GNN-based Reconstruction for LHCb Upgrade II ECAL

Felipe Luan Souza de Almeida

20/11/2025

XVII CPAN DAYS

COMCHA
Computing challenges

PicoCal

UNIVERSITAT DE BARCELONA

erc European Research Council
Established by the European Commission

ERC GA: 101077940

LHCb

ICCUB Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA

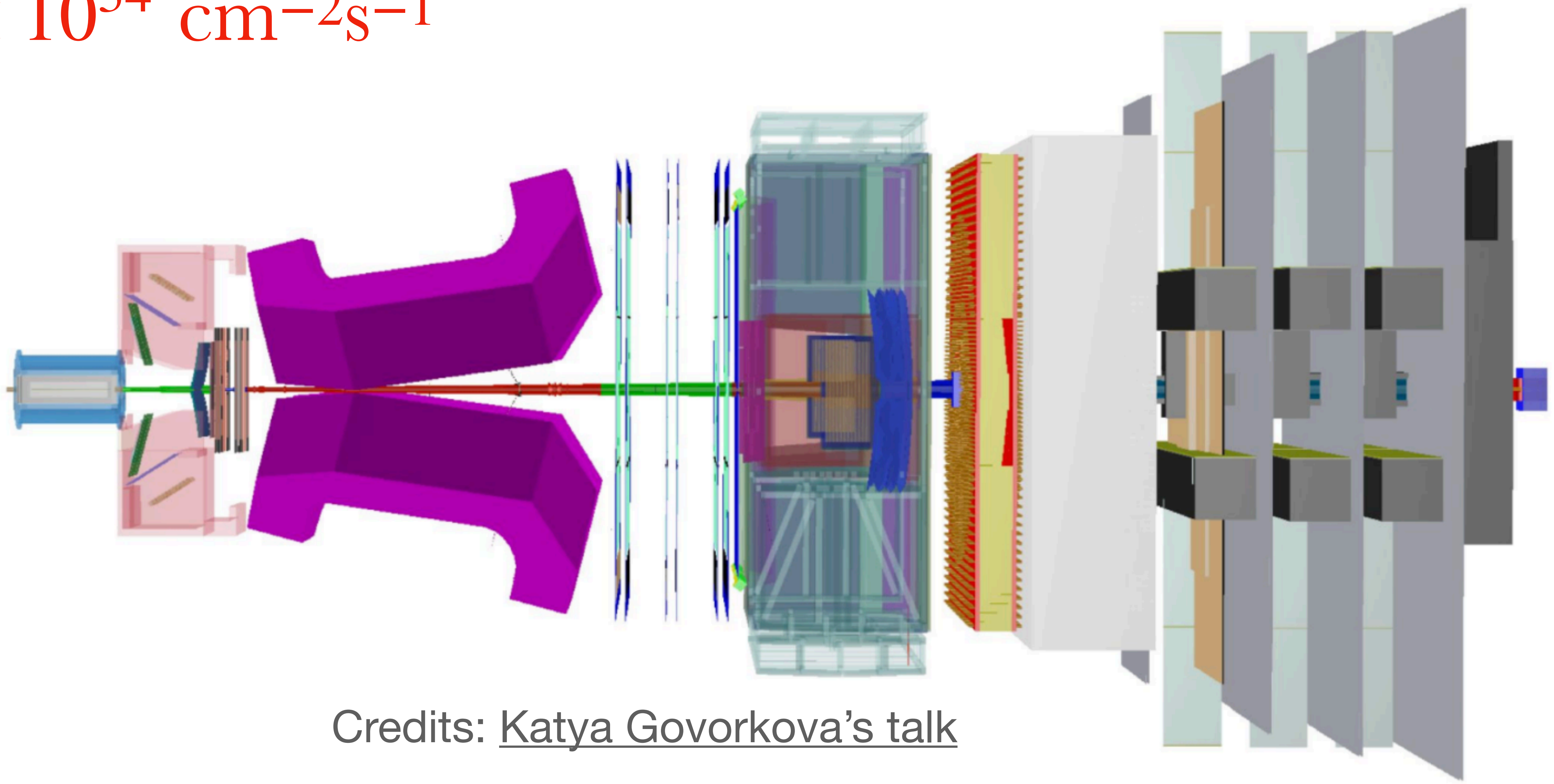
EXCELENCIA MARÍA DE MAEZTU
04/2025-03/2031

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

AGENCIA ESTATAL DE INVESTIGACIÓN

The LHCb Upgrade II

- The LHCb Upgrade II redesigns the LHCb to operate at a much higher instantaneous luminosity: $1.5 \times 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$

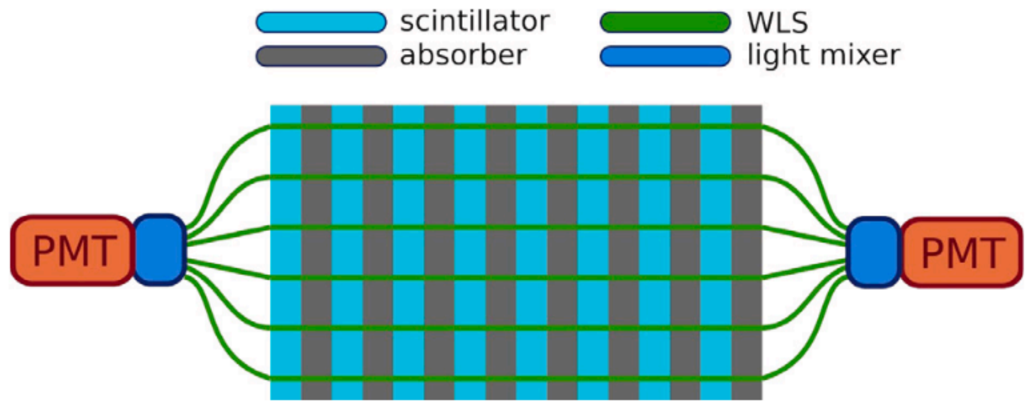


Credits: Katya Govorkova's talk

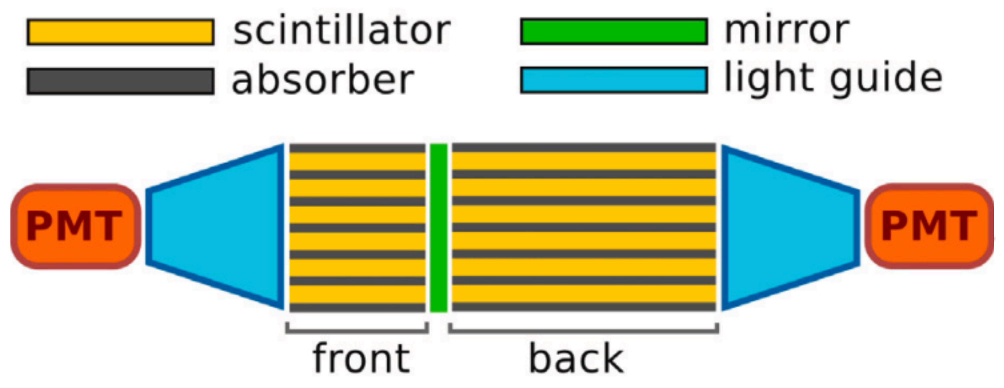
- The PicoCal is the next generation of the electromagnetic calorimeter
  - γ, e^-, π^0 reconstruction
  - Includes timing information with O(10) ps precision

The PicoCal detector

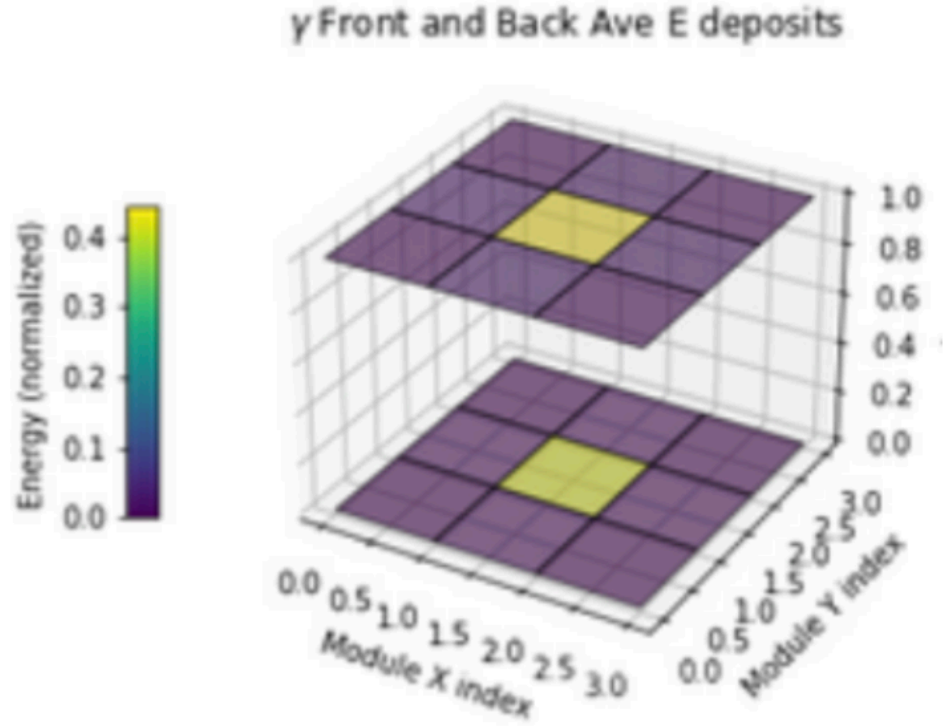
- Consists of thousands of modules combining absorber material with scintillating fibers, either plastic or crystal
  - Shashlik and SpaCal module types foreseen
  - Each module is subdivided into a fixed number of readout cells (ranging from 1 to 64)
  - The layout is further segmented in front and back sections

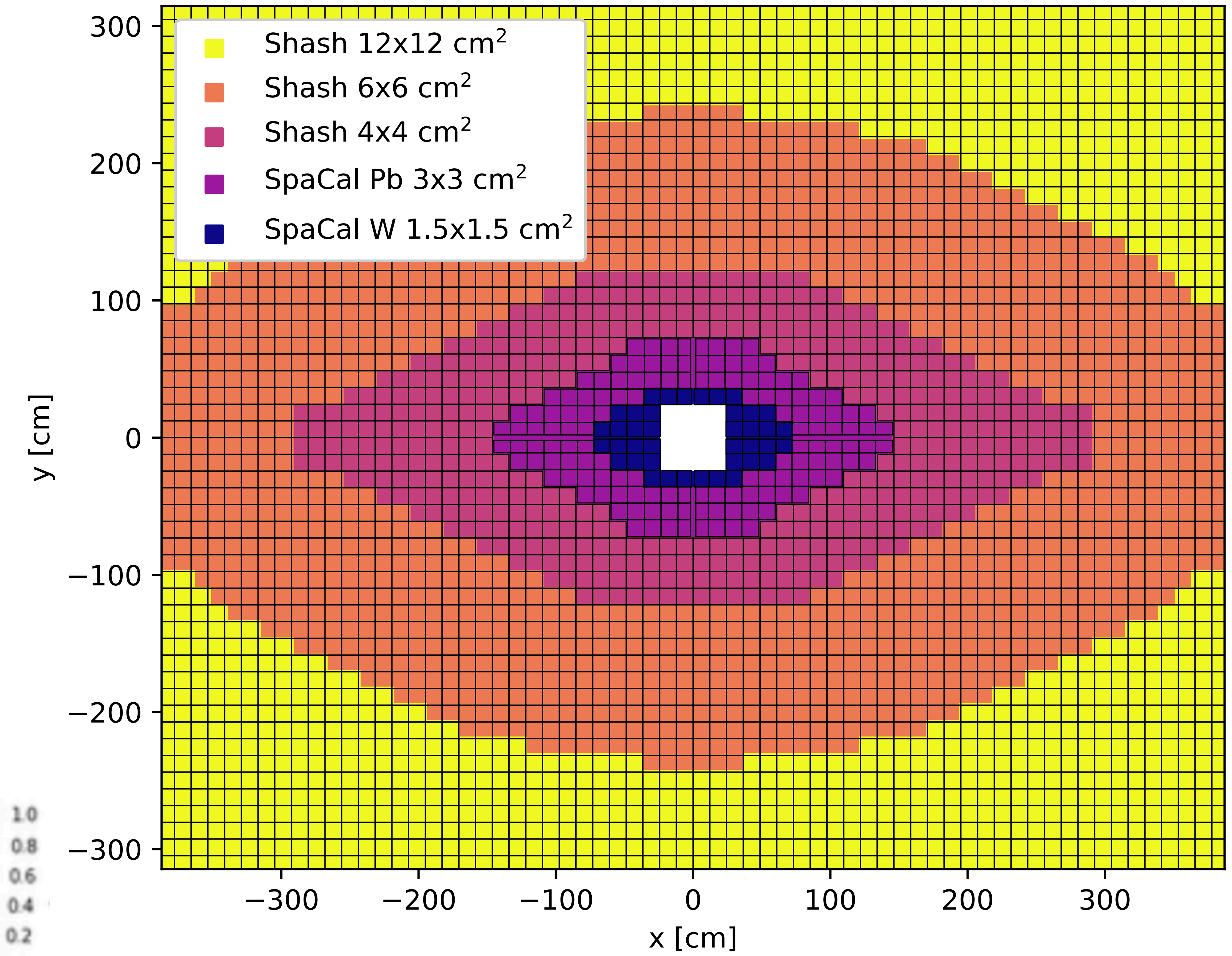


SpaCal



Shashlik





Five regions of increasing granularity, segmented in front and back sections

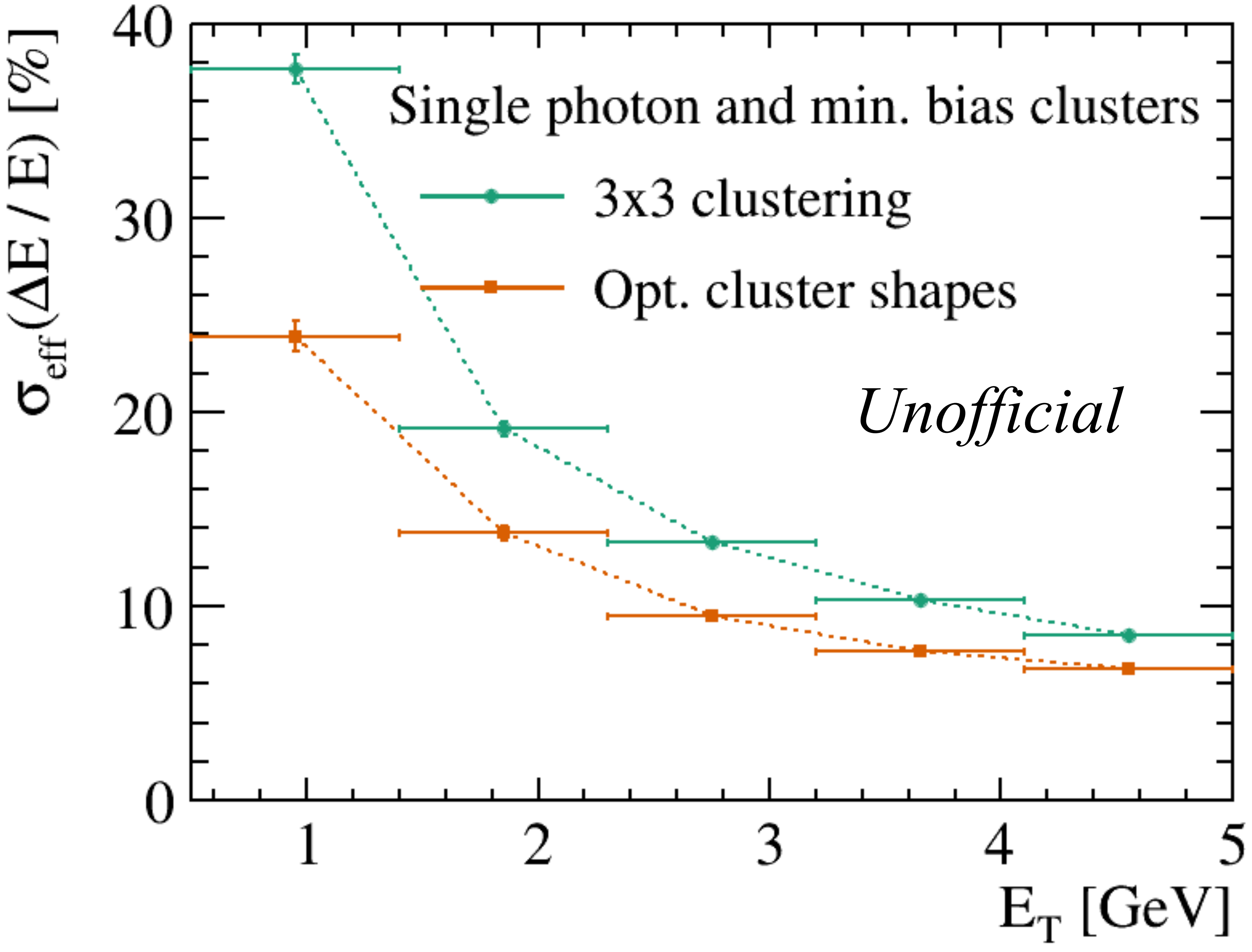
Detector modelling, standard reconstruction

- Detector modelling
  - Geant4 + Hybrid MC framework for detector response
  - Tools for pulse-shape analysis and timing extraction
- "Standard reconstruction approach" based on the Celular Automaton clustering model
  - Used in LHCb Runs 1 & 2
  - Find local maximum (seed) → form 3x3 cluster → resolve overlaps → apply corrections
- Improved variant:
  - Optimized cluster shapes per ECAL region (e.g. cross-shape)

The problem

- In the harsh environment of LHCb Upgrade II, the pile-up contribution severely degrades the energy resolution
  - With the standard reconstruction algorithms and the baseline detector configuration, the energy resolution remains poor, specially for low E_T
  - Interesting physics at low E_T
    - $D^0 \to \pi^-\pi^+\pi^0$, spectroscopy with neutrals, …
    - Modes with π^0 for CKM γ, penguin pollution studies for CKM β, …
  - How to improve? Better detector or better reconstruction!
    - Better detector → more money!

Energy resolution on simulated single photons in Run 5 conditions



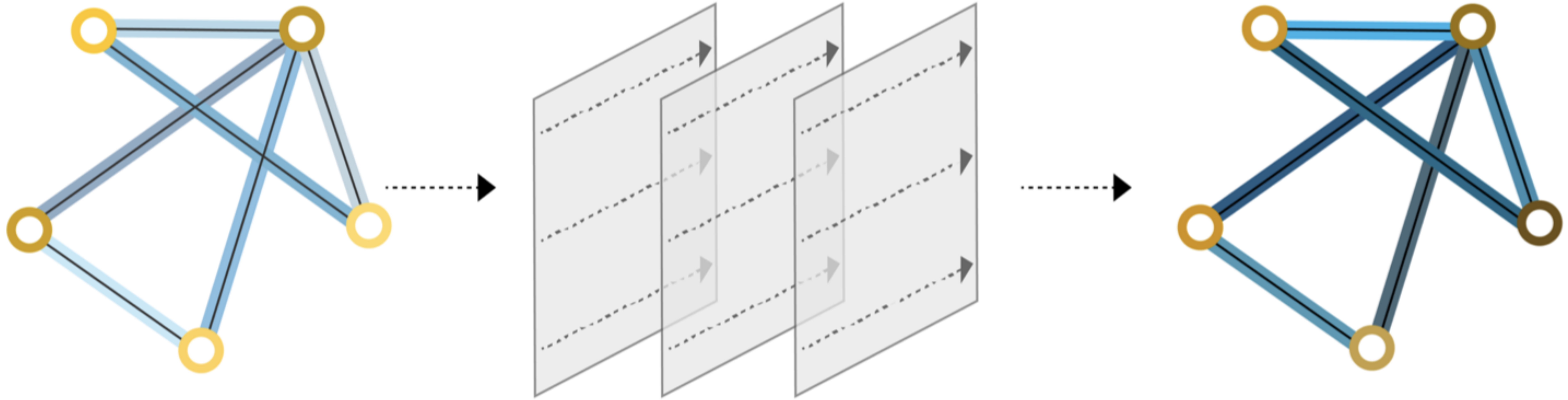
σ_{eff}: half-width of the 68% central interval

ML approaches

- Idea: explore ML-based approaches for the reconstruction
  - Predict energy, position, timing and particle ID of incoming particles hitting the ECAL
  - Supervised regression problem using features extracted from the ECAL cells
- Previous studies by other LHCb groups explored BDTs, DNNs, CNNs, …
  - But each of these require concessions to handle irregular geometries
  - E.g. CNNs: treats energy reconstruction as image recognition problem — how to deal with clusters of cells of different sizes? Merge smaller cells? Split larger cells?
  - Eg. Feed-forward NN: ordering matters — should we order the cells by deposited energy? Dimensionality also matters: each cluster must have the same number of cells

GNNs for calorimeters

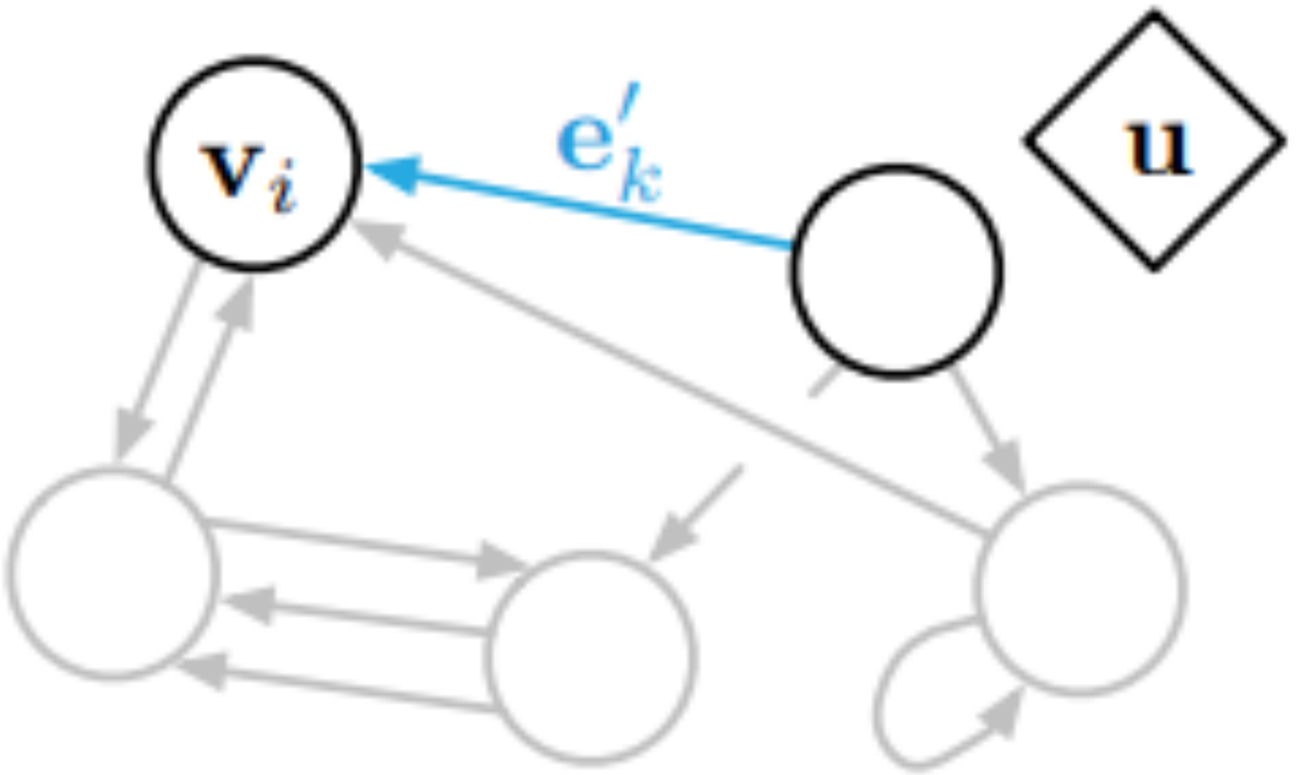
- Inspired by Belle II: GNN reconstruction [Comput Software Big Sci 7, 13 (2013)]
  - GNNs do not require the same concessions to handle irregular geometries
  - Clusters represented as graphs: cells as nodes, edges describing relationships among cells (e.g. spatial distance), and global features for the cluster (e.g. total energy)



The message passing algorithm

- Nodes, edges, and globals are updated through aggregation and MLPs

BLUE updated by BLACK (not utilizing GREY)



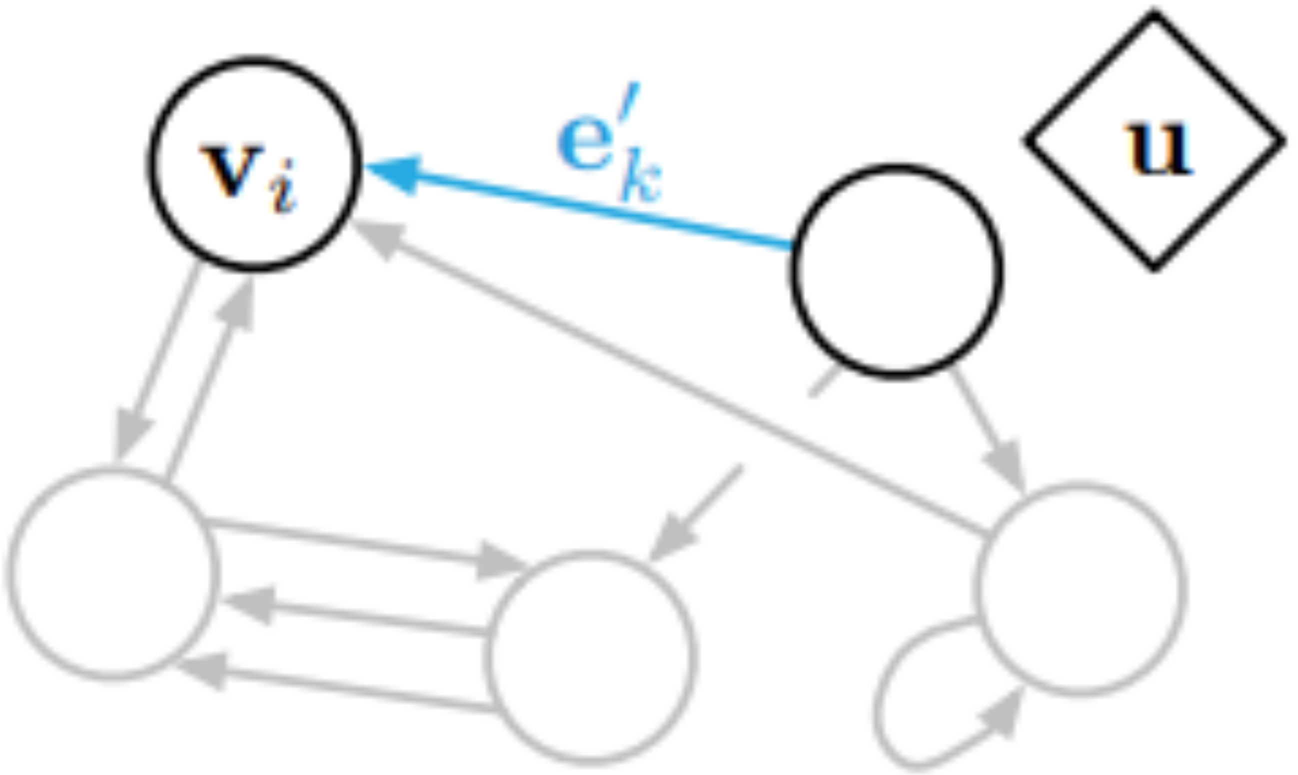
(a) Edge update

Update edges using
connected nodes and globals

The message passing algorithm

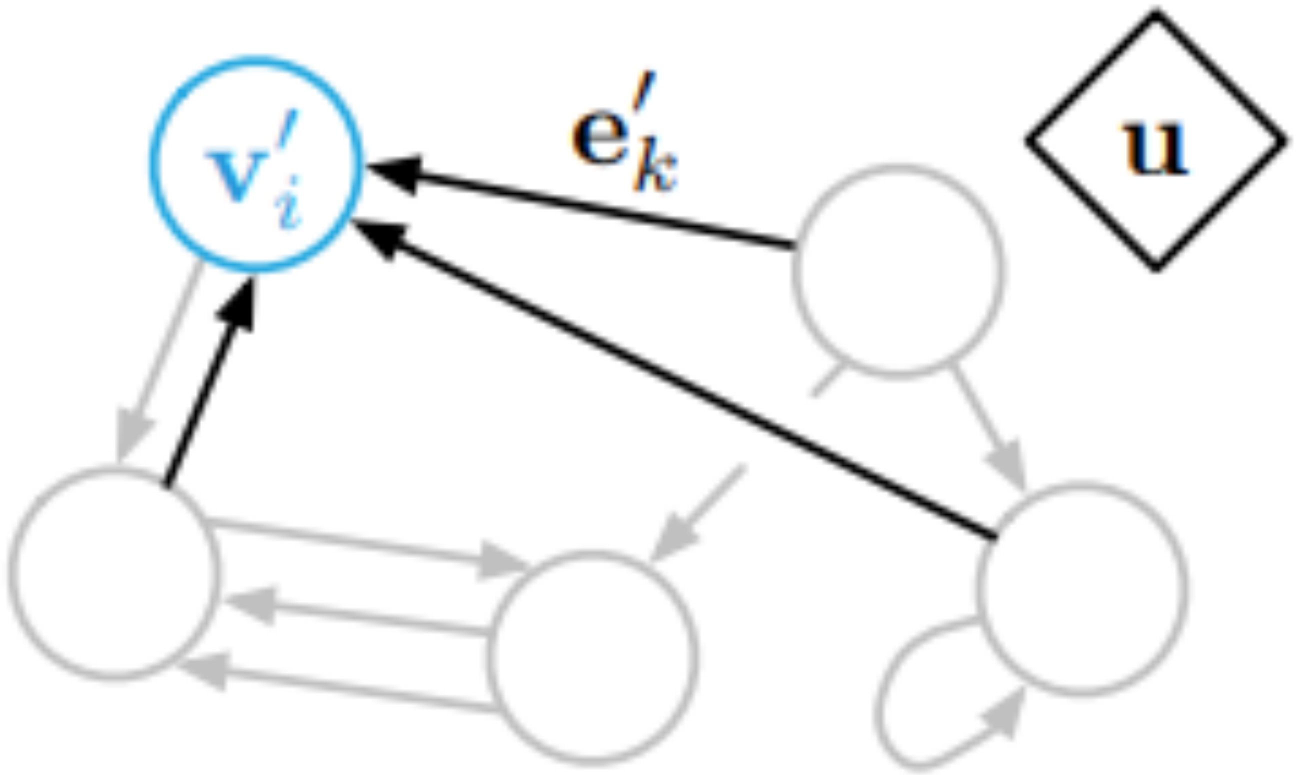
- Nodes, edges, and globals are updated through aggregation and MLPs

BLUE updated by BLACK (not utilizing GREY)



(a) Edge update

Update edges using
connected nodes and globals



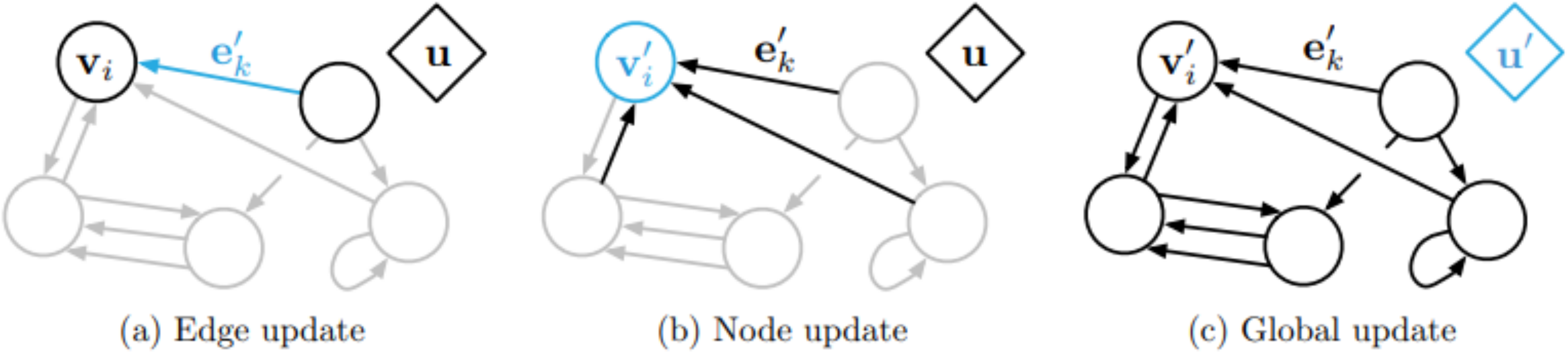
(b) Node update

Update nodes
using (new) edges
and globals

The message passing algorithm

- Nodes, edges, and globals are updated through aggregation and MLPs

BLUE updated by BLACK (not utilizing GREY)



Update edges using connected nodes and globals

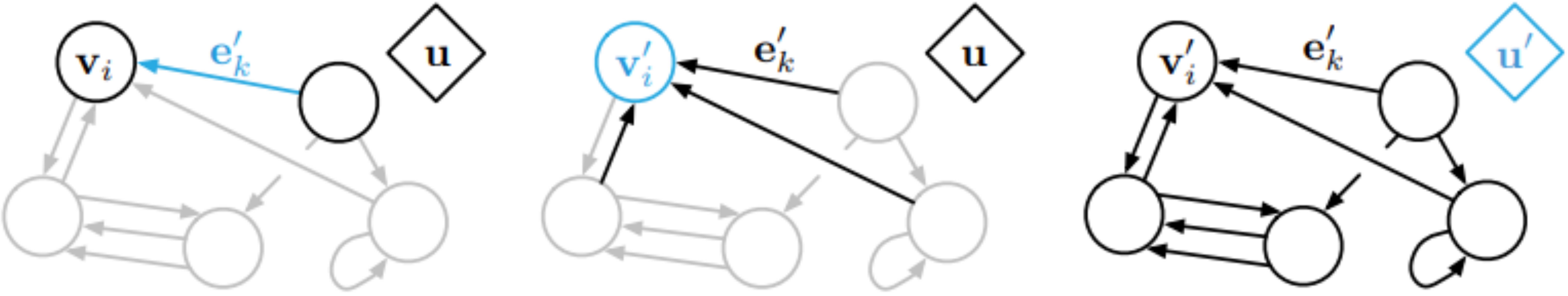
Update nodes using (new) edges and globals

Update globals using (new) nodes and (new) edges

The message passing algorithm

- Nodes, edges, and globals are updated through aggregation and MLPs

BLUE updated by BLACK (not utilizing GREY)



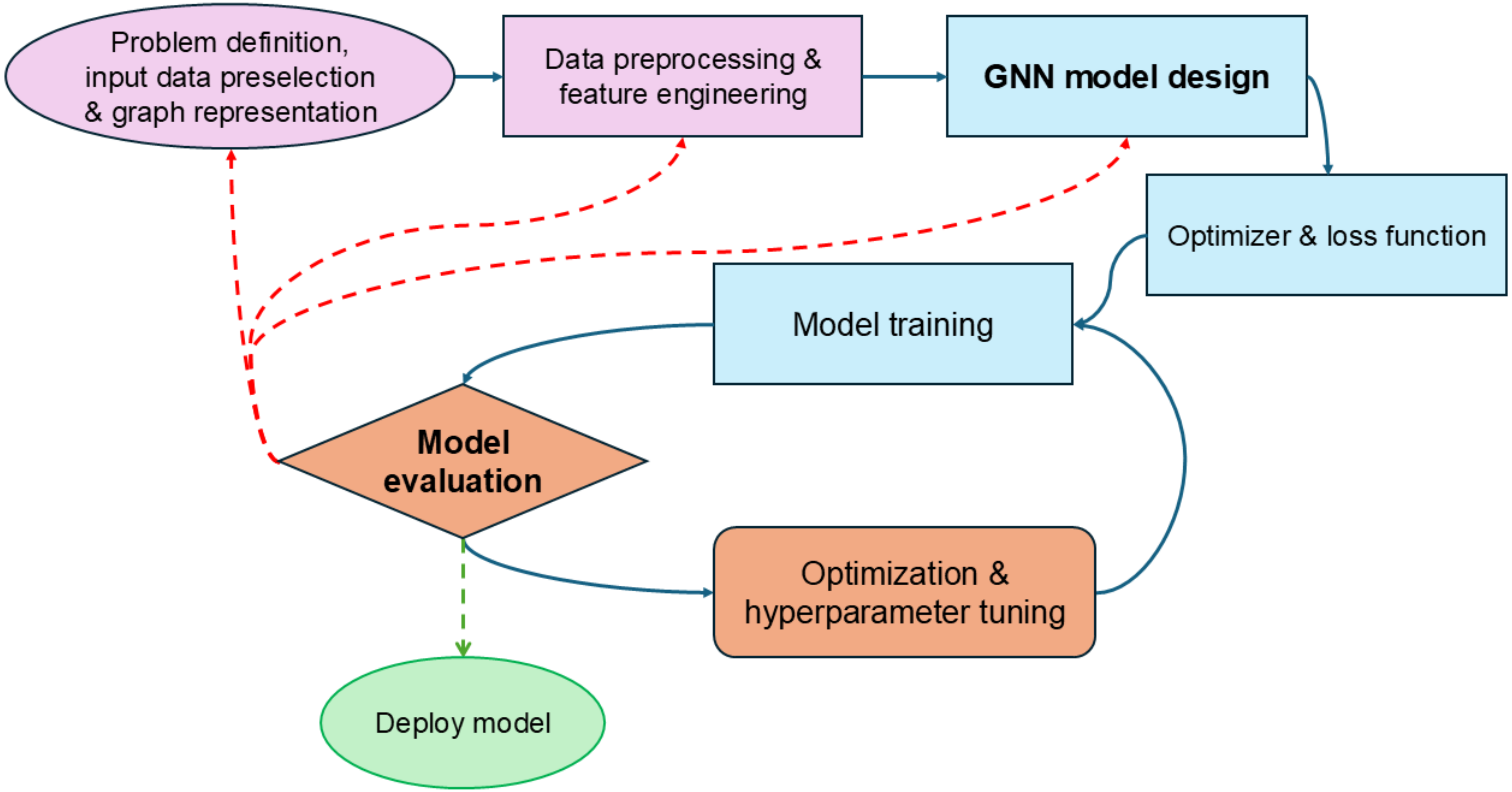
(a) Edge update — Update edges using connected nodes and globals

(b) Node update — Update nodes using (new) edges and globals

(c) Global update — Update globals using (new) nodes and (new) edges

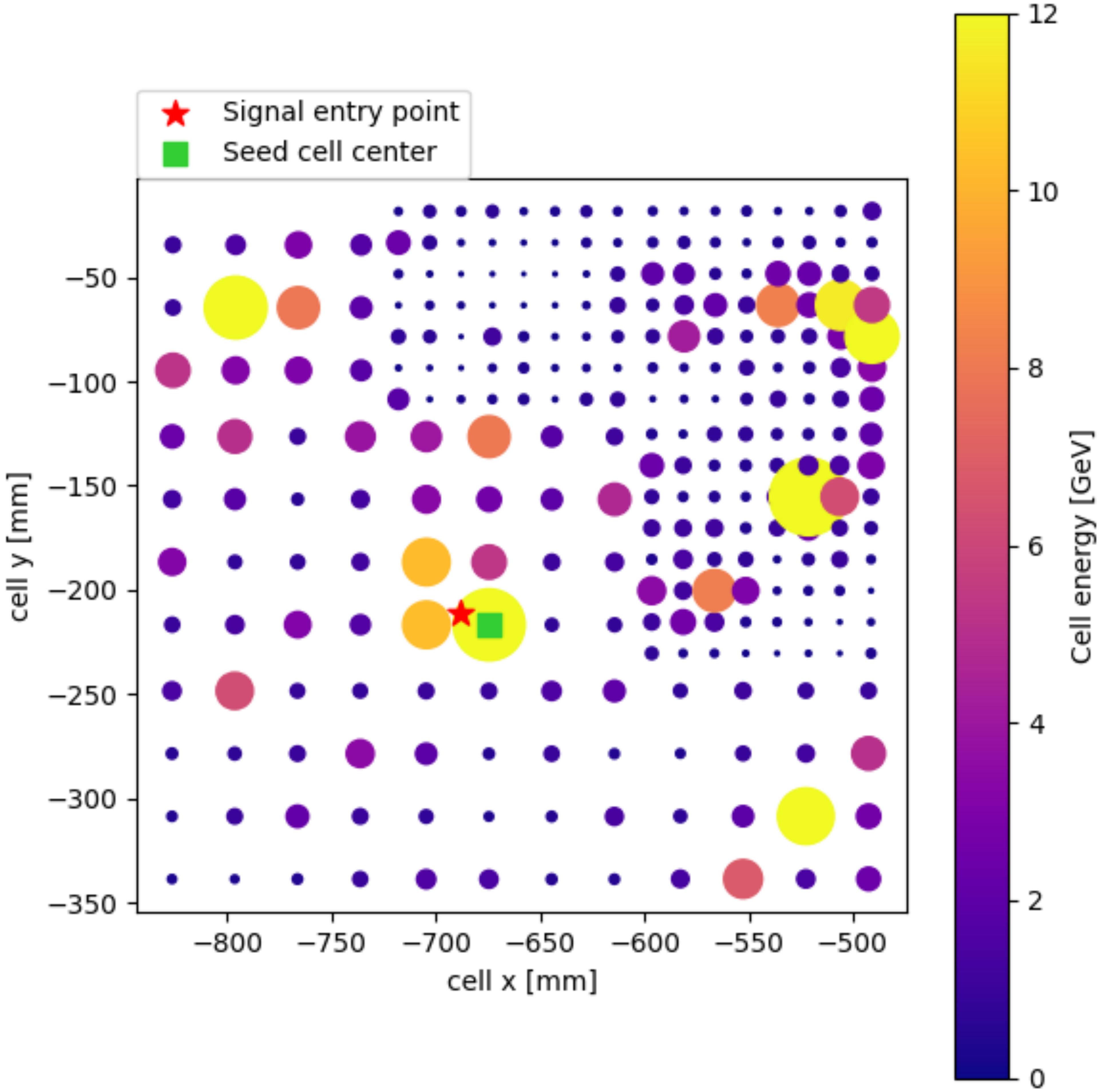
This process can be repeated many times
Each iteration can have a unique set of NNs

Workflow


Problem definition, input data preselection & graph representation
Data preprocessing & feature engineering
GNN model design
Optimizer & loss function
Model training
Model evaluation
Optimization & hyperparameter tuning
Deploy model

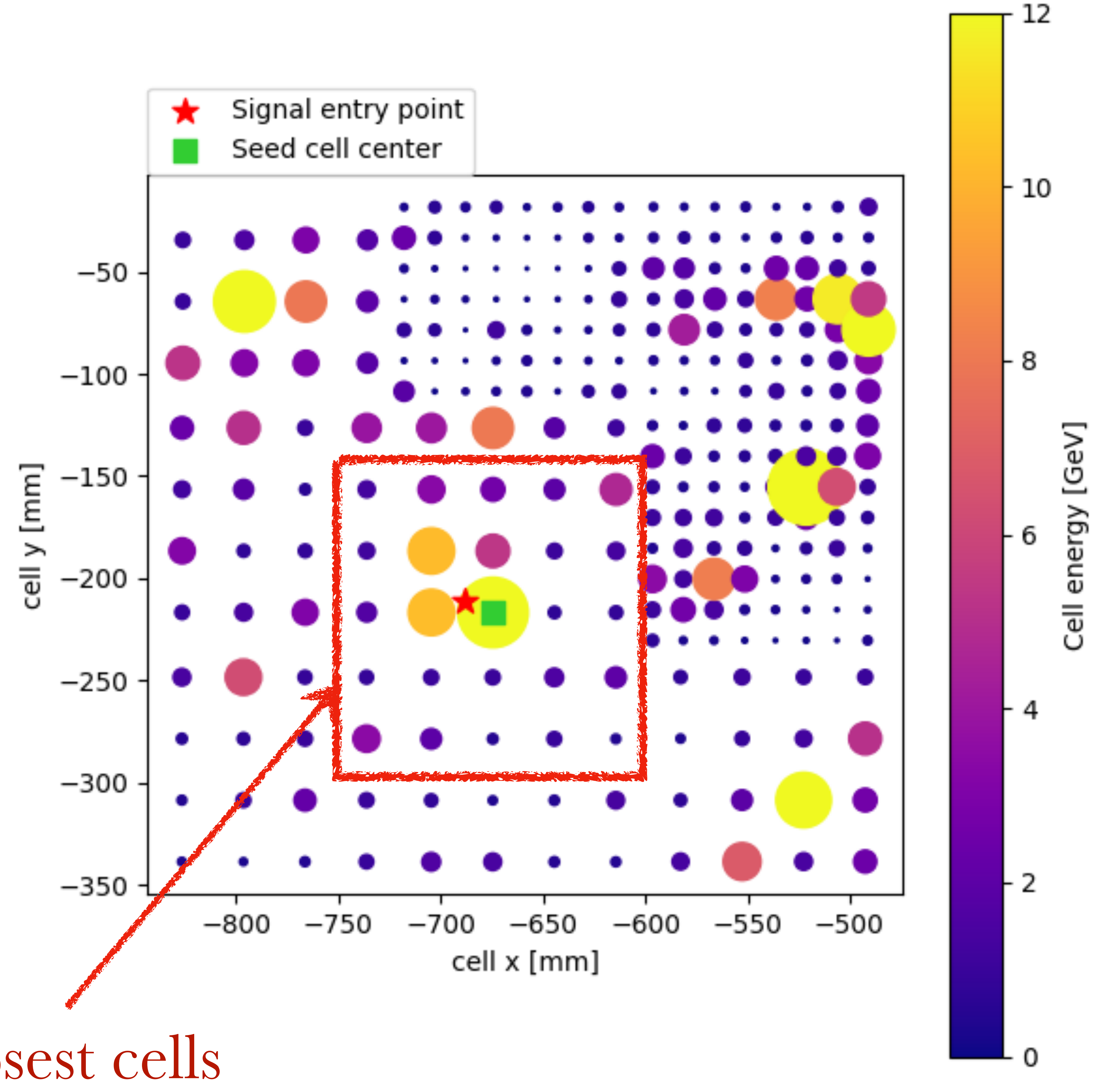
Our approach

- Training sample
  - Single photons in the particle-collision conditions expected for LHCb Run 5
  - $0.5 \leq E_T \leq 5$ GeV photons
- Preselection
  - Photons produced within 100 mm of interaction point
  - Clusters with seed in the central innermost region of the ECAL
  - 5x5 modules around the seed module, reduced to 25 central cells

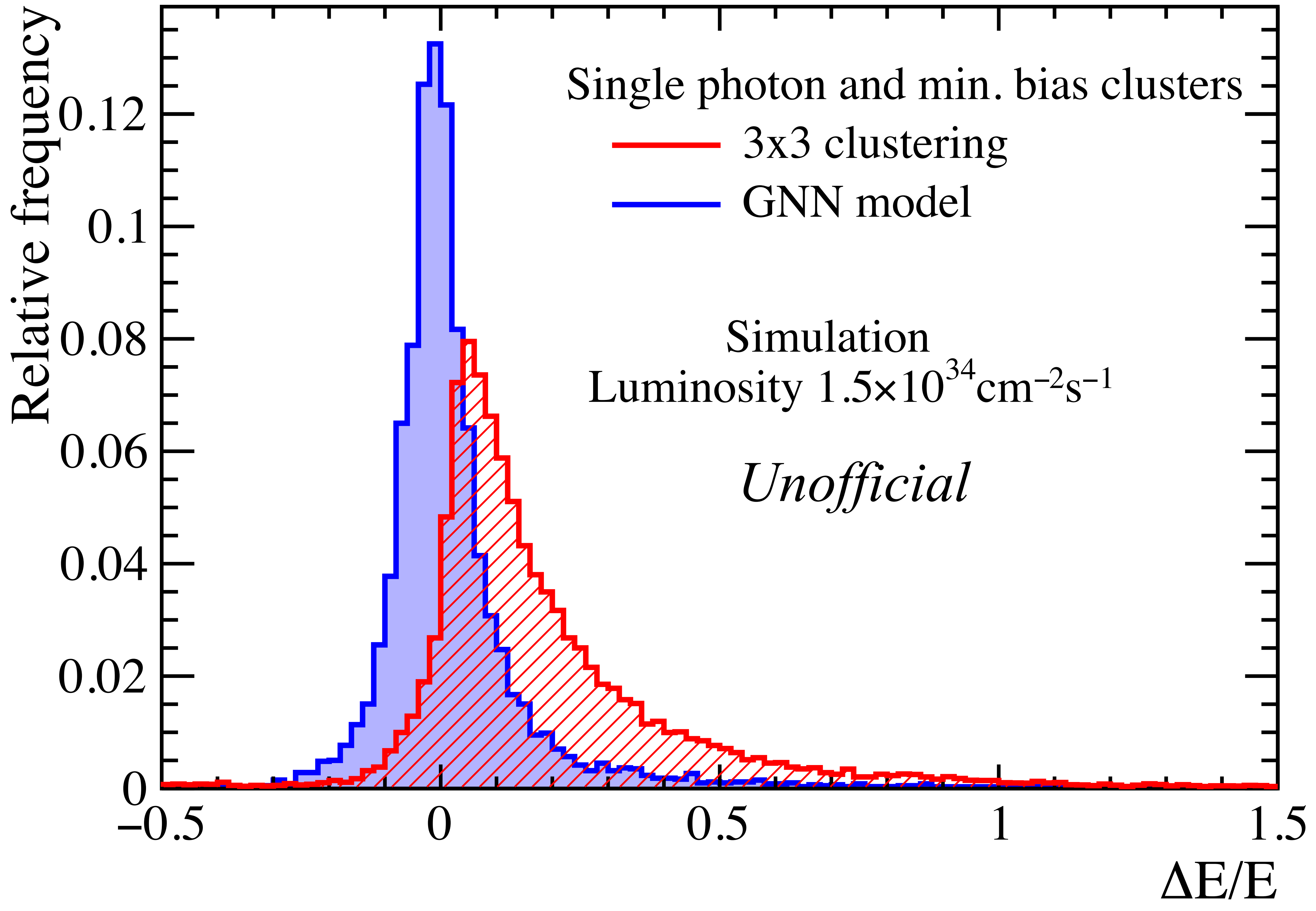


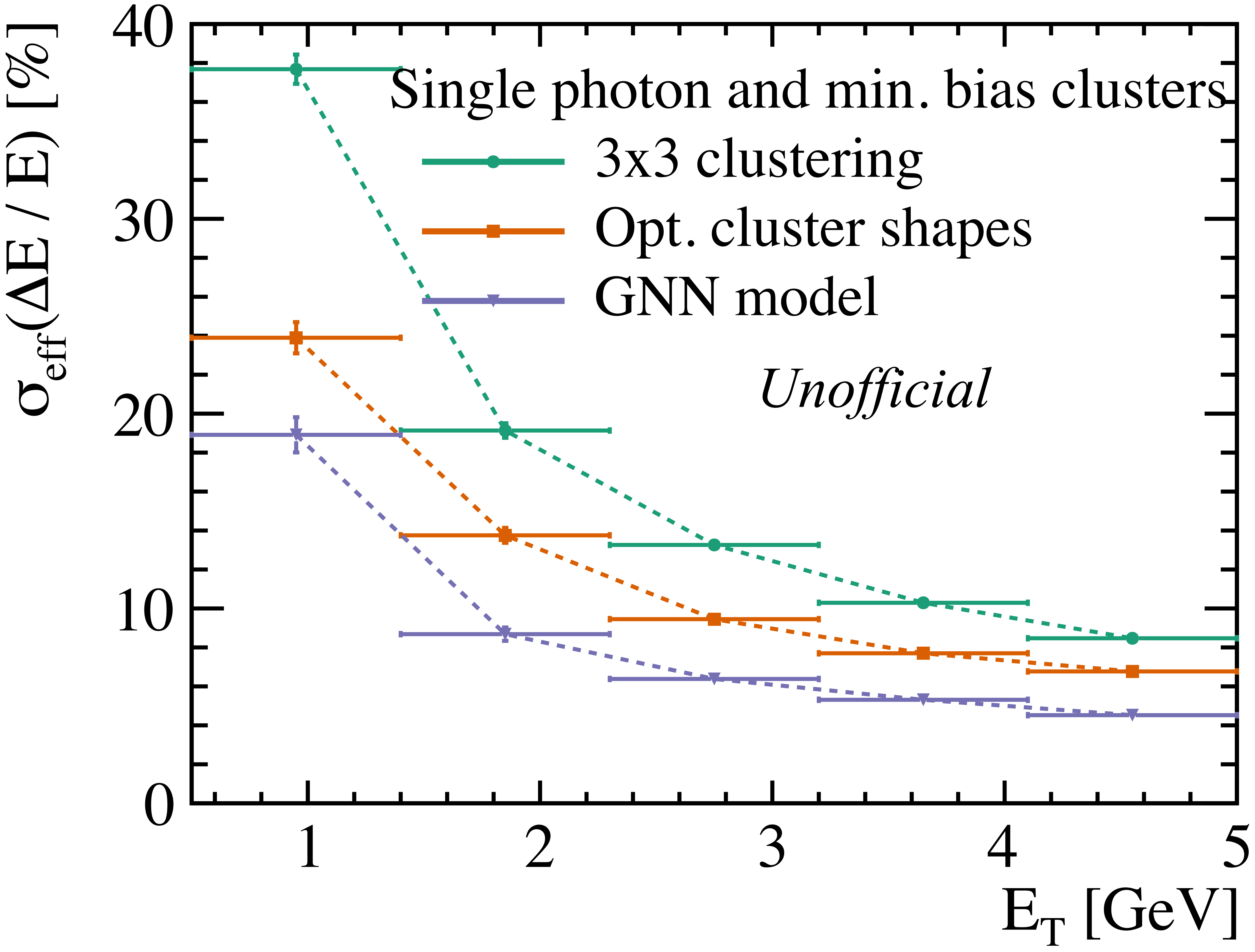
Our approach

- Training sample
  - Single photons in the particle-collision conditions expected for LHCb Run 5
  - $0.5 \leq E_T \leq 5$ GeV photons
- Preselection
  - Photons produced within 100 mm of interaction point
  - Clusters with seed in the central innermost region of the ECAL
  - 5x5 modules around the seed module, reduced to 25 central cells



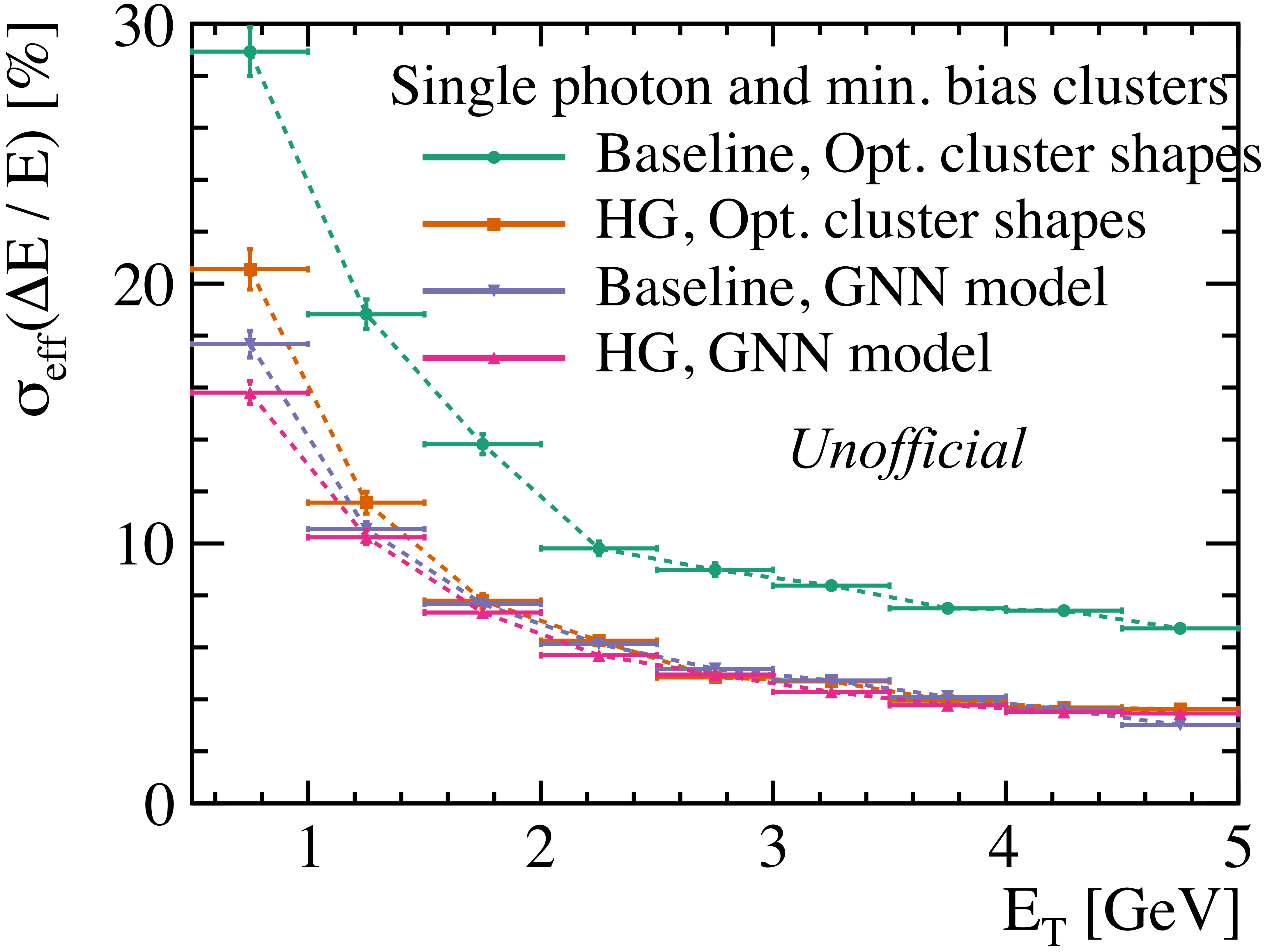
Results





The GNN outperforms the standard approaches over the full E_T range

Higher-granularity alternative



- The GNN achieves better performance than the standard approach even with lower granularity
- Not much improvement observed in the GNN with higher transverse granularity configuration
- But it further demonstrates that the GNN can handle irregular geometries
- Further improvements might be achievable with geometry-aware features

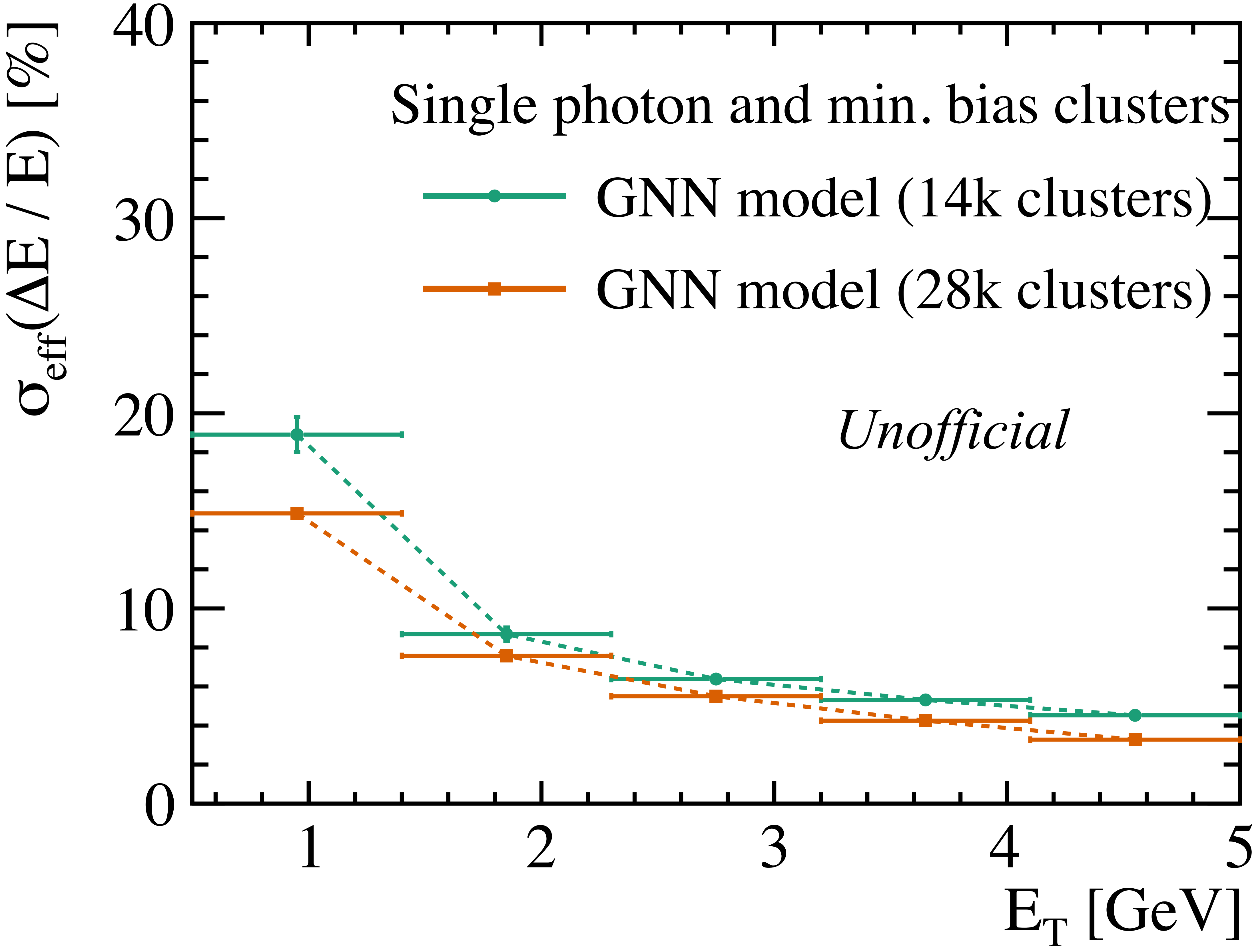
Conclusions

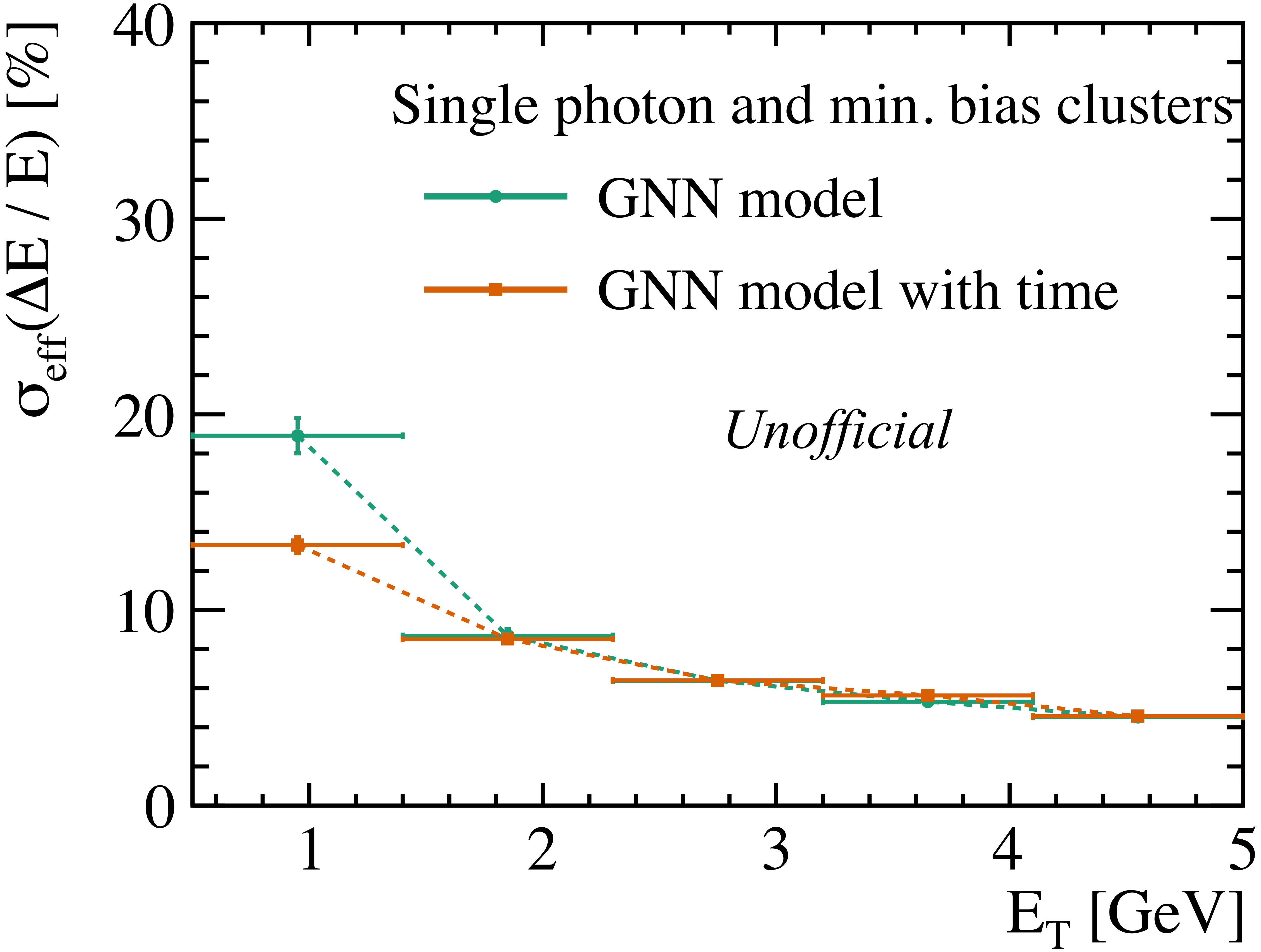
- We successfully developed novel GNN architectures for PicoCal reconstruction
  - Challenging yet highly promising development
  - The approach already outperforms standard reconstruction methods
  - Further improvements expected as new models, features and strategies are explored
  - Active ongoing research
- Flexible PyTorch implementation
  - Currently being tested for LHCb Runs 3&4
    - Further development required for real-time deployment — Check Uzziel's talk!

Thank you!

Backup

Further improvements

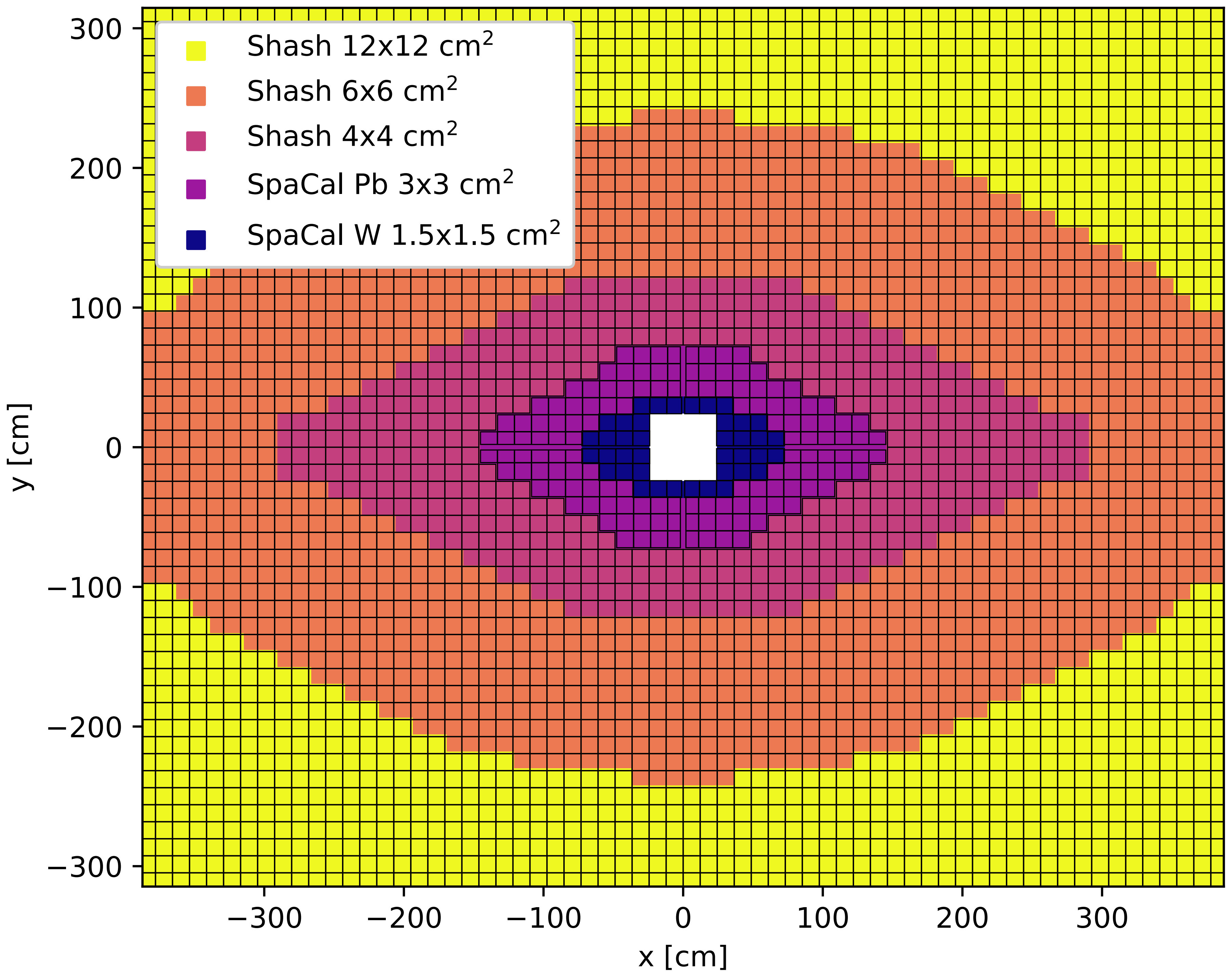




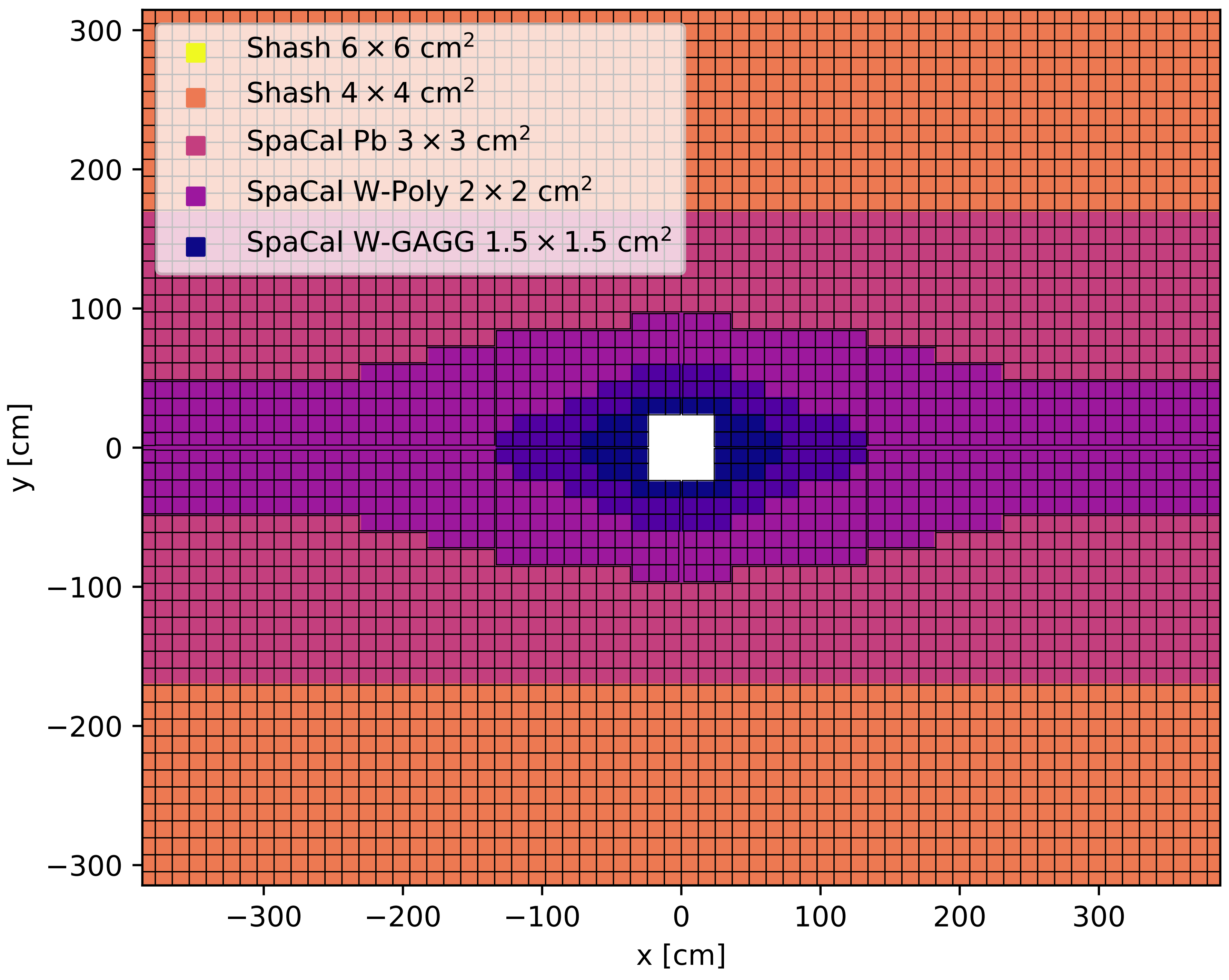
The GNN performance can still be improved with more data, and timing information can also help

High-granularity single-sided configuration

- Designed to keep # of channels ~constant
  - Reduce overal occupancy



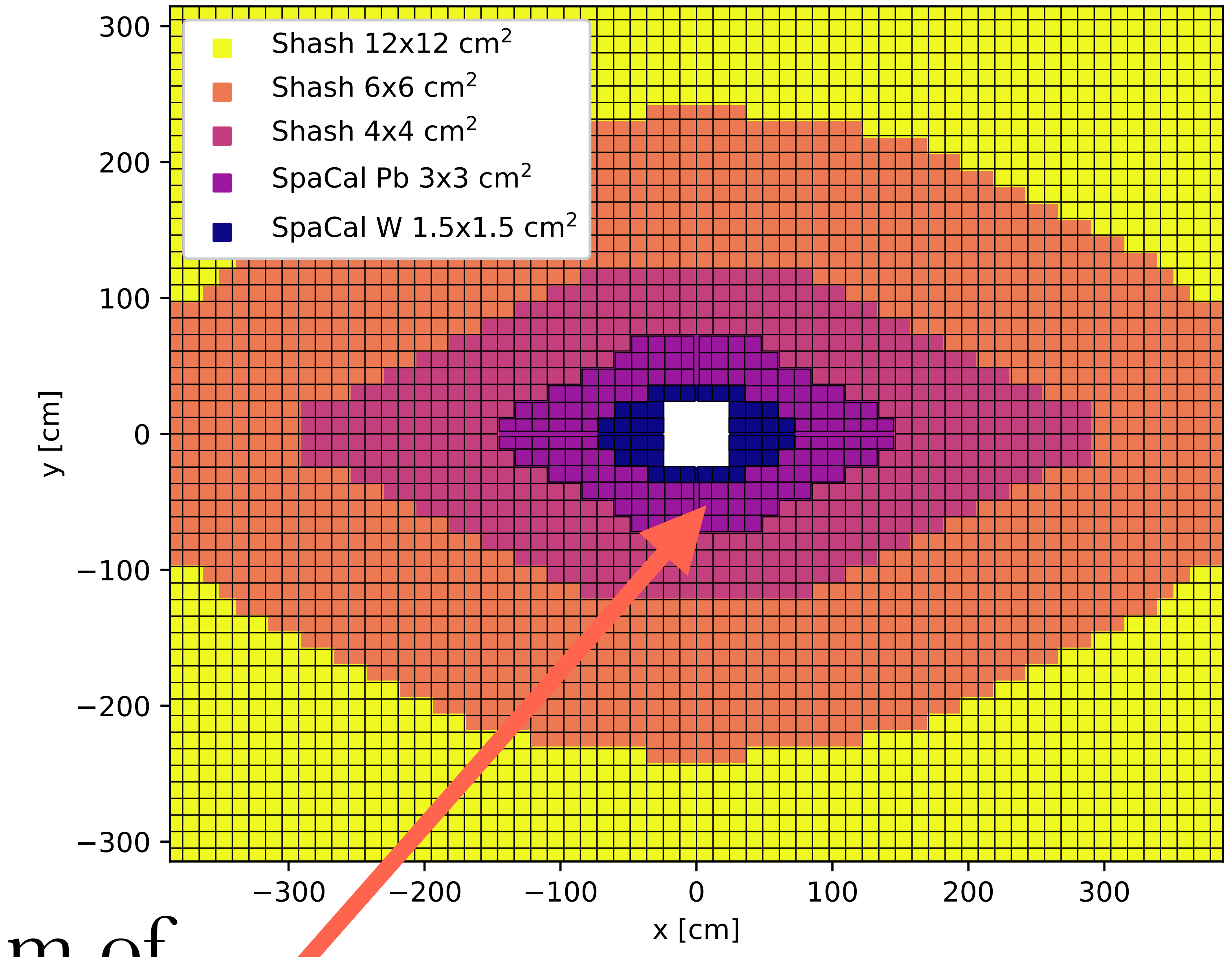
Baseline



HG

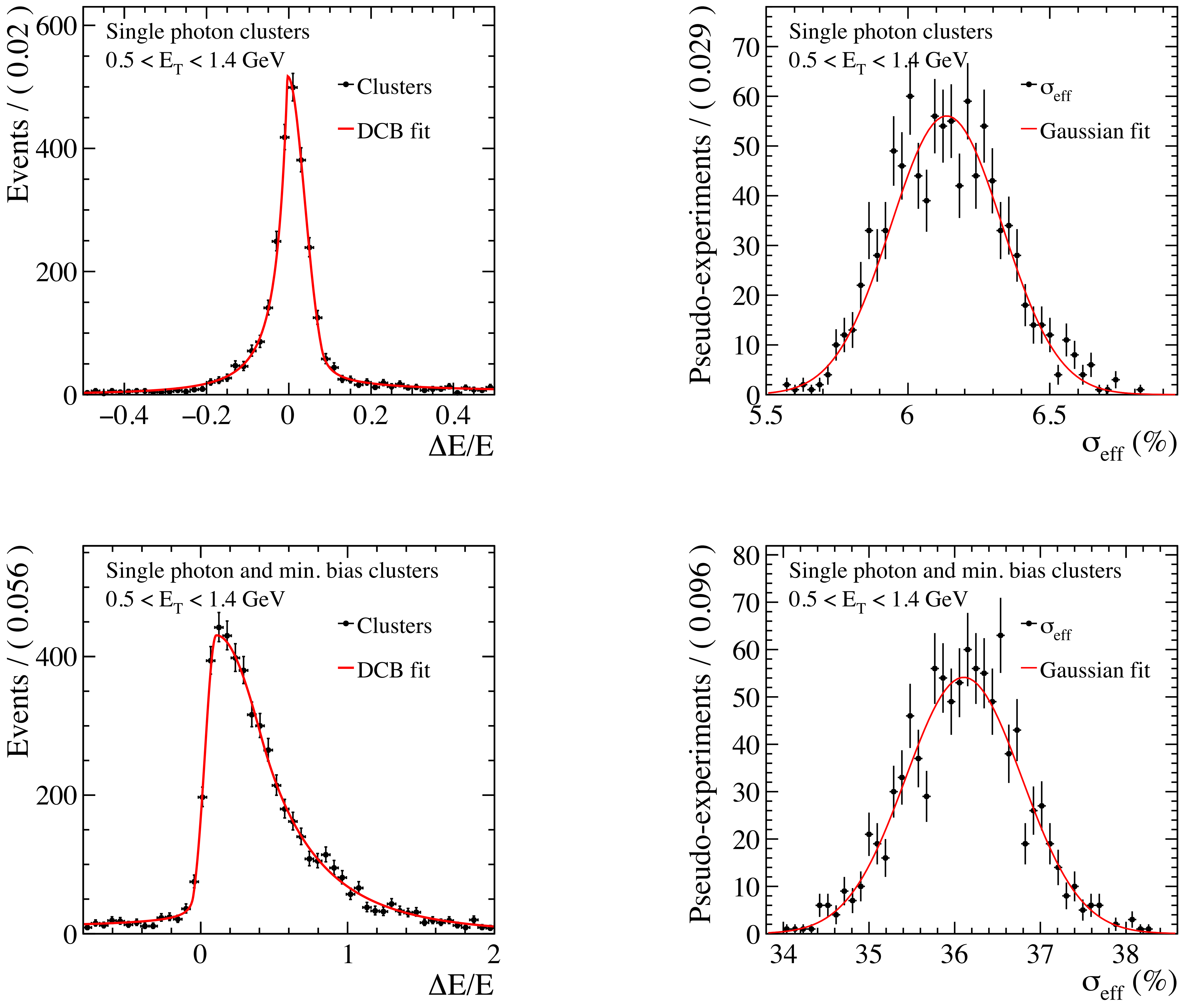
Our approach

- Nodes
  - Deposited energy (front, back)
  - Cell index in module
  - Cell position (x, y)
  - Distance to seed cell
- Edges
  - ΔE, Δx, Δy, total distance
- Globals
  - Total energy
  - Seed cell position (x, y)
- Loss
  - MSE
- Target
  - True energy
- Preselection
  - Photons produced within 100 mm of interaction point
  - Clusters with seed in SpaCal-Pb region
  - 3x3 modules around the seed module, reduced to 25 central cells



How to evaluate the performance?

- Asymmetric distributions due to pile-up: long right tails
  - Defining the 'resolution' is not straightforward
  - We define the resolution (σ_{eff}) as the half-width of the central 68% interval
- Fits to the $\Delta E/E$ distributions
  - Generate toys from the p.d.f., recompute metric
  - Gaussian fit to distribution of metric to extract mean and width (uncertainty)

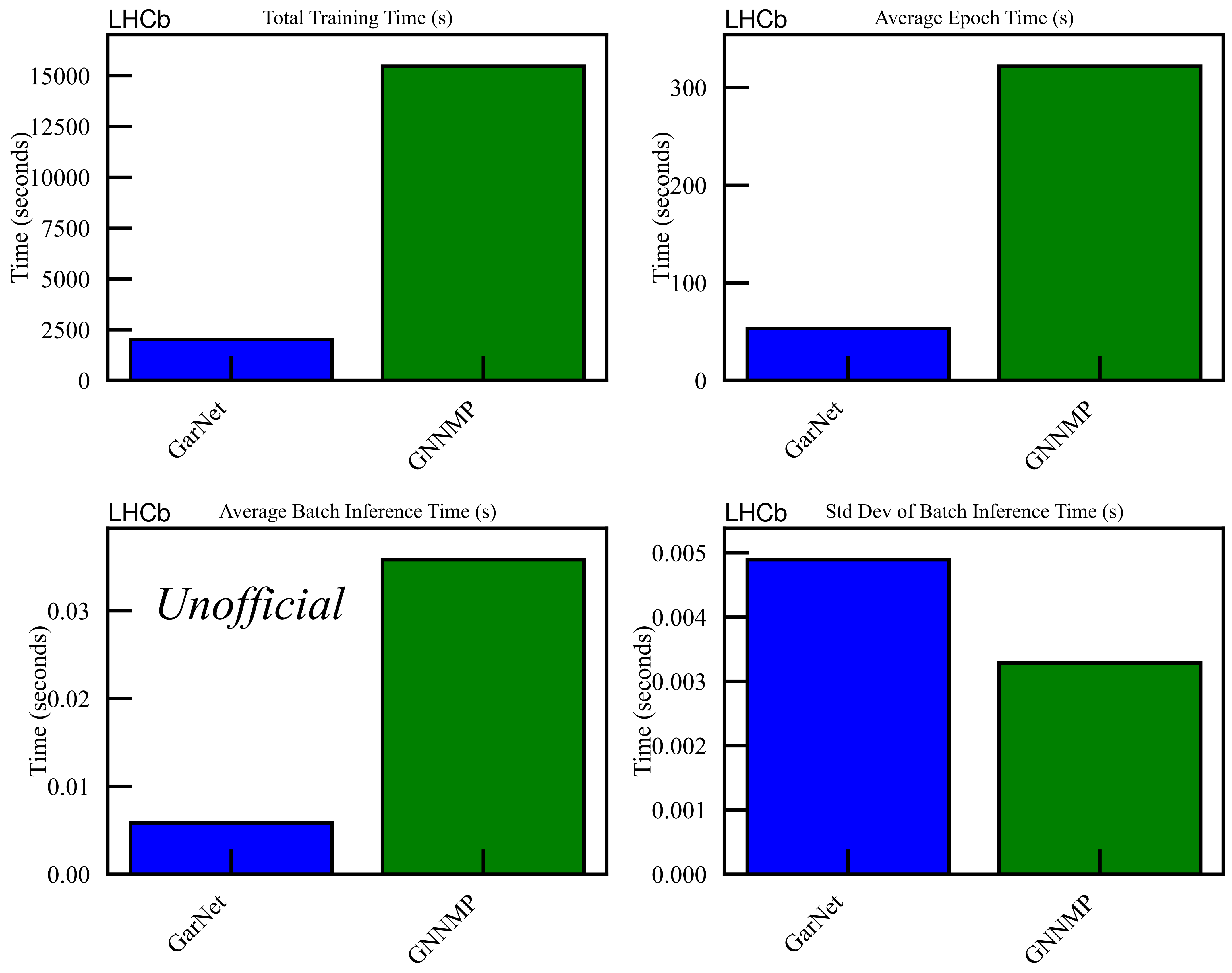


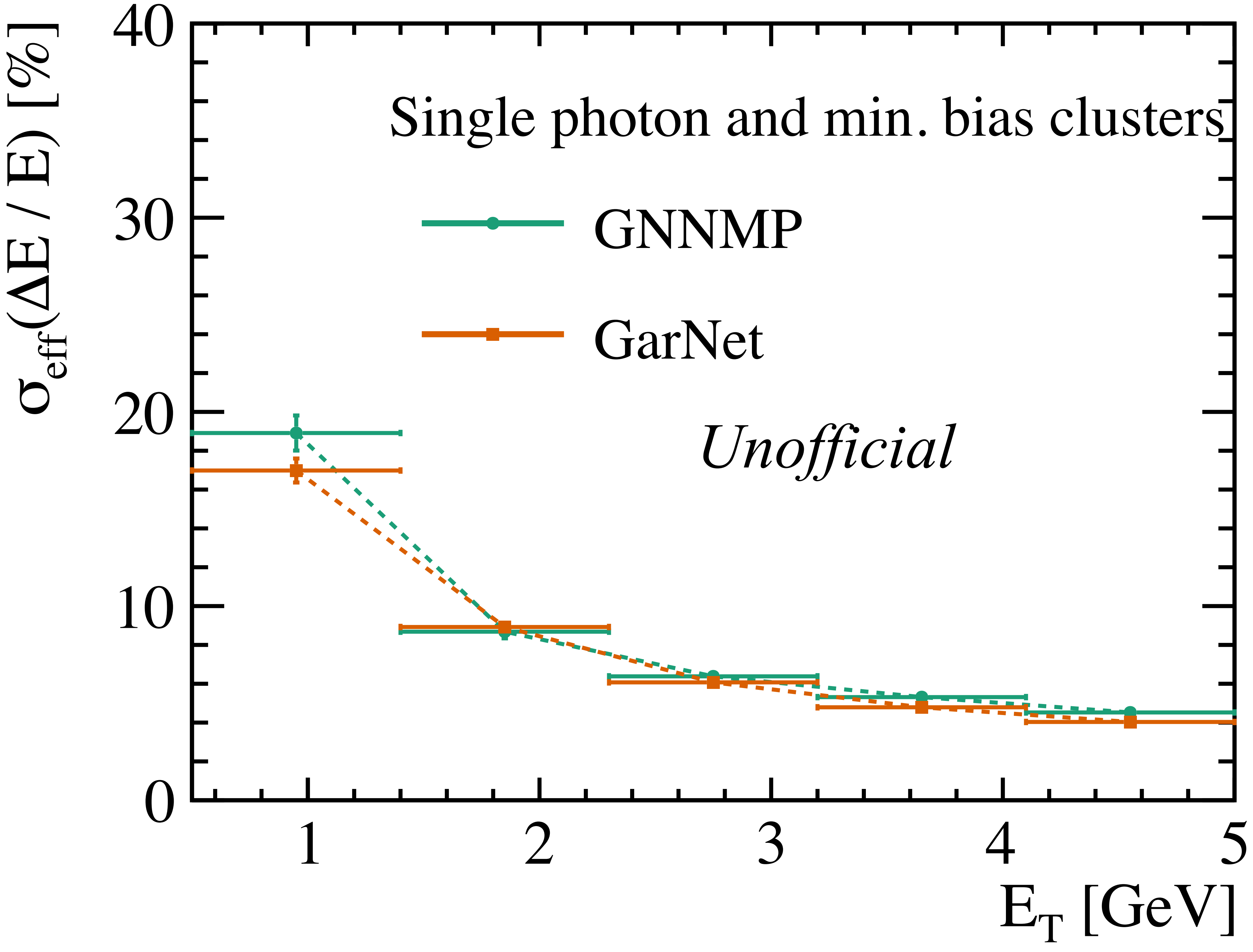
GarNet

- A lightweight, attention-enhanced variant GNN architecture
  - Introduced for real-time particle reconstruction at CMS by Iiyama et al. (Eur. Phys. J. C 79, 608 (2019); Front. Big Data 3, 598927 (2021))
  - Explicit edges are replaced by learnable aggregators connecting nodes through latent vertices
    - This reduces the number of mathematical operations, decreasing training and inference time
- GarNet has a know ONNX implementation for FPGA deployment

GarNet results

- GarNet has similar performance, but is 6-7x faster!





Ongoing studies

- Comparison for performance in and out of boundary regions
- Alternative strategies for connecting nodes in the full GNNMP
- Spatial symmetries
- Node-level targets
- Curriculum learning
- GarNet:
  - Distillation, quantization-aware training, post quantization
  - FPGA implementation, GPU parallelization

Multitarget regression

Simultaneous regression of energy, time, and position

- Each target contribution must be properly weighted in the loss
- Uncertainty-aware losses can help the model dynamically balance target contributions
  - However, regularization may be needed to prevent degenerate minima and stabilize training
- Strategies range from simple weighted MSE to multivariate Gaussian NLL with full covariance and Cholesky decomposition

Multi-target MSE (predefined weights)

$$\mathcal{L} = \sum_i w_i (y_i - \hat{y}_i(x))^2$$

learn one uncertainty per target

Homoscedastic (learned)

$$\mathcal{L} = \sum_i \left(\frac{1}{2\hat{\sigma}_i^2} (y_i - \hat{y}_i(x))^2 + \frac{1}{2} \log \hat{\sigma}_i^2 \right)$$

make uncertainty input-dependent

Heteroscedastic (diagonal)

$$\mathcal{L} = \sum_i \left(\frac{1}{2\hat{\sigma}_i^2(x)} (y_i - \hat{y}_i(x))^2 + \frac{1}{2} \log \hat{\sigma}_i^2(x) \right)$$

include target correlations

Multivariate Gaussian NLL

$$\mathcal{L} = \frac{1}{2} (\boldsymbol{y} - \hat{\boldsymbol{y}}(x))^\top \hat{\boldsymbol{\Sigma}}(x)^{-1} (\boldsymbol{y} - \hat{\boldsymbol{y}}(x)) + \frac{1}{2} \log \det \hat{\boldsymbol{\Sigma}}(x)$$