



First Studies of a GNN based ECL Trigger on Heterogeneous Versal SoCs for the Belle II Upgrade

Belle II Germany Meeting 2025

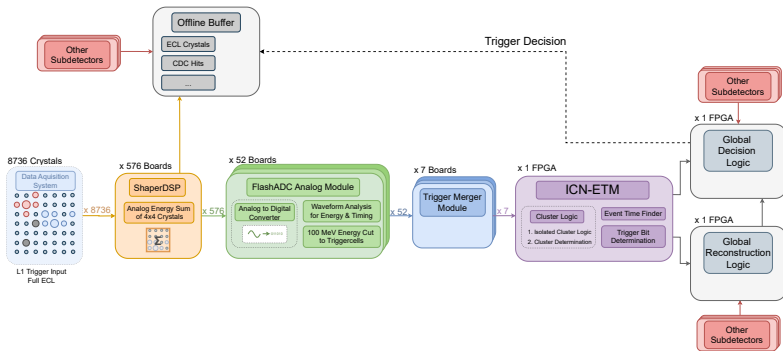
Thomas Lobmaier* (thomas.lobmaier@student.kit.edu), Isabel Haide*, Frank Baptist, Marc Neu*, Fabio Papagno*, Jürgen Becker*, Torben Ferber*

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

09/09/2025

Current ECL Trigger Pipeline

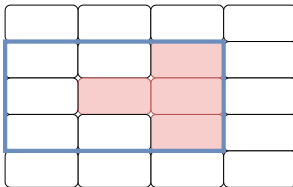
1. Crystal input
2. Forming of 4x4 TCs
3. Determine TC timing and energy
4. Apply 100 MeV cut on TCs
5. ICN-ETM reconstructs cluster objects with position and energy prediction
6. Based on clusters create trigger bits (e.g. total energy in ECL, 3 clusters above certain threshold, ...)
7. Pass the trigger bits to Global Decision Logic
8. Apply prescaling on certain trigger bits and make trigger decision



ICN-ETM

1. Apply pattern matching on all active TCs
2. Take up to 6 TCs fulfilling this pattern
3. shift 3x3 window towards highest energetic TC as center
4. Energy and position of the ICN-Cluster is the energy sum of the 3x3 window and the position of the central TC

Triggercells with Energy Depositions

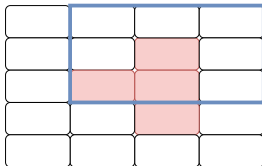


Isolated Cluster Logic

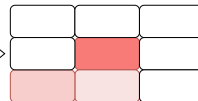
	1	
2	0	
3	4	

0 & !(1 or 2) & !(3 and 4)

Triggercells with Energy Depositions



Cluster Energy and Position



Cluster Energy = Sum of All Energies in Window

Cluster Position = Position of Highest Energetic TC

The Belle II upgrade

Belle II Upgrade

- Higher luminosity
- Higher background
- Too high trigger rates



Possible measures to handle the new conditions:

- higher TC energy threshold
- higher cluster energy threshold
- higher multiplicity triggers



Reduced performance for:

- Muon pair ECL endcap trigger (golden mode for Chiral Belle)
- inclusive energy sums (τ 1-vs-1, LFV $\tau \rightarrow e\ell\ell$)
- low multiplicity events

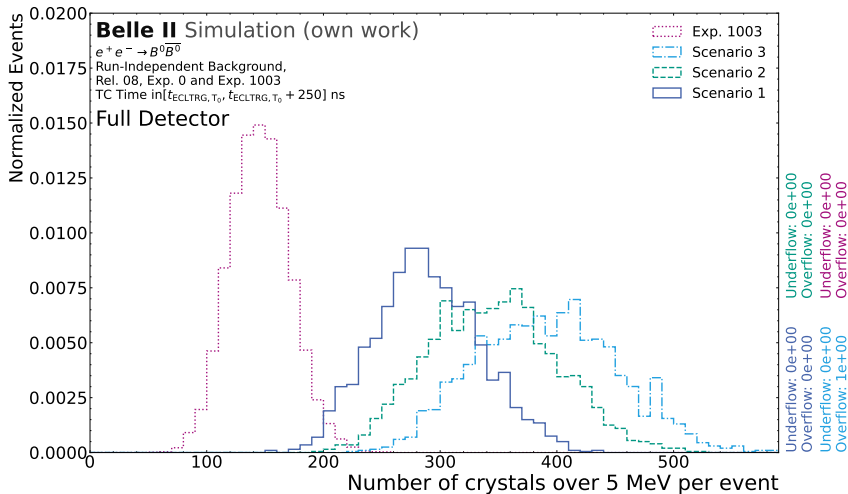
Belle II Upgrade - Background Conditions

3 different extrapolated scenarios:

- Scenario 1: "optimistic"
- Scenario 2: "realistic"
- Scenario 3: "pessimistic"

Based on conceptual design report

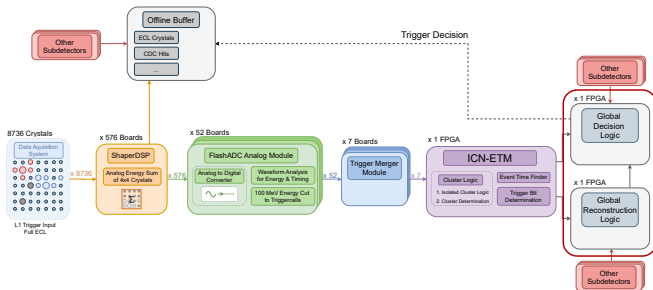
<https://arxiv.org/abs/2406.19421>



Possible Adjustment points in the ECL Trigger Pipeline to reduce the Trigger Rate

Possibilities to reduce trigger rate:

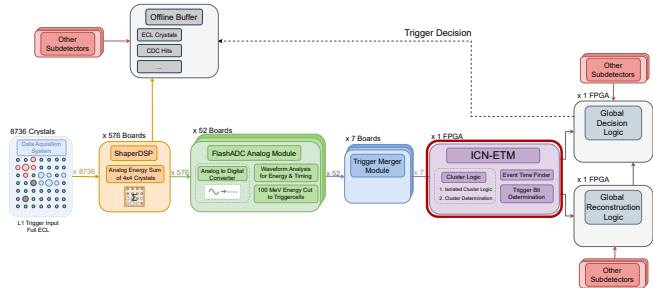
1. Add prescaling to trigger bits
2. Higher threshold for energy trigger bits
3. Higher multiplicity for cluster counting trigger bits
4. Replace ICN-ETM
5. Increase energy cut for TCs
6. Decrease timing window of 250ns
7. Higher granularity TCs



Possible Adjustment points in the ECL Trigger Pipeline to reduce the Trigger Rate

Possibilities to reduce trigger rate:

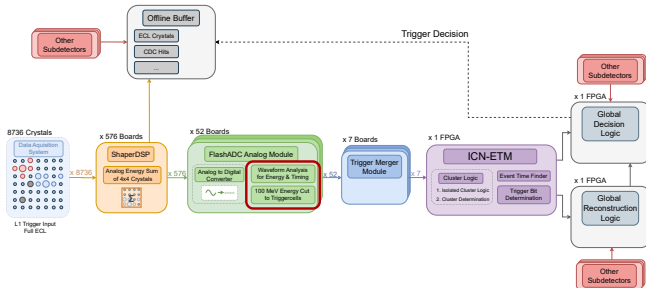
1. Add prescaling to trigger bits
2. Higher threshold for energy trigger bits
3. Higher multiplicity for cluster counting trigger bits
4. Replace ICN-ETM
5. Increase energy cut for TCs
6. Decrease timing window of 250ns
7. Higher granularity TCs



Possible Adjustment points in the ECL Trigger Pipeline to reduce the Trigger Rate

Possibilities to reduce trigger rate:

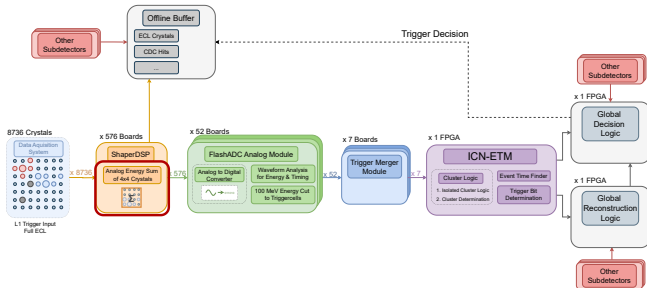
1. Add prescaling to trigger bits
2. Higher threshold for energy trigger bits
3. Higher multiplicity for cluster counting trigger bits
4. Replace ICN-ETM
5. Increase energy cut for TCs
6. Decrease timing window of 250ns
7. Higher granularity TCs



Possible Adjustment points in the ECL Trigger Pipeline to reduce the Trigger Rate

Possibilities to reduce trigger rate:

1. Add prescaling to trigger bits
2. Higher threshold for energy trigger bits
3. Higher multiplicity for cluster counting trigger bits
4. Replace ICN-ETM
5. Increase energy cut for TCs
6. Decrease timing window of 250ns
7. Higher granularity TCs

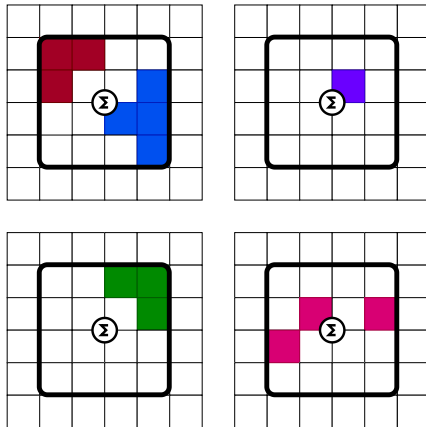


Higher Granularity

Adapting the granularity will influence the rest of the ECL Trigger Pipeline!

Potential of higher granularity trigger:

- cluster separation
- muons in ECL endcap
- improved position resolution
- shower shape analysis for background rejection



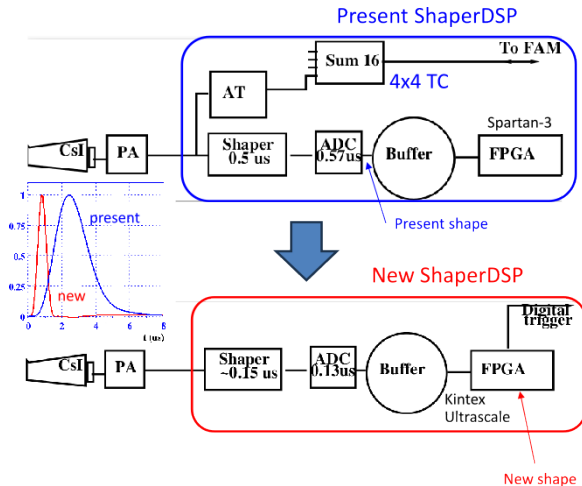
ECL (TRG) Upgrade Plans

ECL Upgrade Plans:

- Upgrade of shaper boards for single crystal shaping (TRG and DAQ)
- Preshower detector in front of the ECL
- Complementing the PiN diode photosensors with avalanche photodiodes
- Complementing the PiN diode photosensors with Silicon Photomultipliers (SiPM)

Additionally: Longer time window of $10\mu\text{s}$ for full system buffer

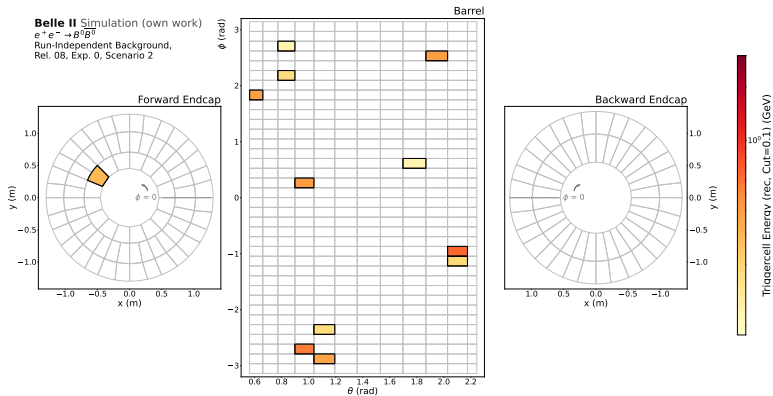
<https://arxiv.org/abs/2406.19421>



Graphic by A. Kuzmin

EventDisplay: Low Granularity

- Current input for the ECL trigger module
- Energy sum to 4x4 TCs
- Timing window of 250ns

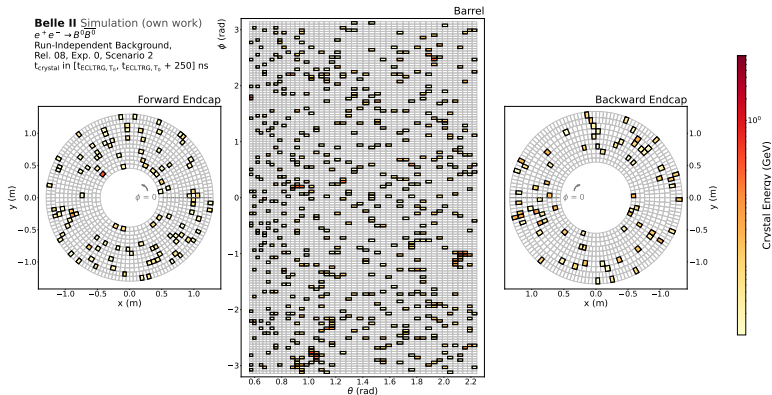


EventDisplay: High Granularity (in same 250ns)

Much more input!

Possibilities to reduce:

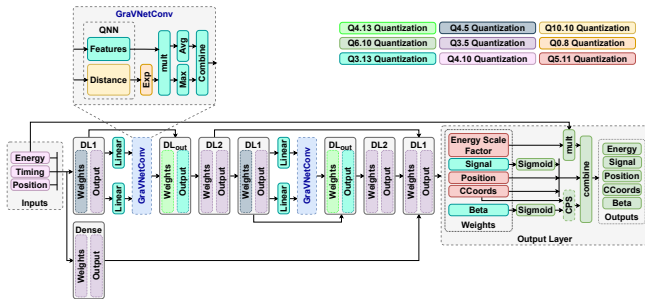
1. Coarser granularity (e.g. 2x2)
2. Segmentation of the detector
3. "low" energy cut on crystals
4. "high" energy cut on crystals with region of interest around them



GNN-ETM

GNN-based clustering algorithm for the current system (more detail in I. Haides Talk).

- Transformation into latent space + features
- GravNet: The learnt features are weighted and aggregated
- Input and output are concatenated and passed further on
- Output: classic predictions + CCoords and Beta



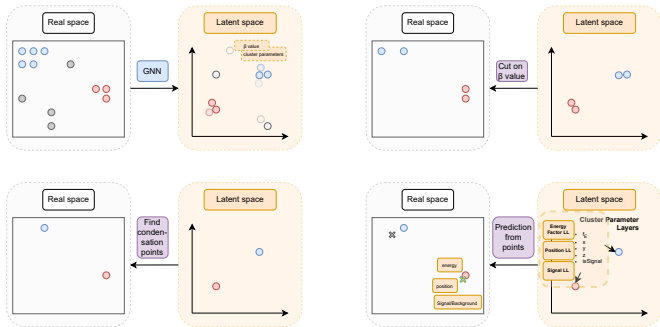
Object Condensation

Object Condensation:

- Apply a cut on the β -value, to remove background
- Sort by β -value and apply isolation criteria in order on the latent space representation
- The remaining points are Condensation Points
- The prediction of the Condensation Point represents the whole real space object

Advantages:

1. Arbitrary amount of cluster predictions possible
2. Inherent Background reduction
3. Reduced accuracy requirement for every individual input
4. Improved separation of clusters in latent space compared to real space



Object Condensation can also be used for track finding in the CDC <https://arxiv.org/abs/2411.13596>

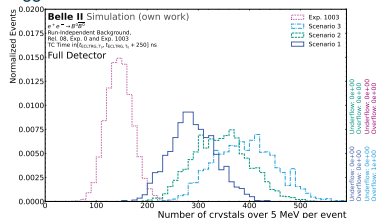
GNN-ETM Utilization of system resources on the current trigger board

(Already shown in I. Haides talk)

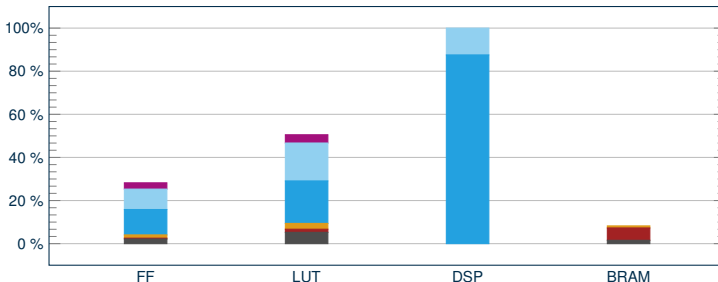
Configuration:

- AMD Ultrascale XCVU190 FPGA
- up to 32 inputs
- Latency of less than $3.1\mu\text{s}$
- Throughput of 8MHz

Problem: Not possible to increase the model size to 128 or 256 inputs on current trigger board



- Base Firmware
- B2Link Subsystem
- Preprocessing Stage
- Dense
- GraVNetConv
- Condensation Point Selection



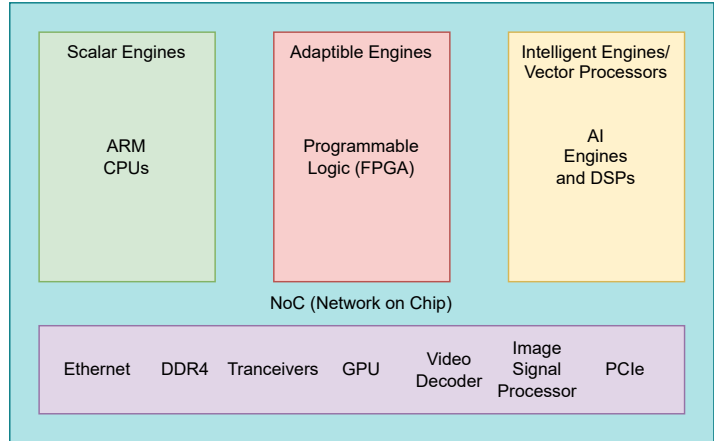
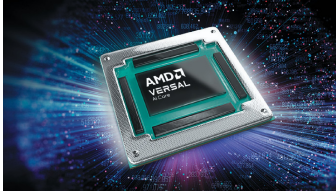
The Tasks at Hand

Adapt working GNN-ETM for high granularity input

Implement a version of the GNN-ETM with 128 or 256 inputs

Versal

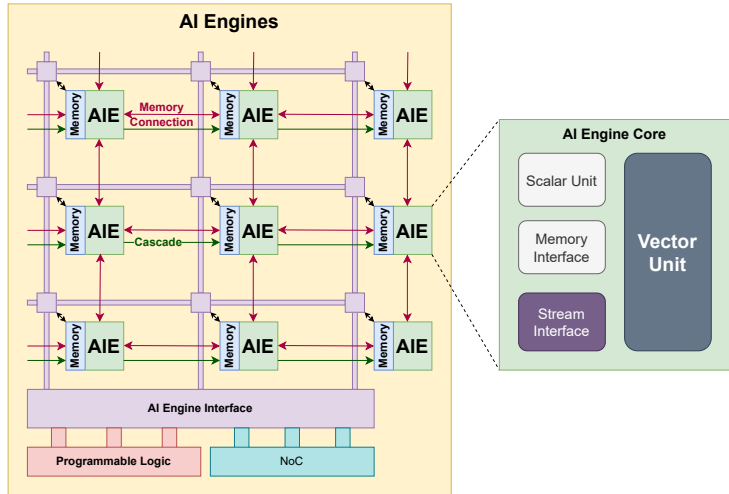
- Combination of CPUs, FPGA and vector processors on single chip design
- CPU irrelevant for our use case
- Combined advantages of classic FPGAs and ISA-based VLIW processors ("GPU-like")
- significantly reduced overhead compared to real GPUs



AI Engines

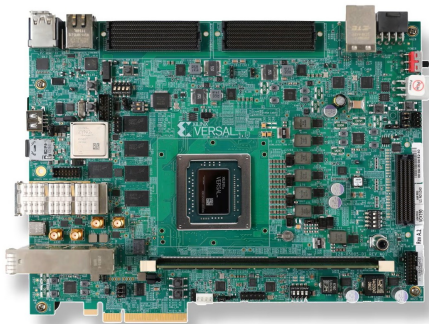
Good at matrix multiplications

- Initialization via NoC
- Direct streaming from PL to AI Engines via PLIOs
- Latency overhead in transitions between PL and AI Engines
- PL data reordering kernels required
- Very complex programming, timing, routing, ...



Versal VCK-190

The VCK-190 is not the next to next generation trigger board, but a test board to validate the potential of Versal boards.



Component	Currently deployed Trigger Boards	next generation Trigger Boards	VCK-190*
System Logic Cells	1M-2M	4M-7M	2M
LUTs	500k-1M	1.7-3.3M	900k
DSP Engines	700-1.8k	7k-14k	2k
AI Engines	0	0	400

* Rather small FPGA, but 400 AI Engines

<https://www.amd.com/en/products/adaptive-socs-and-fpgas/evaluation-boards/vck190.html#tabs-a75507a83a-item-df61ba4d87-tab>

Proof of Concept: GNN-ETM on Versal

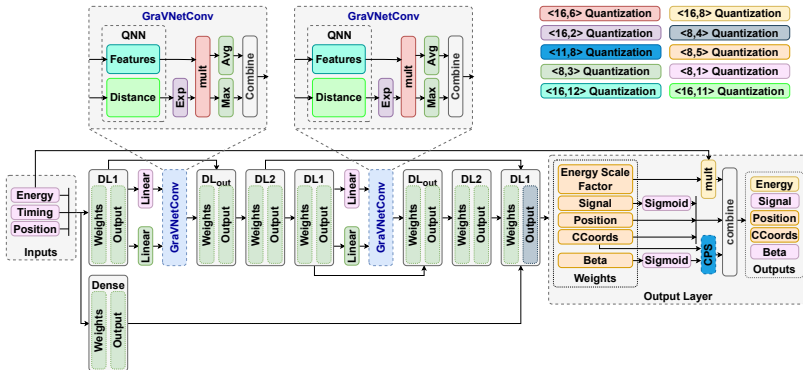
Design Decisions:

- 128 inputs
- 10 μ s latency
- 8-bit AI Engine Design
- 2 GravNet Blocks
- 7 nearest Neighbours

Limiting Factors:

- AI Engine: No dynamic data type conversion possible. float (not feasible), 16-bit fixed or 8-bit fixed (newer models also 4-bit)
- Domain transitions of the GravnetBlocks ~500ns in each direction

Implementation nearly finished!



Proof of Concept: GNN-ETM on Versal

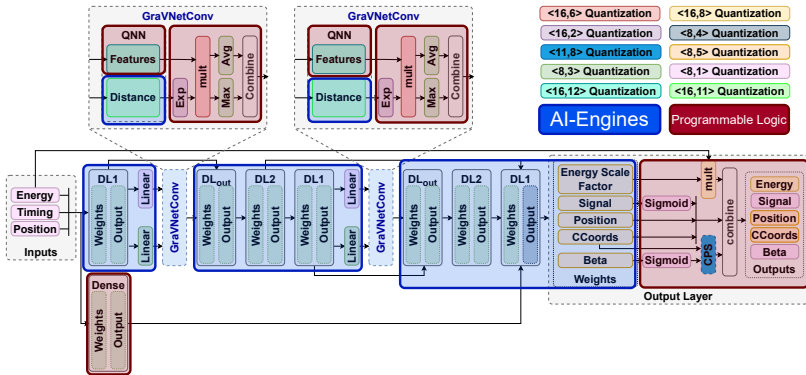
Design Decisions:

- 128 inputs
- 10 μ s latency
- 8-bit AI Engine Design
- 2 GravNet Blocks
- 7 nearest Neighbours

Limiting Factors:

- AI Engine: No dynamic data type conversion possible. float (not feasible), 16-bit fixed or 8-bit fixed (newer models also 4-bit)
- Domain transitions of the GravnetBlocks ~500ns in each direction

Implementation nearly finished!



Training

We train on low beam background conditions, to start this studies. We know that on low beam background and with 4x4 TCs GNN-ETM is a valid alternative to ICN-ETM (as shown by I. Haide).

Training Sample:

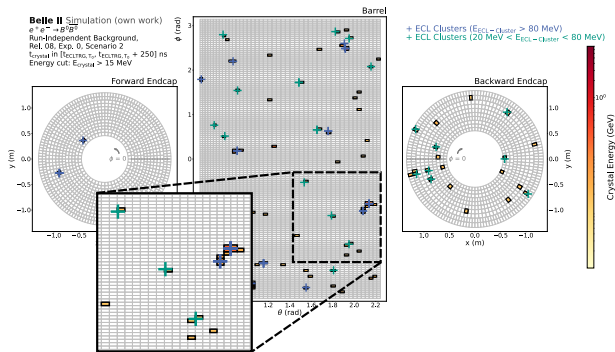
1. Exp. 1003 MC Photon sample: with 1-6 energetically uniformly distributed photons and additionally a numerically Poisson distributed background, with an exponential energy.
2. Second sample with same properties and additional added photon pair with an opening angle smaller than 0.2 rad.

Training Targets:

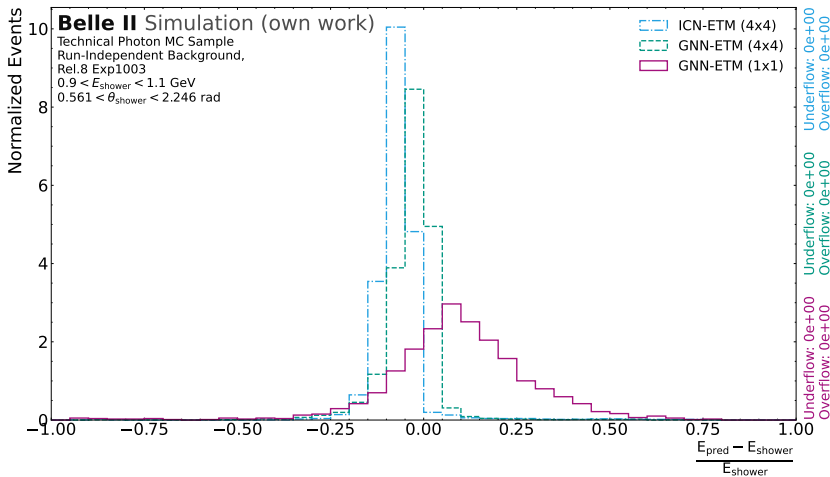
- ECL cluster objects from Basf2 reconstruction, with in the 250ns trigger window, above 80 MeV

Input reduction:

- Crystal energy cut of 15 MeV



Preliminary Model Performance



Outlook and Conclusion

Belle II Upgrade

- Higher Luminosity
- Higher Background
- Higher ECL Trigger Granularity

1x1 ECL Trigger Algorithm

- Optimization of first model
- Higher background performance
- Other model architectures and input reduction methods

Implementation on Versal

- Implementation almost finished
- Increased Inputsize
- Much Information Gain for future codesign