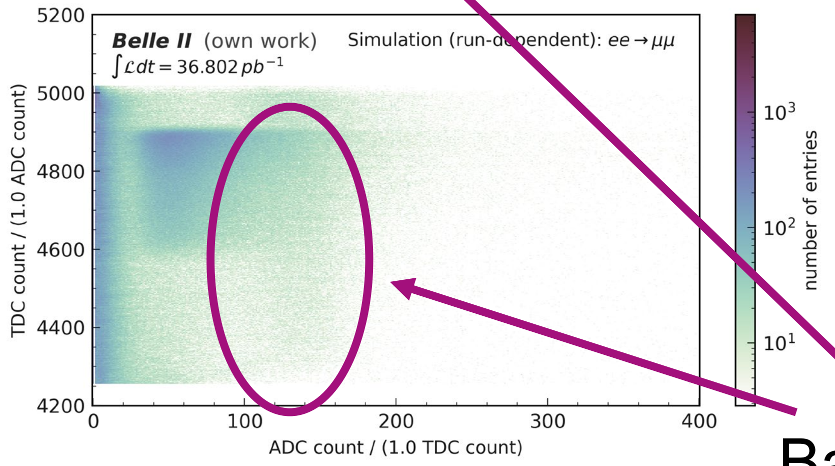


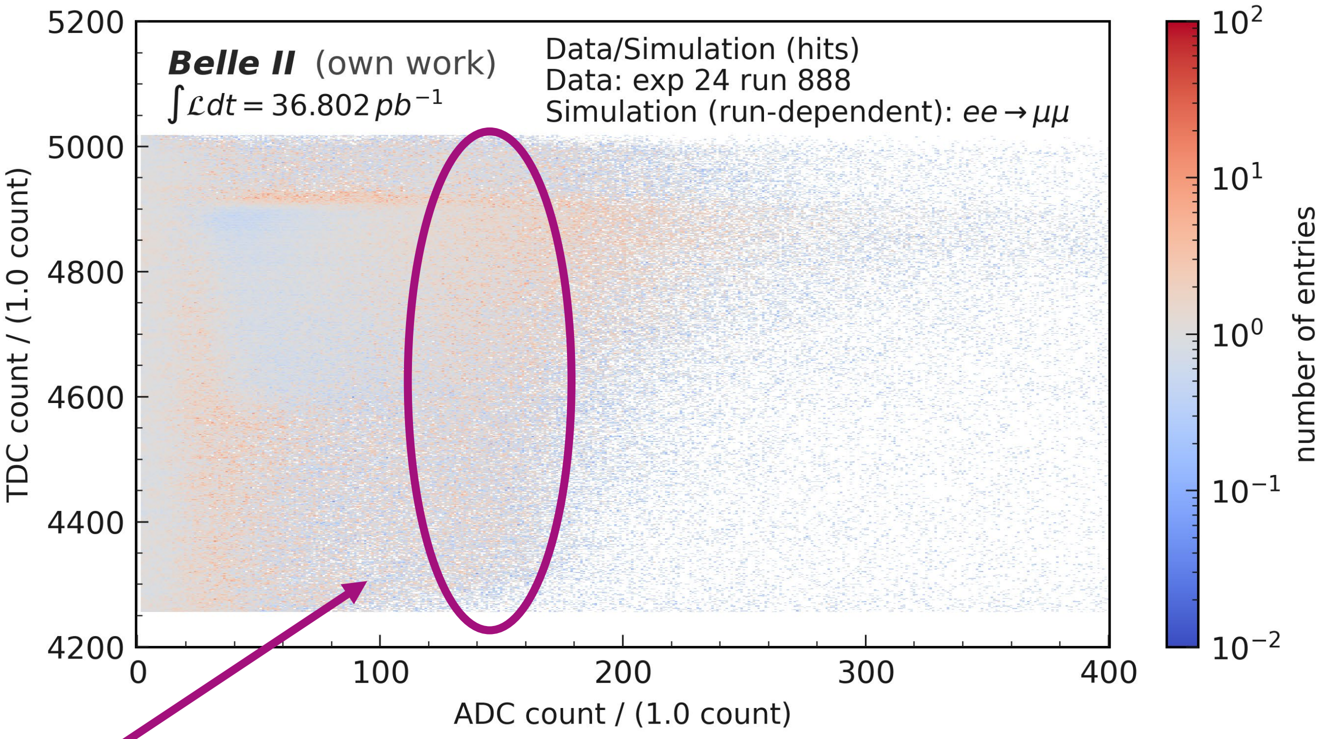




ADC and TDC 2D Comparison



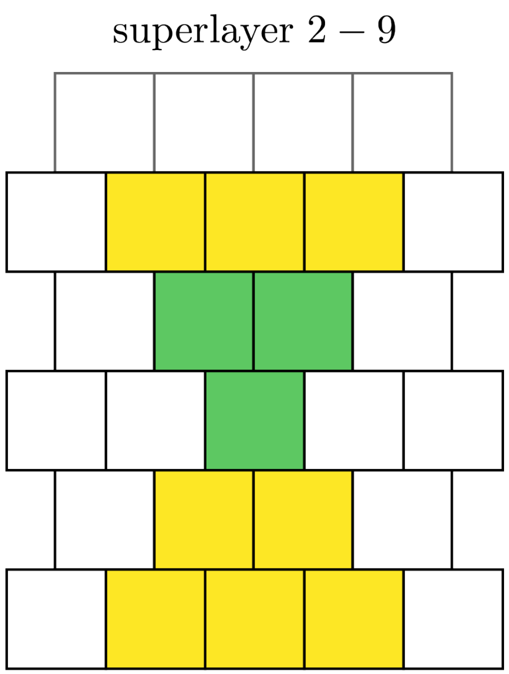


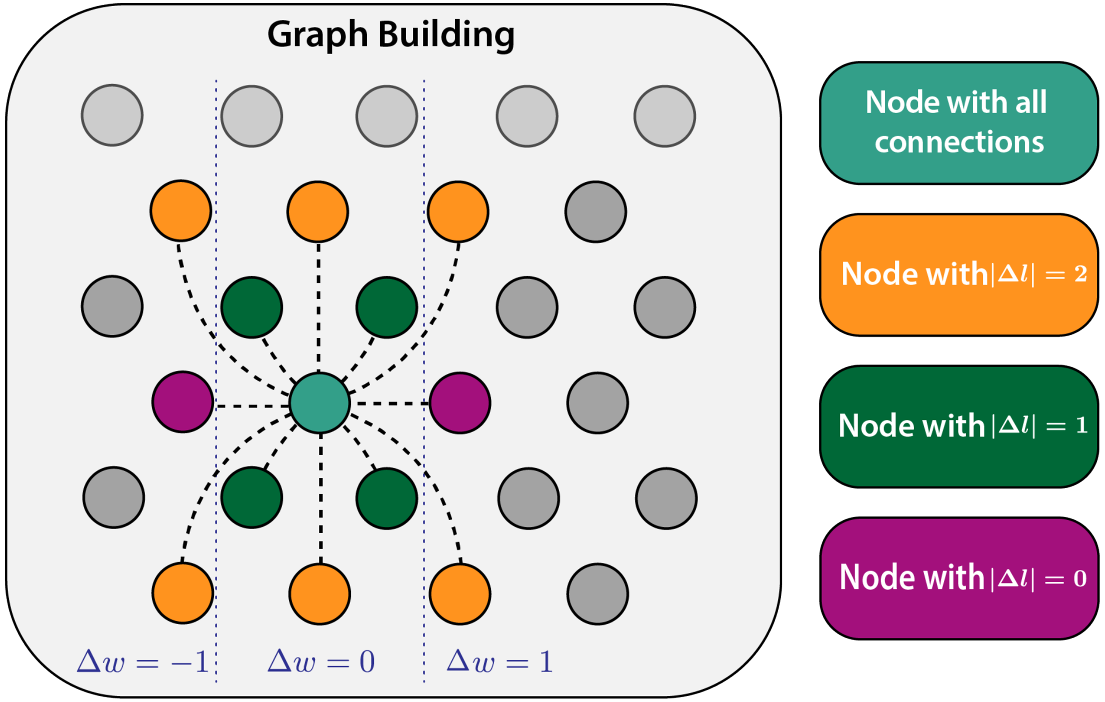


Background region

Patterns For Displaced Vertices

- Two parameters define the patterns in the CDC
  - layer distance Δl
  - wire distance Δw





Online Application – Limitations

Precision:

PyG employs high-precision (32 Bit) operations.

For FPGA, lower precision arithmetic may be needed, potentially affecting network accuracy.

Resource Consideration:

Space multiplexing, used over time multiplexing, increases FPGA resource usage.

The size of the IN and input graphs directly impacts resource usage.

Memory Limitations:

FPGA memory is limited and reliance on external memory can increase latency.

Optimizing input graphs and neural network model sizes is key for low latency.

Input Resolution:

Anticipated reduction in input feature resolution may require modification in data processing techniques and consideration of relative information.

Limitations when comparing PyTorch Geometric to the implementation of FPGAs

Online Application – Estimate Computational Complexity

model	total FLOPs (MFLOP)
01	20.23 ± 0.05
02	33.94 ± 0.08
03	24.79 ± 0.06
04	34.03 ± 0.08
05	38.51 ± 0.09
06	47.53 ± 0.12
07	37.11 ± 0.08
08	46.15 ± 0.10
09	50.53 ± 0.11
10	59.97 ± 0.14

- Floating Point Operations (FLOPs) measure computational workload
  - They may not reflect actual FPGA performance!
- Model 10 shows slightly better node efficiency but comes with highly increased computational complexity