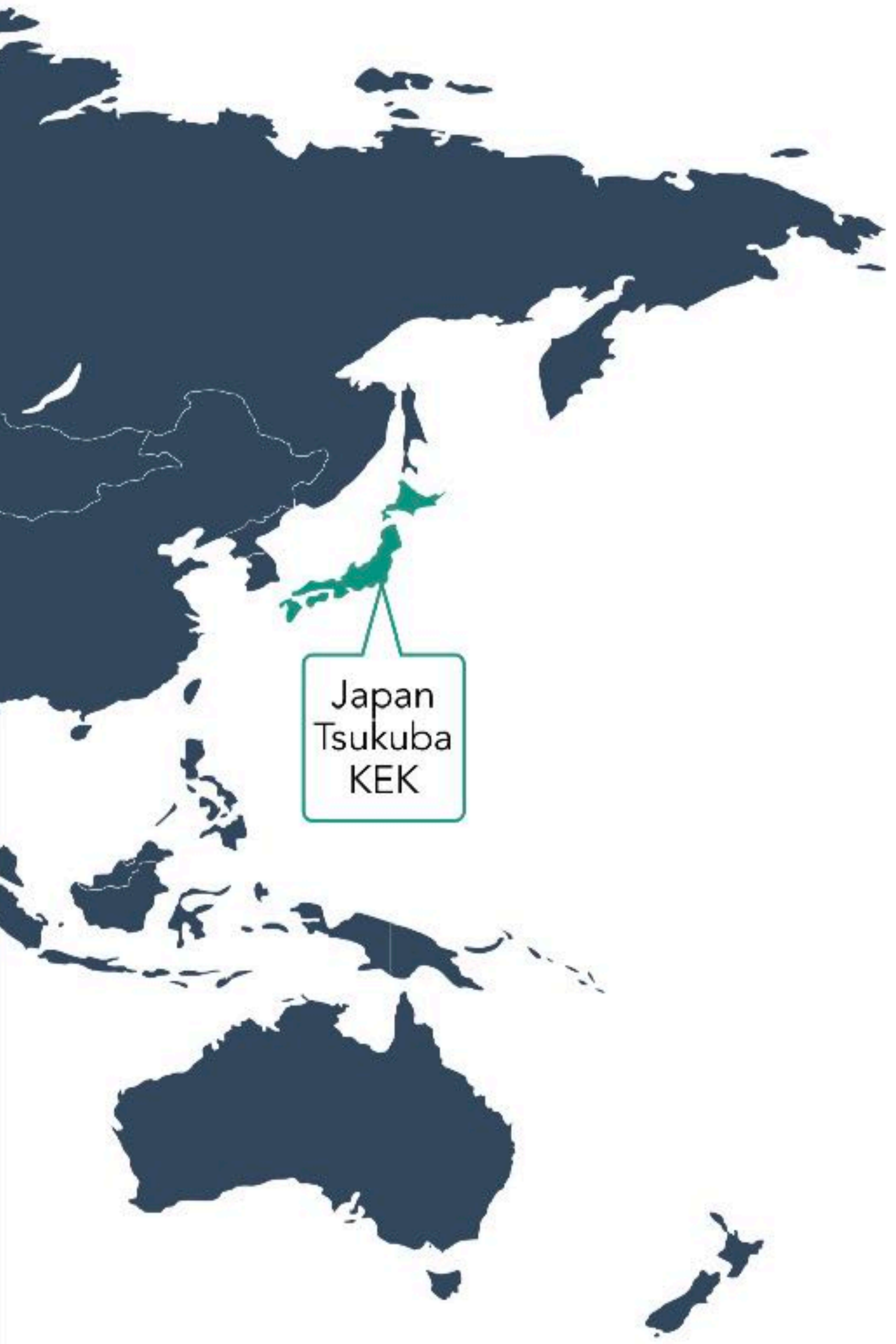
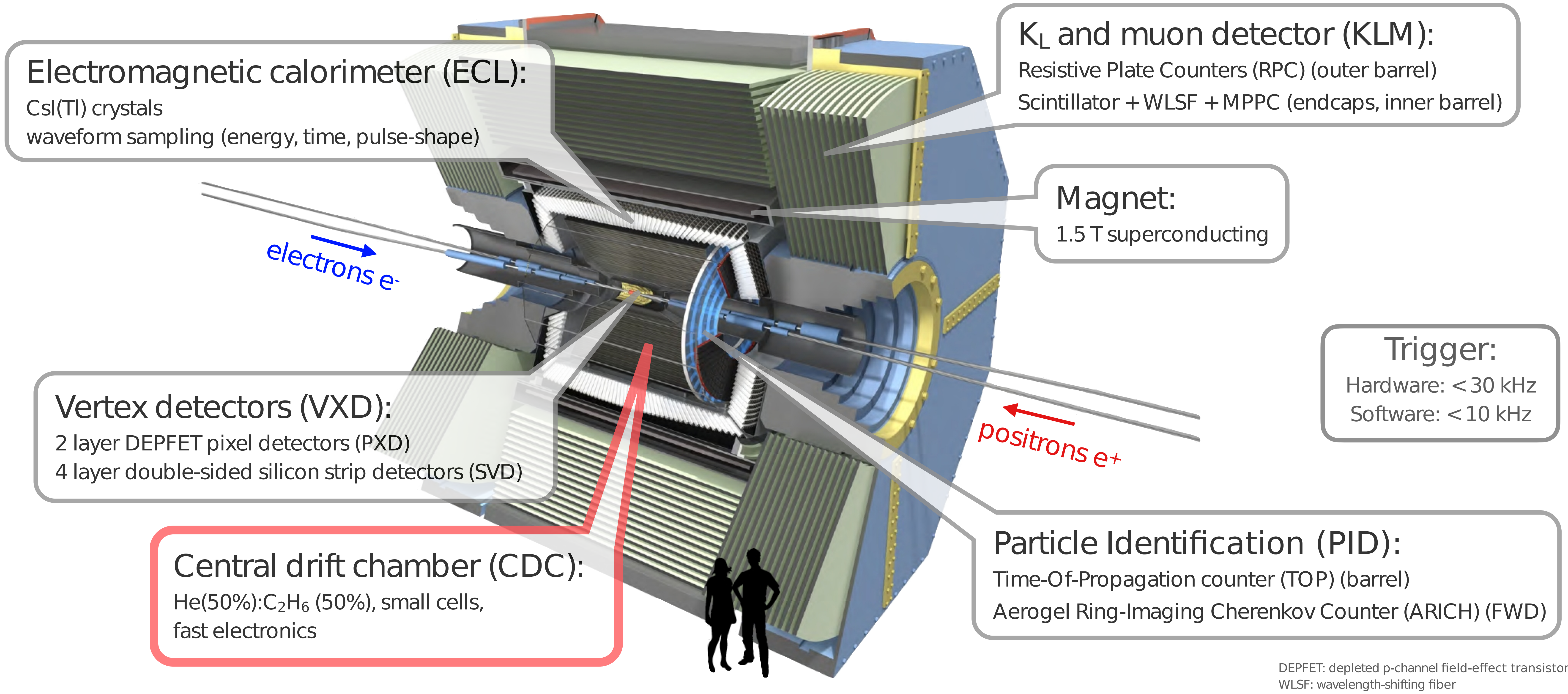
End-to-End Multi-Track Reconstruction using Graph Neural Networks at Belle II

Glühwein-Workshop 2024 KIT

Lea Reuter, Giacomo De Pietro, Slavomira Stefkova, Torben Ferber

Belle II Experiment

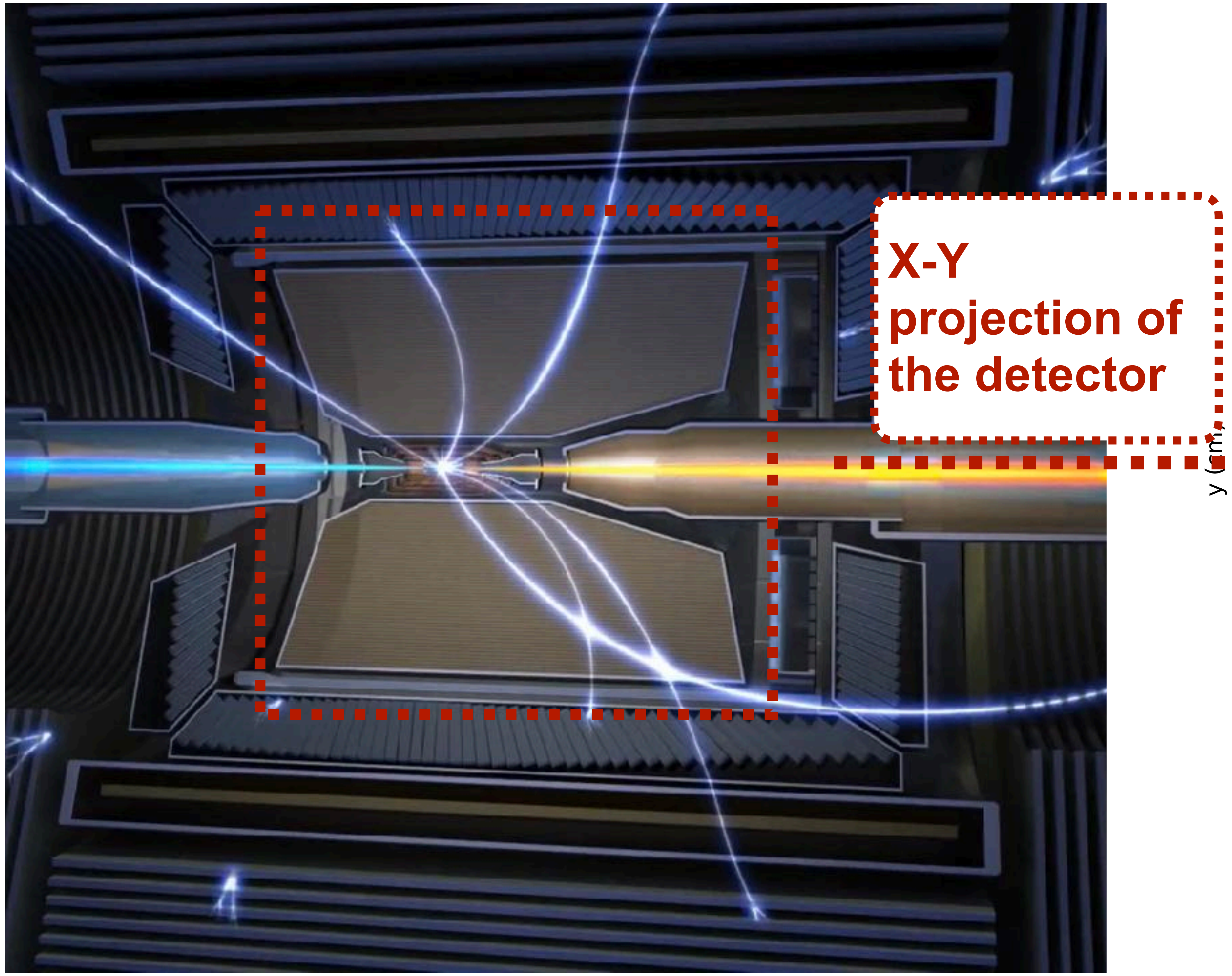


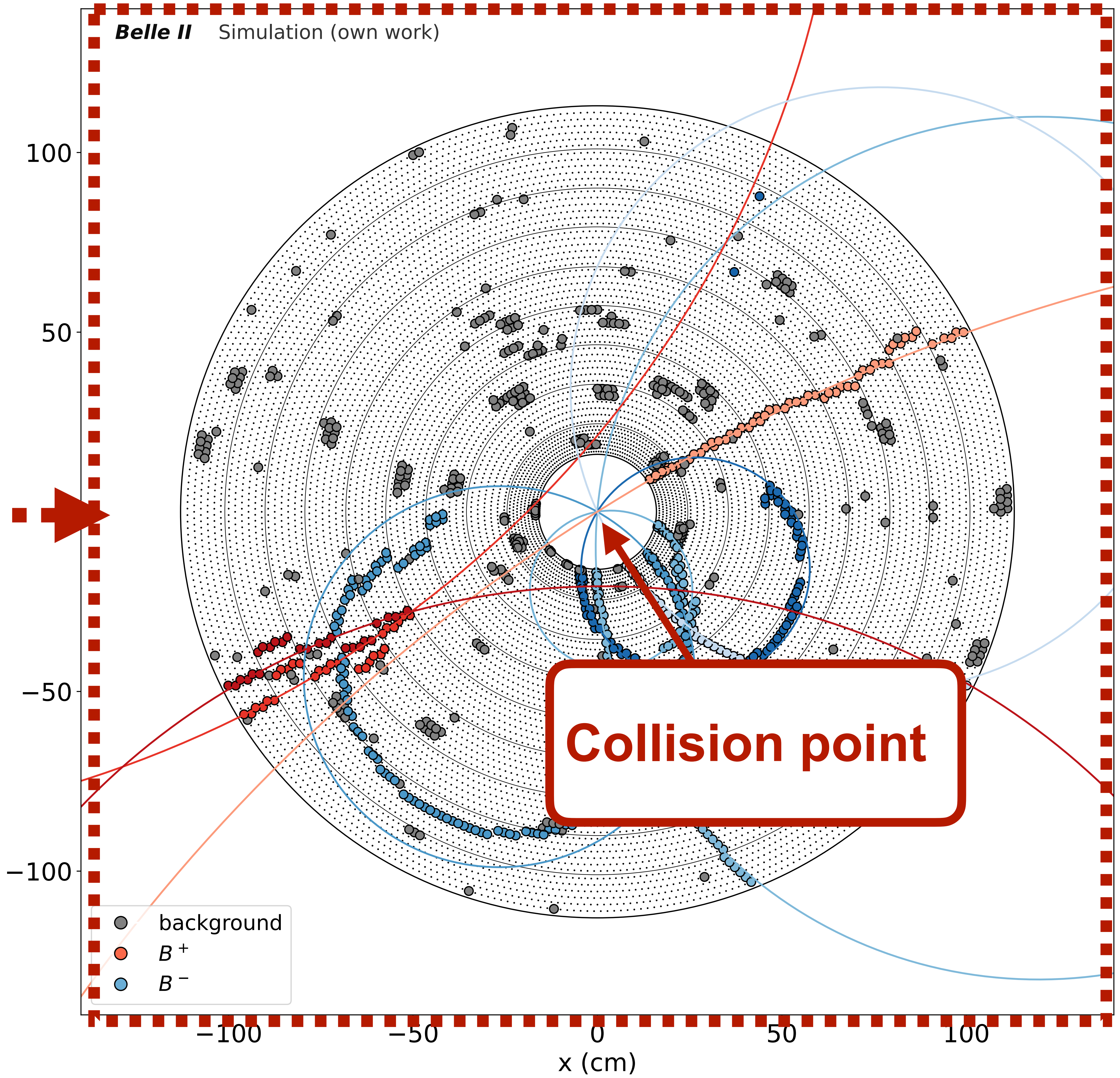


Trigger:
Hardware: < 30 kHz
Software: < 10 kHz

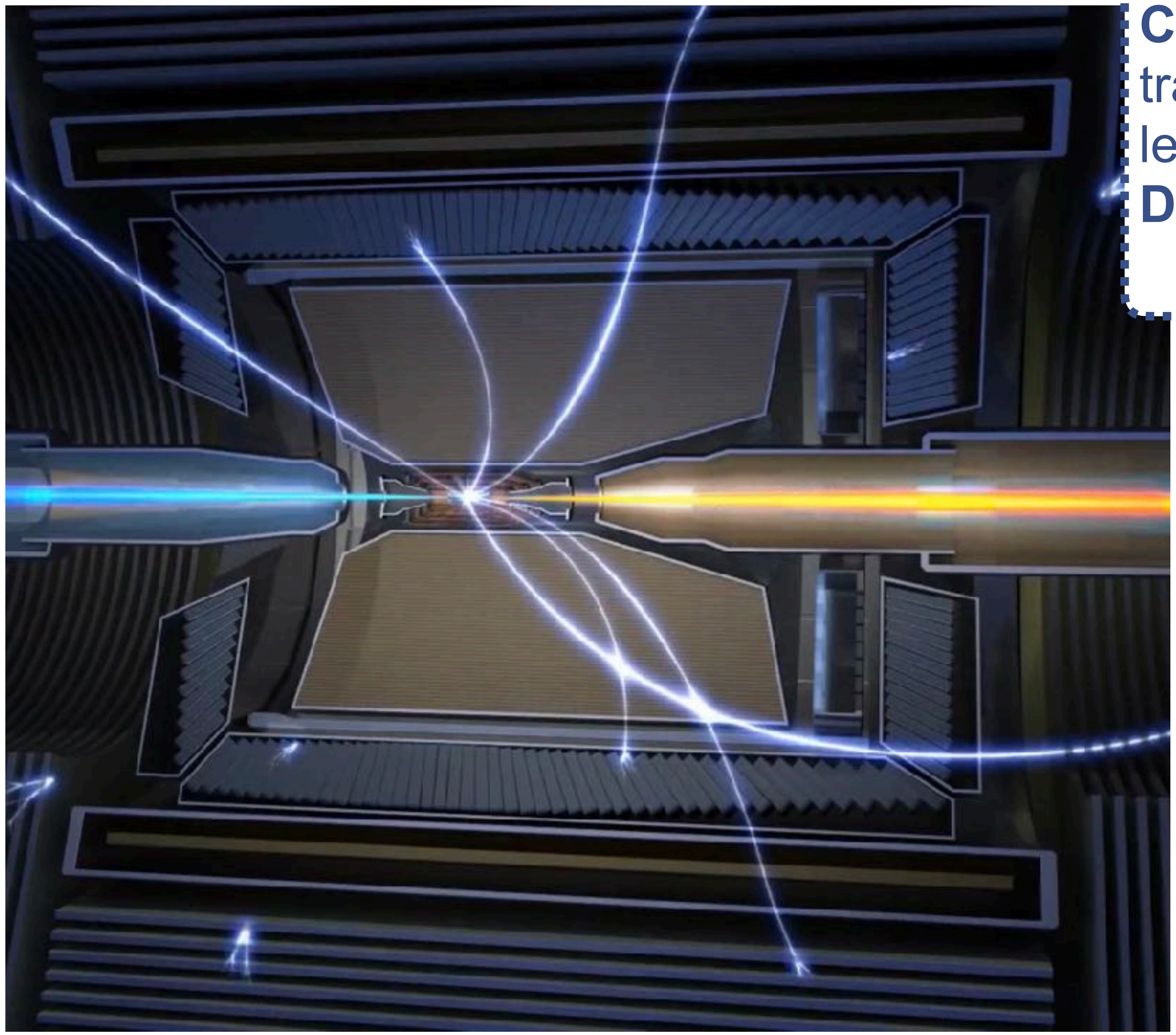
DEPFET: depleted p-channel field-effect transistor
WLSF: wavelength-shifting fiber
MPPC: multi-pixel photon counter

Main Tracking Detector

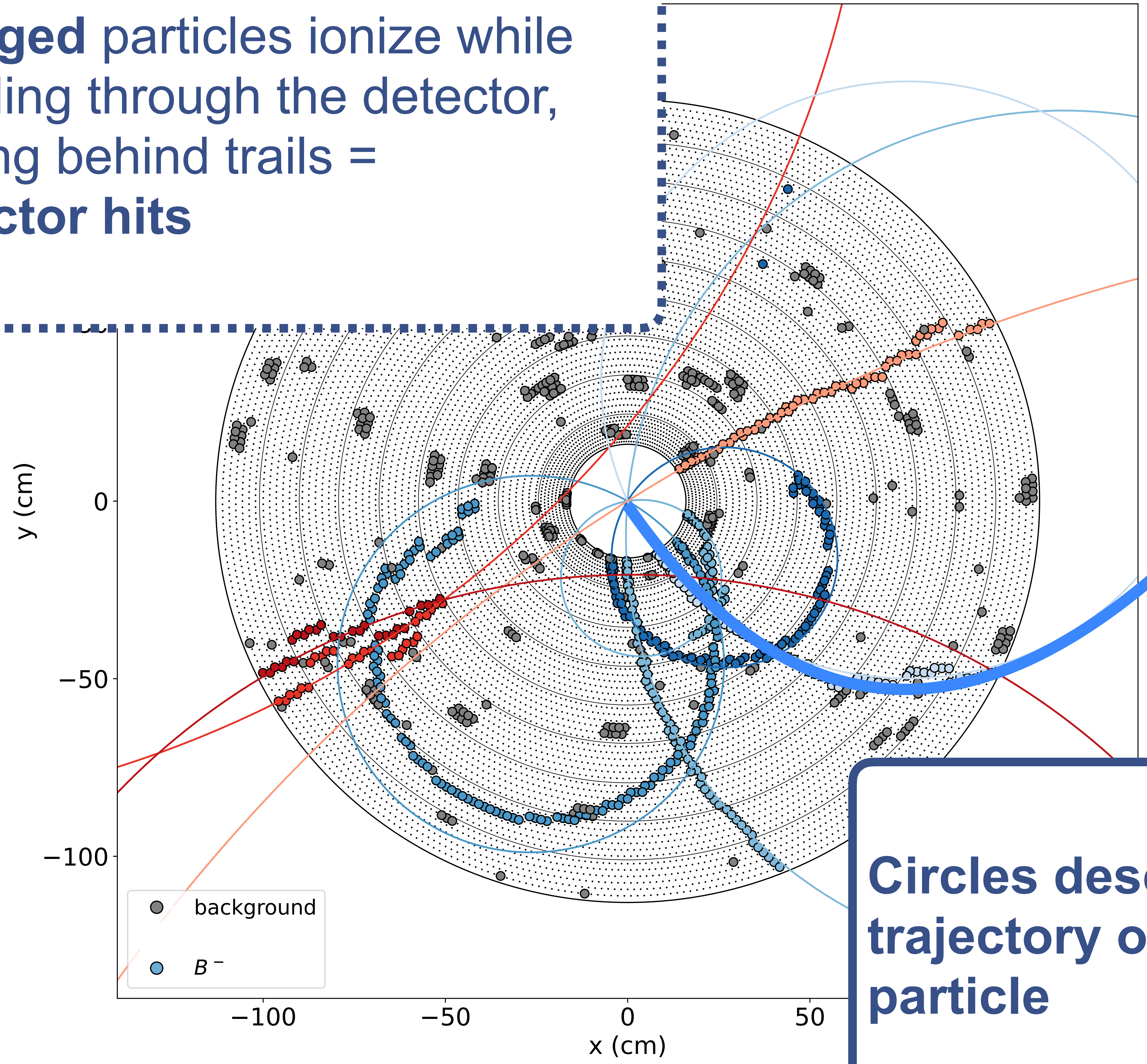




Main Tracking Detector

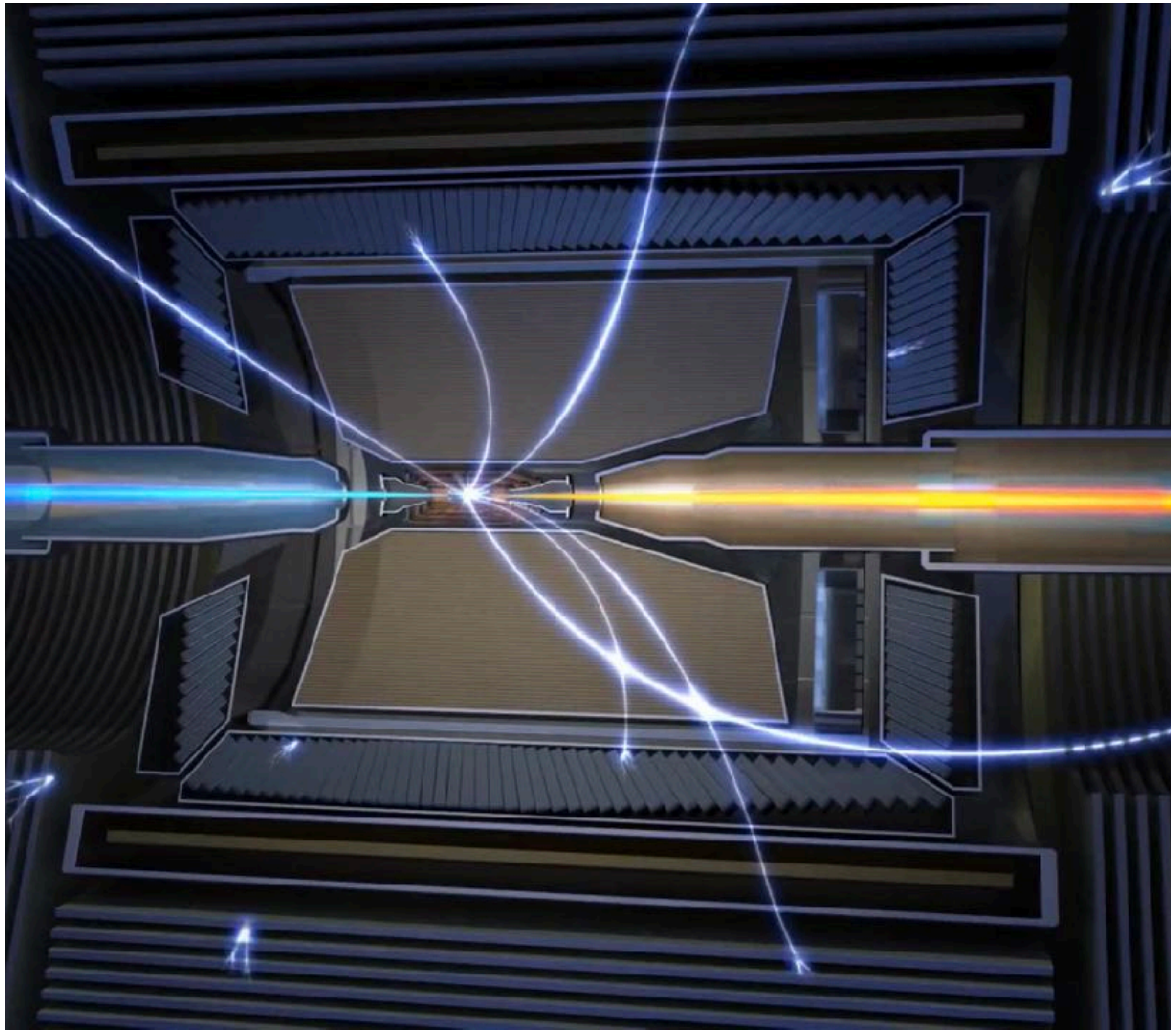


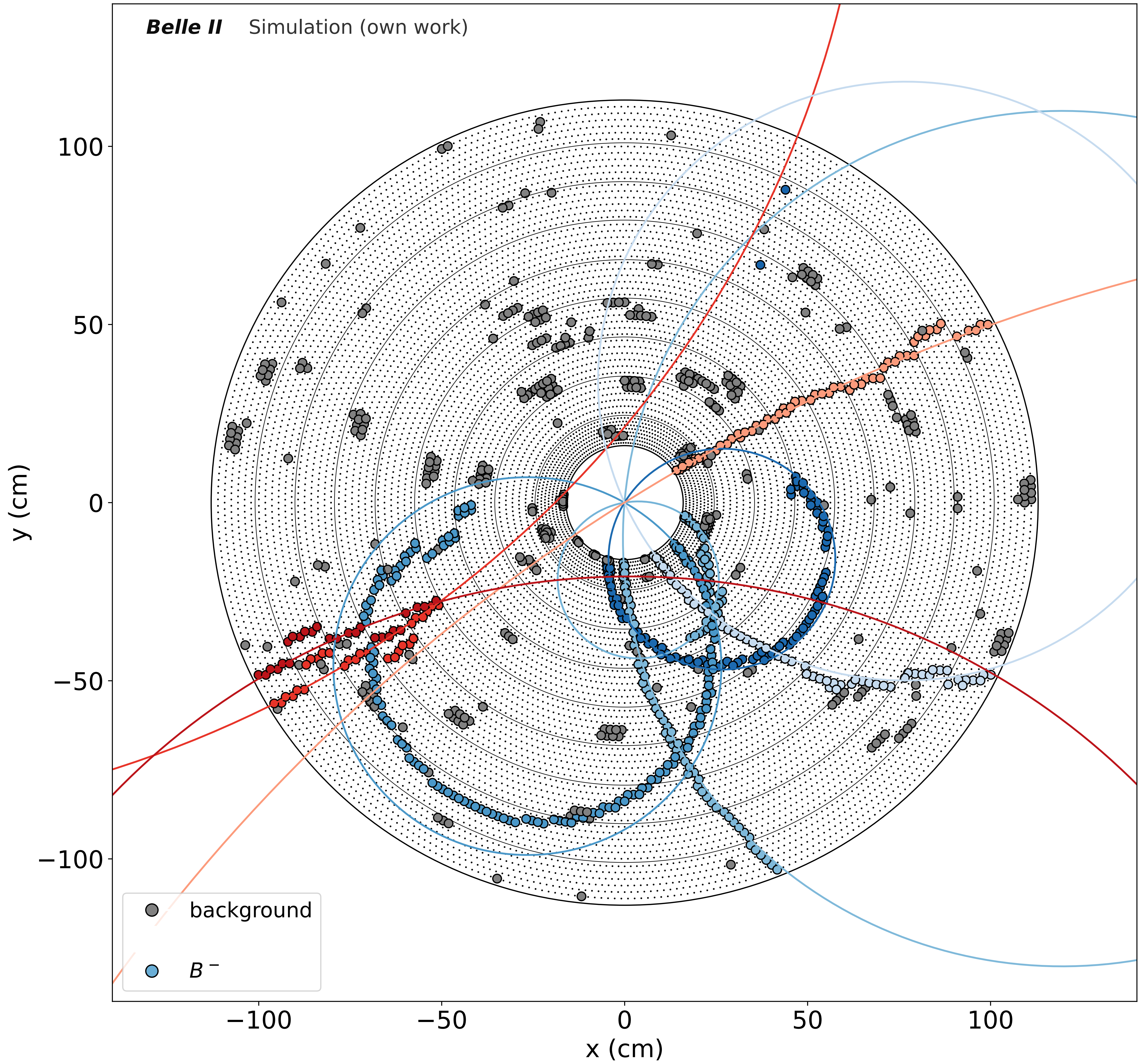
Charged particles ionize while traveling through the detector, leaving behind trails = **Detector hits**



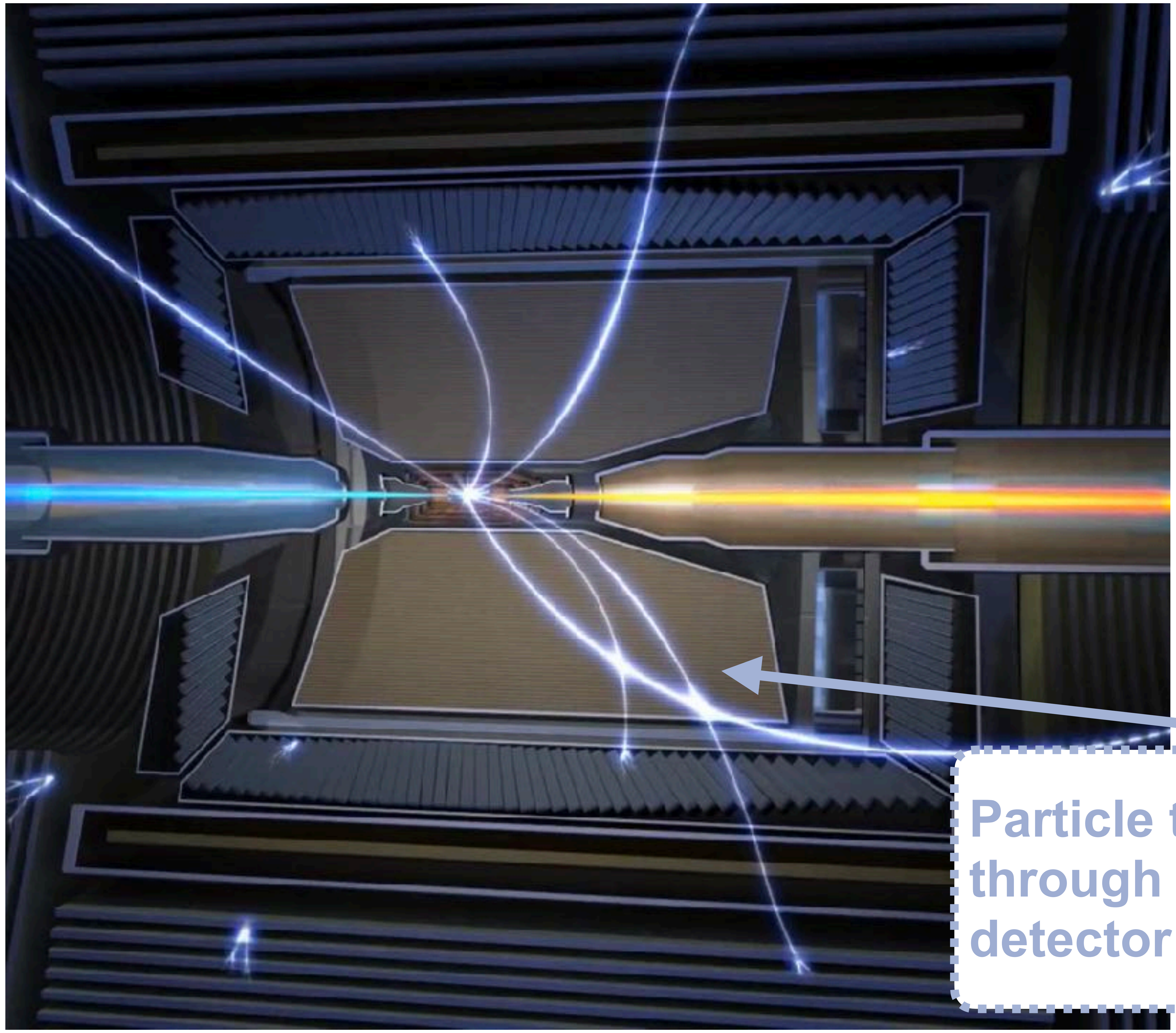
Circles describe trajectory of the particle

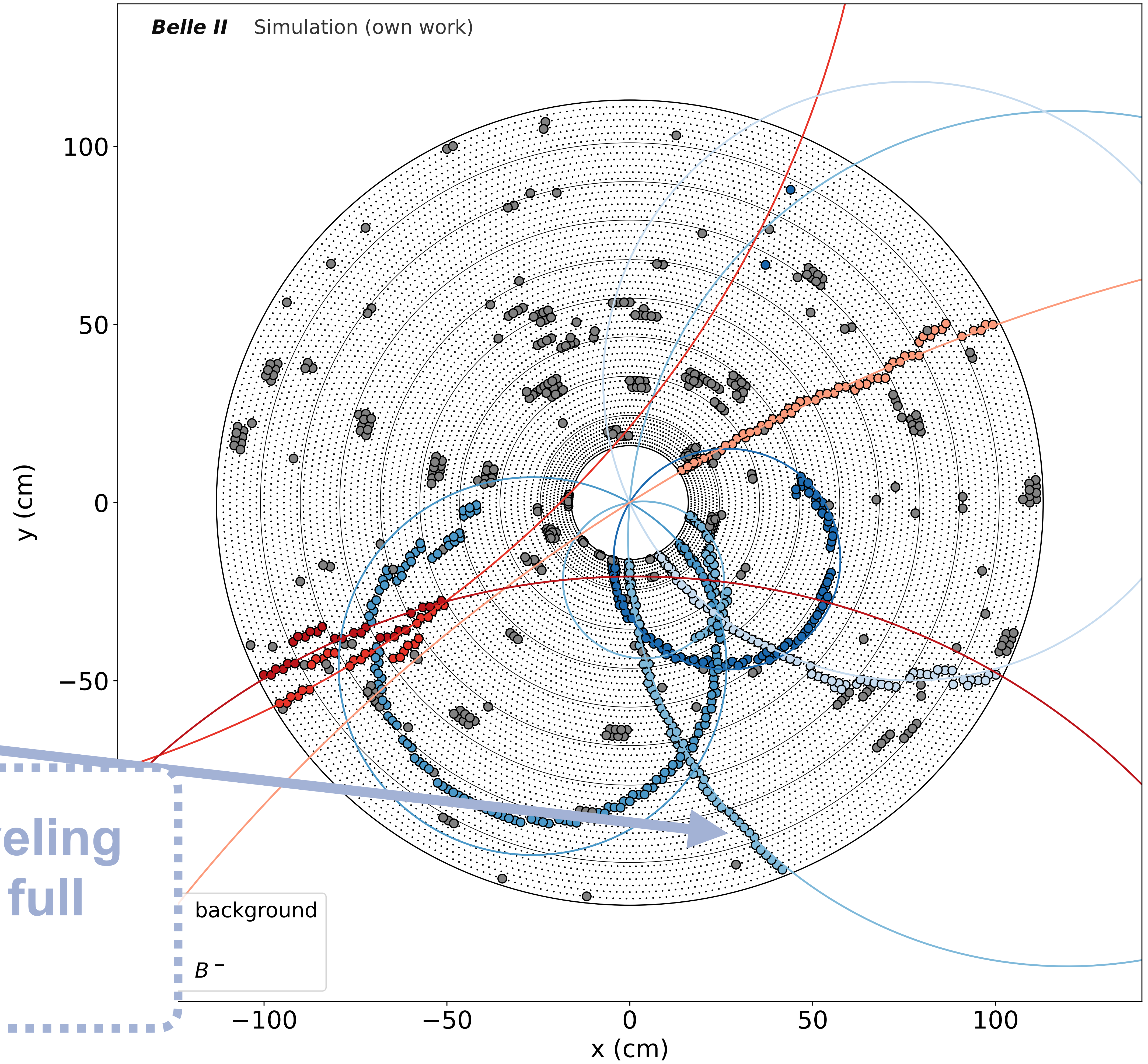
Main Tracking Detector





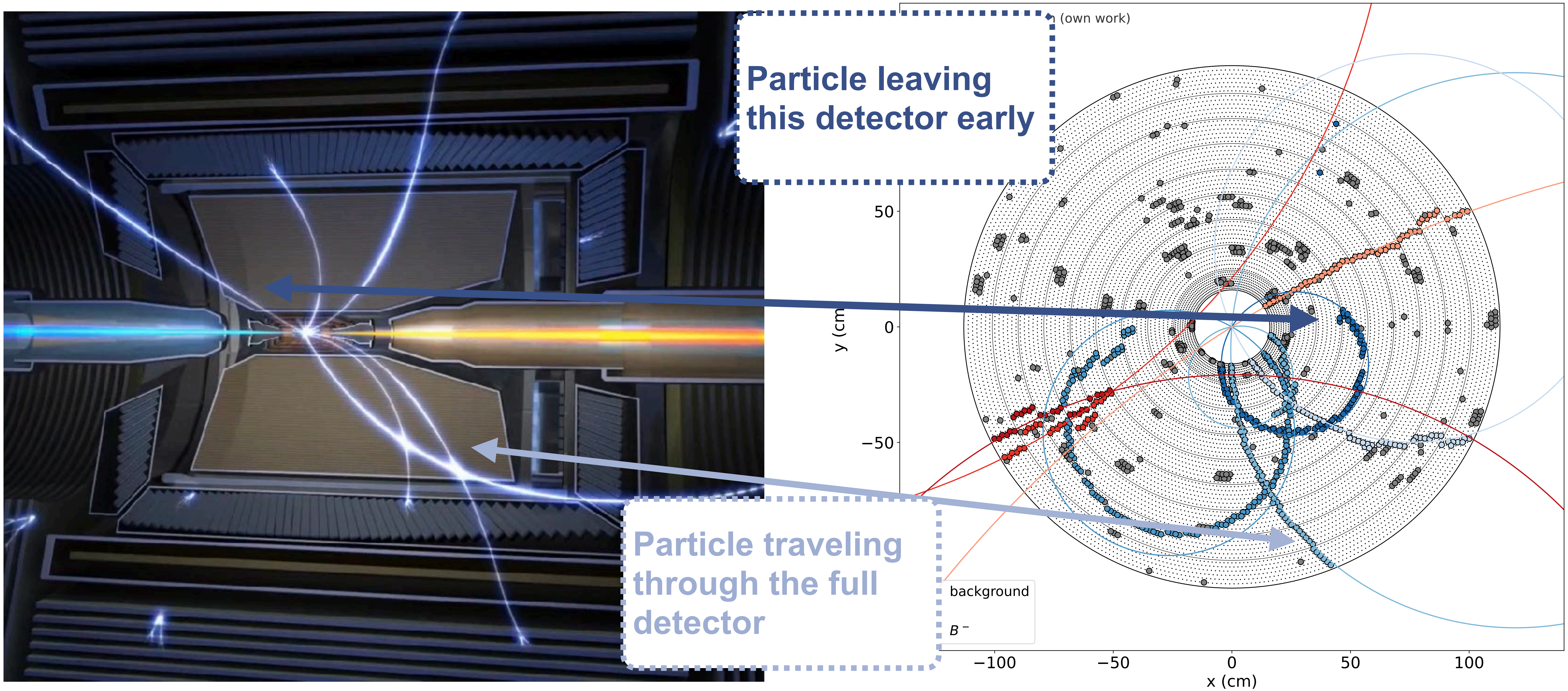
Main Tracking Detector



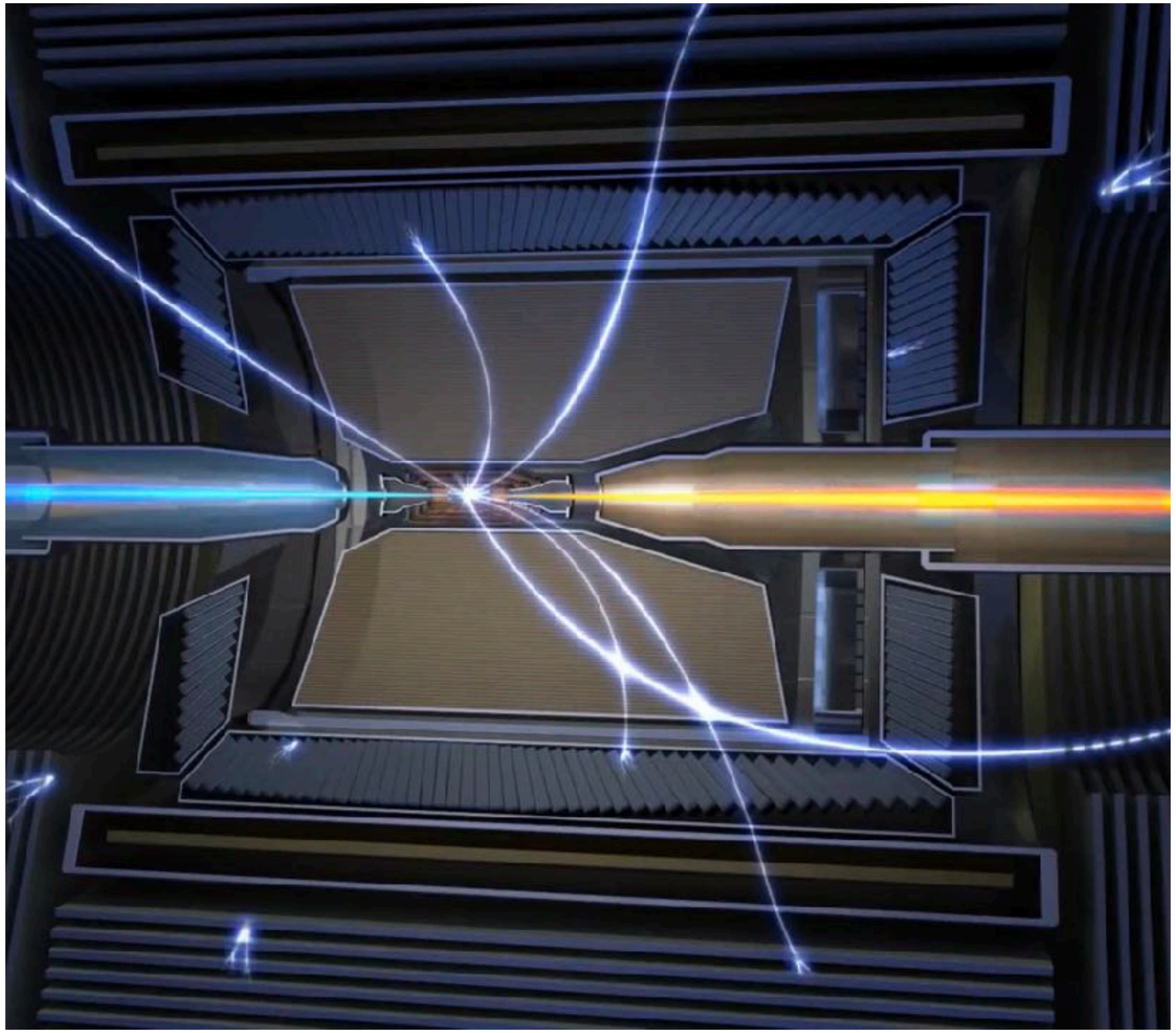


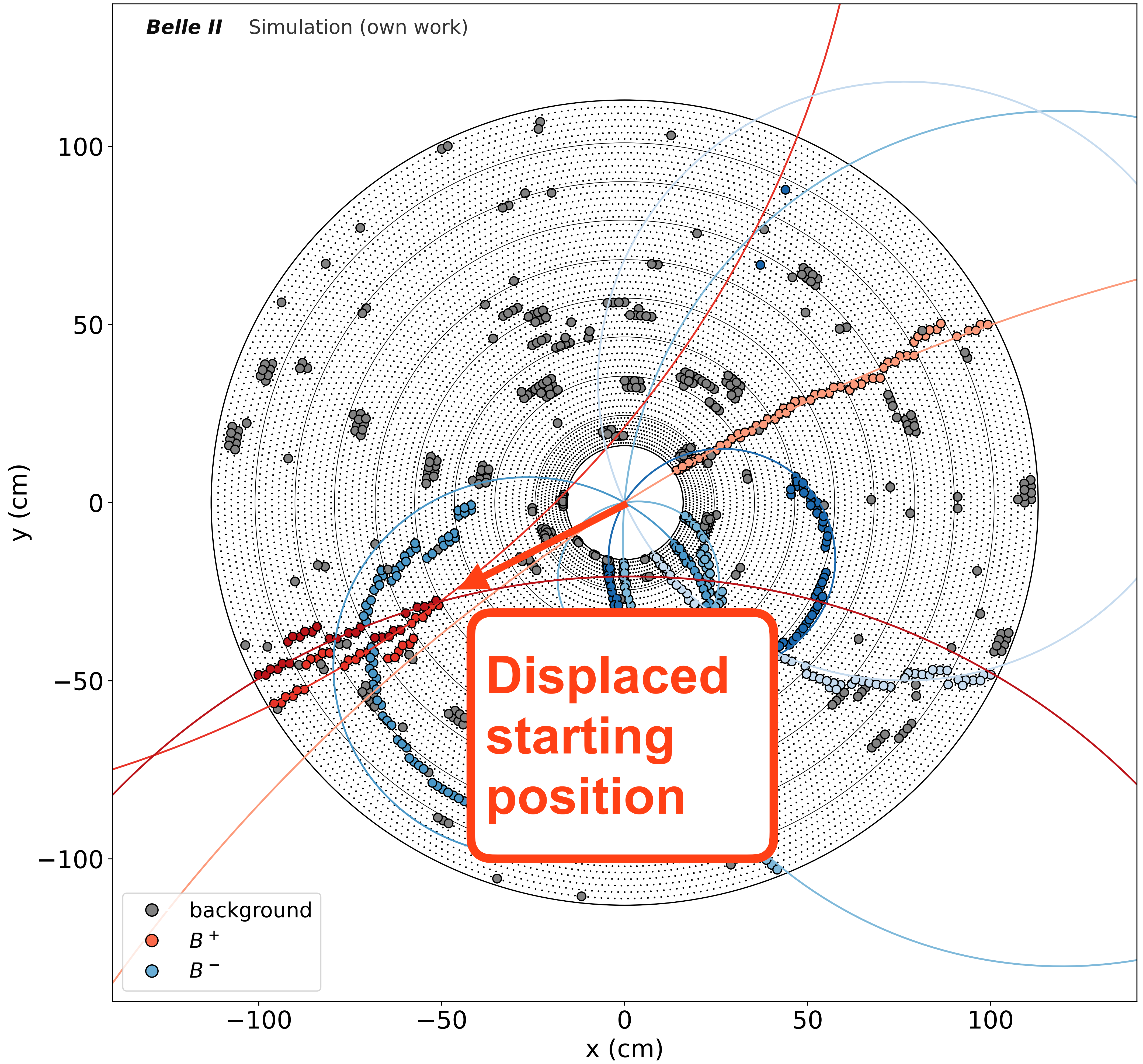
Particle traveling through the full detector

Main Tracking Detector



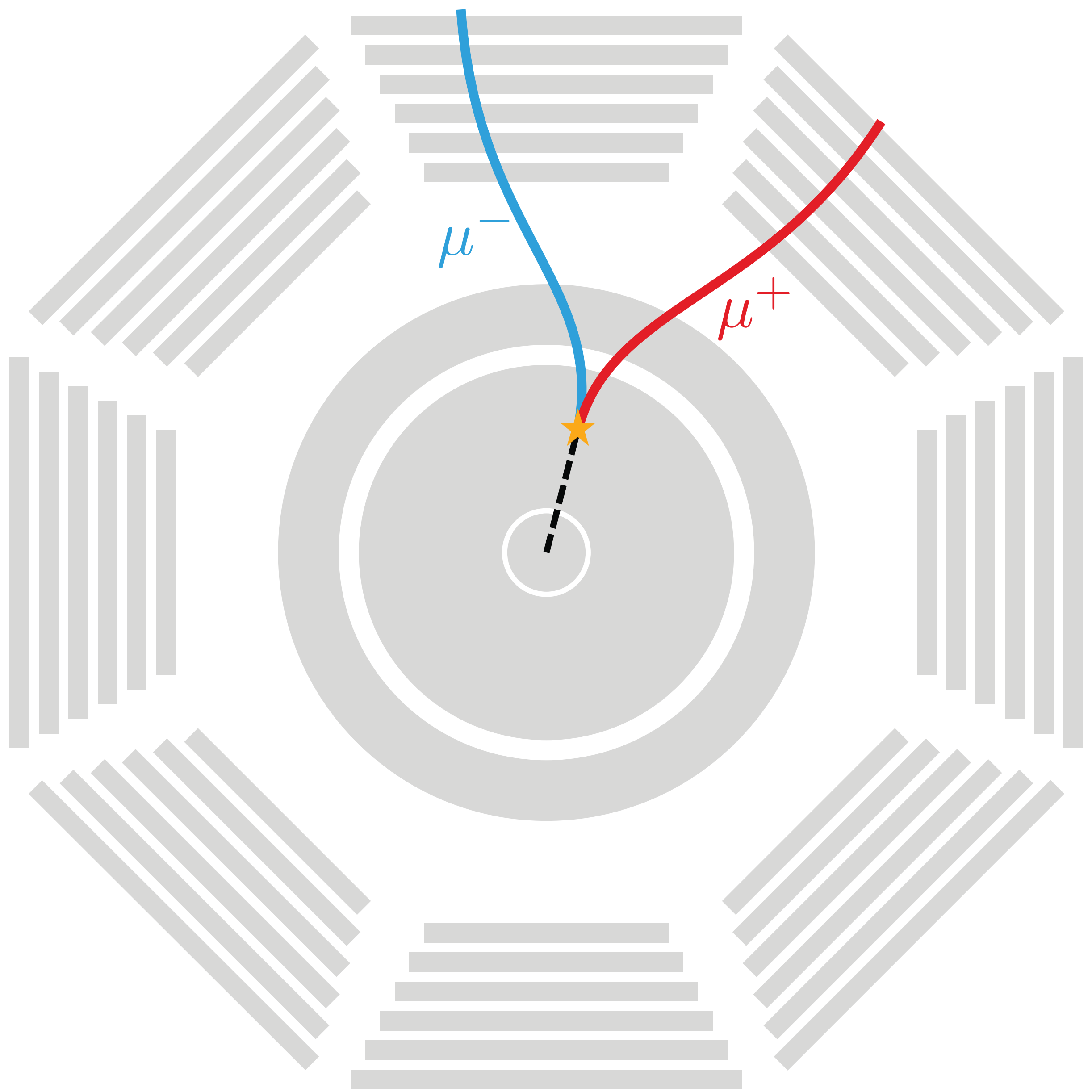
Main Tracking Detector

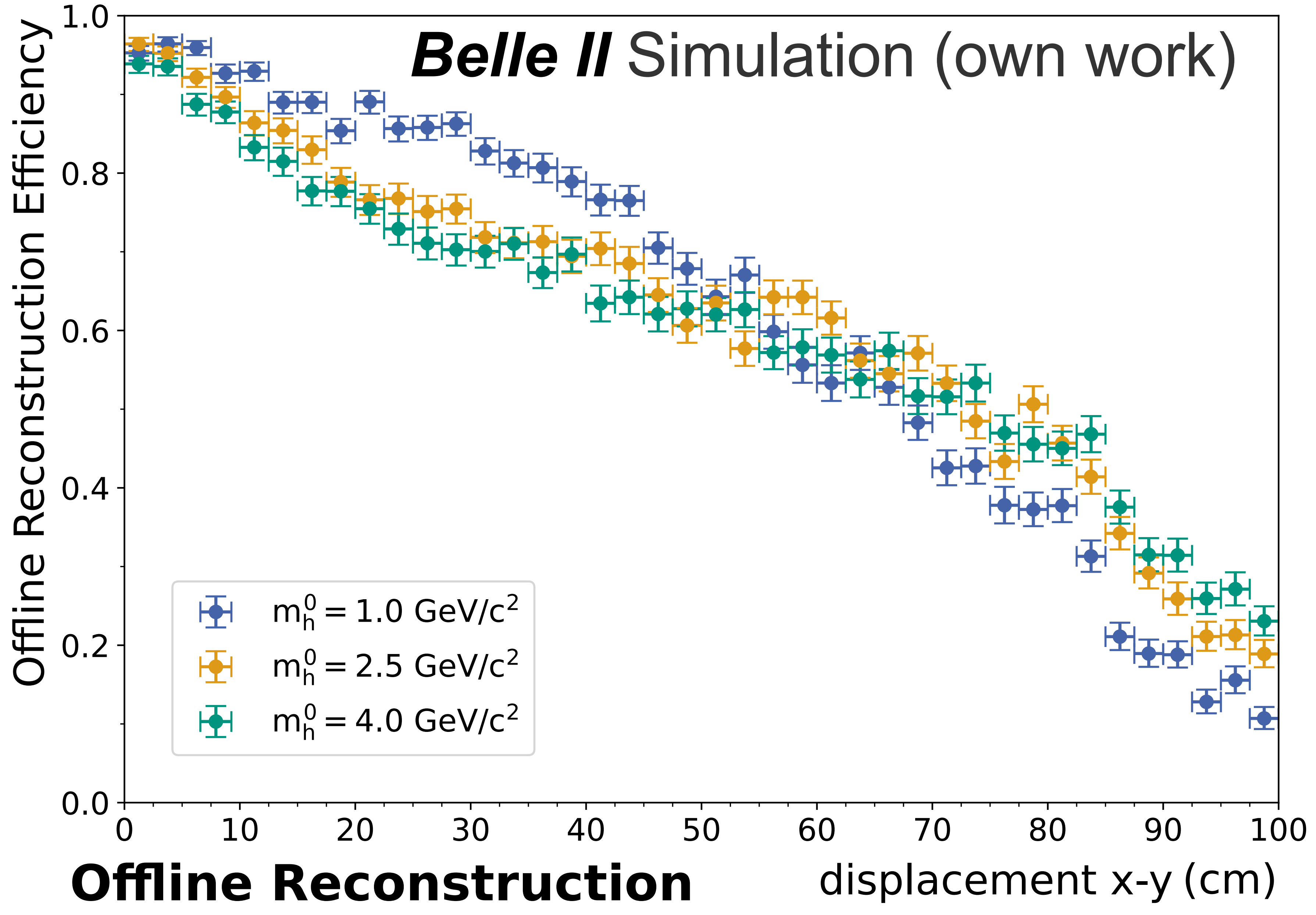




Motivation - Displaced Vertices

Displaced vertices important signature in searches for new physics
(for example [1] [2] and arXiv:2012.08595, arXiv:202.03452, arXiv:1911.03176)





Efficiency decreases depending on displacement (K_S^0, Λ^0, Dark Sector searches)

[1] Search for a long-lived spin-0 mediator in b →s transitions at the Belle II experiment (arXiv:2306.02830)

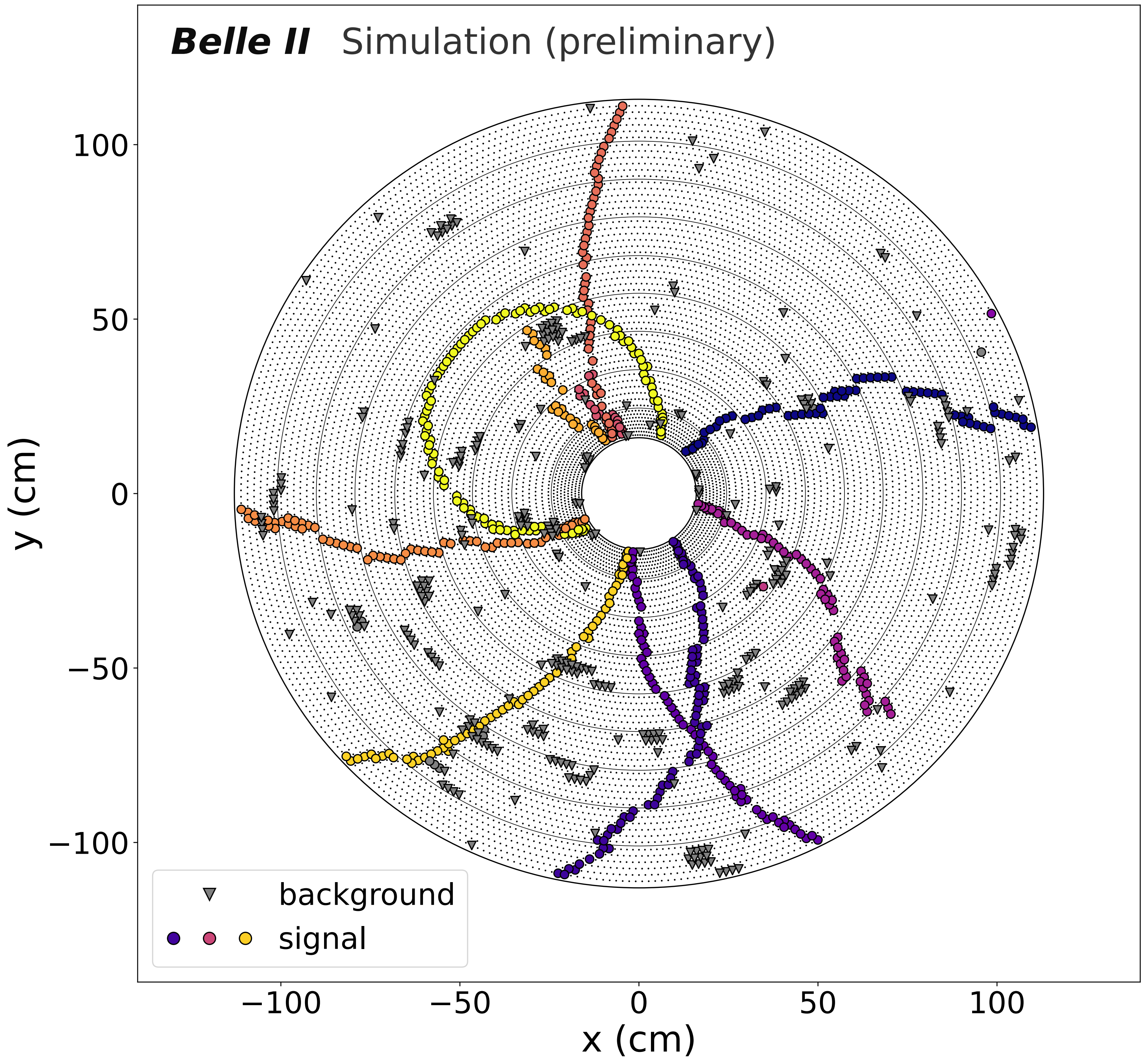
[2] Search for Inelastic Dark Matter produced in association with a Dark Higgs (LHC DM WG workshop 2024)

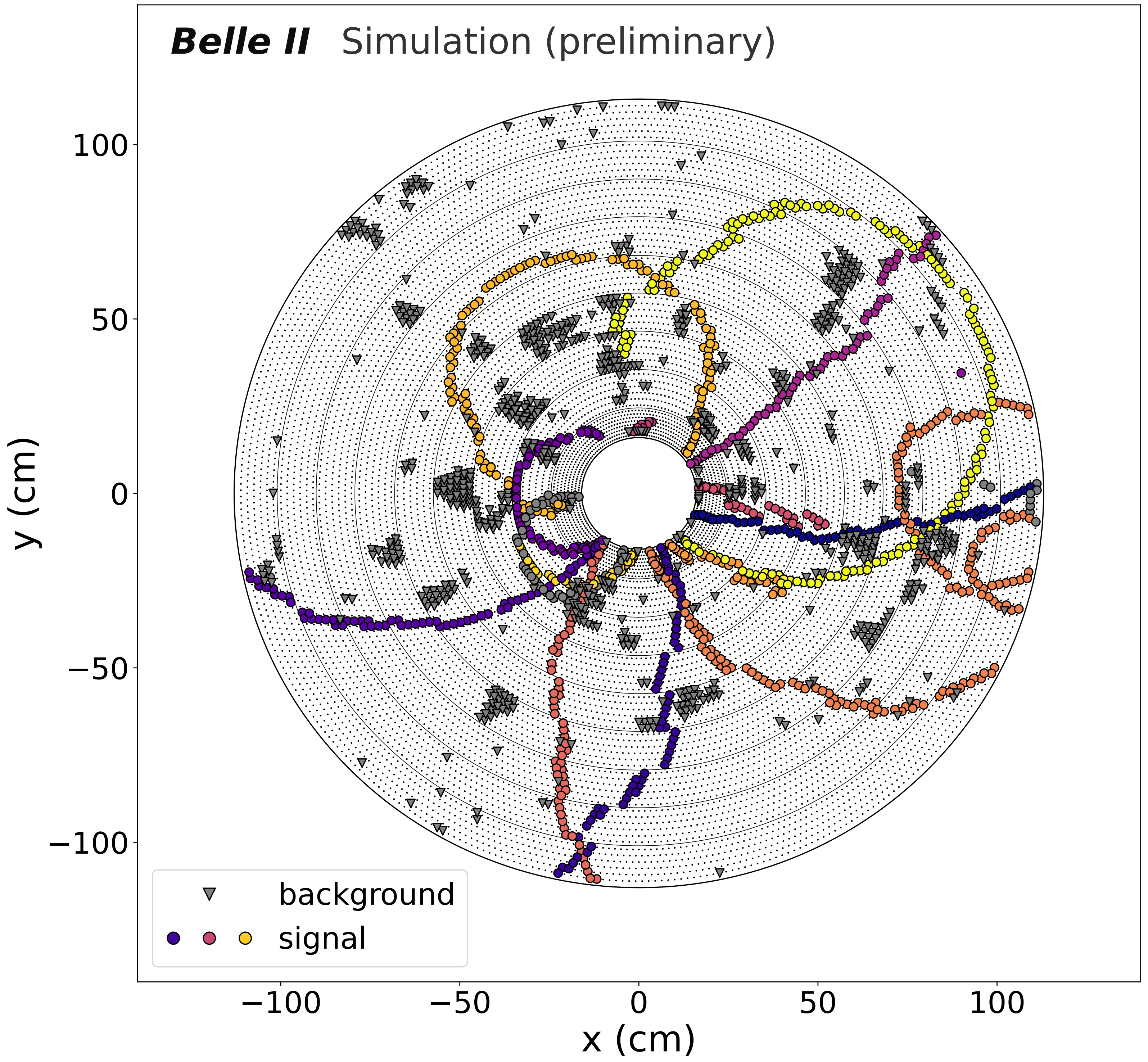
Motivation - High Backgrounds

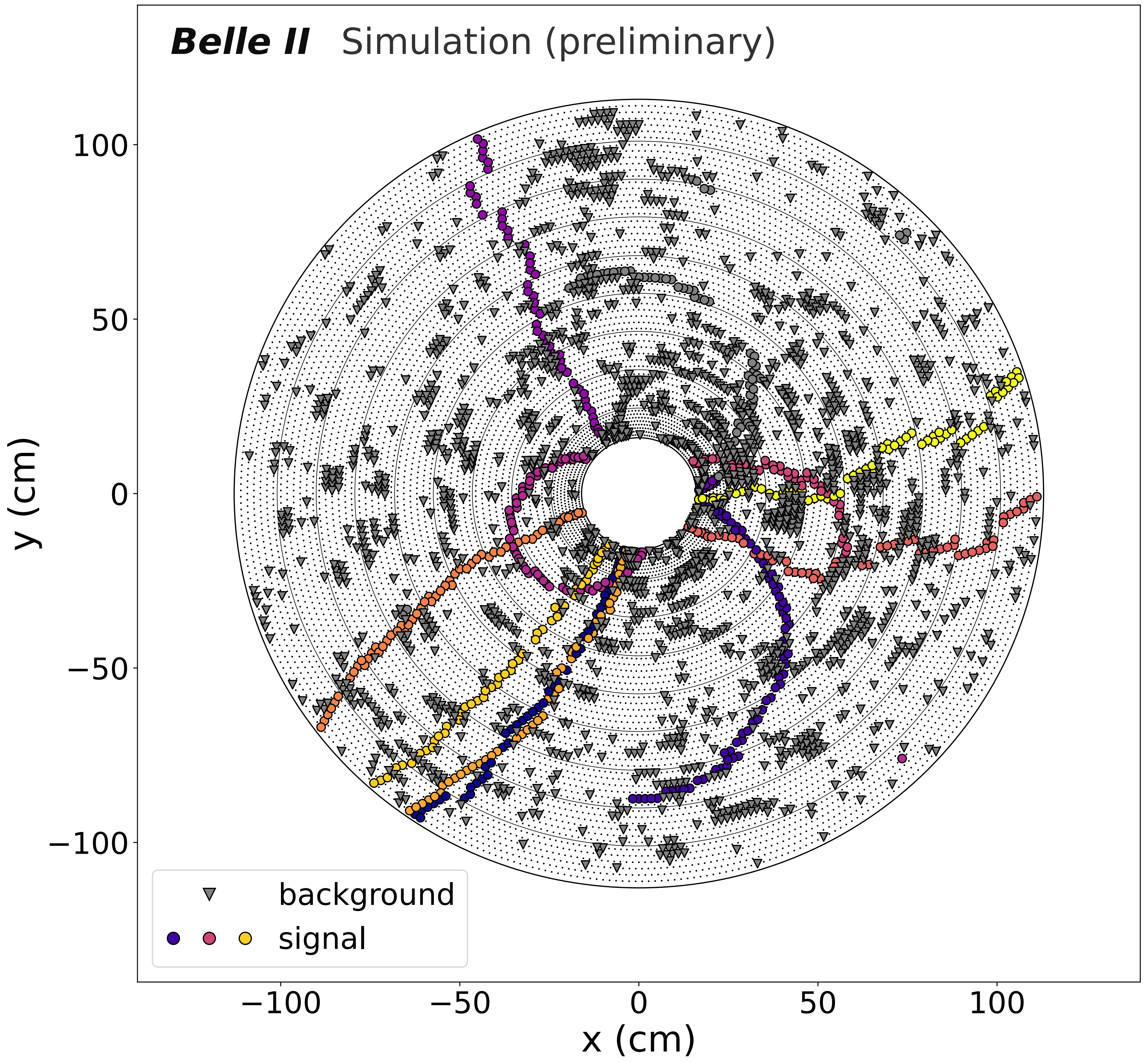
Beginning of 2021
360 background CDC hits on average

End of 2022
1280 background CDC hits on average

Expected 2030
3000 background CDC hits on average



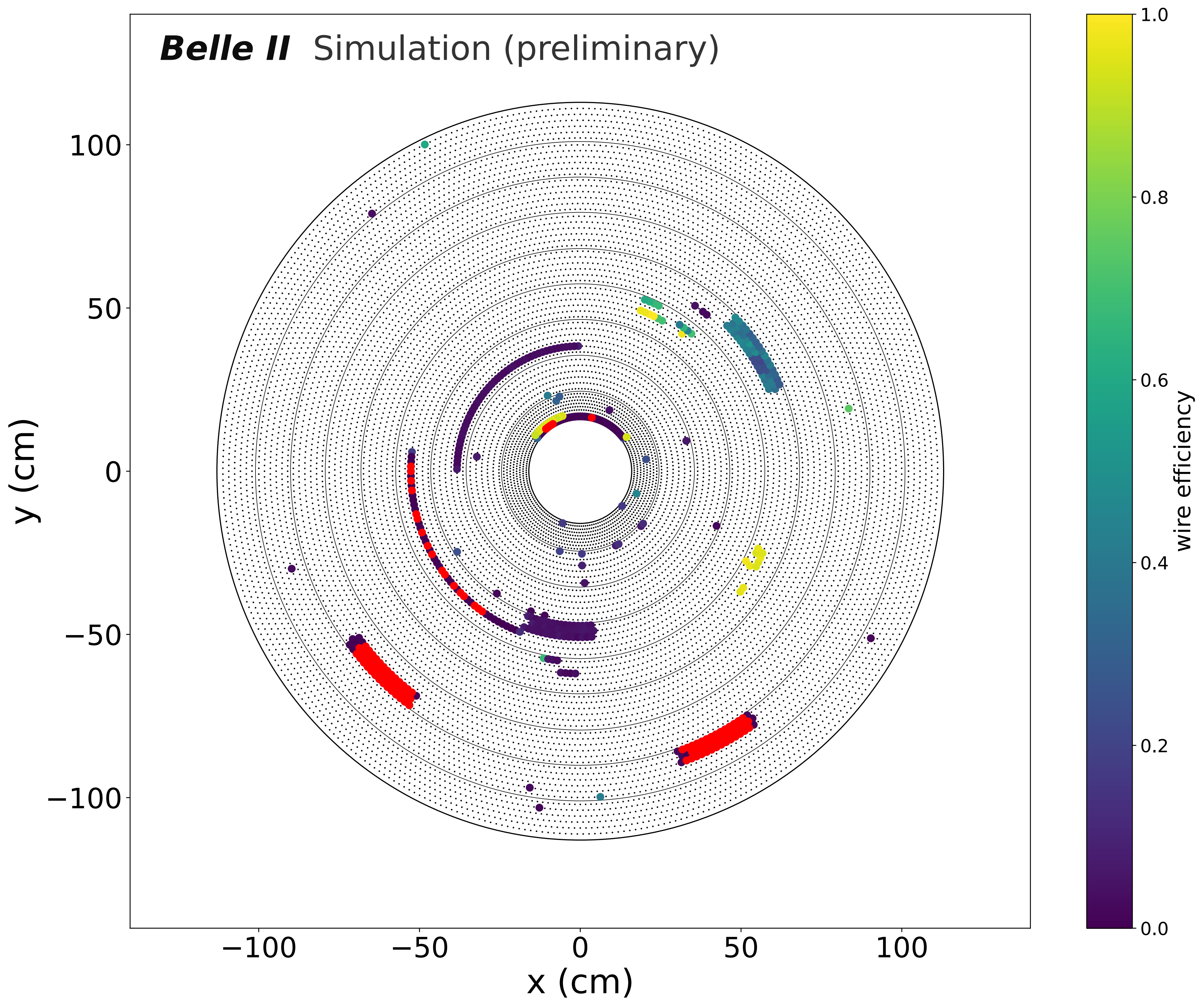




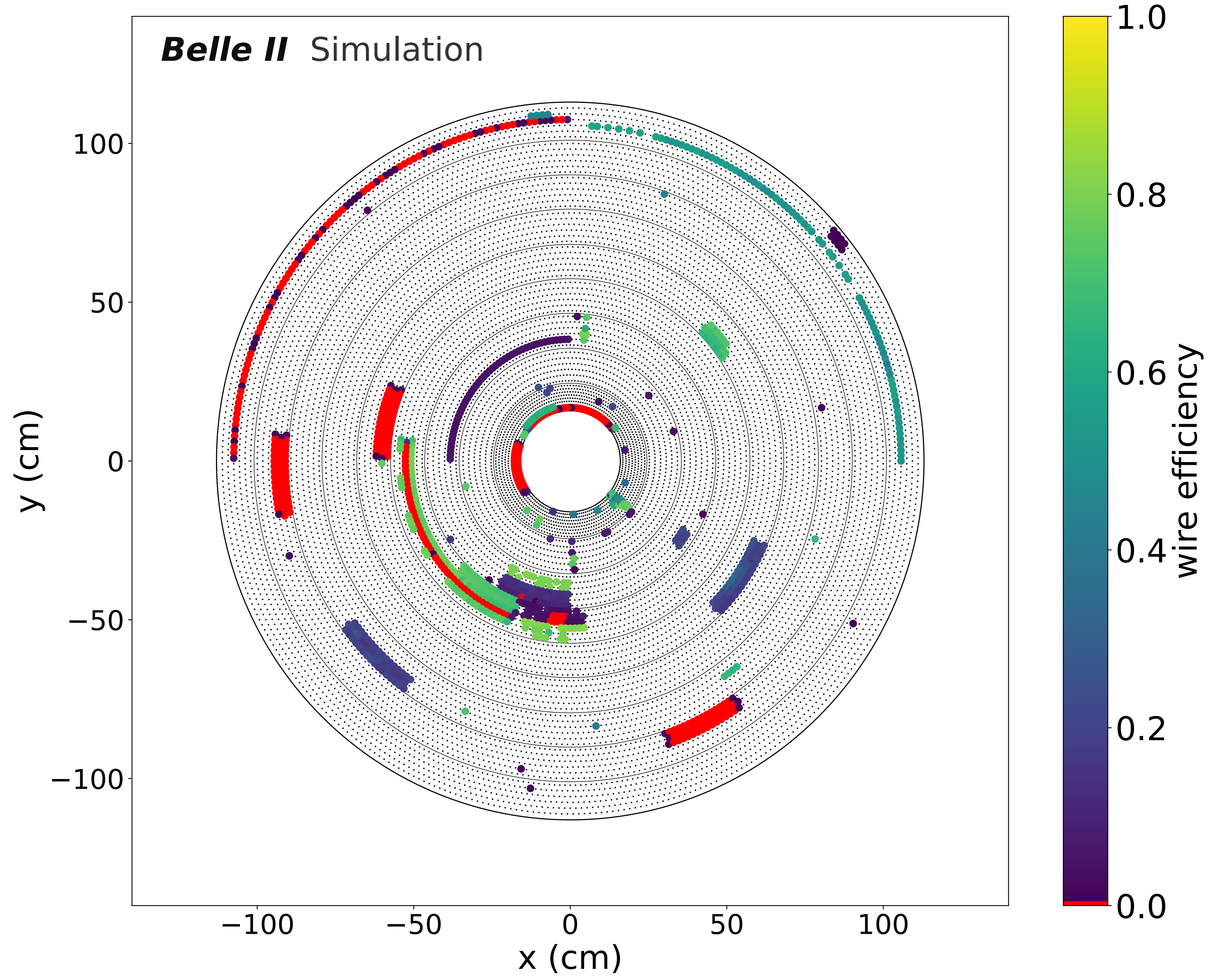
Backgrounds are getting higher, harder for tracking

Motivation - CDC Wire Inefficiencies

Beginning of 2021
50 wires off, 368 decreased efficiency
Total of 3% of the CDC wires



End of 2022
168 wires off, 809 decreased efficiency
Total of 7% of the CDC wires

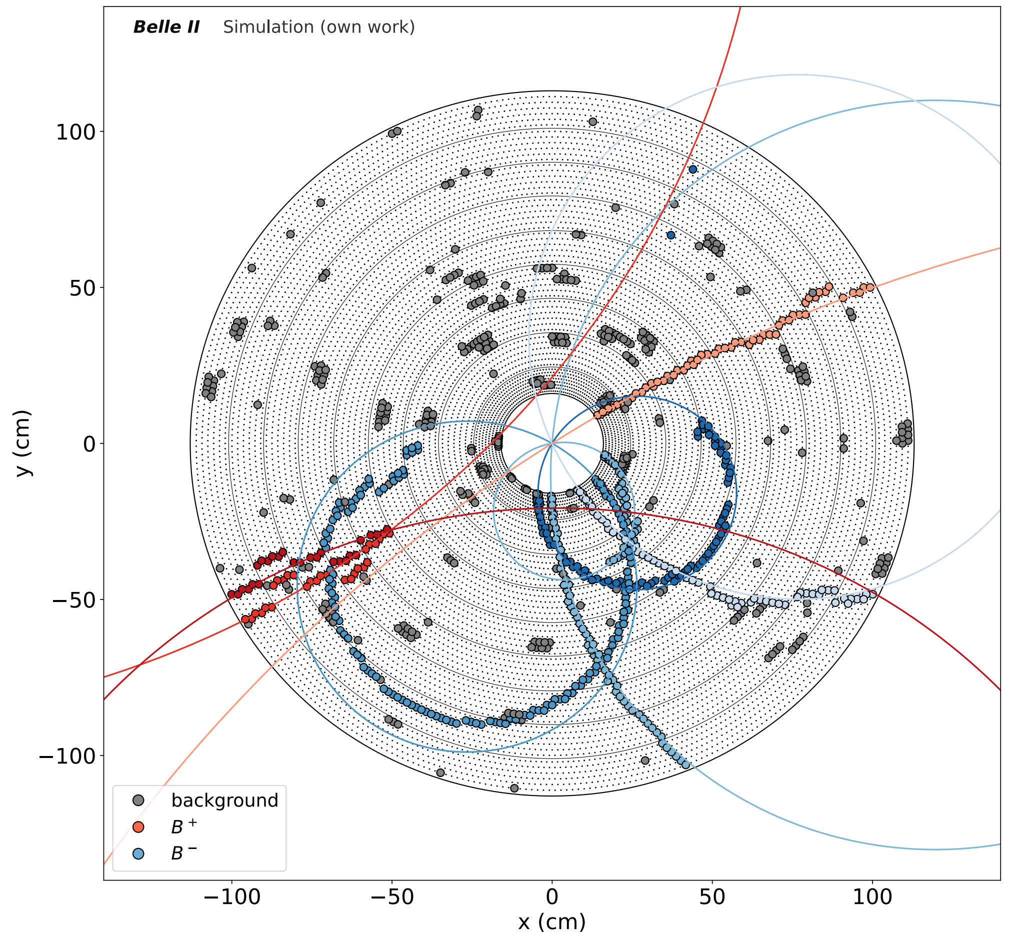


New Machine Learning Tracking Algorithm

Framing the Problem:

Have: Hits in our tracking detector
Want: Tracks

- Starting position
- Momentum (starting direction and curvature)
- Charge
- All hits belonging to the track

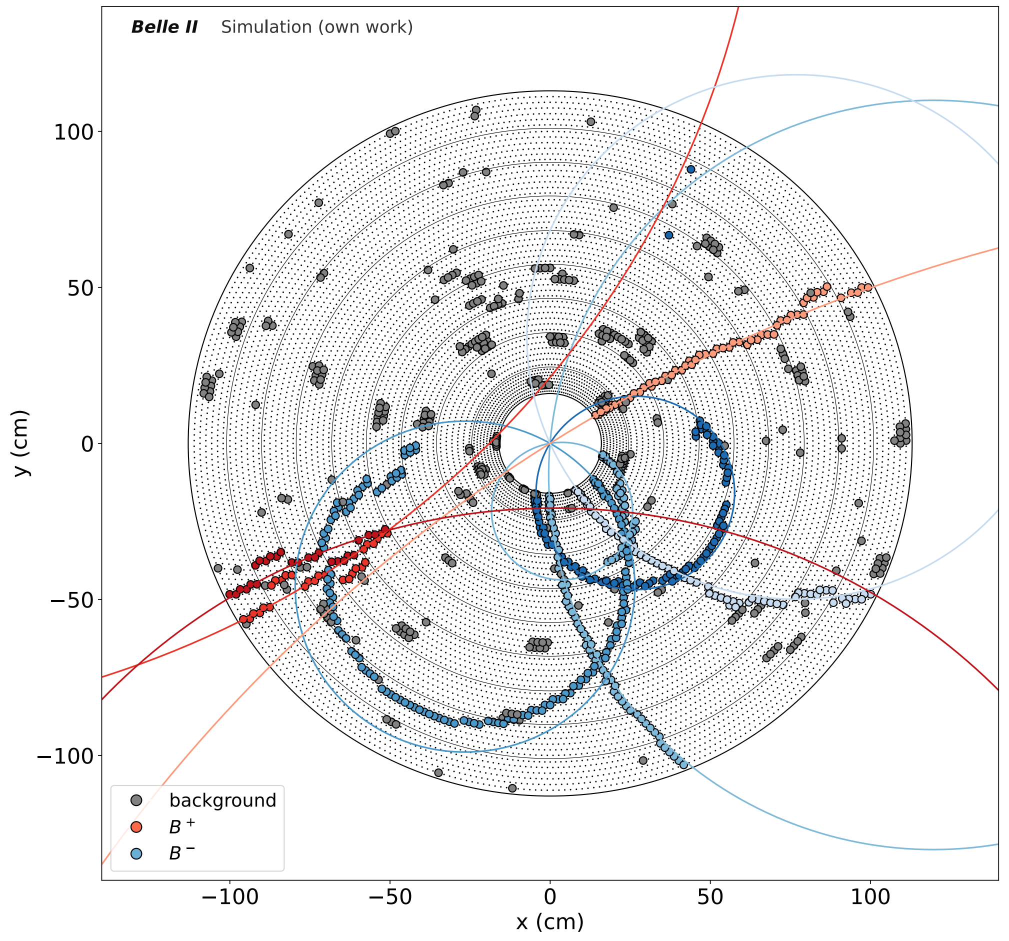


New Machine Learning Tracking Algorithm

Framing the Problem:

Have: Hits in our tracking detector
Want: Tracks

- Starting position
- Momentum (starting direction and curvature)
- Charge
- All hits belonging to the track



Challenges

- Unknown number of tracks
  - Very high event efficiency and purity required (find all tracks in each event)
  - Very large dynamic range of momenta (curvature)
  - Overlapping objects
- Large and varying number of sense wire inputs (15000 sense wires with up to 30% occupancy)
  - Varying background conditions
- High event rate leads to a lot of data, computing resource and time constraint

Our Approach

Graph Neural Network based Object Condensation

- Applicable to overlapping objects without clear spatial boundaries
- No assumption on the object size or sorting is needed

Kieseler, Object Condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data (arXiv:2002.03605)

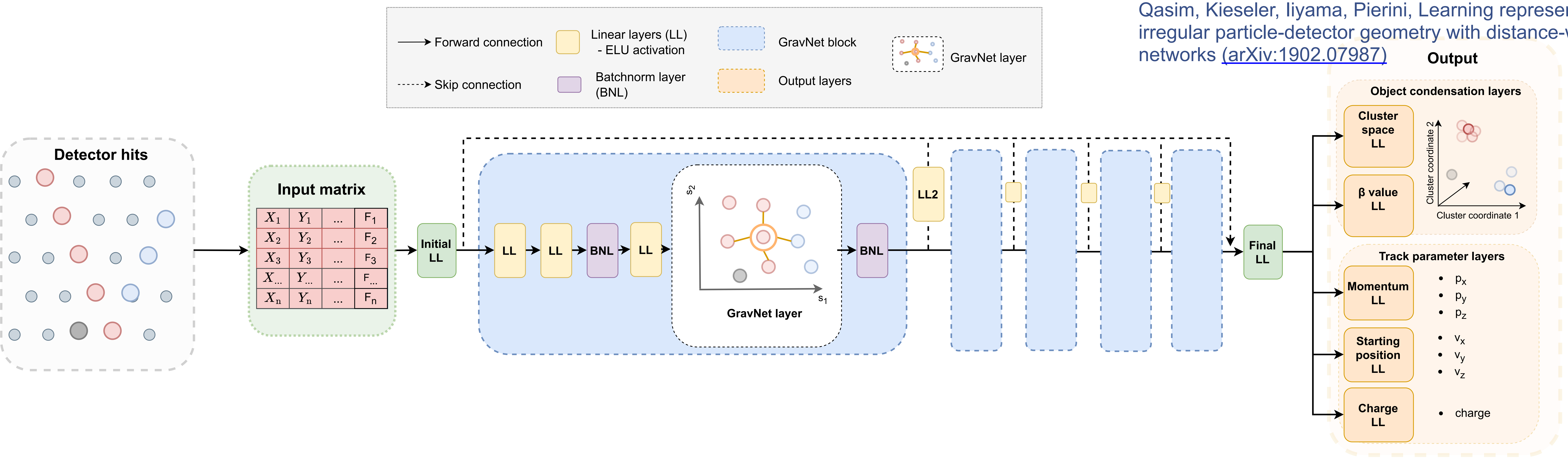
Our Approach

Graph Neural Network based Object Condensation

- Applicable to overlapping objects without clear spatial boundaries
- No assumption on the object size or sorting is needed

Kieseler, Object Condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data (arXiv:2002.03605)

Qasim, Kieseler, Iiyama, Pierini, Learning representations of irregular particle-detector geometry with distance-weighted graph networks (arXiv:1902.07987)



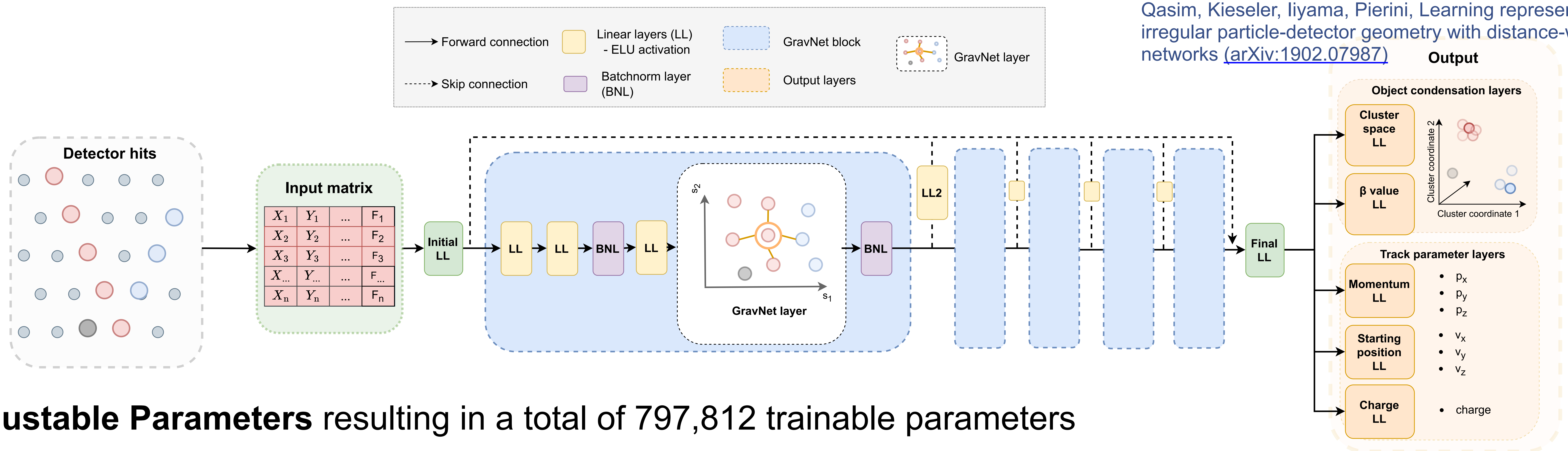
Our Approach

Graph Neural Network based Object Condensation

- Applicable to overlapping objects without clear spatial boundaries
- No assumption on the object size or sorting is needed

Kieseler, Object Condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data [(arXiv:2002.03605)](https://arxiv.org/abs/2002.03605)

Qasim, Kieseler, Iiyama, Pierini, Learning representations of irregular particle-detector geometry with distance-weighted graph networks [(arXiv:1902.07987)](https://arxiv.org/abs/1902.07987)



Adjustable Parameters resulting in a total of 797,812 trainable parameters

General Parameters

- Dimension of Linear Layers
- Number of GravNet Blocks

GravNet Parameters

- Number of k-nearest neighbours in GravNet
- GravNet space dimensions

Output Parameters

- Dimension of Cluster Coordinates
- Number of output layers according to track parameter predictions

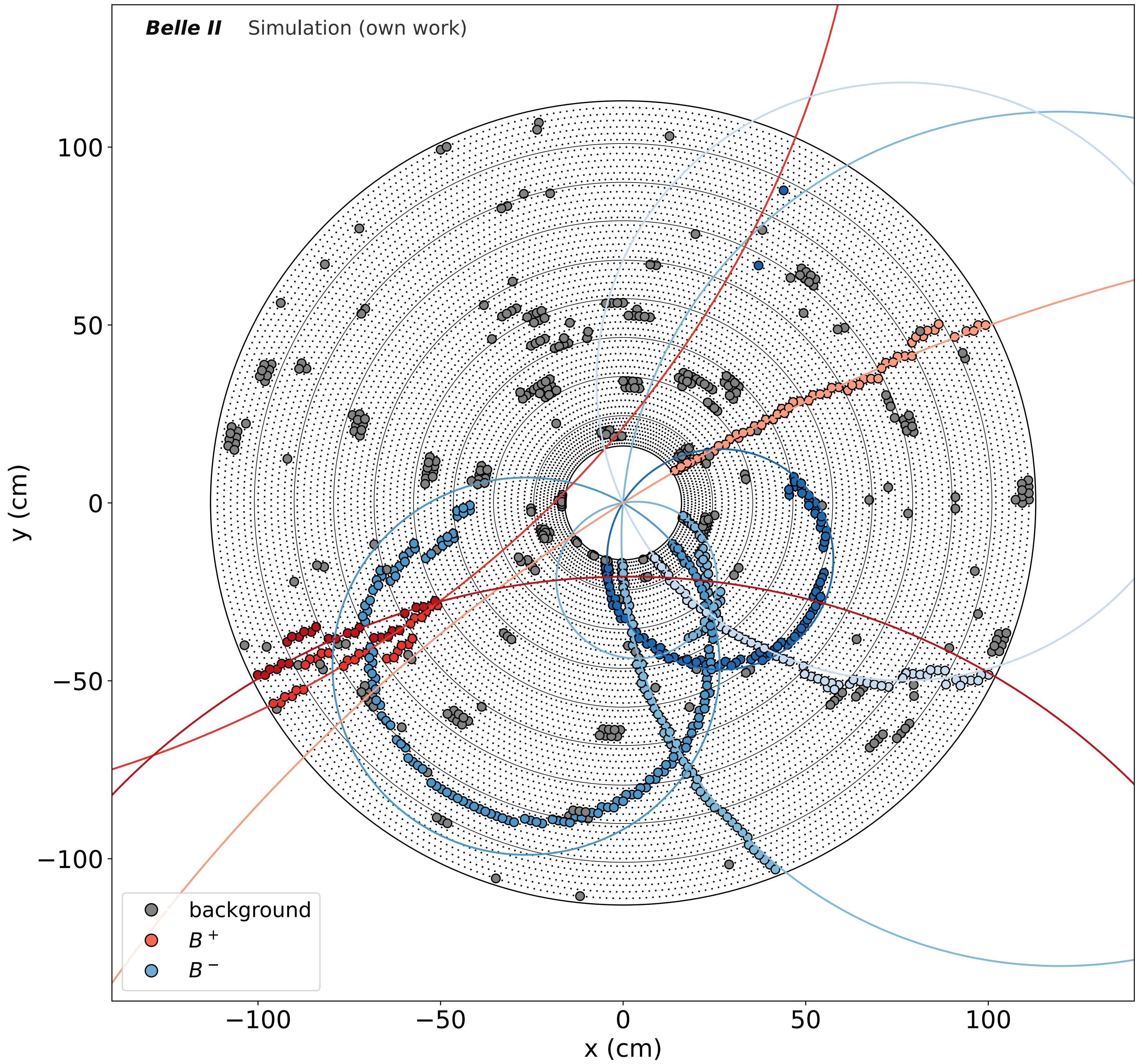
Training samples - Input Features

- Cartesian coordinates for the CDC hits:
  - X - position
  - Y - position

(We don't want polar coordinates due to the flip at 2π and because the radius diverges for high momentum)

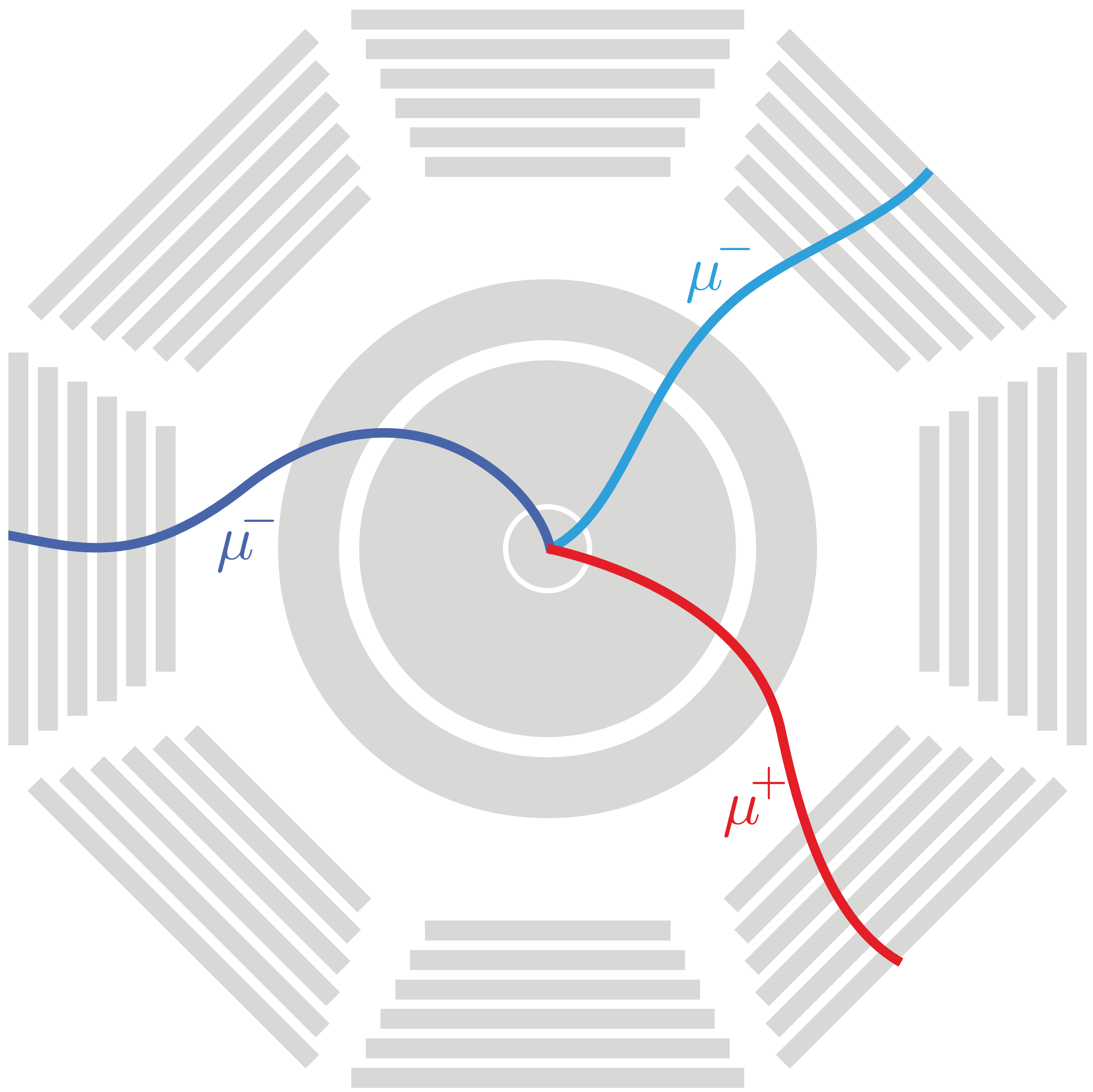
- Wire Signal Measurements
  - Signal height
  - Signal timing
  - Signal time over threshold
- CDC layer information
  - Superlayer
  - Layer within superlayer
  - Total layer

Bachelor thesis about the input feature optimization

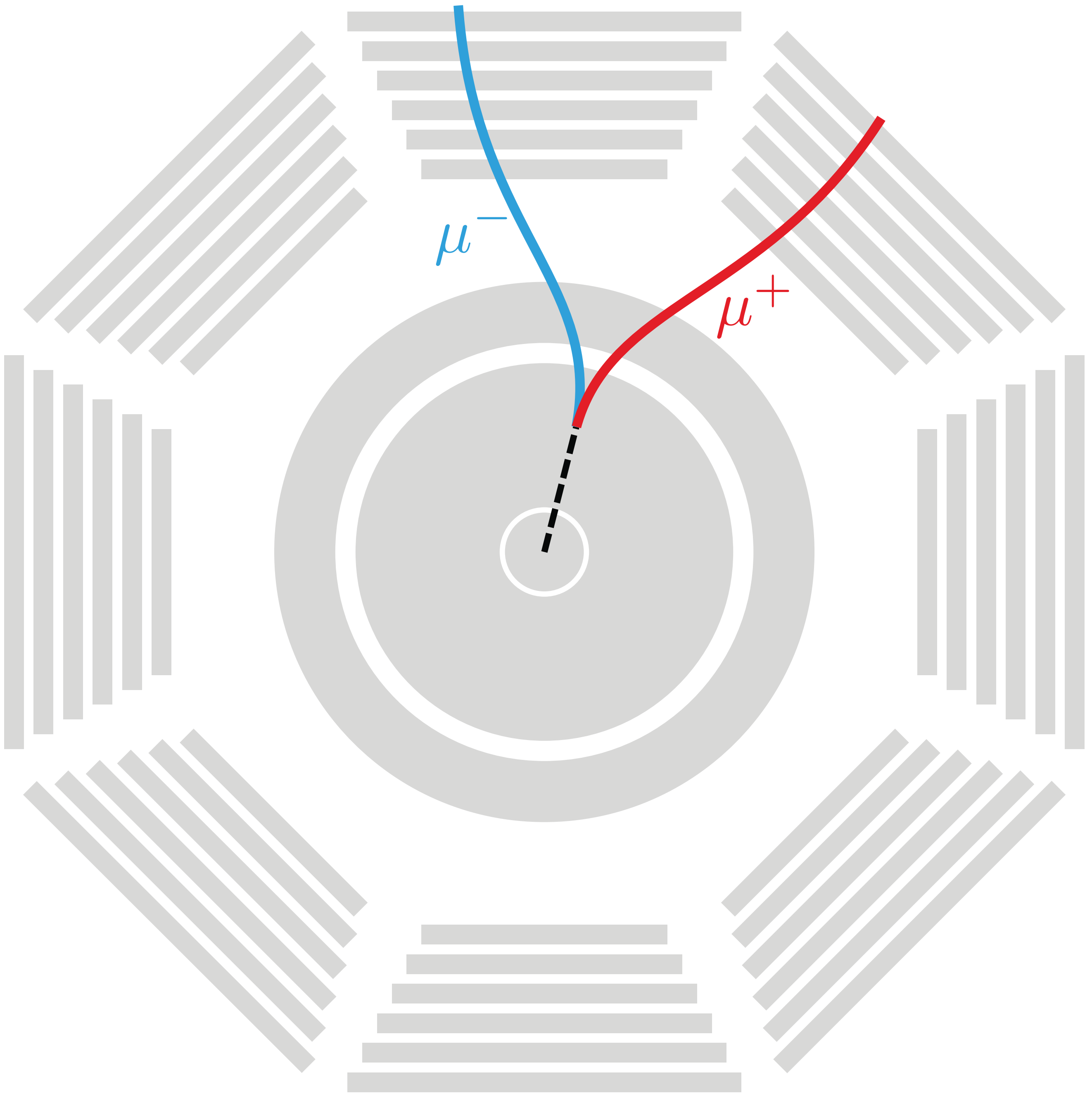


Training Samples - Topologies

Prompt particles from the interaction point

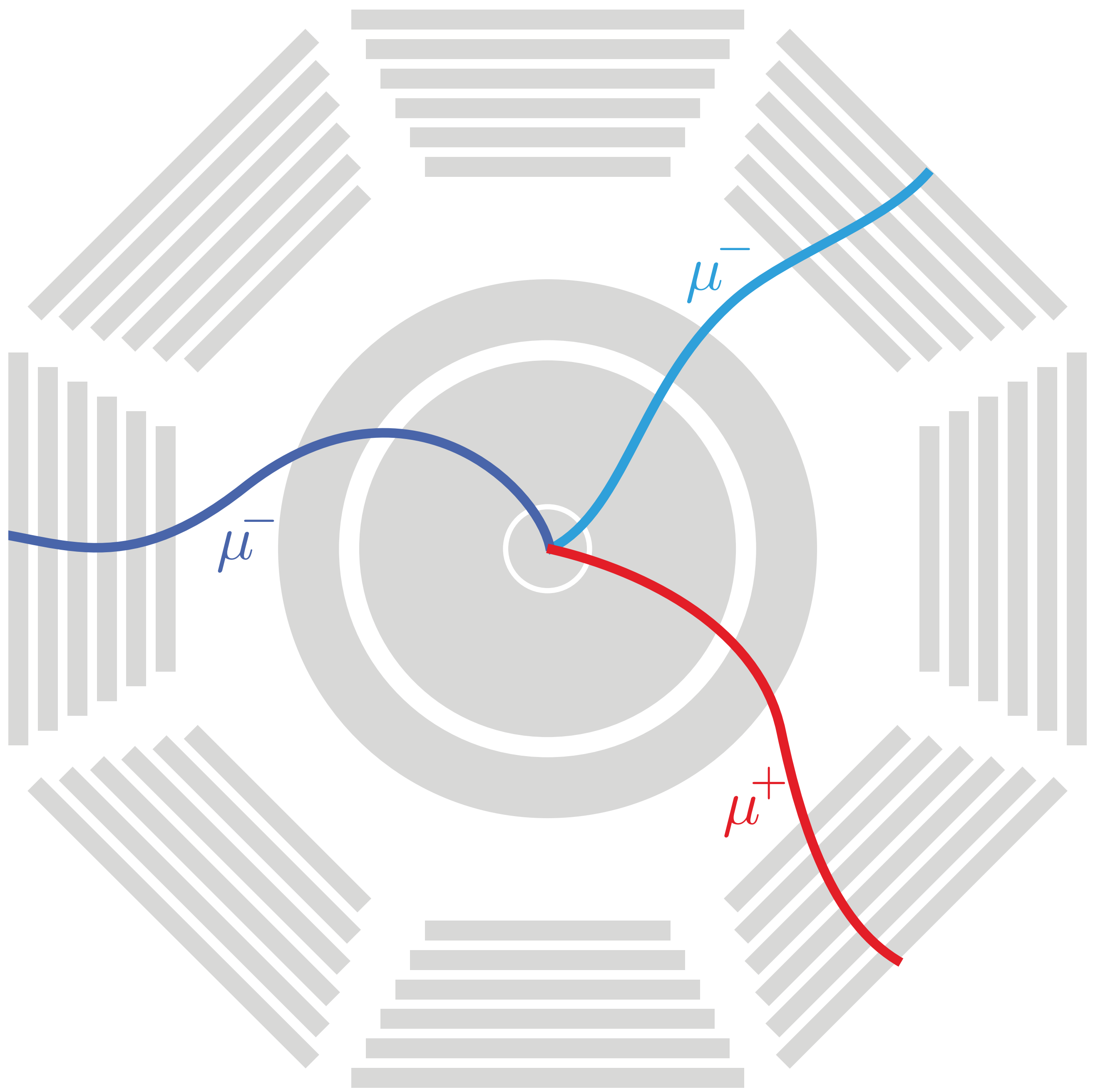


Particles from displaced vertices

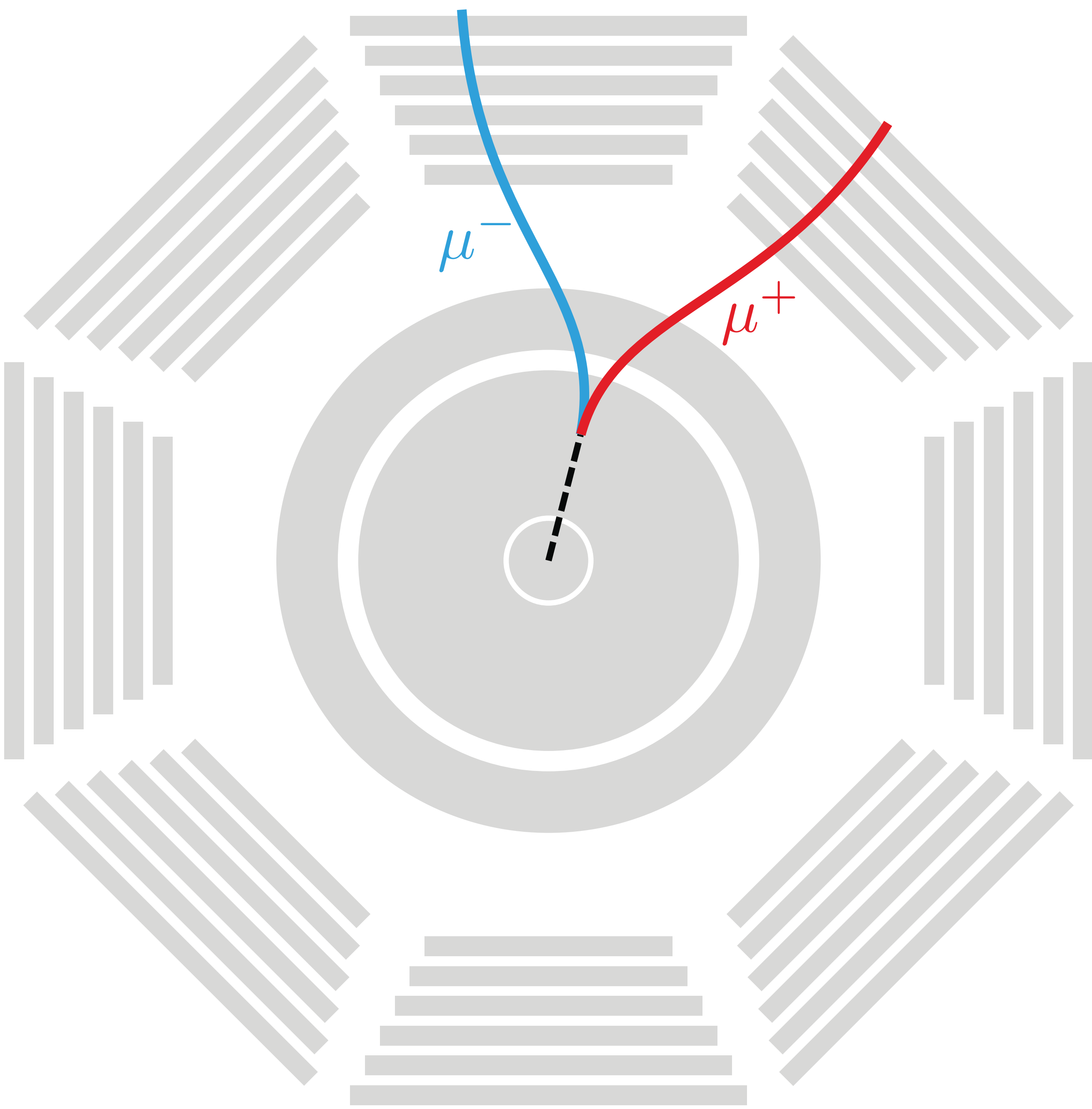


Training Samples - Topologies

Prompt particles from the interaction point



Particles from displaced vertices

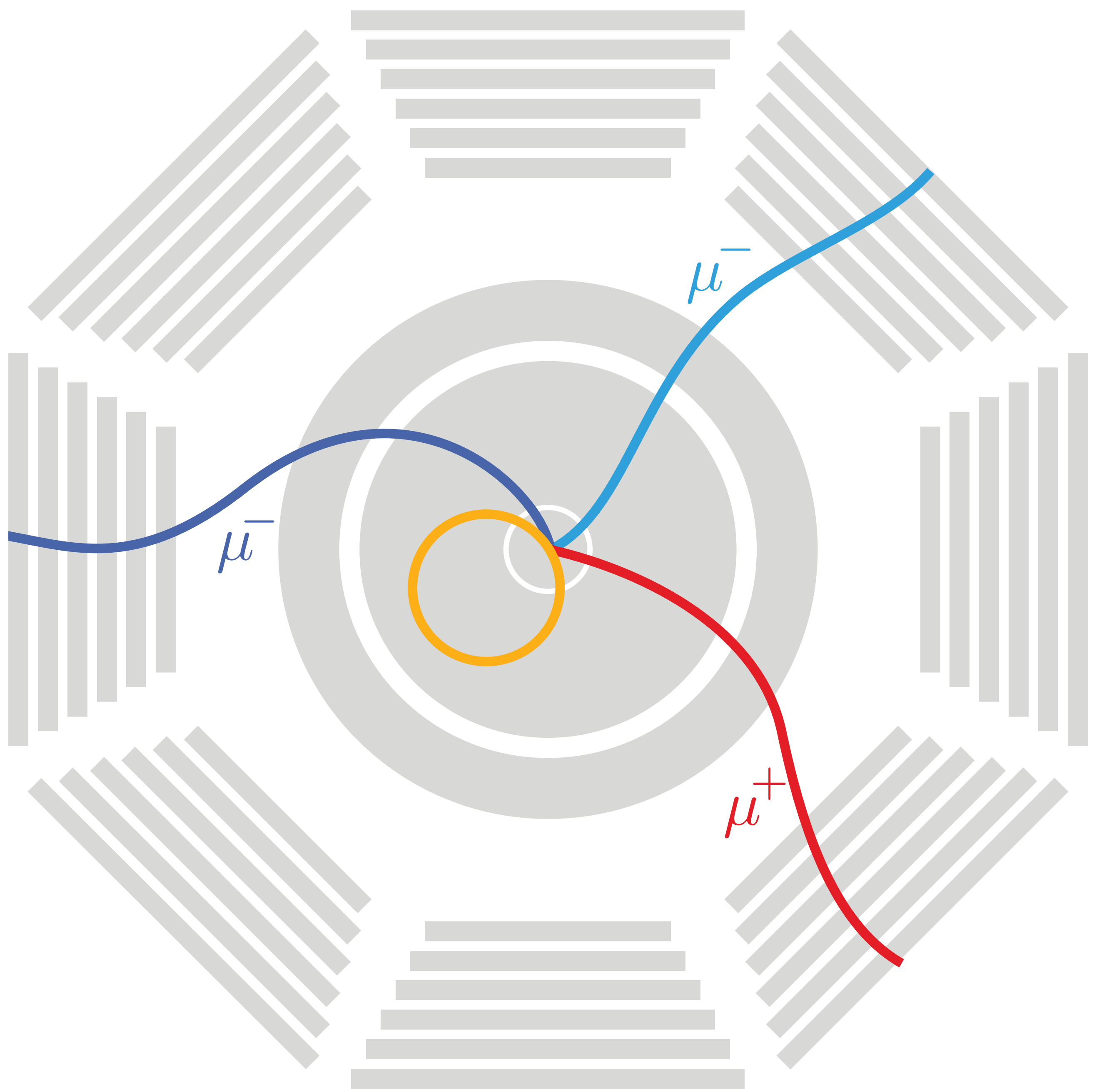


Goal: general model that can be used for every possible physics event

- The model proved to be very good in learning physics dependencies
- To not bias the model towards any physics events, we chose particle gun samples where we sample
  - random positive or negative charge
  - random momentum within detector acceptance (no energy or momentum conservation)
  - random displacement
  - random number of particles per event

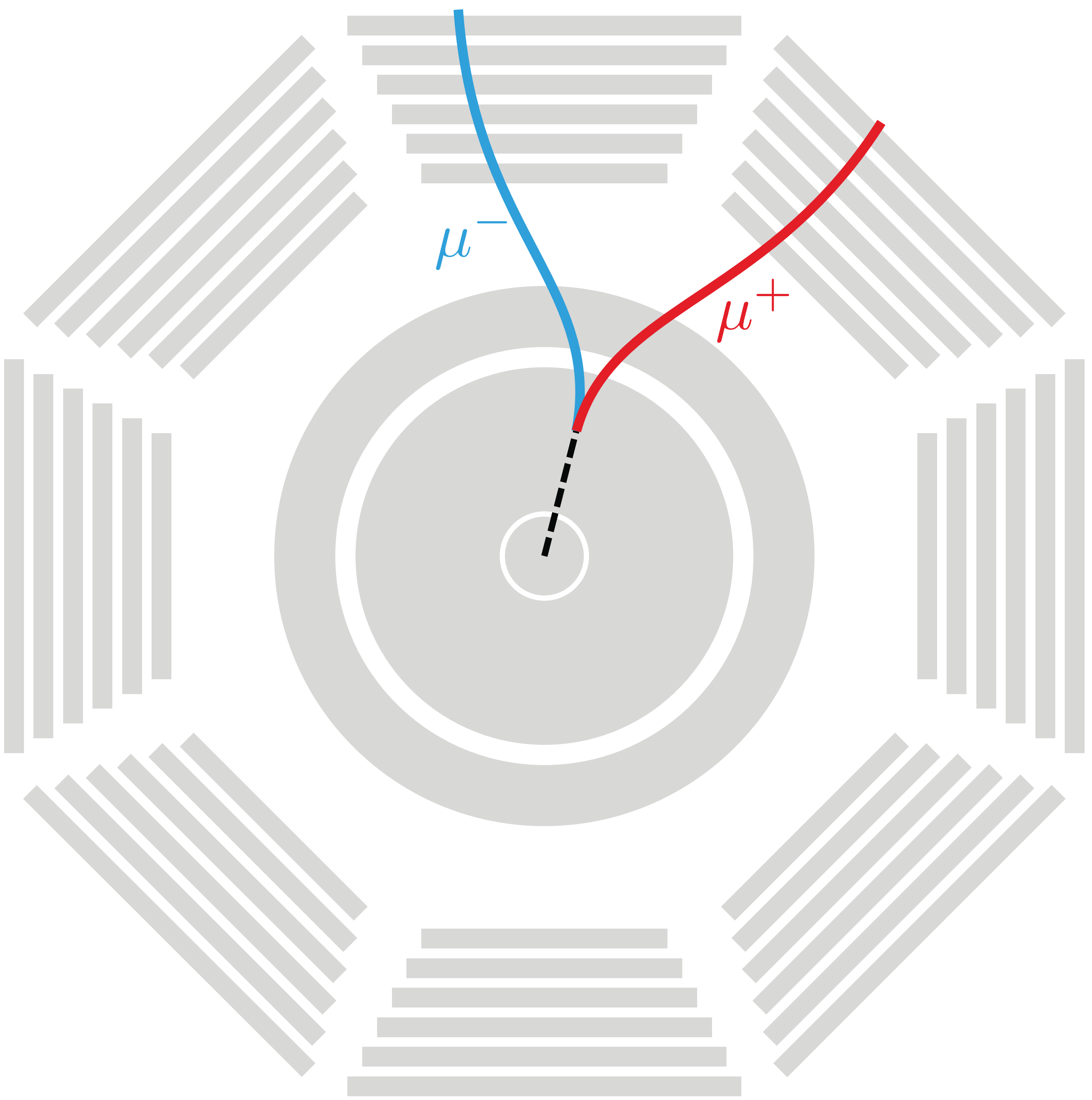
Training Samples - Topologies

Prompt particles from the interaction point



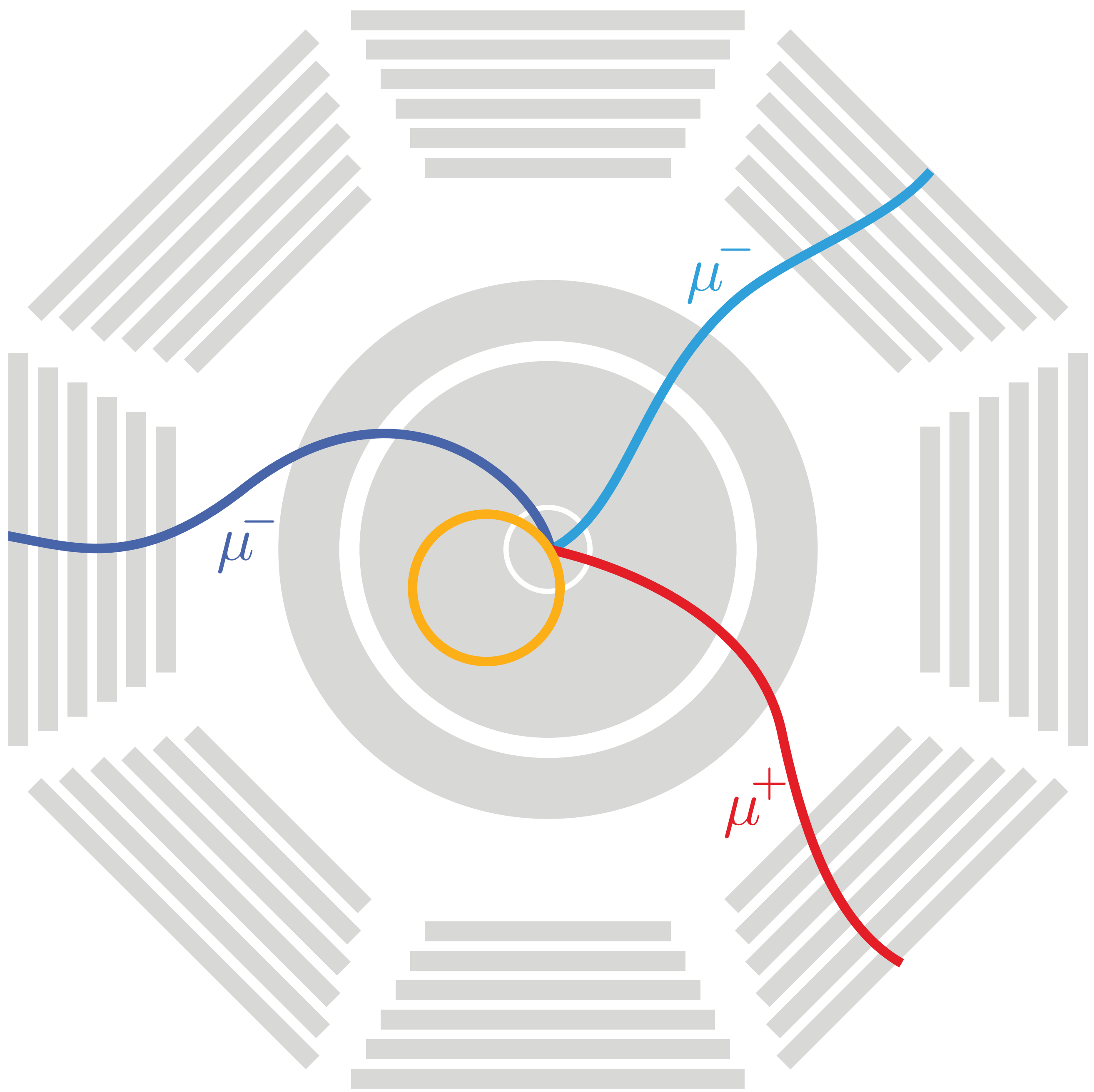
Enrich trainings samples with low momentum particles

Particles from displaced vertices

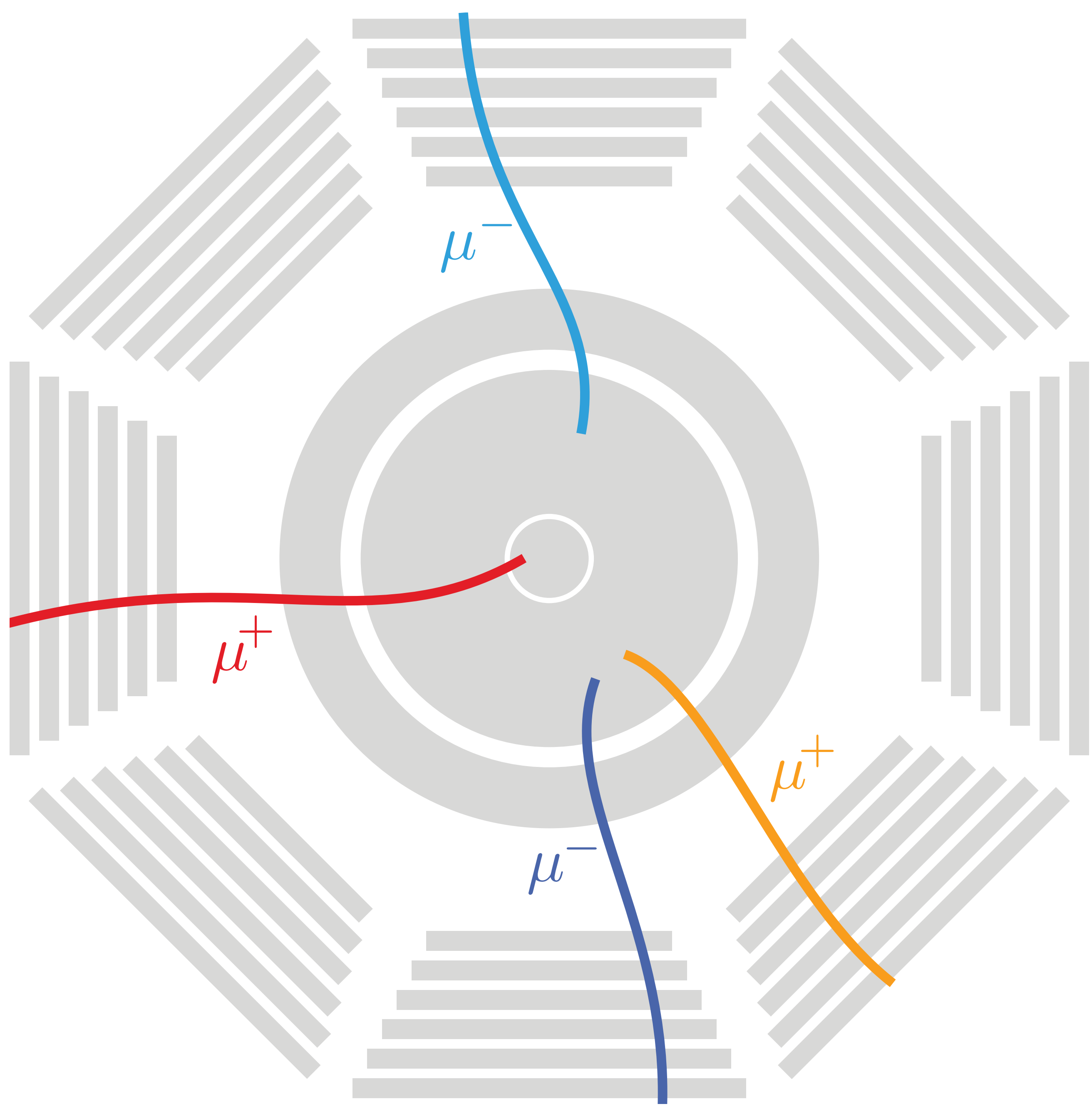


Training Samples - Topologies

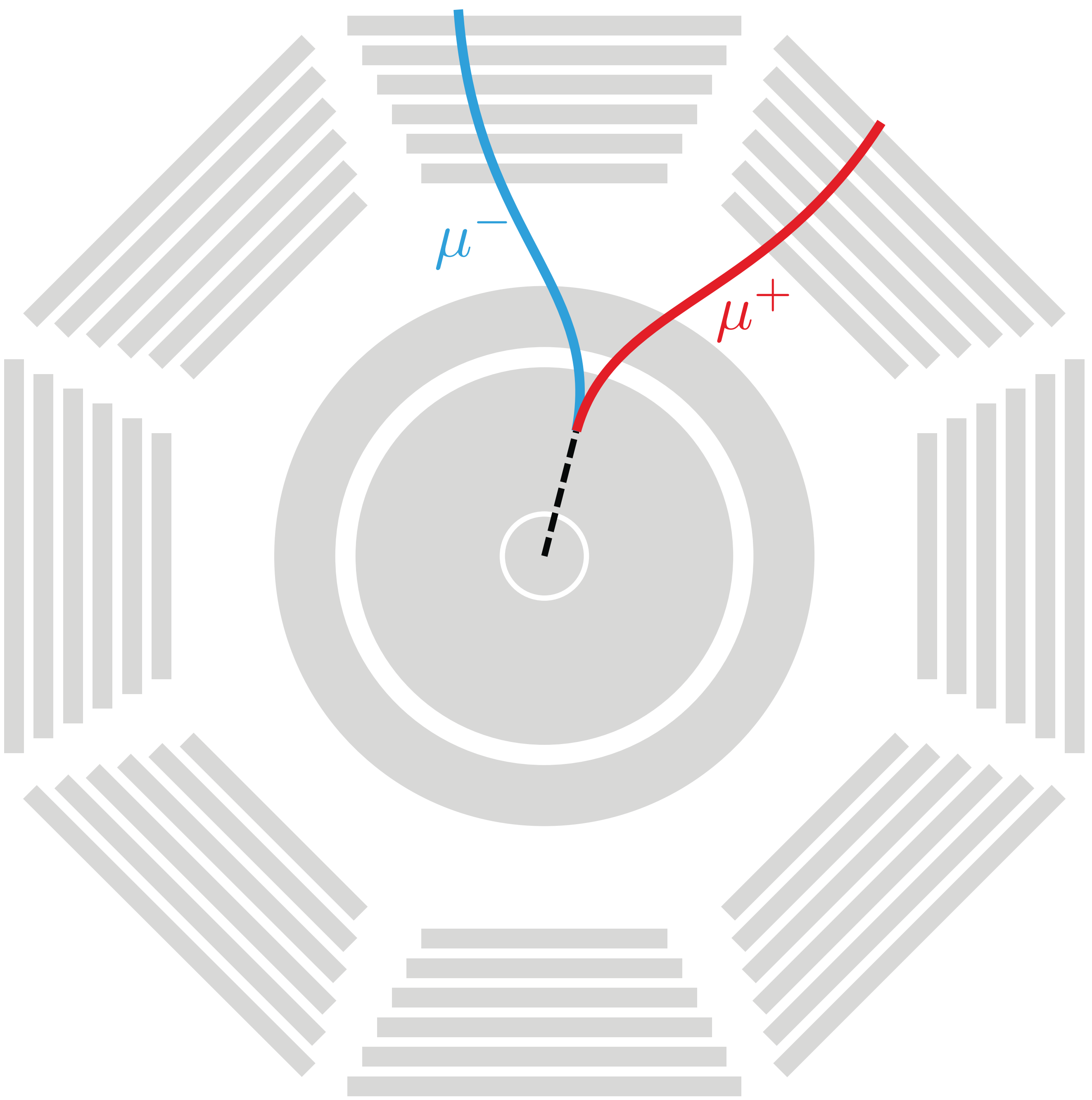
Prompt particles from the interaction point



Adding the transition sample of single displaced particles helped training

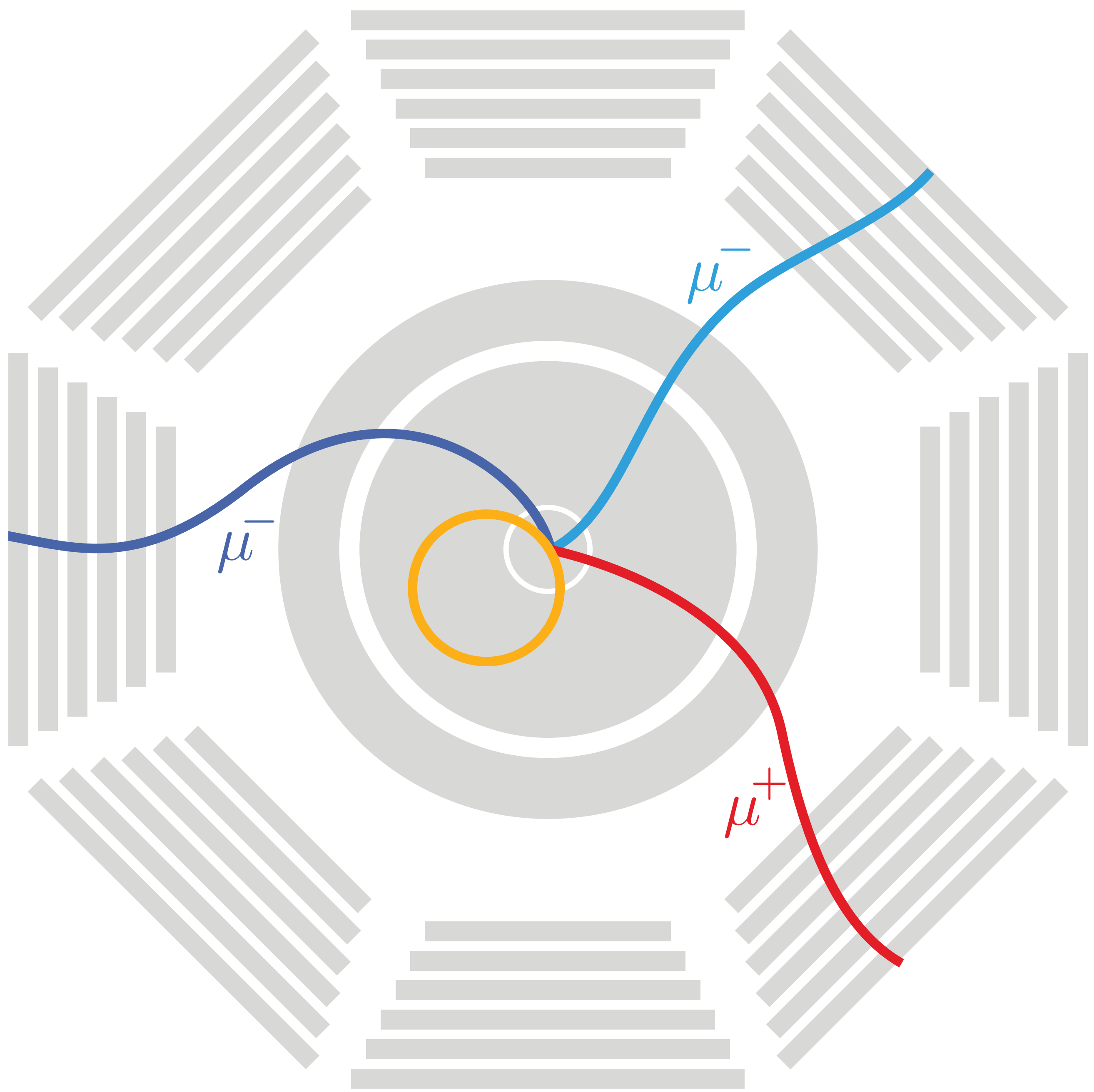


Particles from displaced vertices

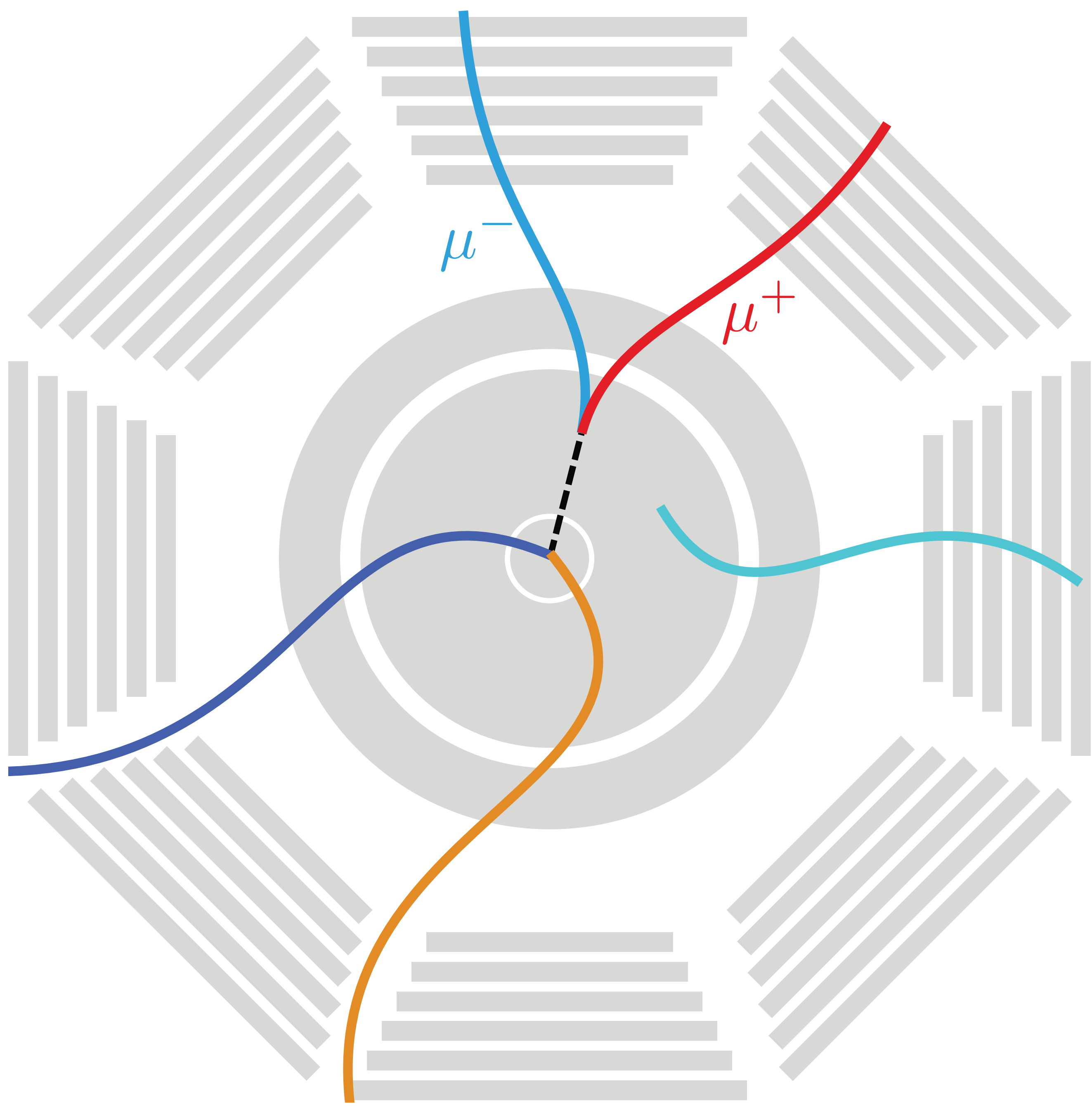


Training Samples - Topologies

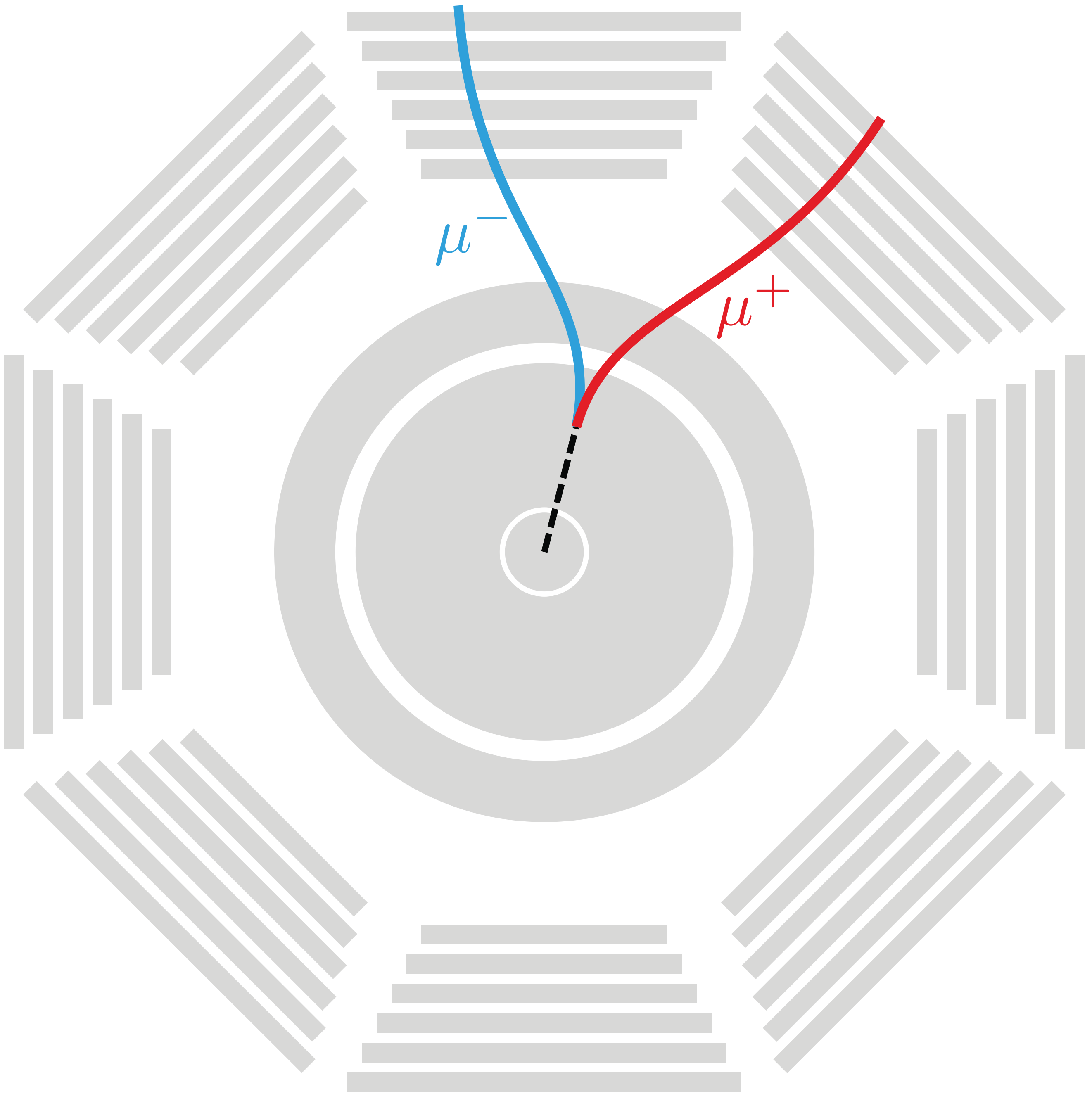
Prompt particles from the interaction point



Mix everything so that the model does not learn topology dependency



Particles from displaced vertices



1 million samples to train on, split 80/20 in training and validation dataset
Equal distribution between all training topologies (prompt, displaced, displaced vertex and mix)

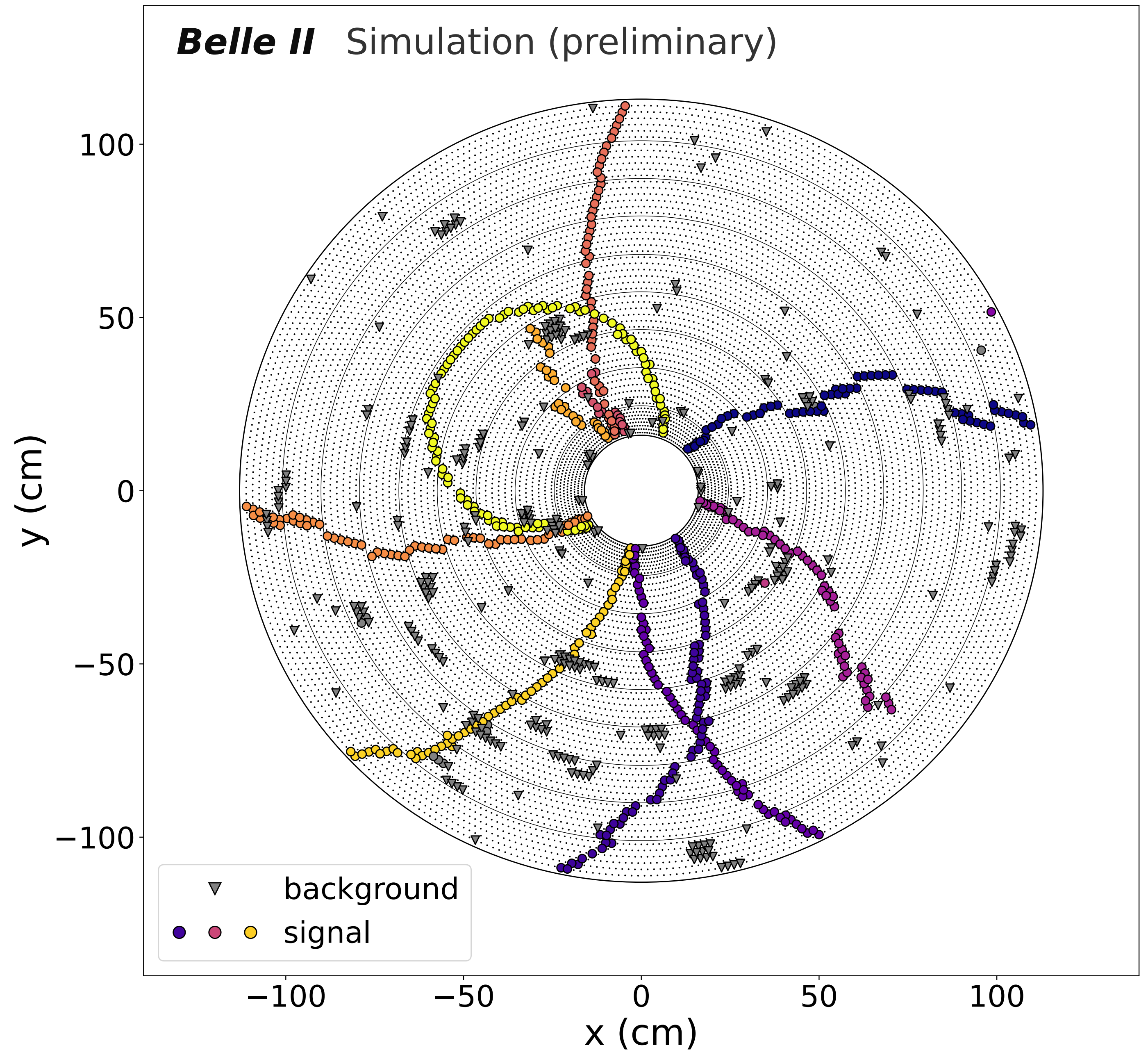
Training Samples - Backgrounds

Extensive background simulations, correlated over multiple detector components, also including noise from read-out chips

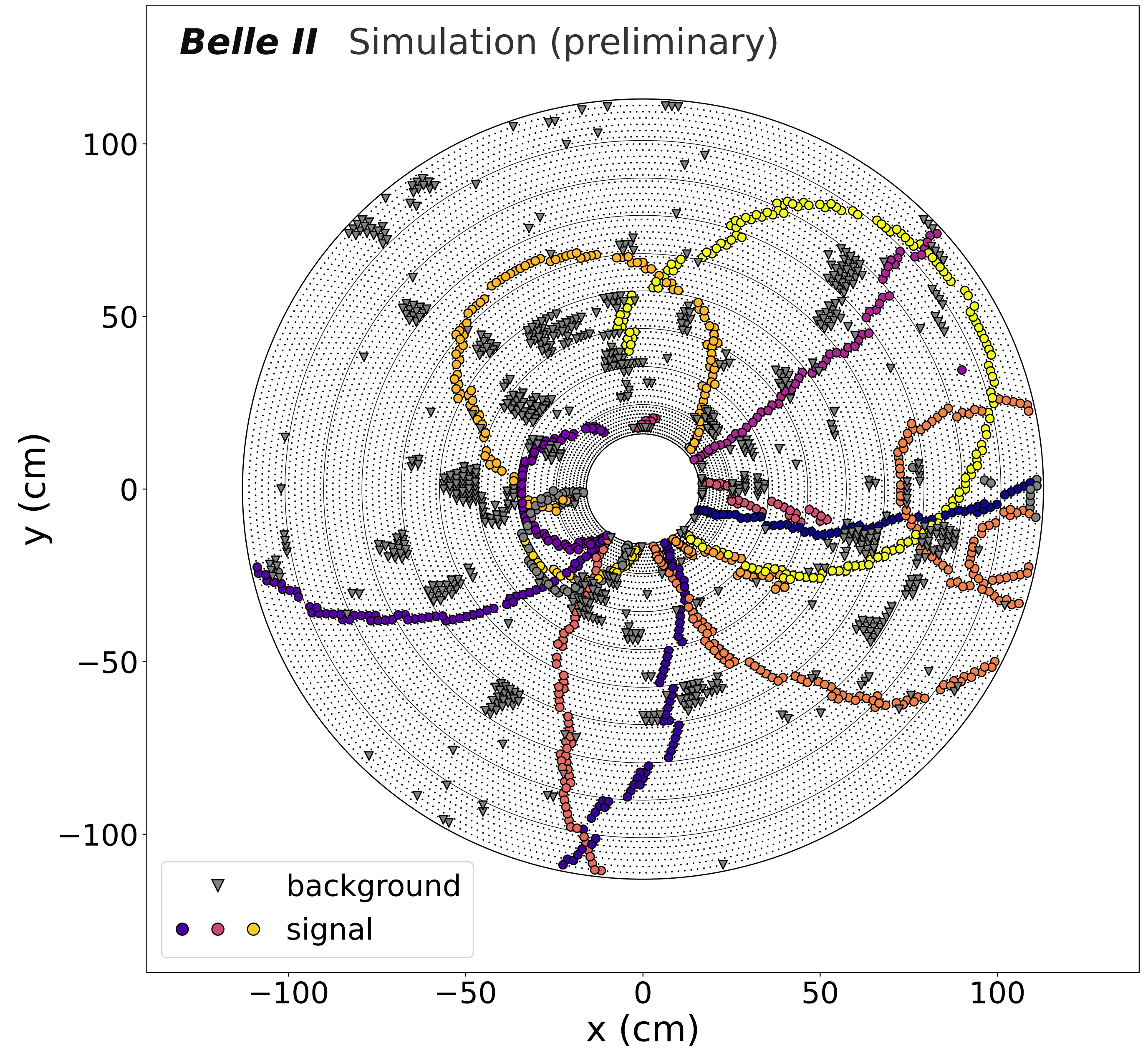
Natochii et al., Beam background expectations for Belle II at SuperKEKB (arXiv:2203.05731)

Liptak et al., Measurements of Beam Backgrounds in SuperKEKB Phase 2 (arXiv:2112.14537)

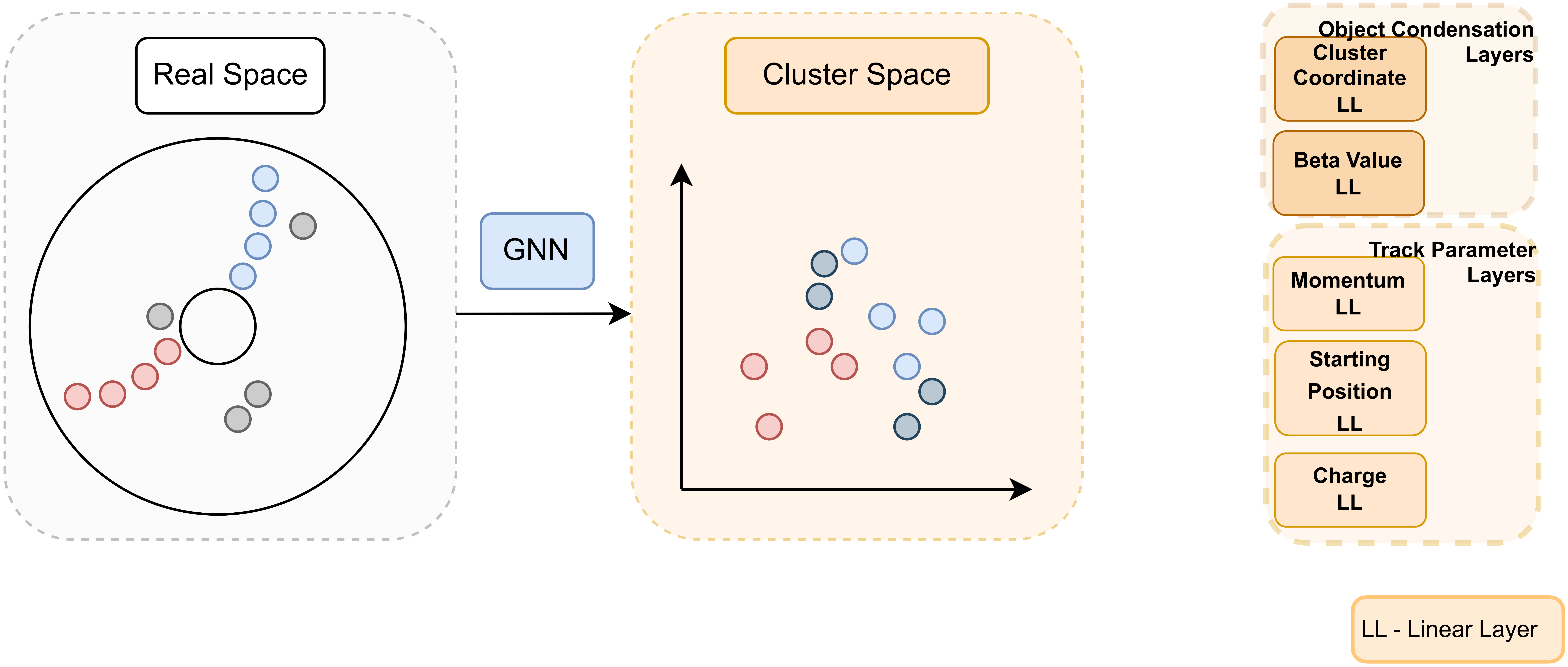
Pretrain model on simulated, low background conditions



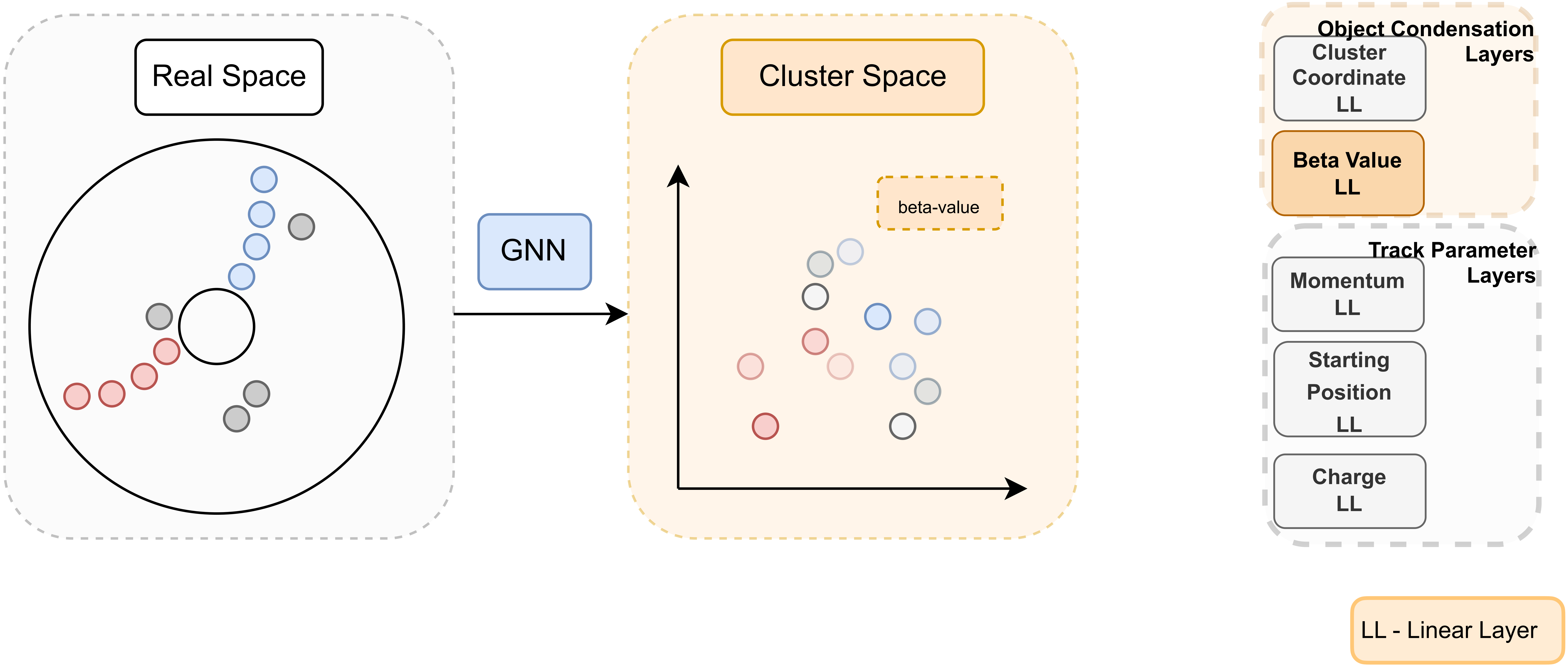
Retrain model on high background conditions taken from data during end of run 1 in 2022



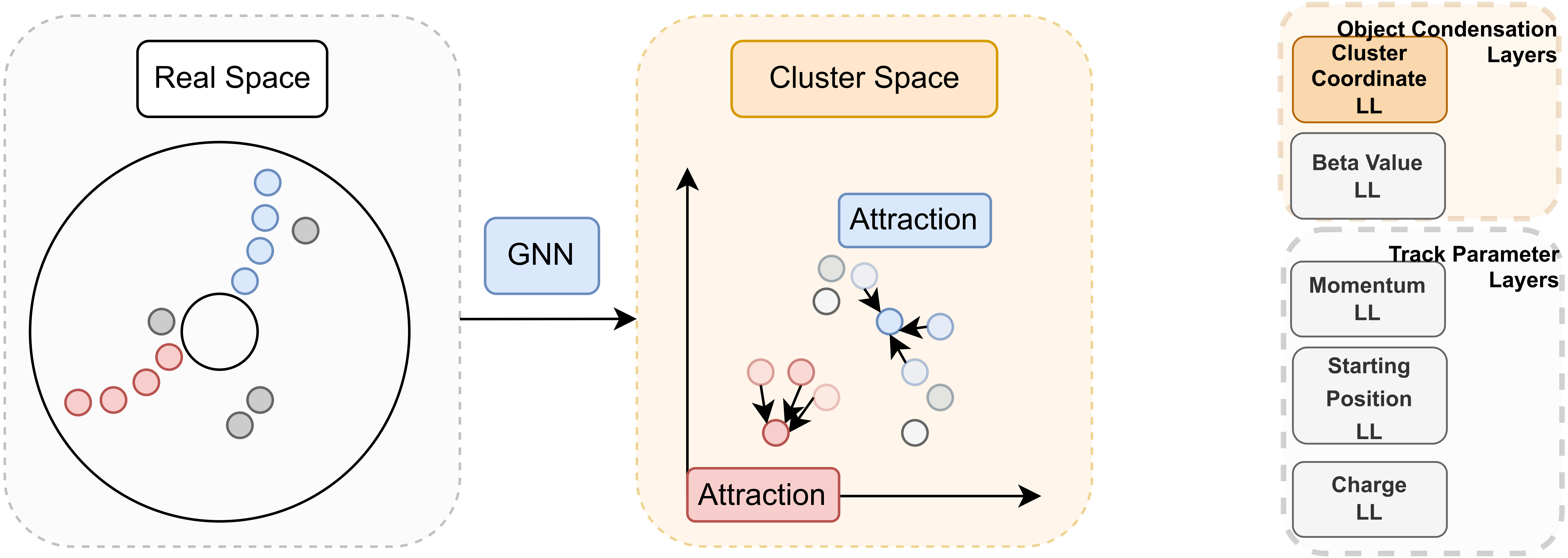
Object Condensation - Training Loss



Object Condensation - Training Loss
β-loss

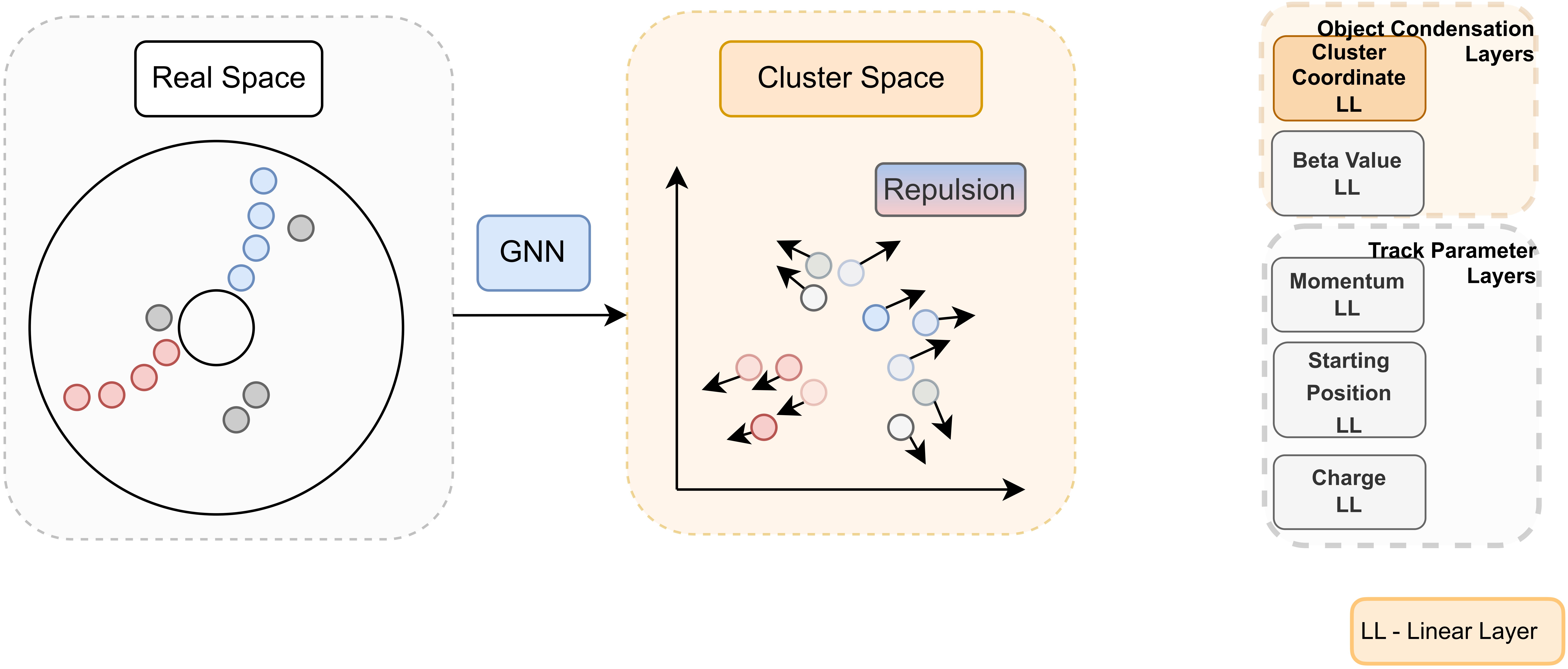


Object Condensation - Training Loss
Attraction loss

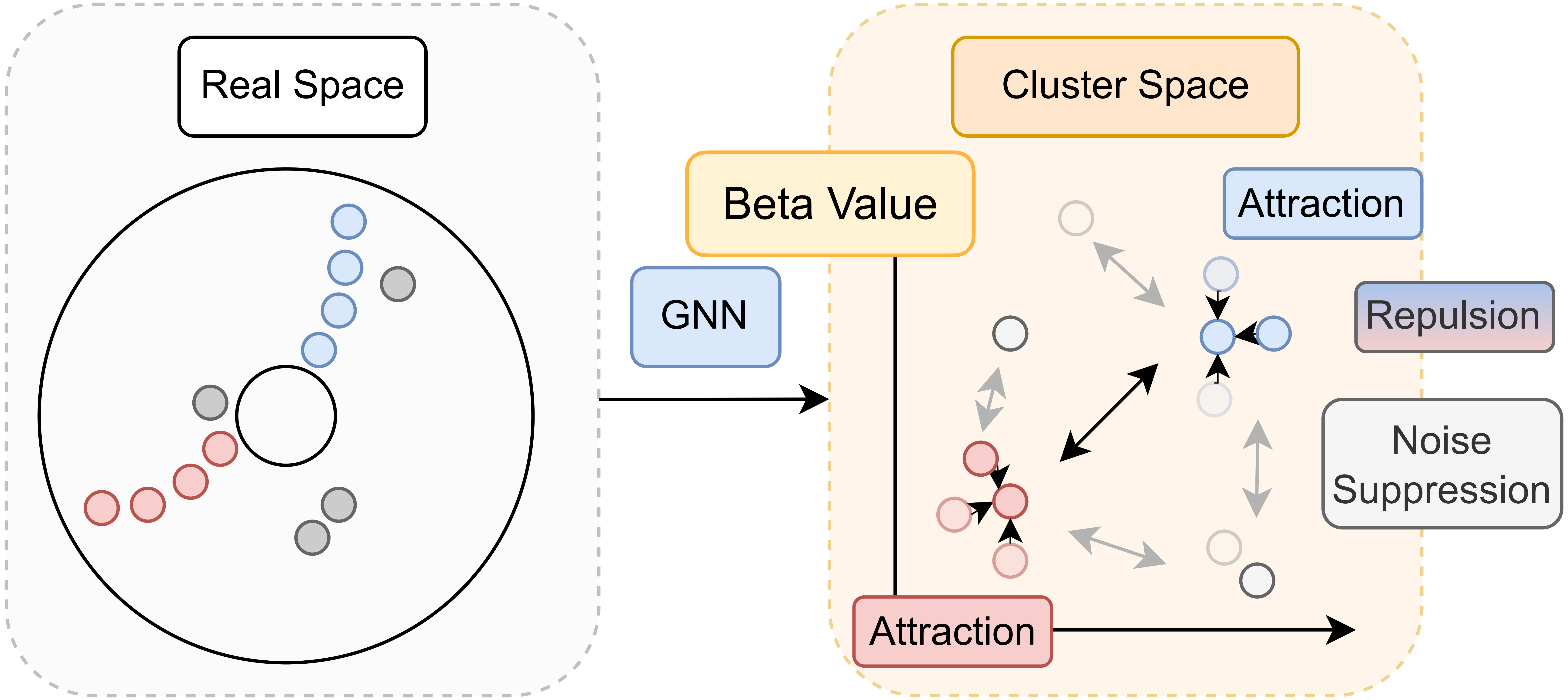


LL - Linear Layer

Object Condensation - Training Loss
Repulsion loss

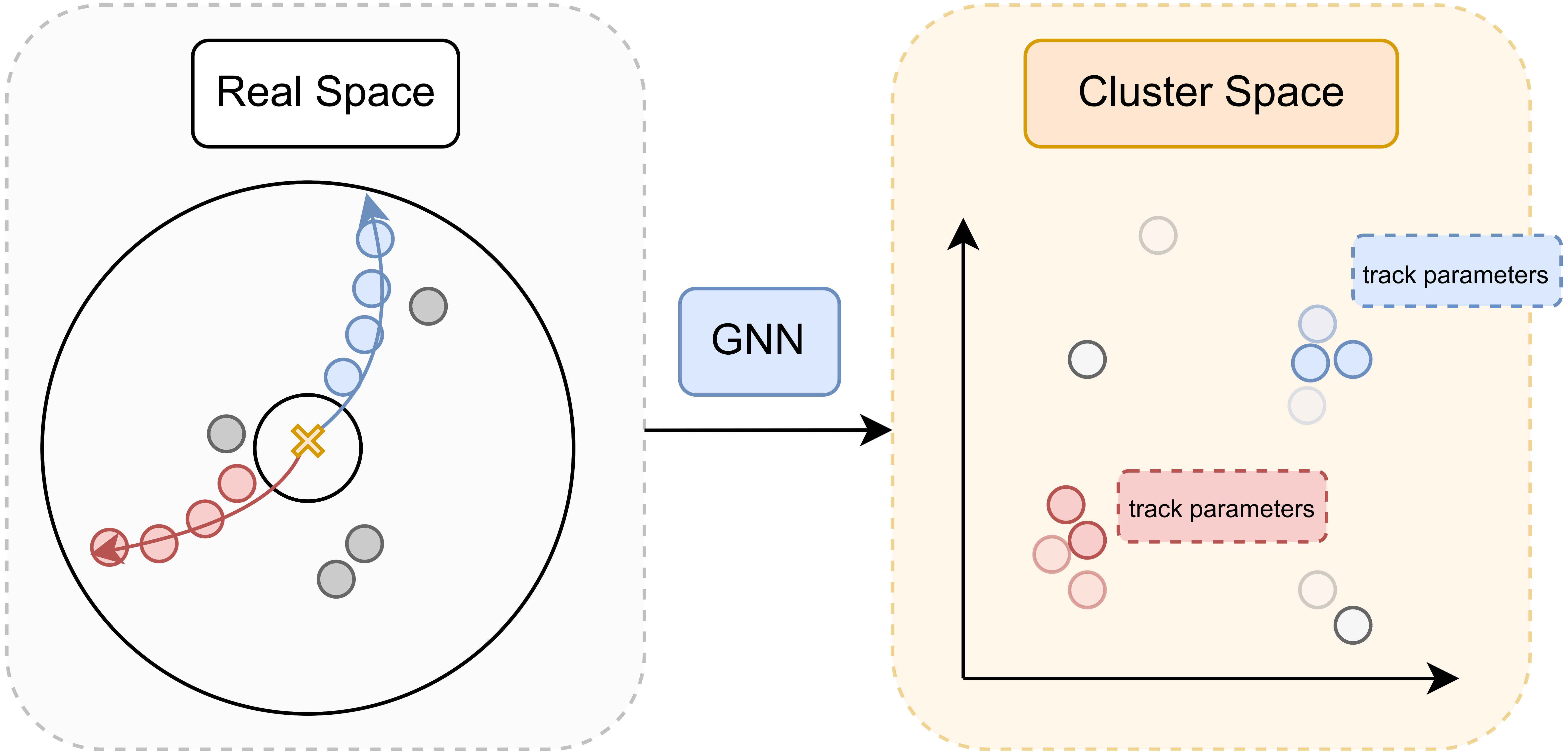


Object Condensation - Training Loss
Noise suppression



4 loss terms for object condensation

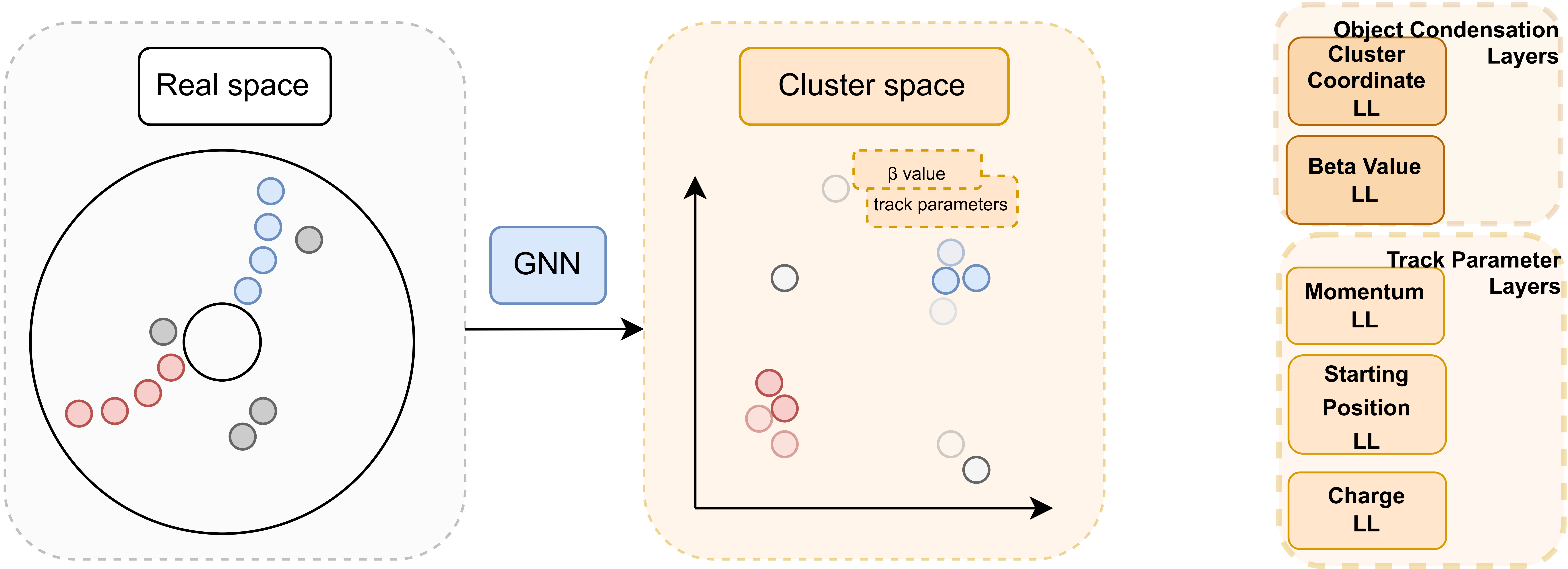
Object Condensation - Training Loss
Parameter loss



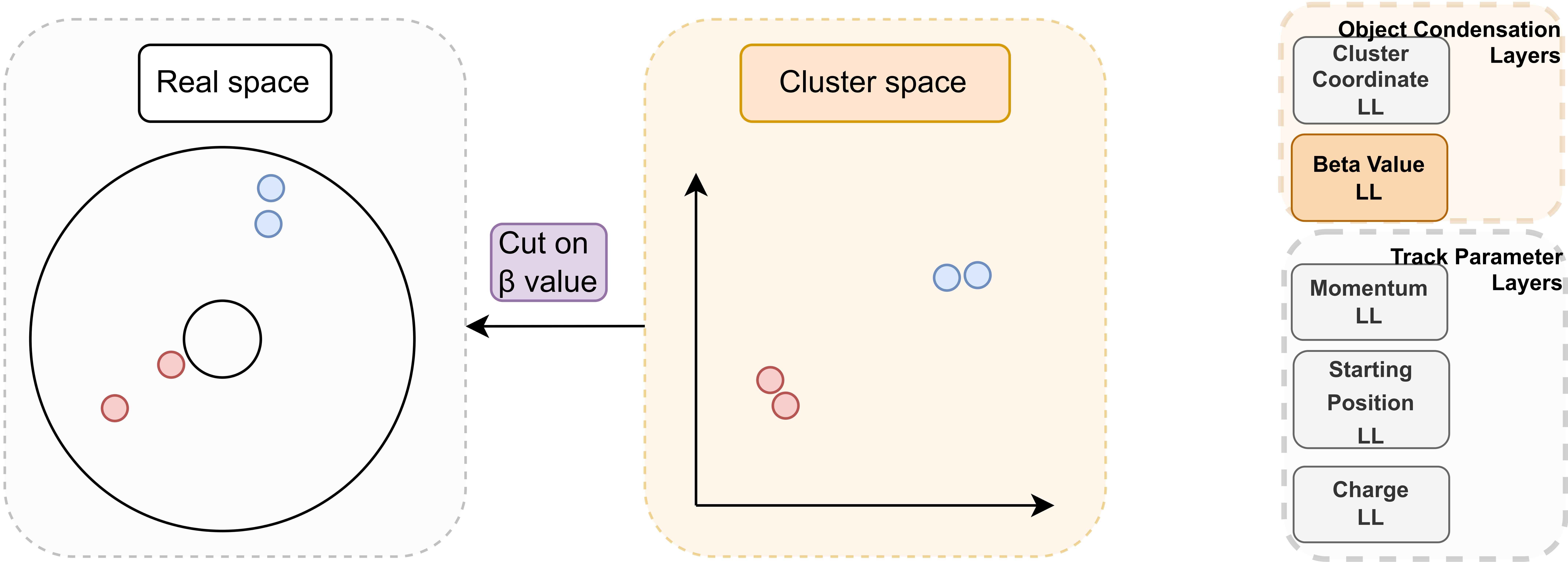
4 loss terms for object condensation

Additional 3 loss terms according to parameter predictions

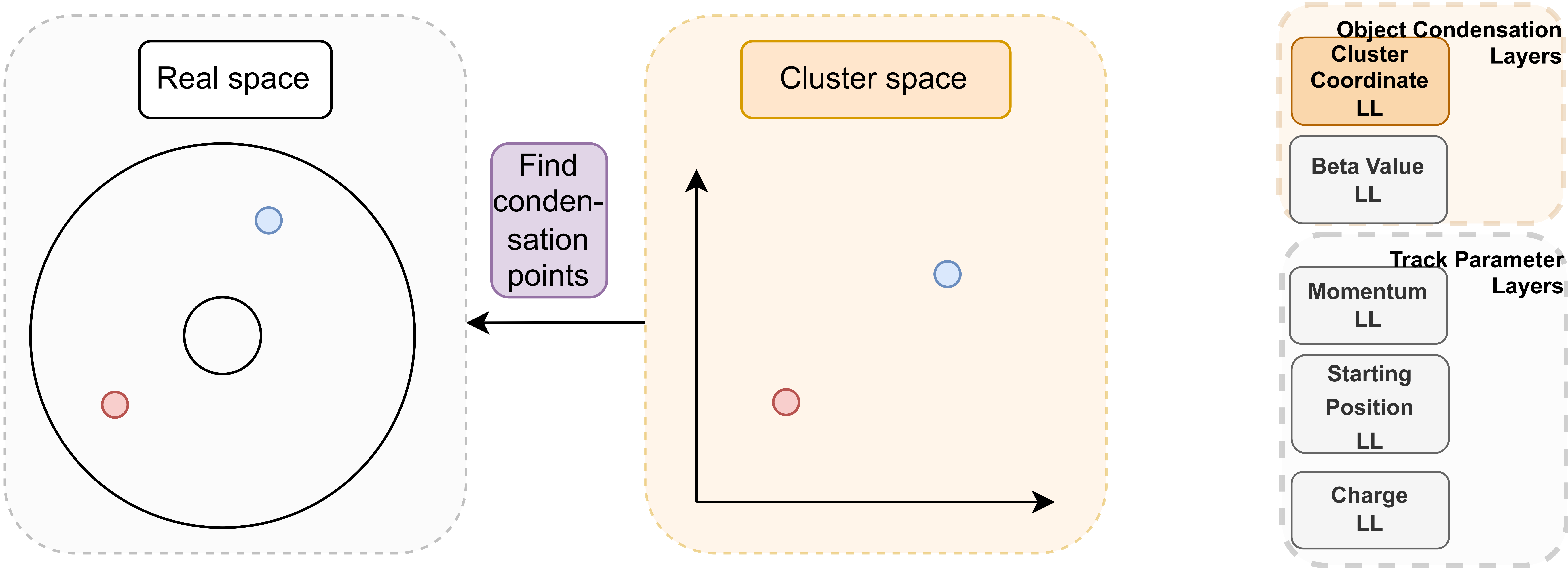
Inference - Translation from real space to cluster space



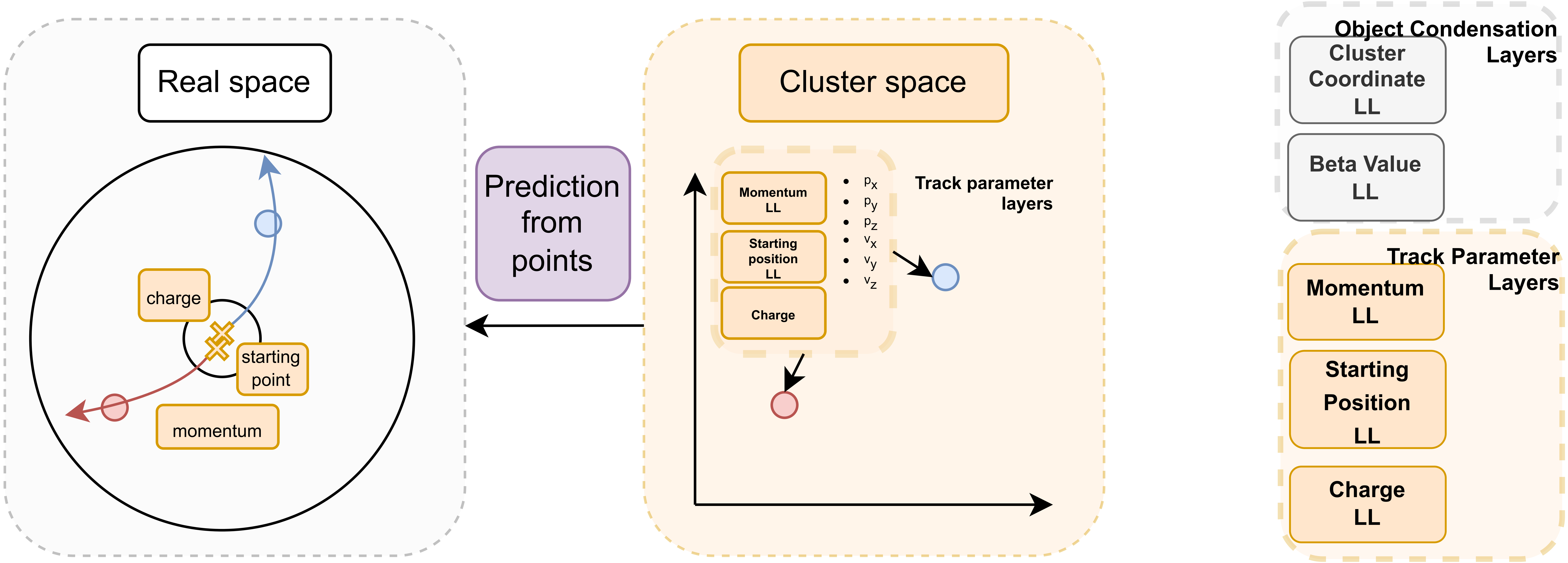
Inference - Condensation point candidate selection based on β threshold



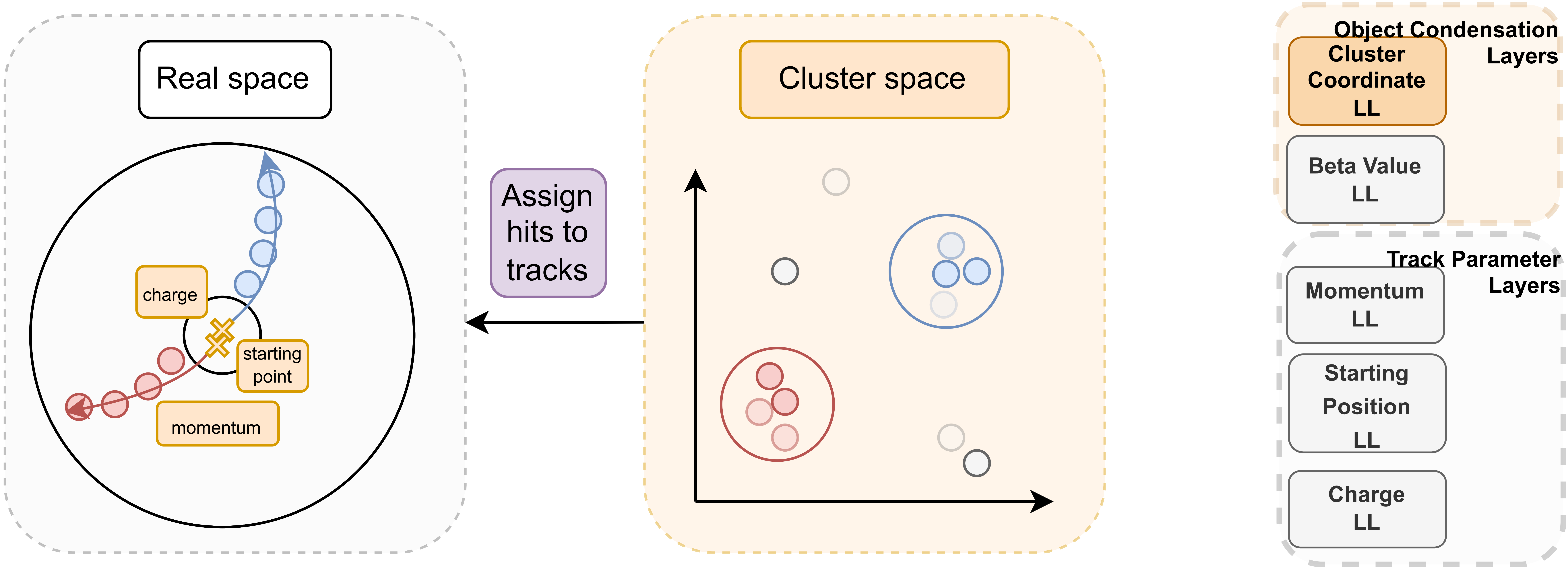
Inference - Condensation point selection based on isolation criteria



Inference - Parameter extraction from condensation points

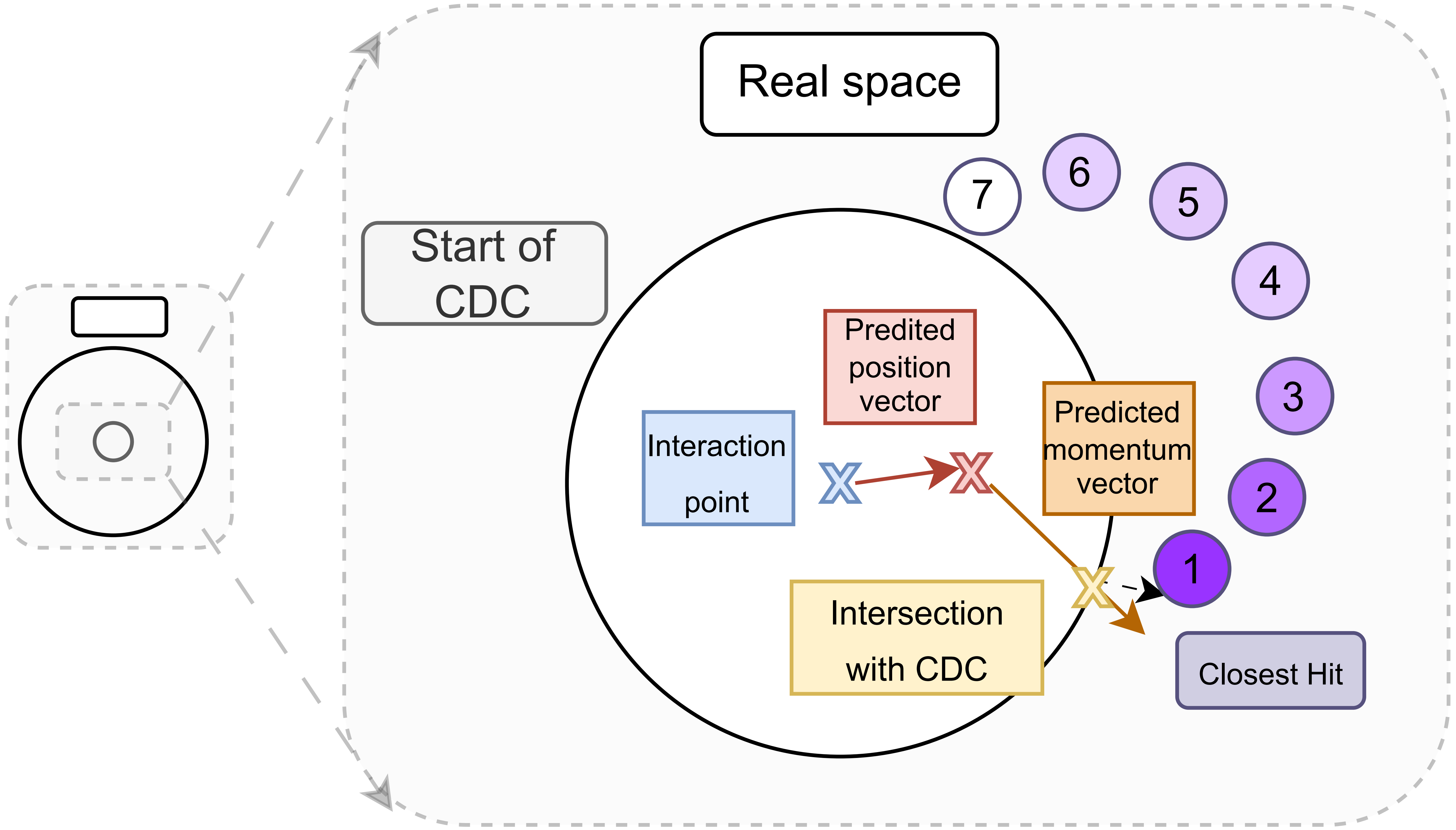


Inference - Clusterisation based on condensation points with hit assignment

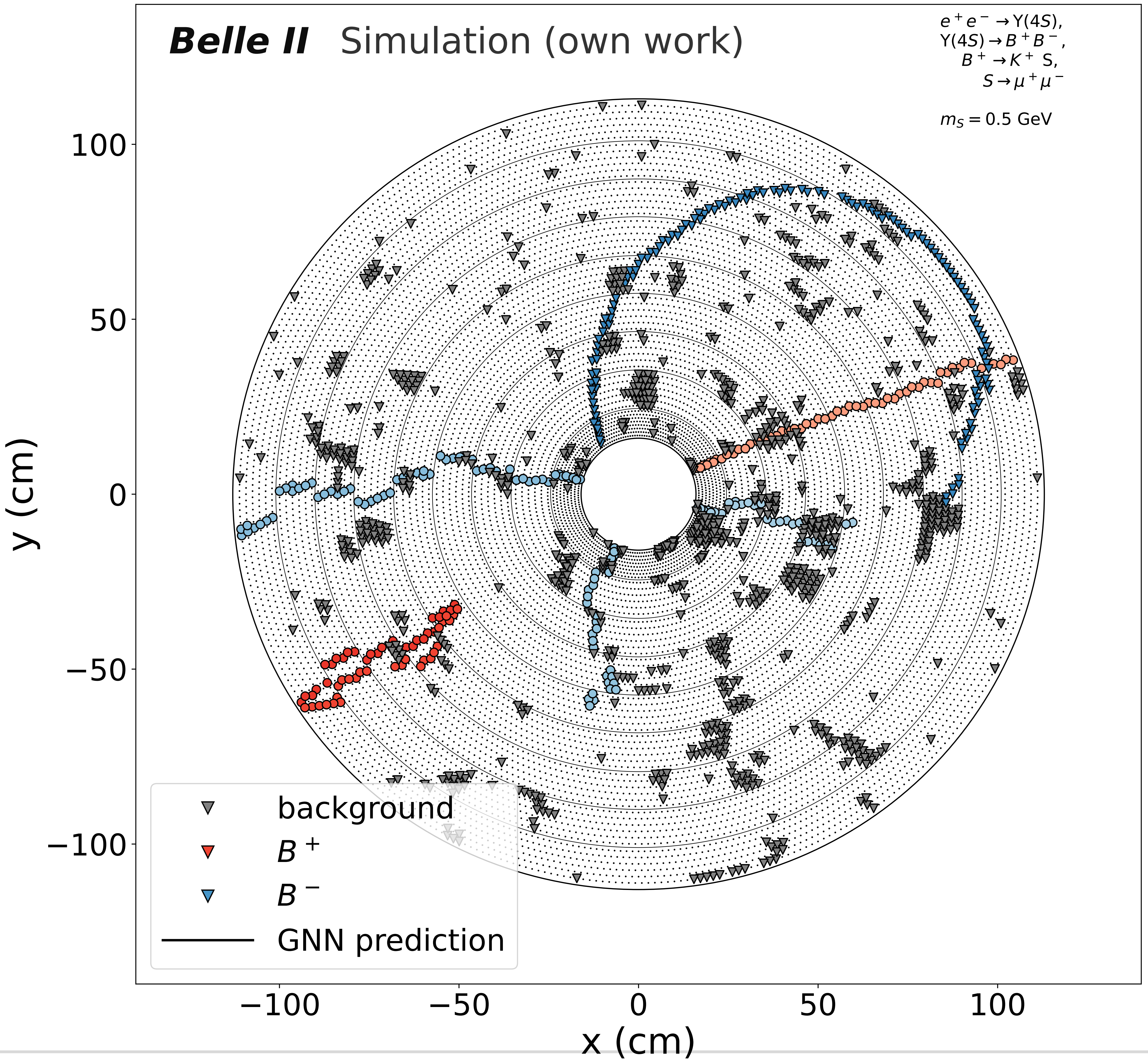


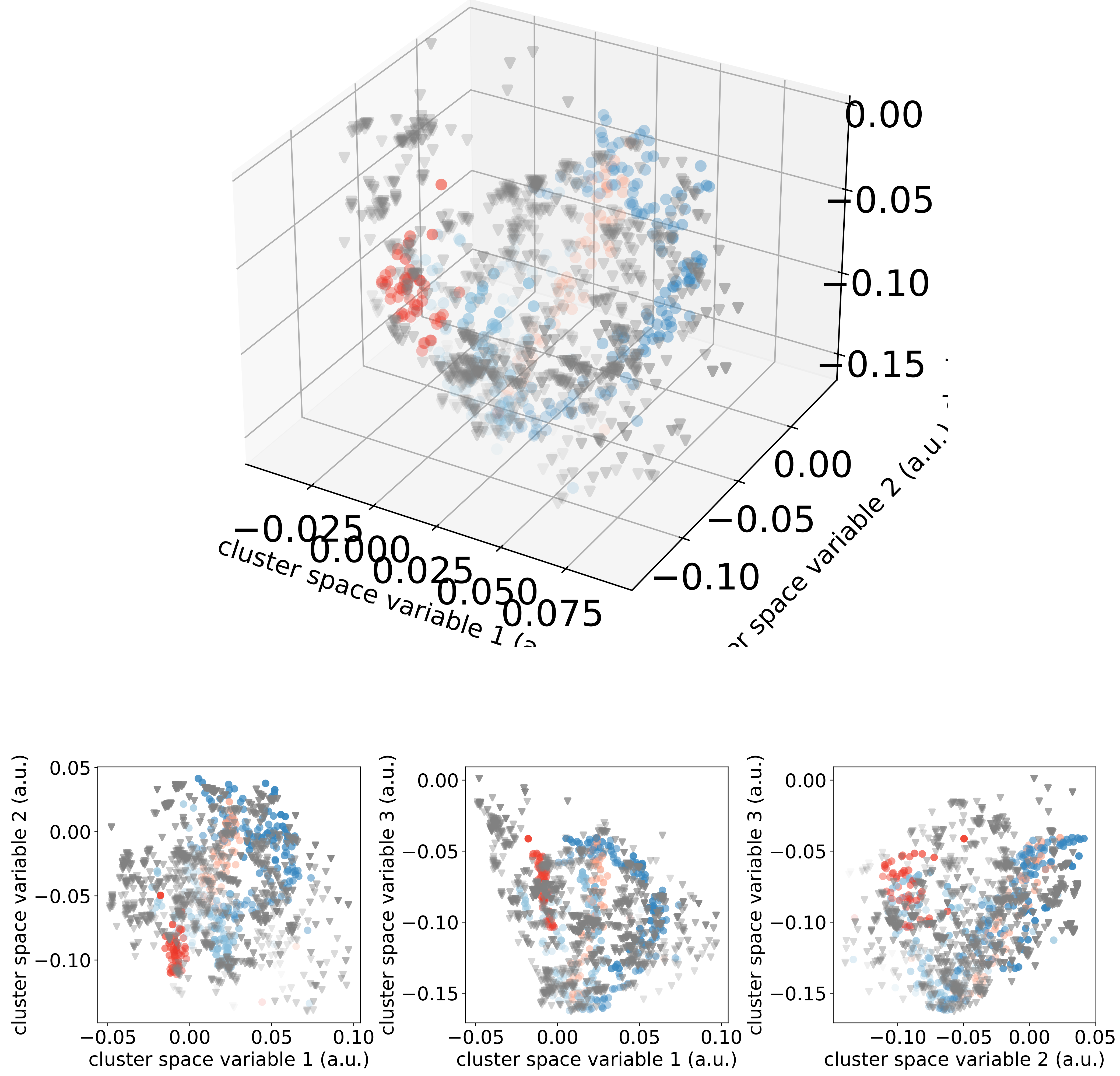
3 additional hyperparameters to tune

Inference - Hit ordering in real space

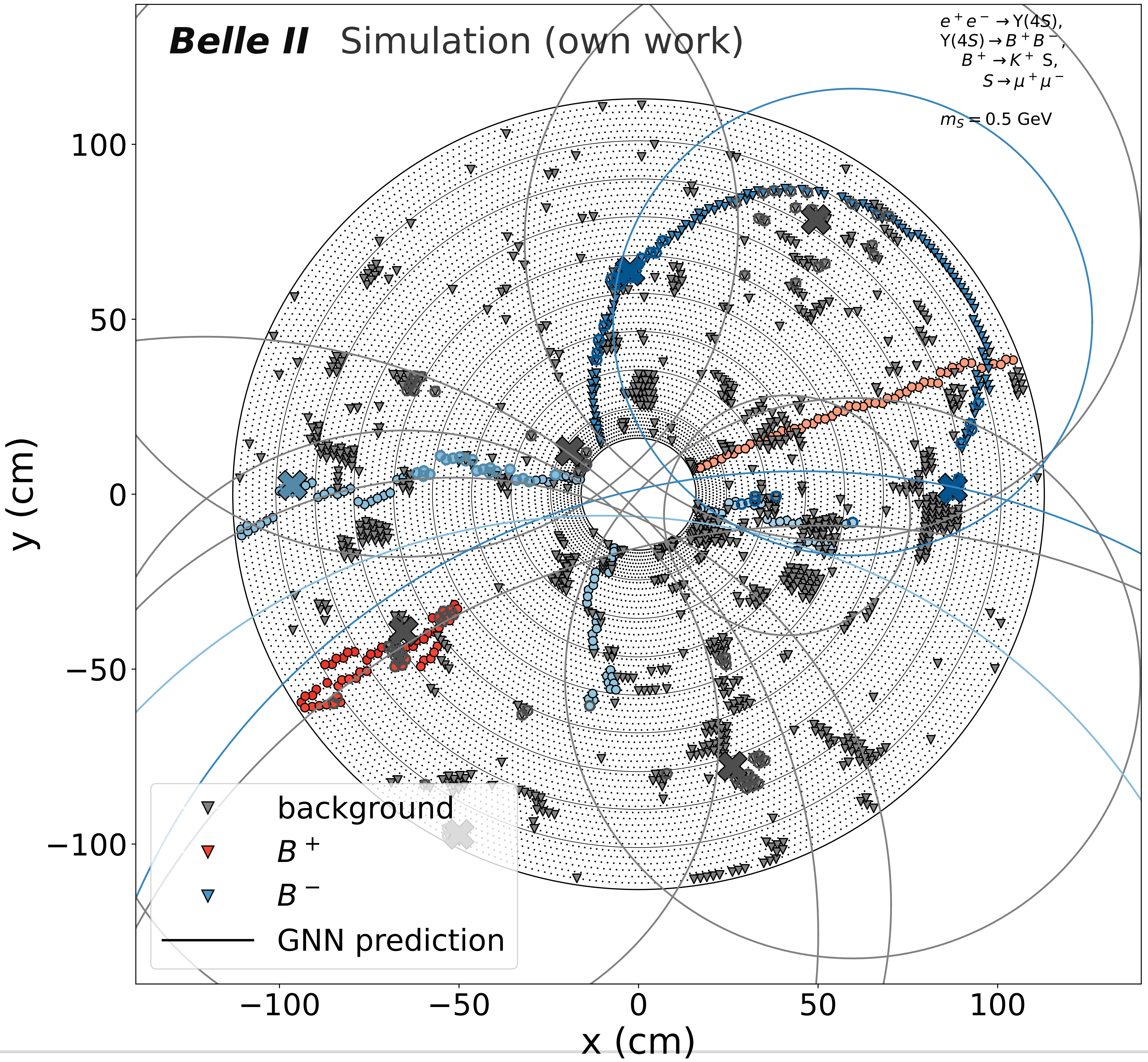


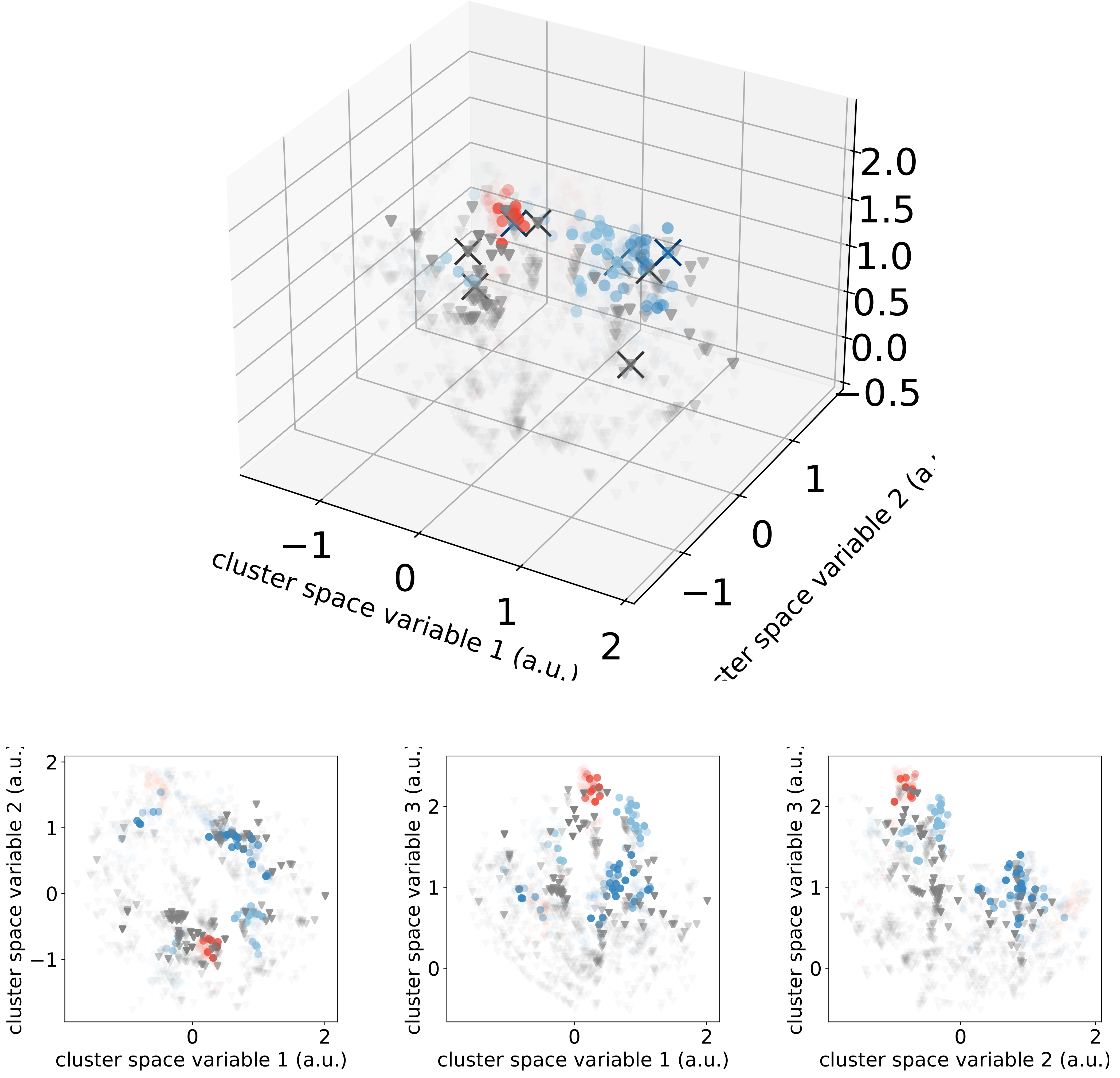
Model Learning



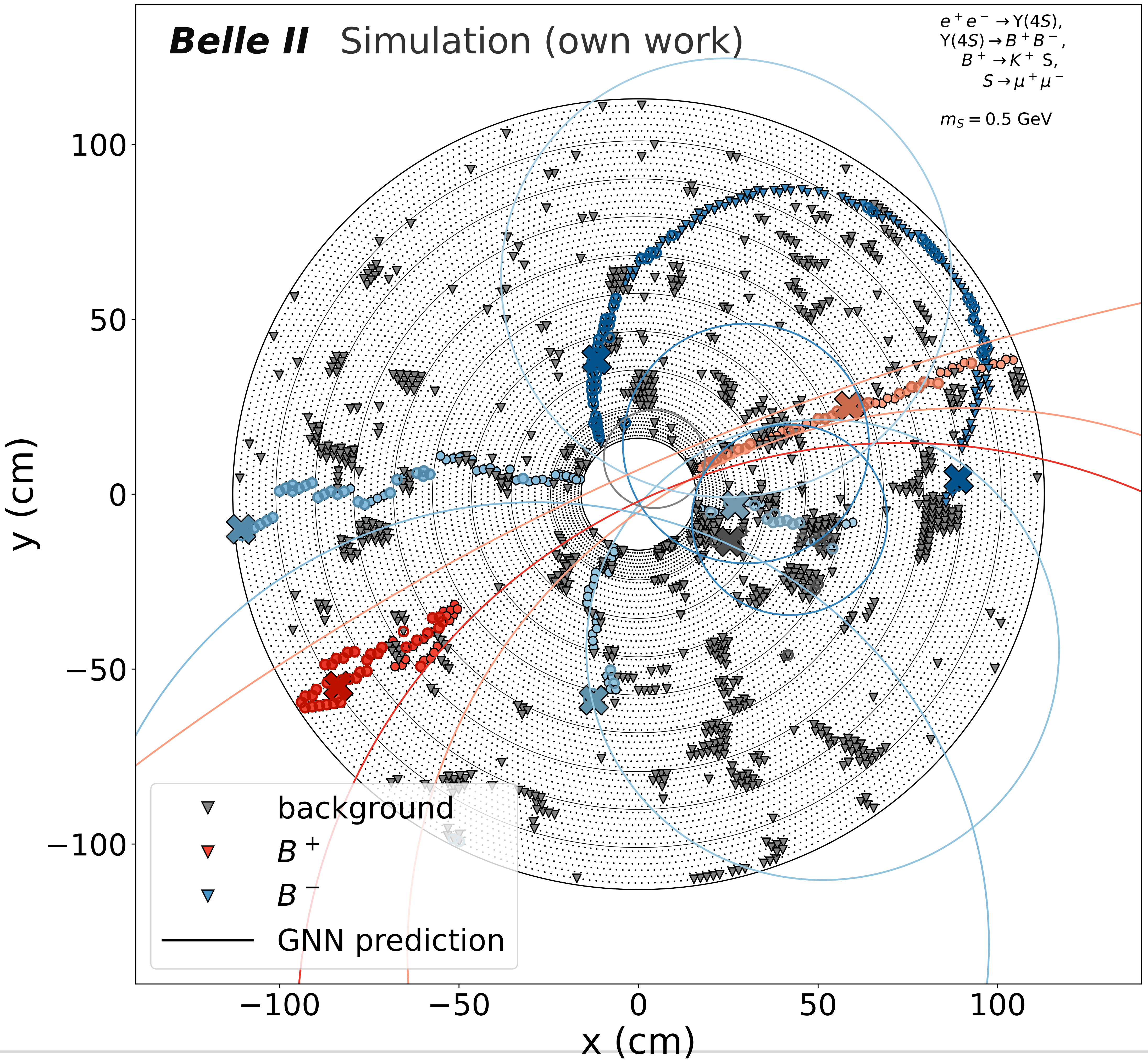


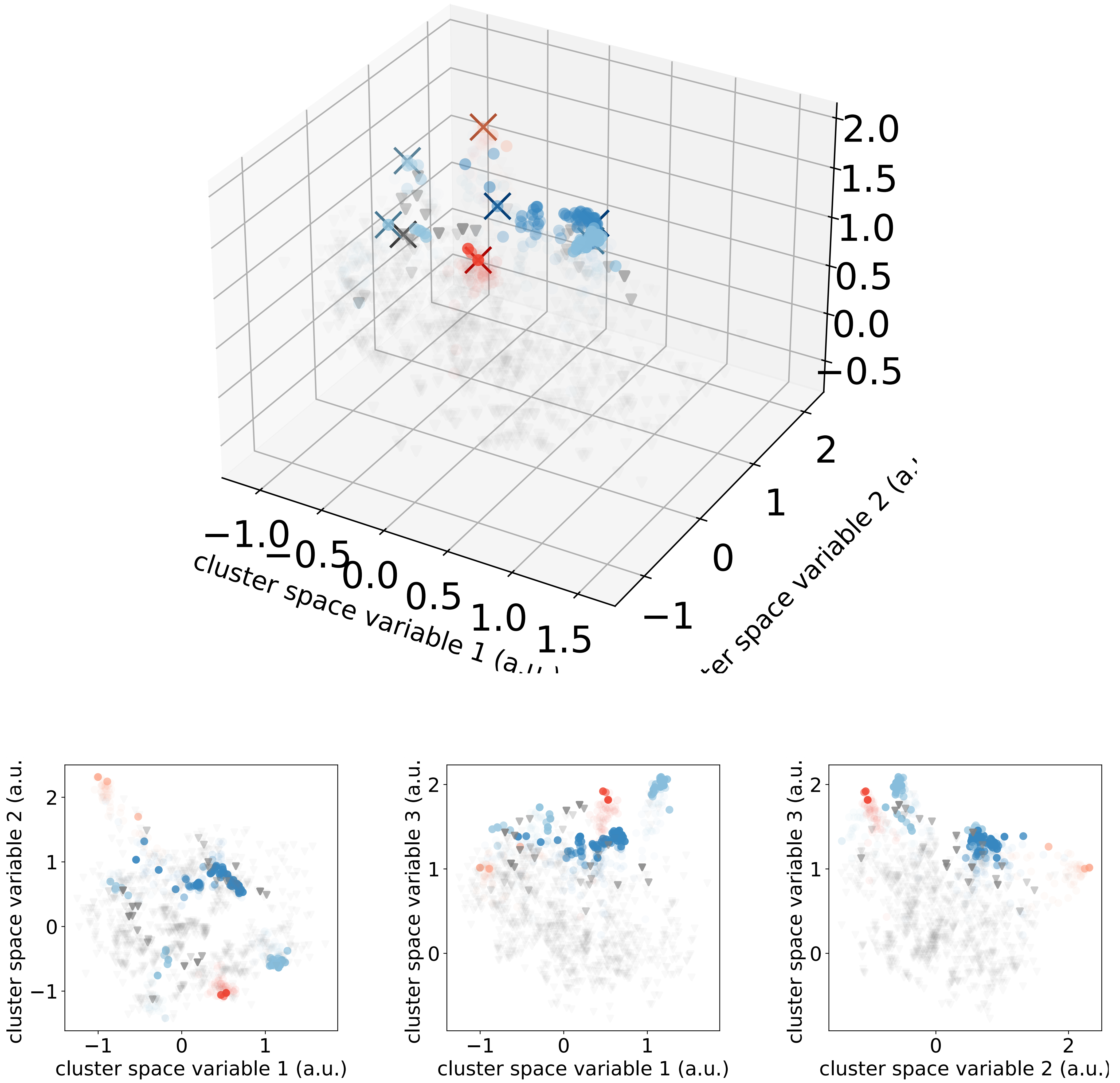
Model Learning



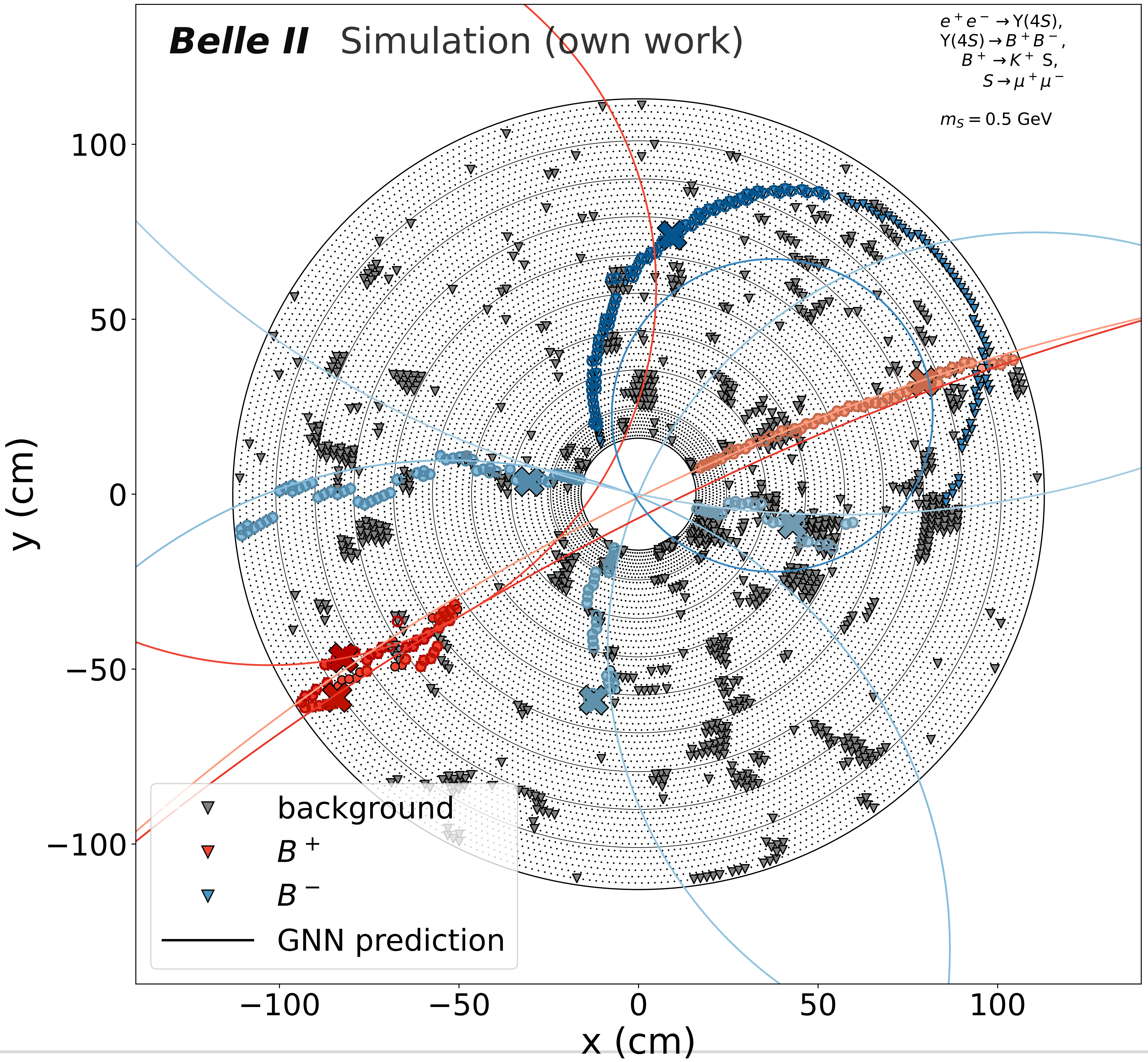


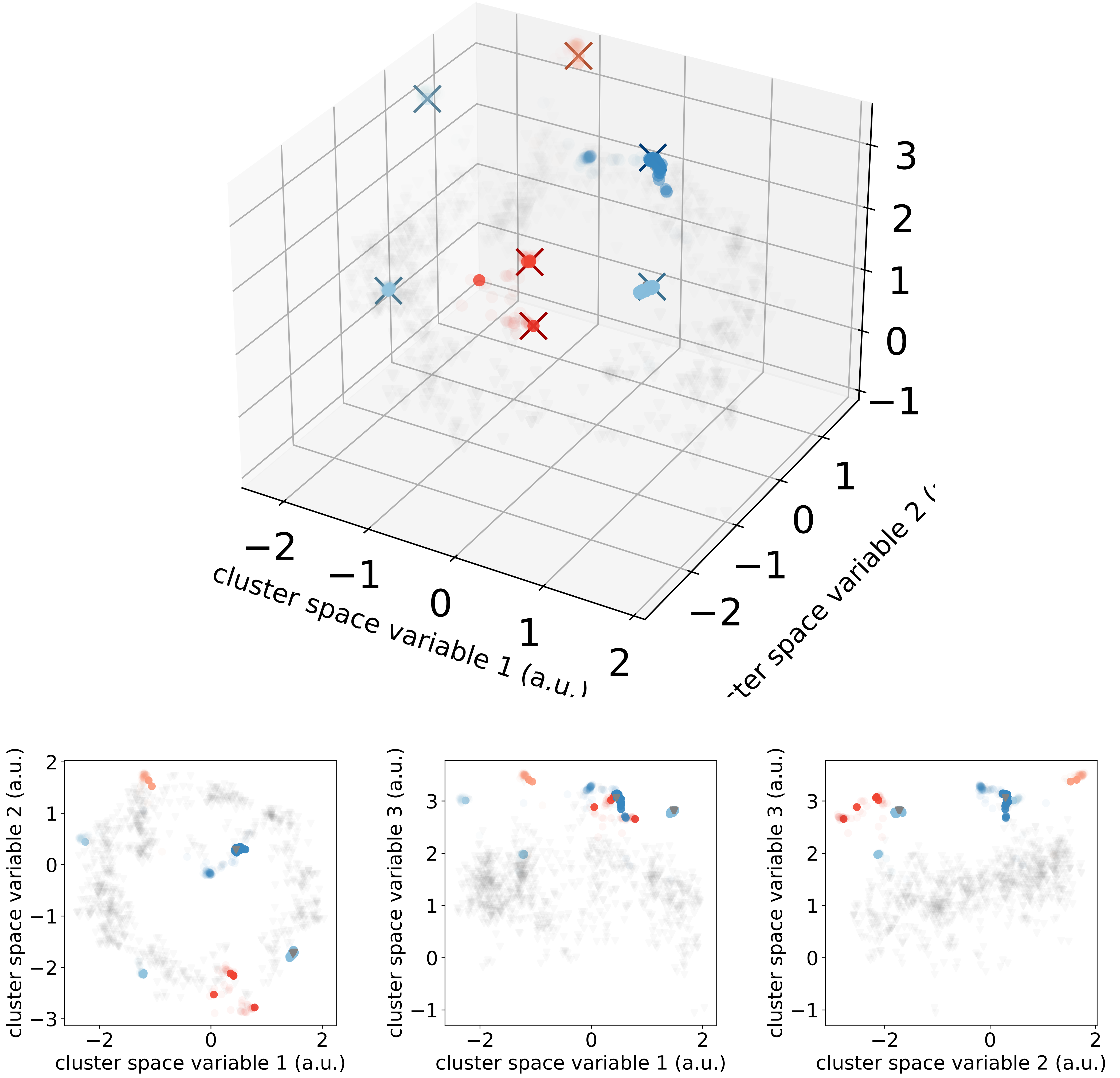
Model Learning



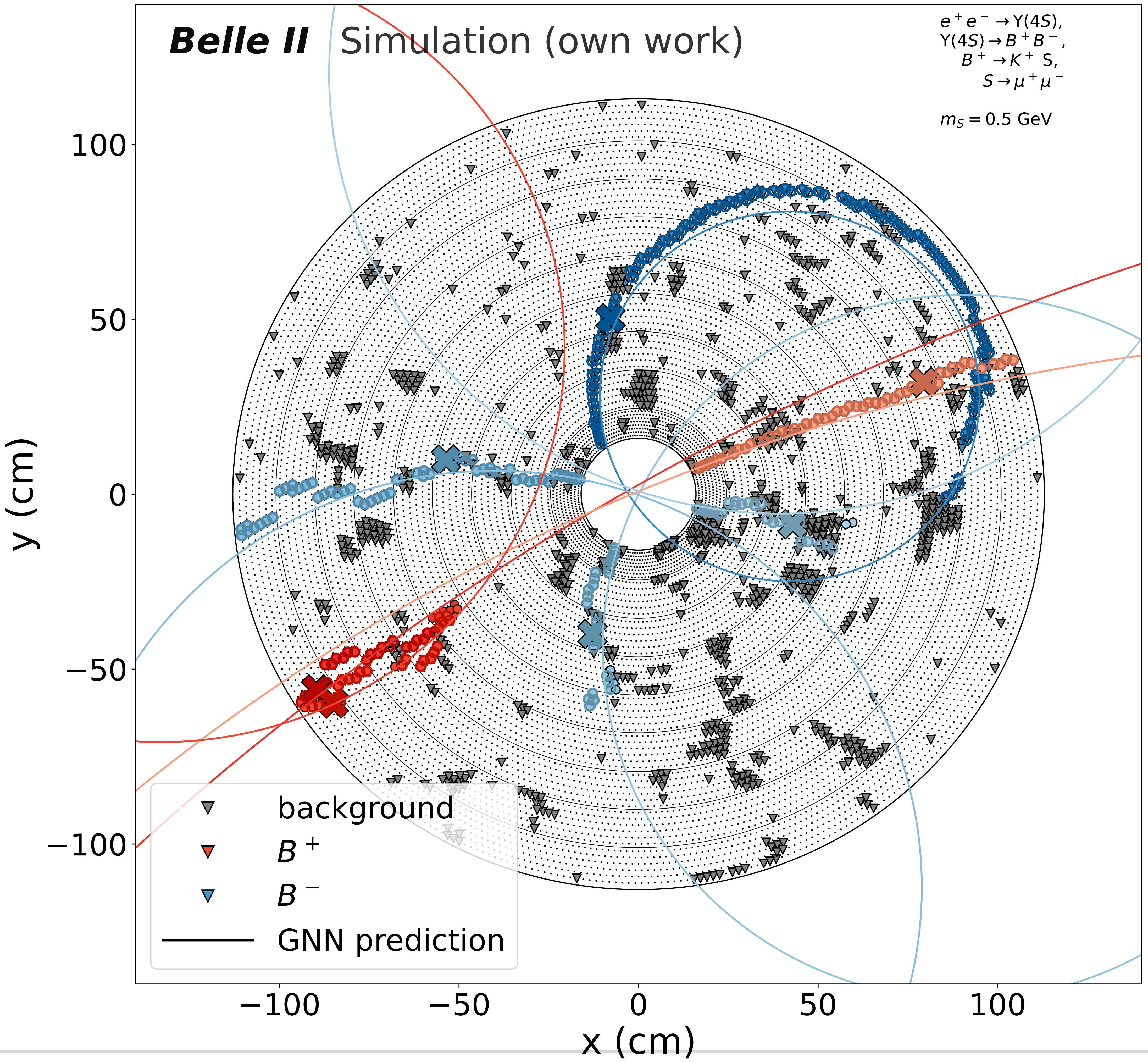


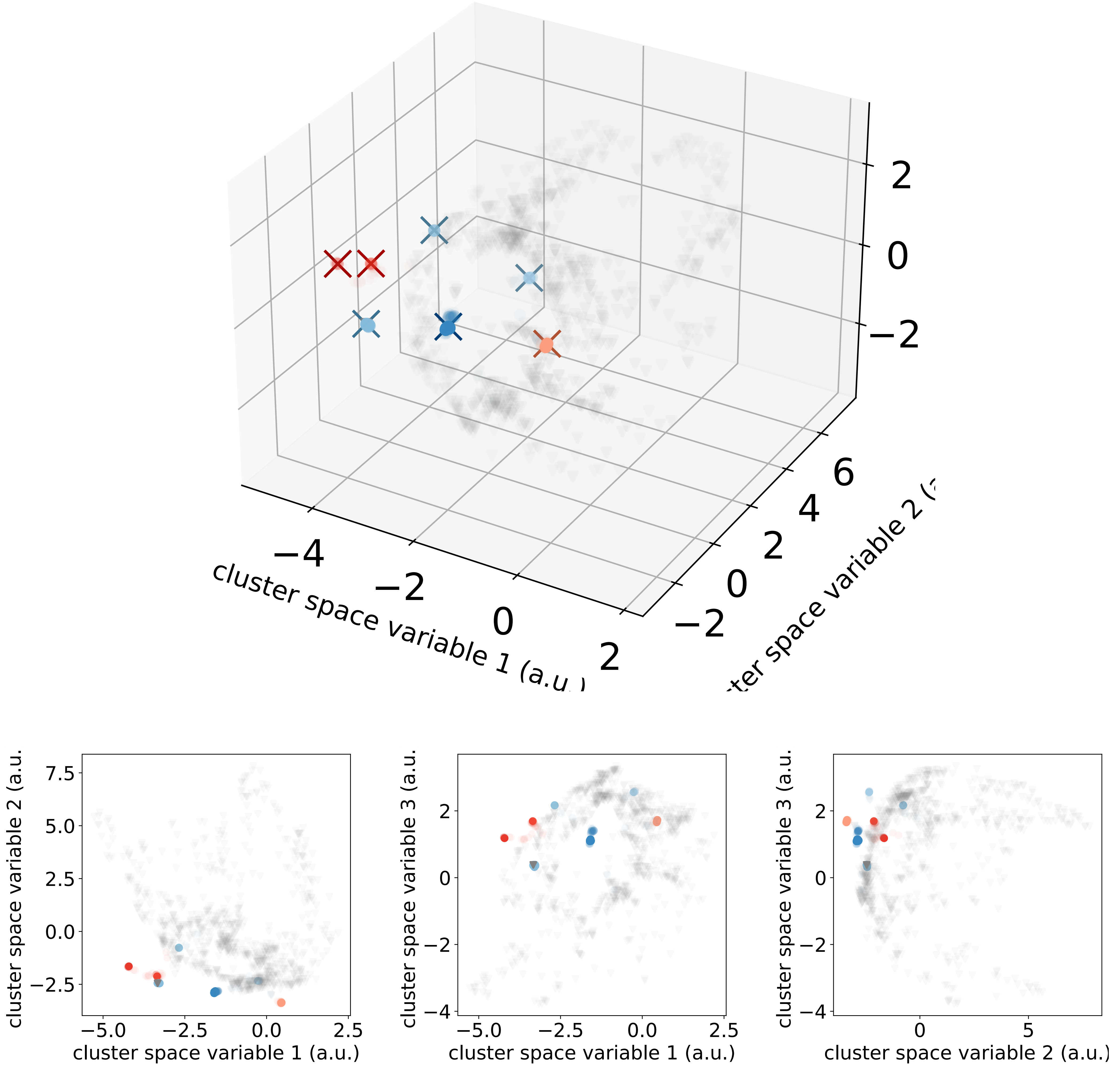
Model Learning



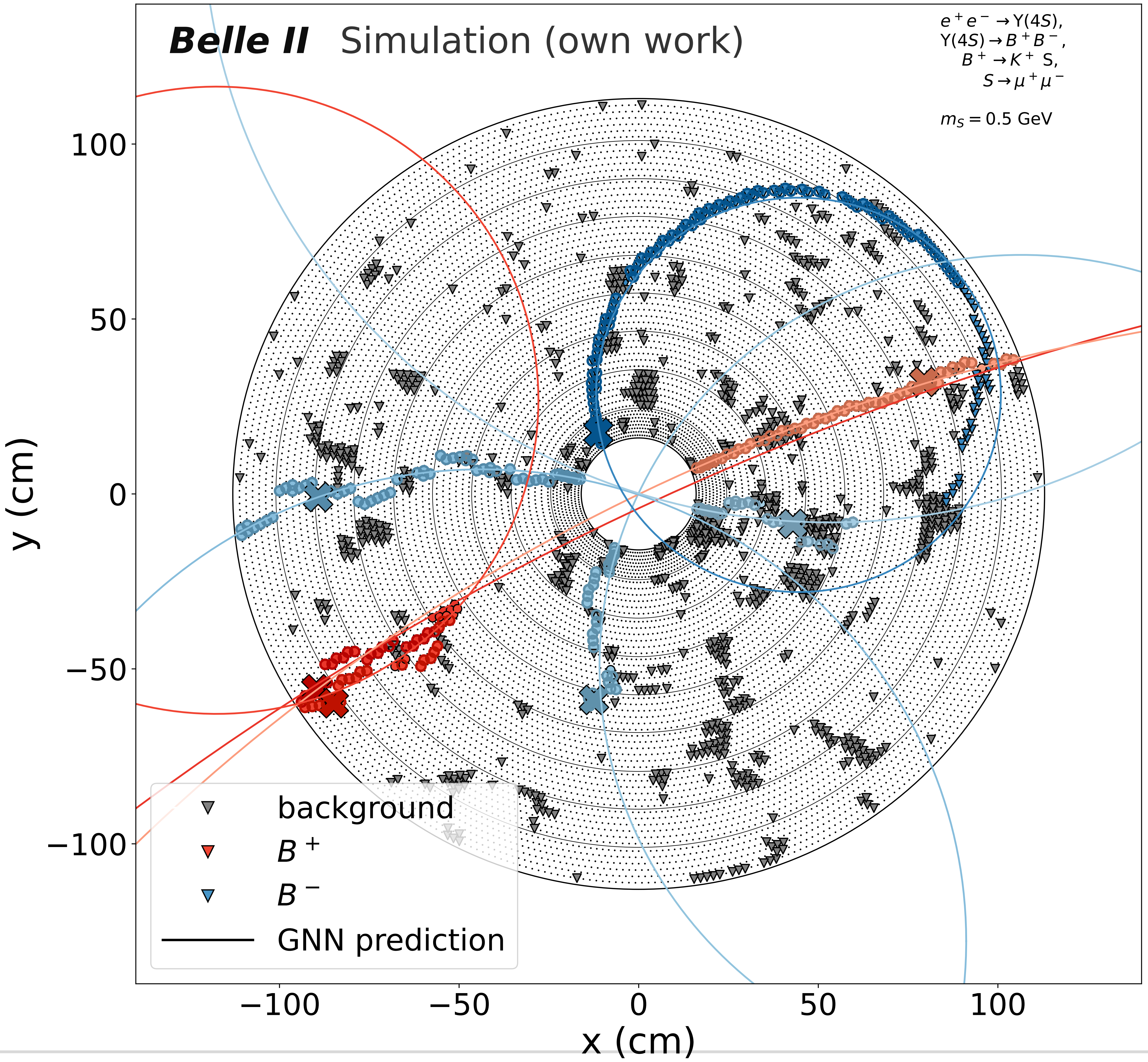


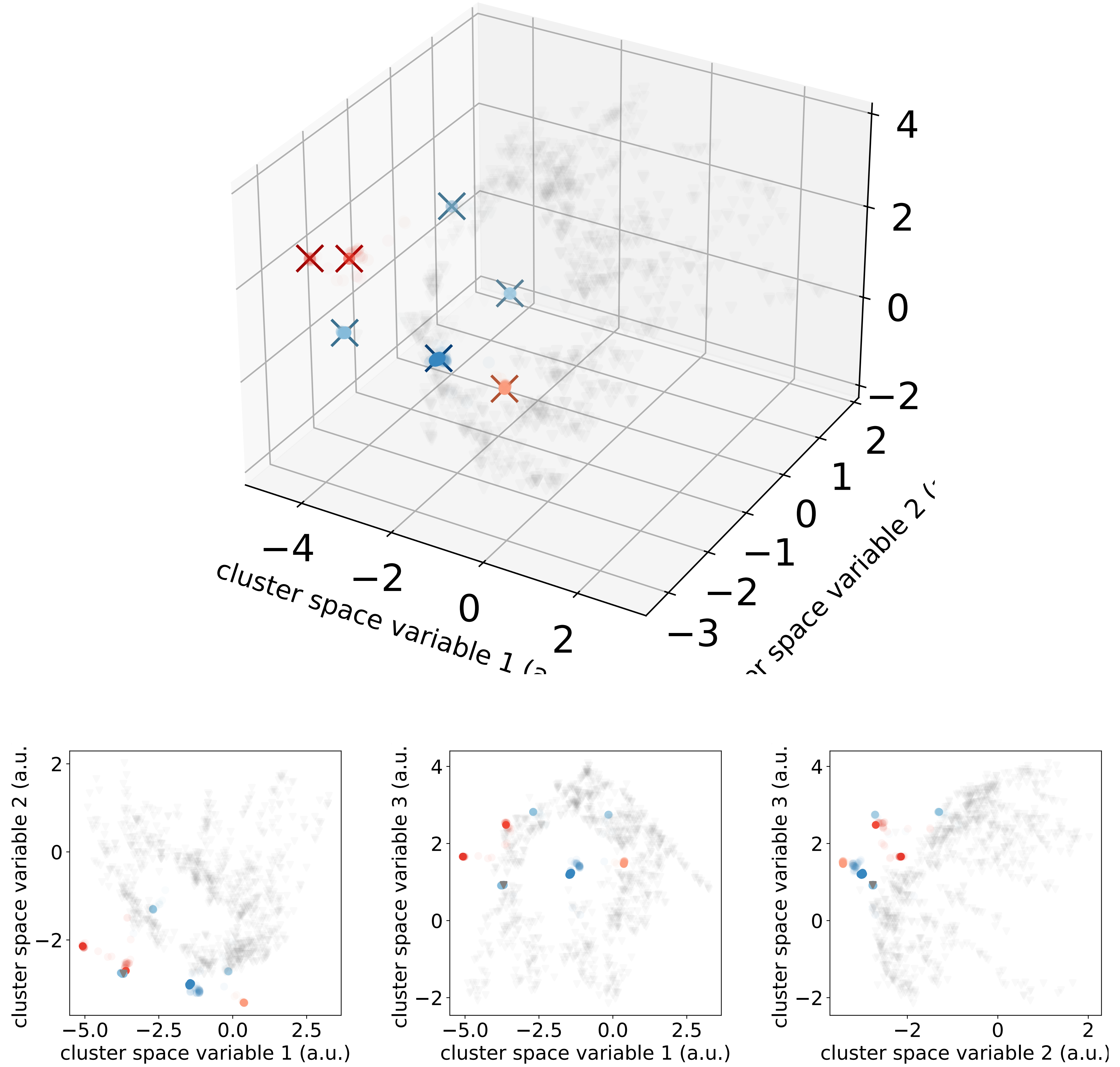
Model Learning





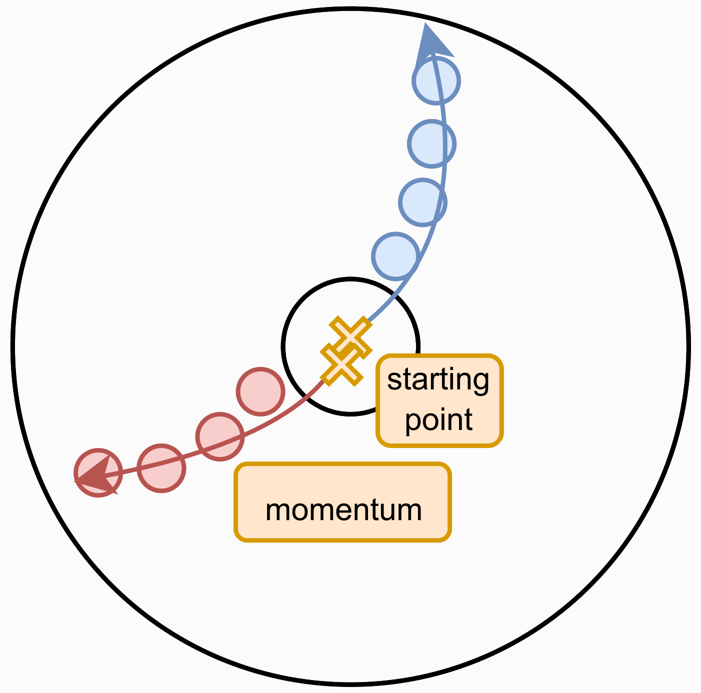
Model Prediction





Implementation in our analysis software for a comparison to Baseline Tracking

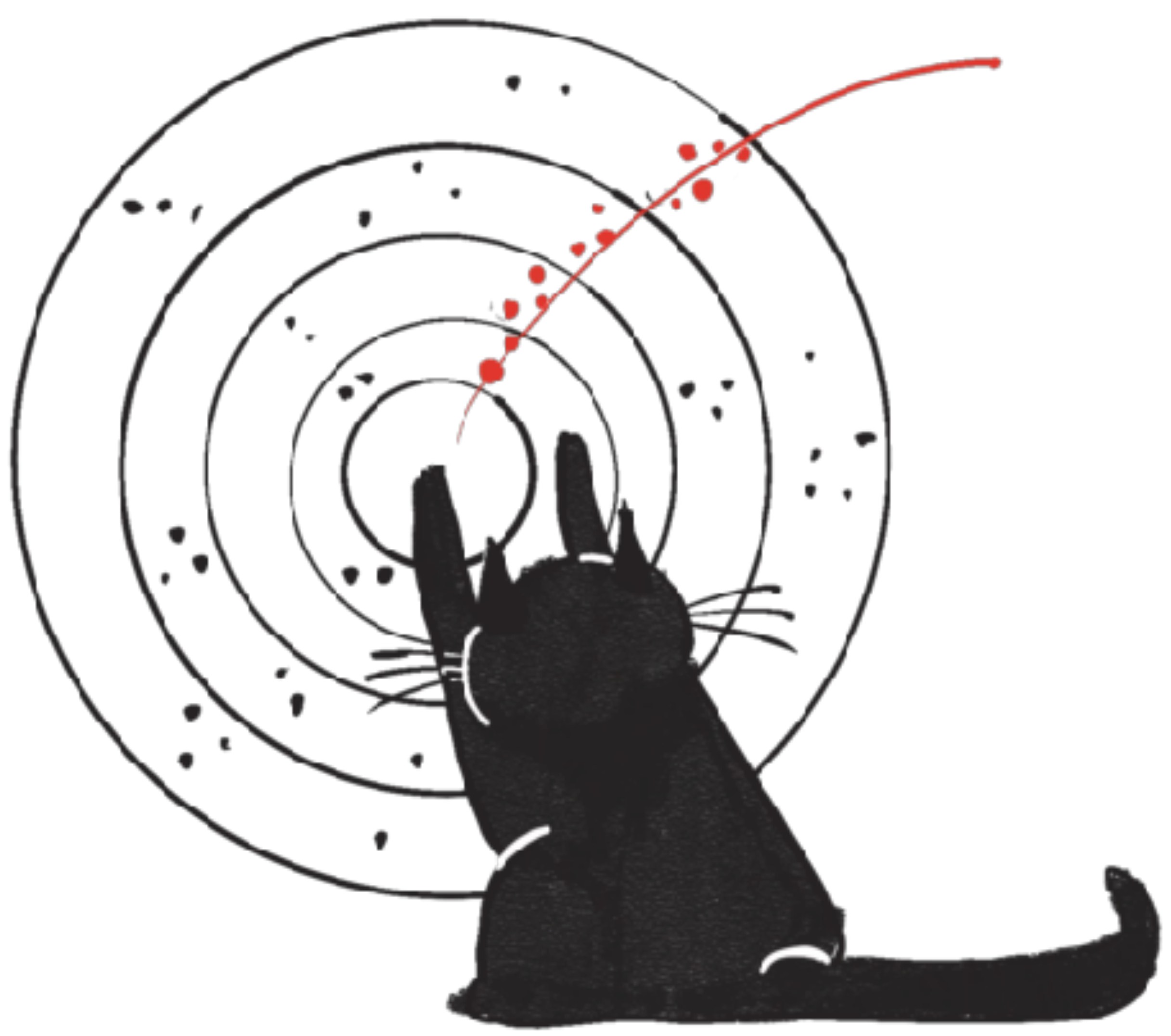
CAT Finder



Belle II Analysis Framework basf2

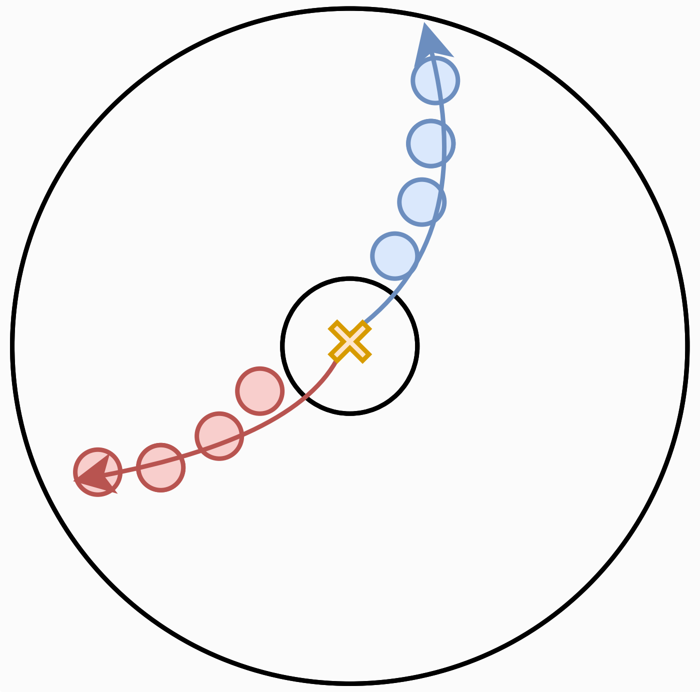
Additional Track Fitting

Analysis Object



Fair comparison for track finding algorithms on analysis level

Current basf2 Finder

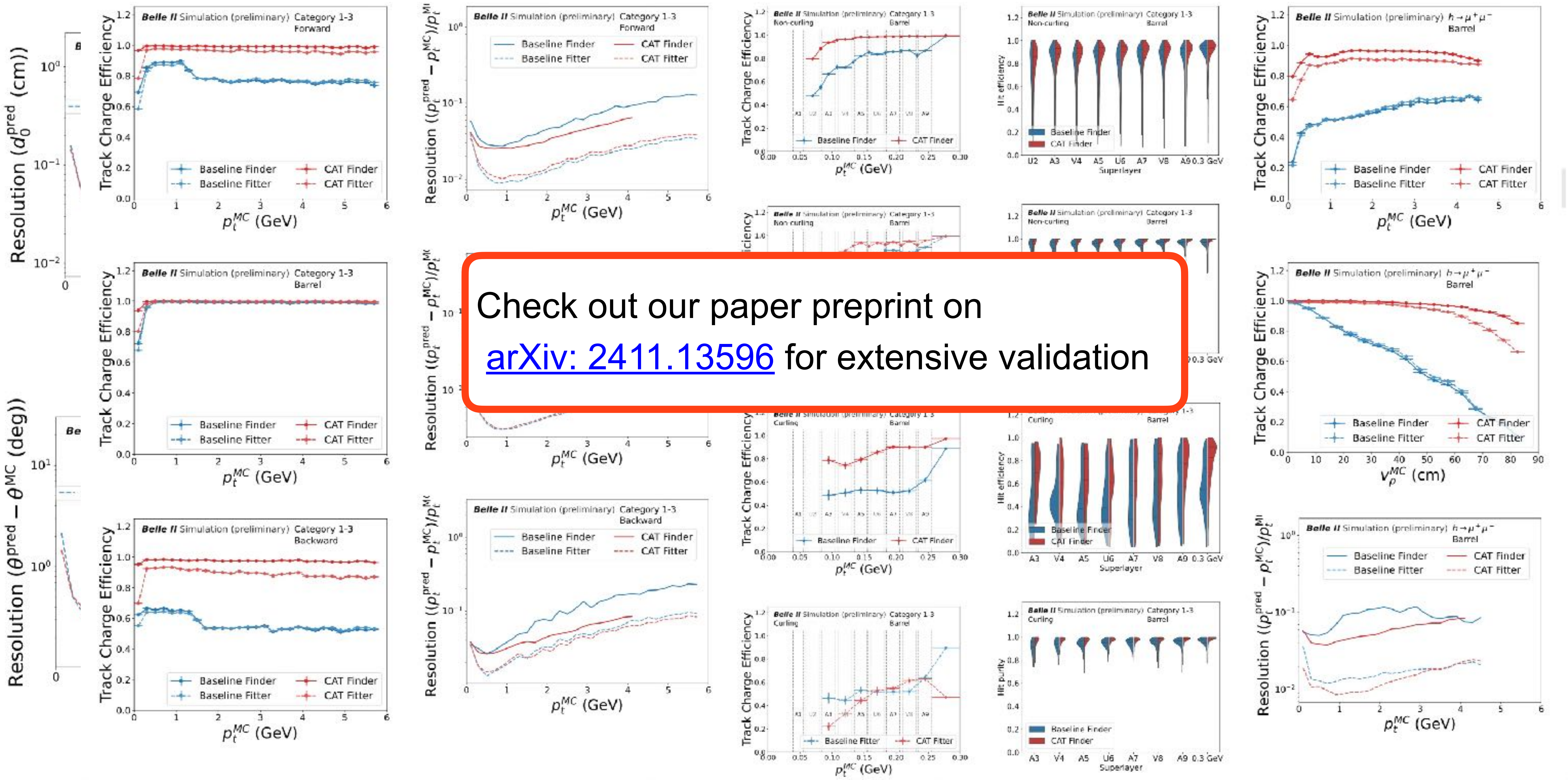


Belle II Analysis Framework basf2

Additional Track Fitting

Analysis Object

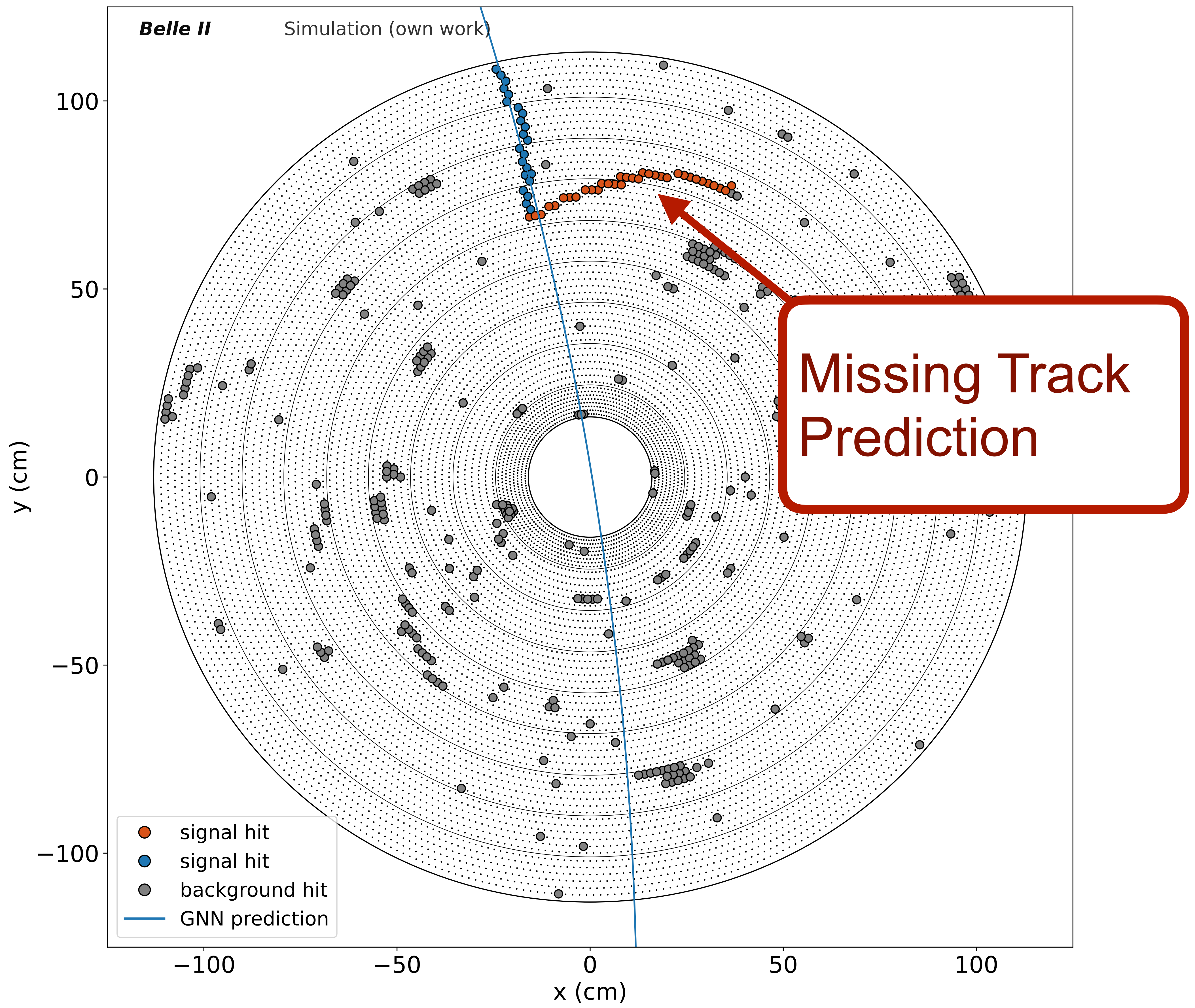
Extensive studies and validation



Check out our paper preprint on [arXiv: 2411.13596](https://arxiv.org/abs/2411.13596) for extensive validation

Important Metrics

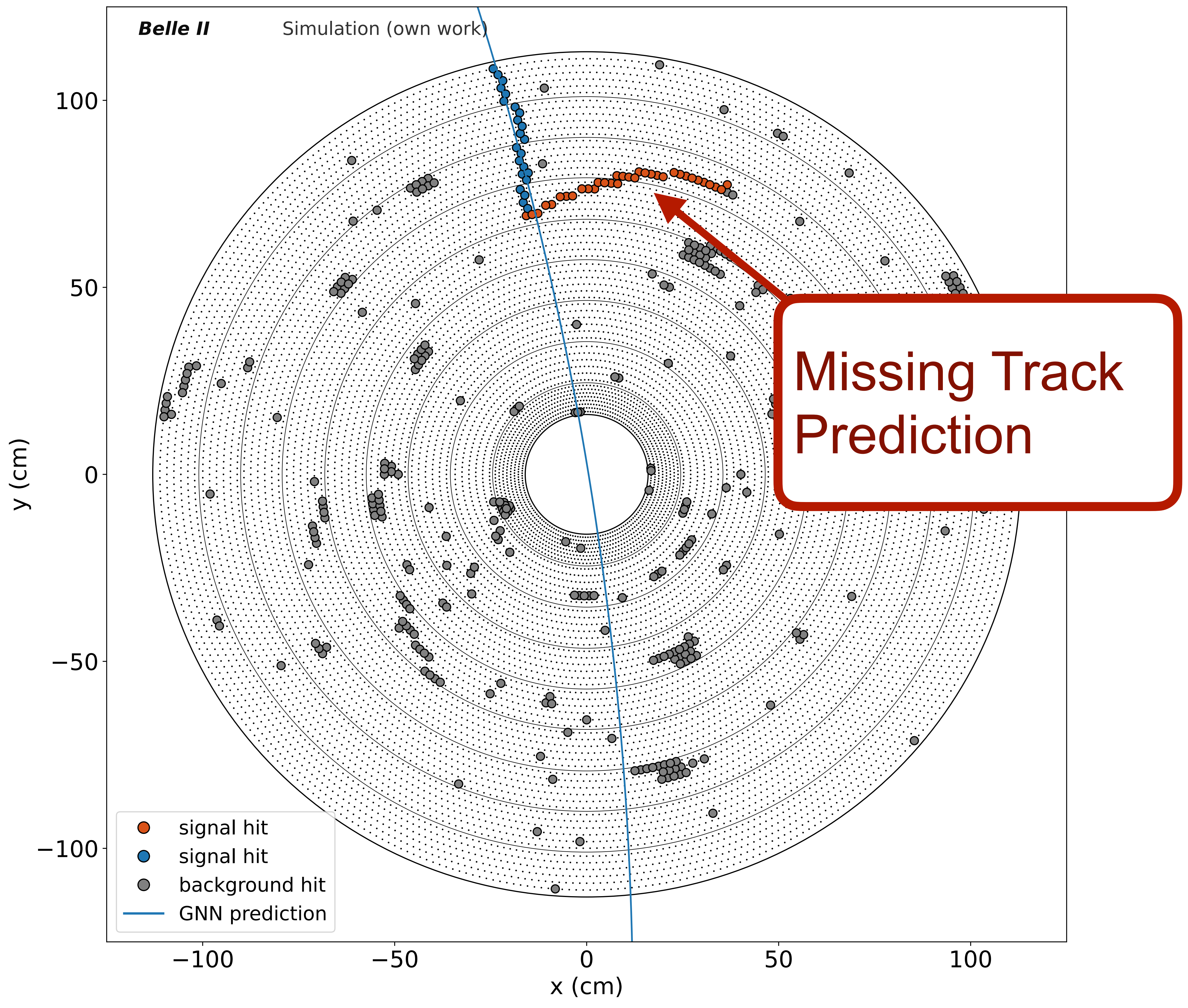
$$\text{Efficiency} = \frac{\text{true tracks}}{\text{all particles}}$$

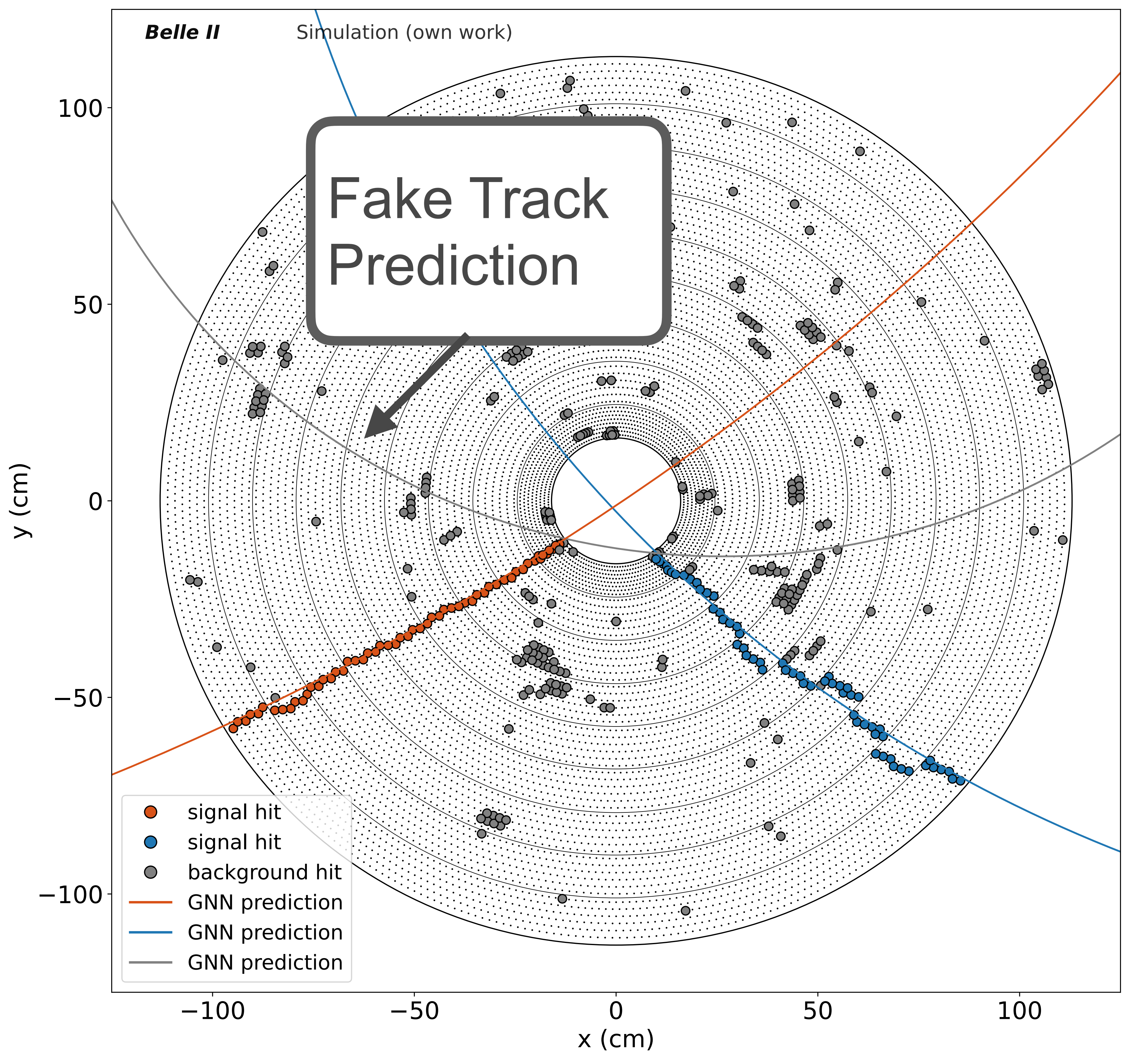


Important Metrics

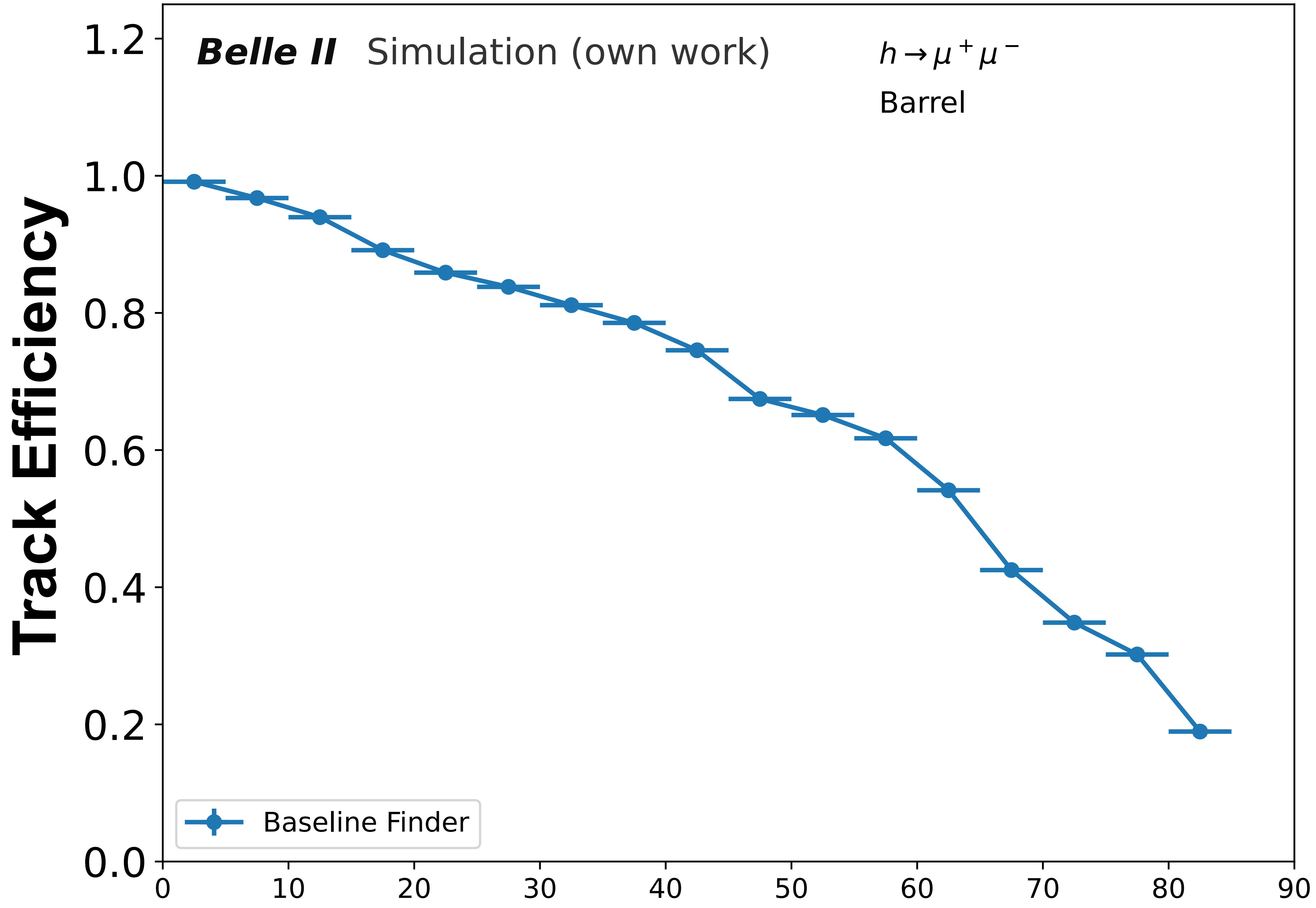
$$\text{Efficiency} = \frac{\text{true tracks}}{\text{all particles}}$$

$$\text{Purity} = \frac{\text{true tracks}}{\text{all tracks}}$$





Performance on displaced particles

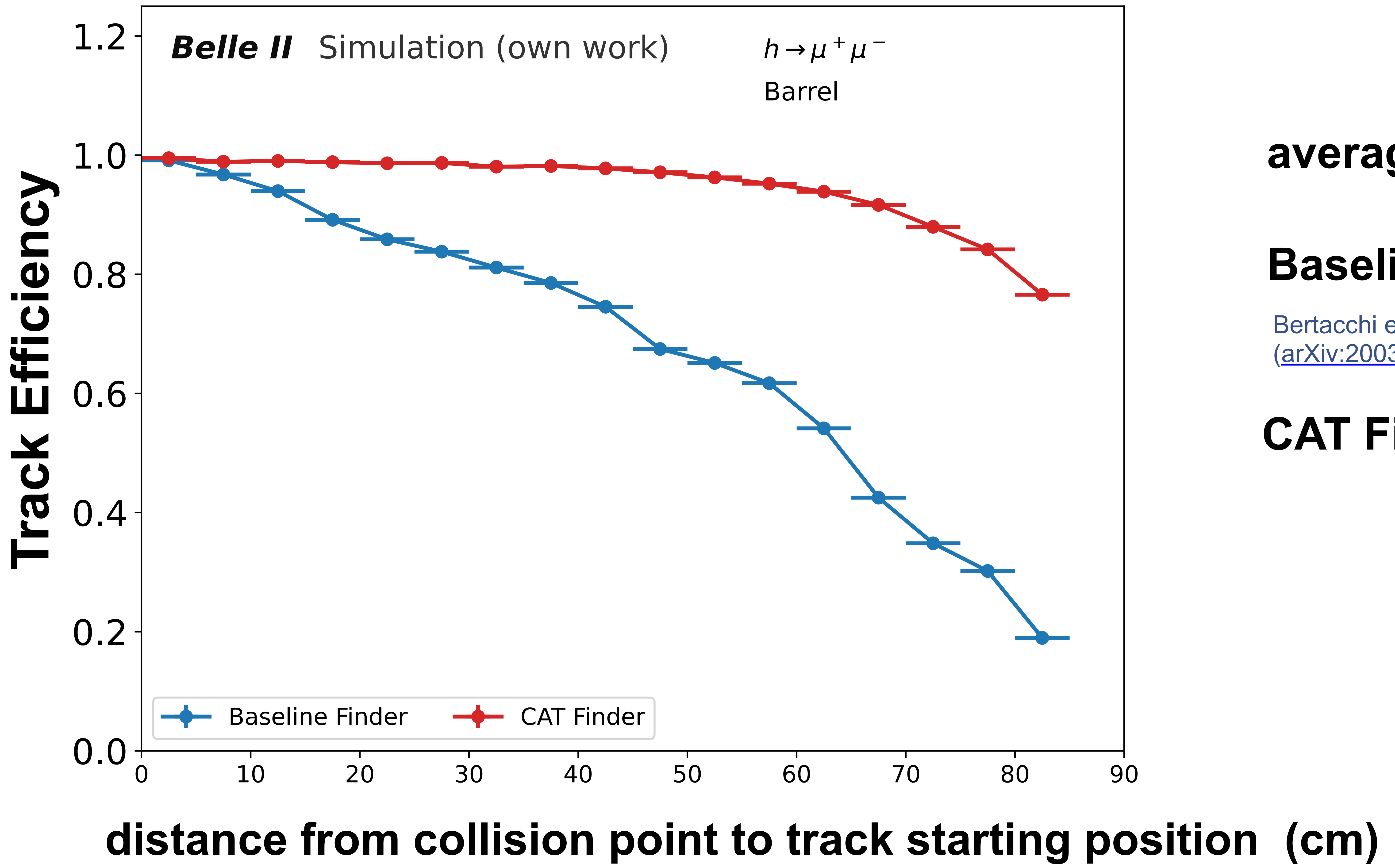


distance from collision point to track starting position (cm)

averaged	Efficiency	Purity
Baseline	$0.574^{+/-0.001}$	$0.964^{+/-0.001}$

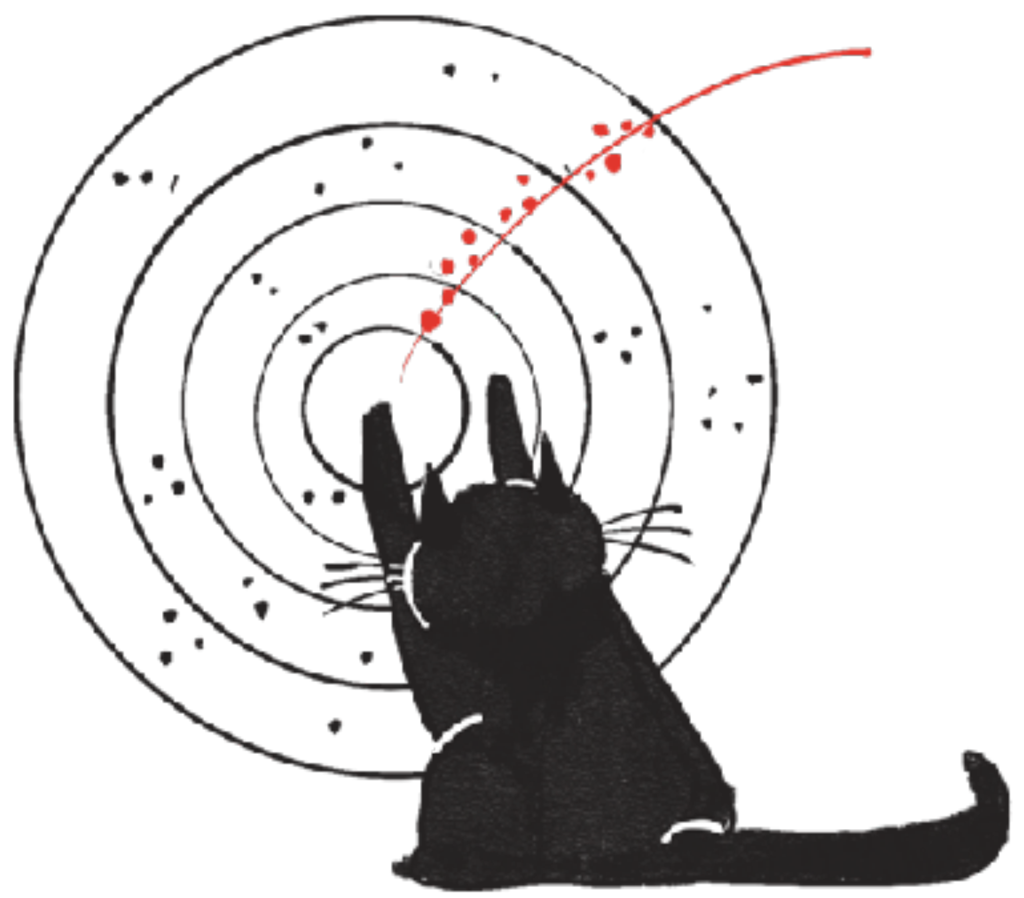
Bertacchi et al., Track Finding at Belle II (arXiv:2003.12466)

Performance on displaced particles

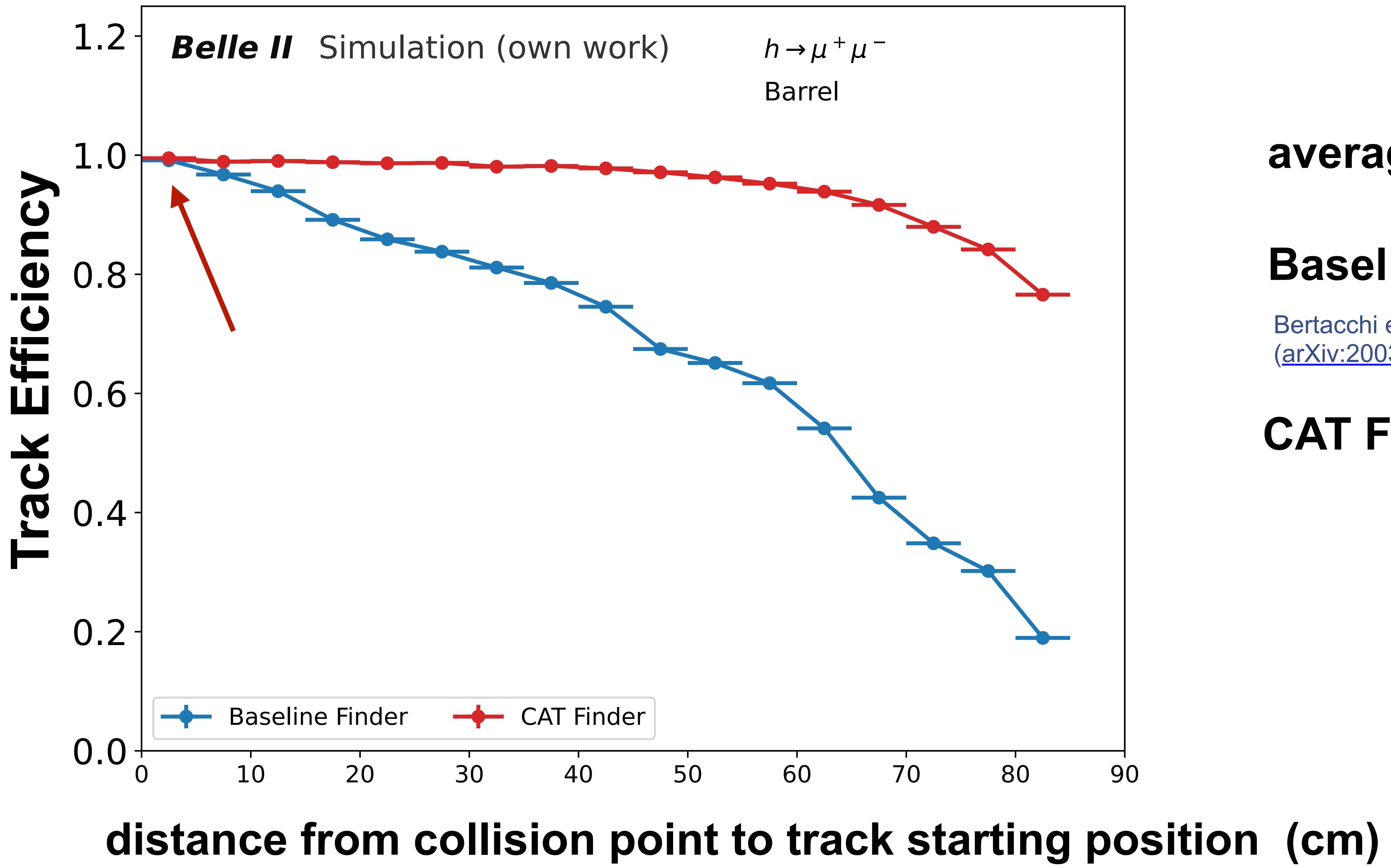


averaged	Efficiency	Purity
Baseline	0.574$^{+/-0.001}$	0.964$^{+/-0.001}$
CAT Finder	0.892$^{+/-0.001}$	0.978$^{+/-0.001}$

Bertacchi et al., Track Finding at Belle II (arXiv:2003.12466)

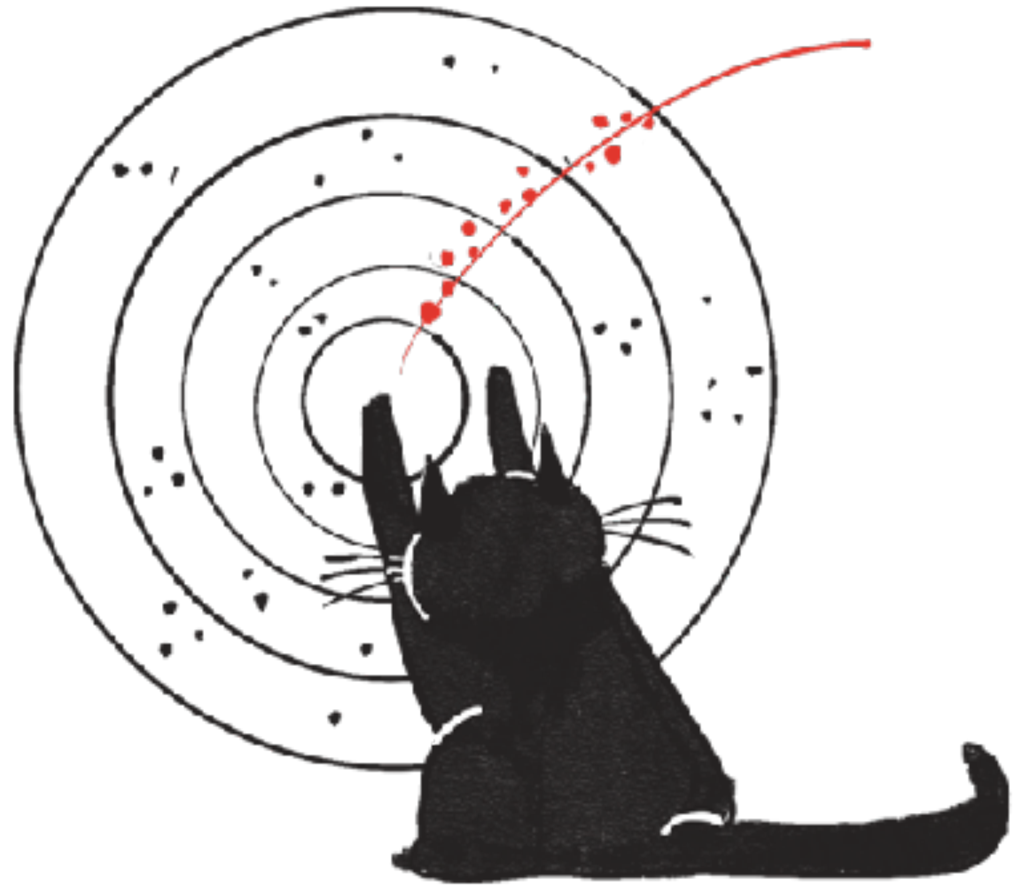


Performance on displaced particles

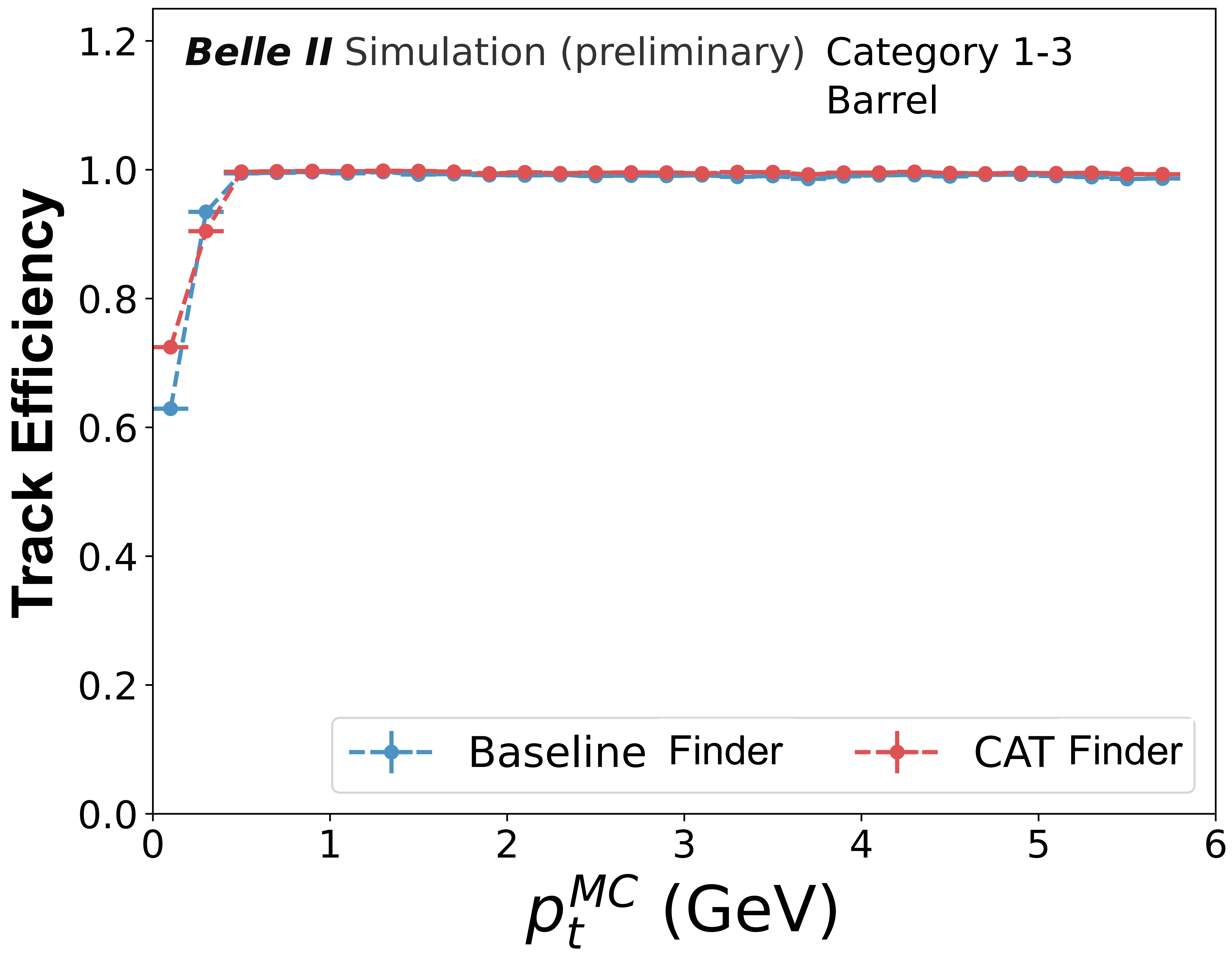


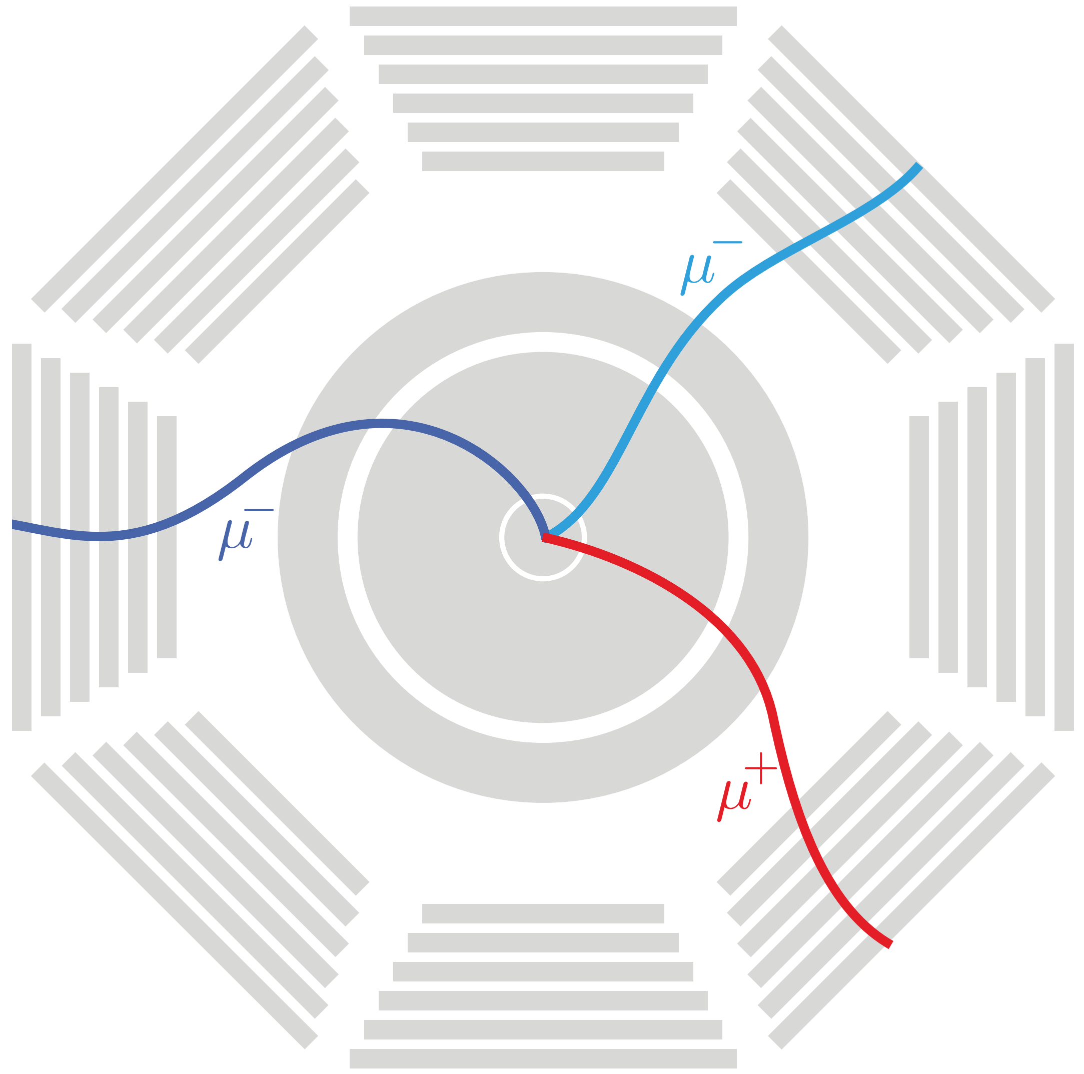
averaged	Efficiency	Purity
Baseline	0.574$^{+/-0.001}$	0.964$^{+/-0.001}$
CAT Finder	0.892$^{+/-0.001}$	0.978$^{+/-0.001}$

Bertacchi et al., Track Finding at Belle II (arXiv:2003.12466)



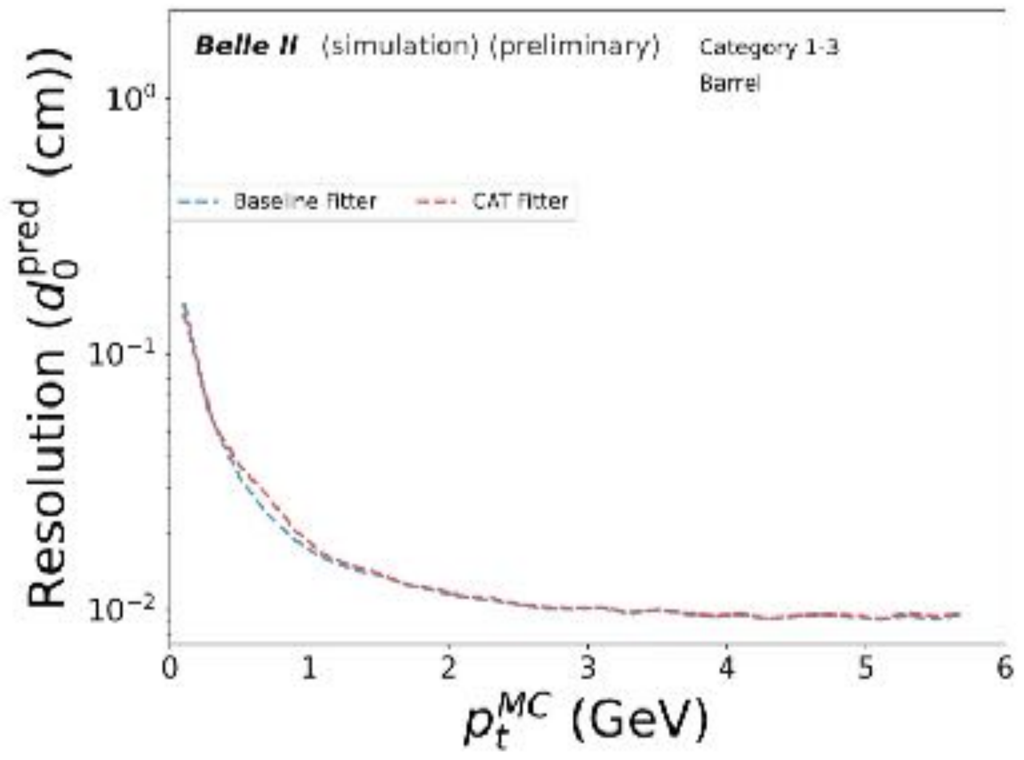
Performance on prompt particles



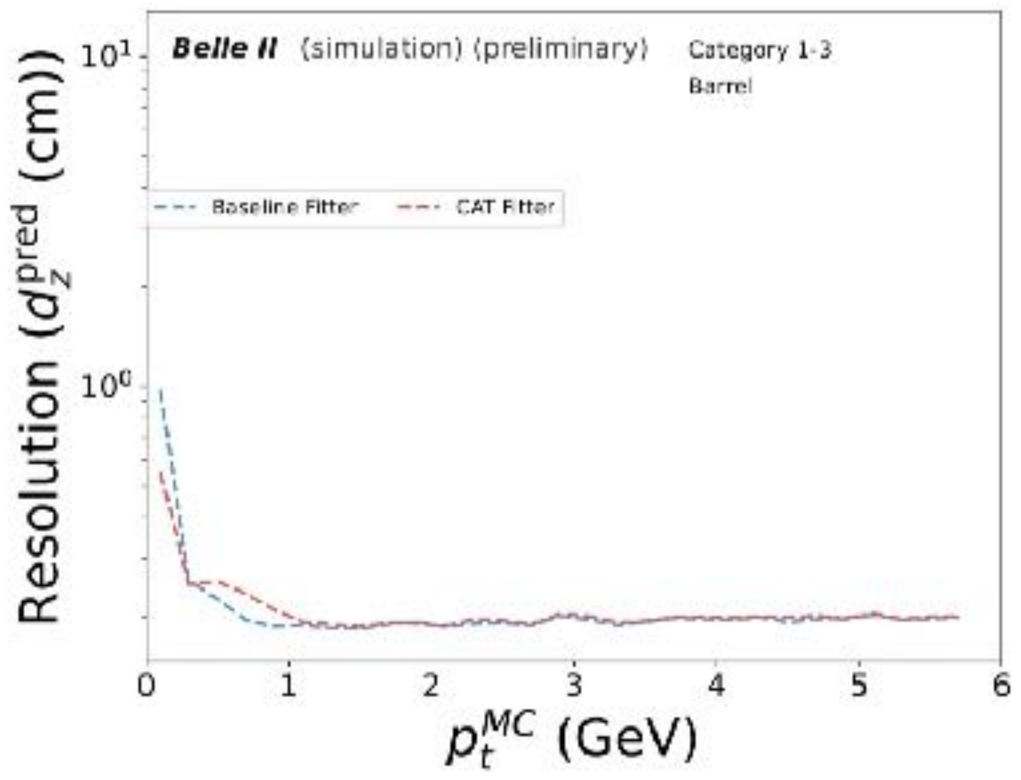


averaged	Efficiency	Purity
Baseline	$0.942^{+/-0.001}$	$0.982^{+/-0.001}$
CAT Finder	$0.951^{+/-0.001}$	$0.987^{+/-0.001}$

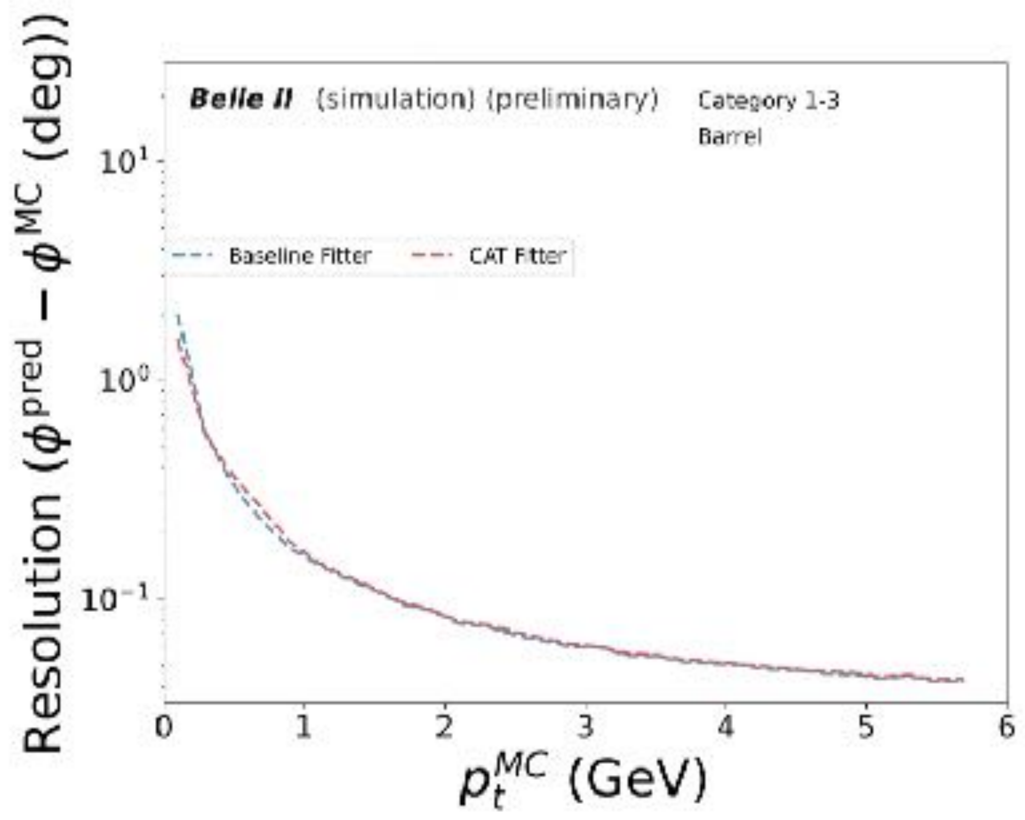
Track parameter resolution



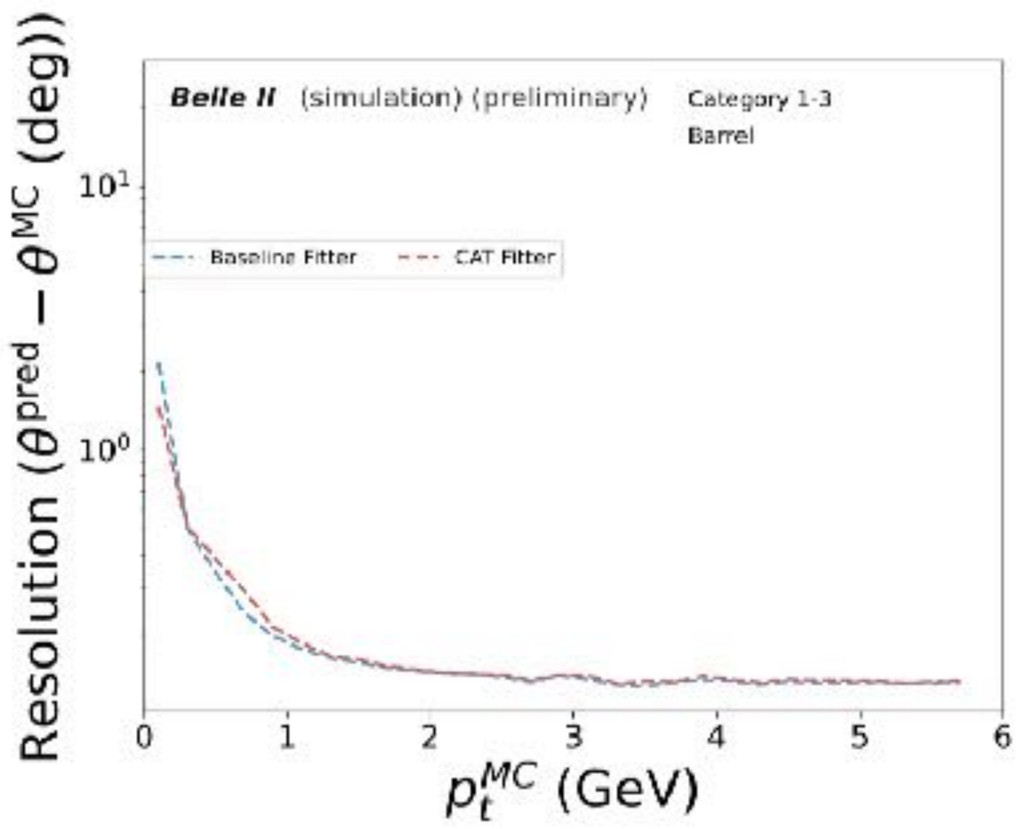
(a) d_0 Resolution.



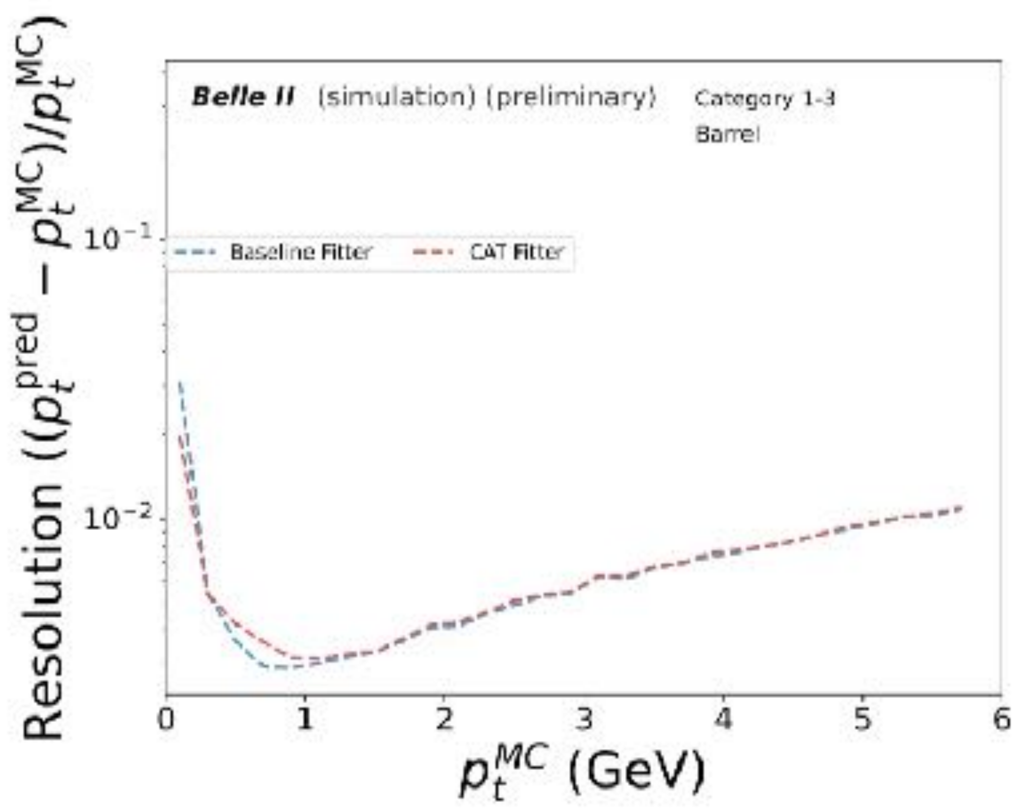
(b) d_z resolution.



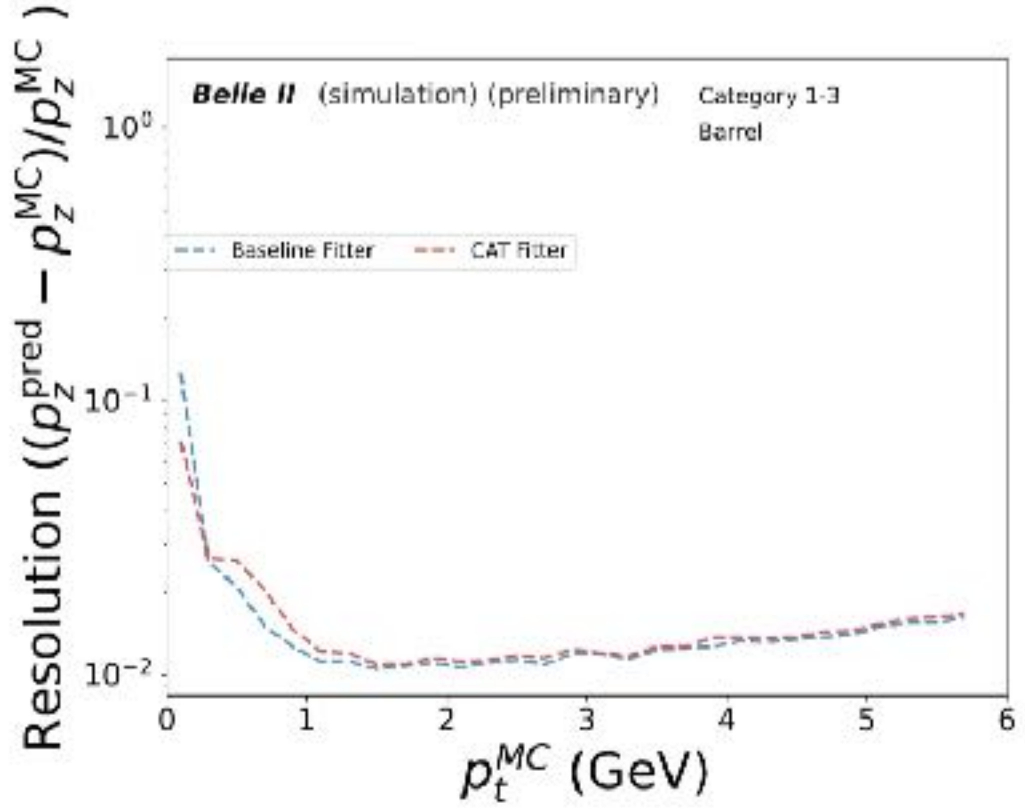
(c) ϕ resolution.



(d) θ resolution.

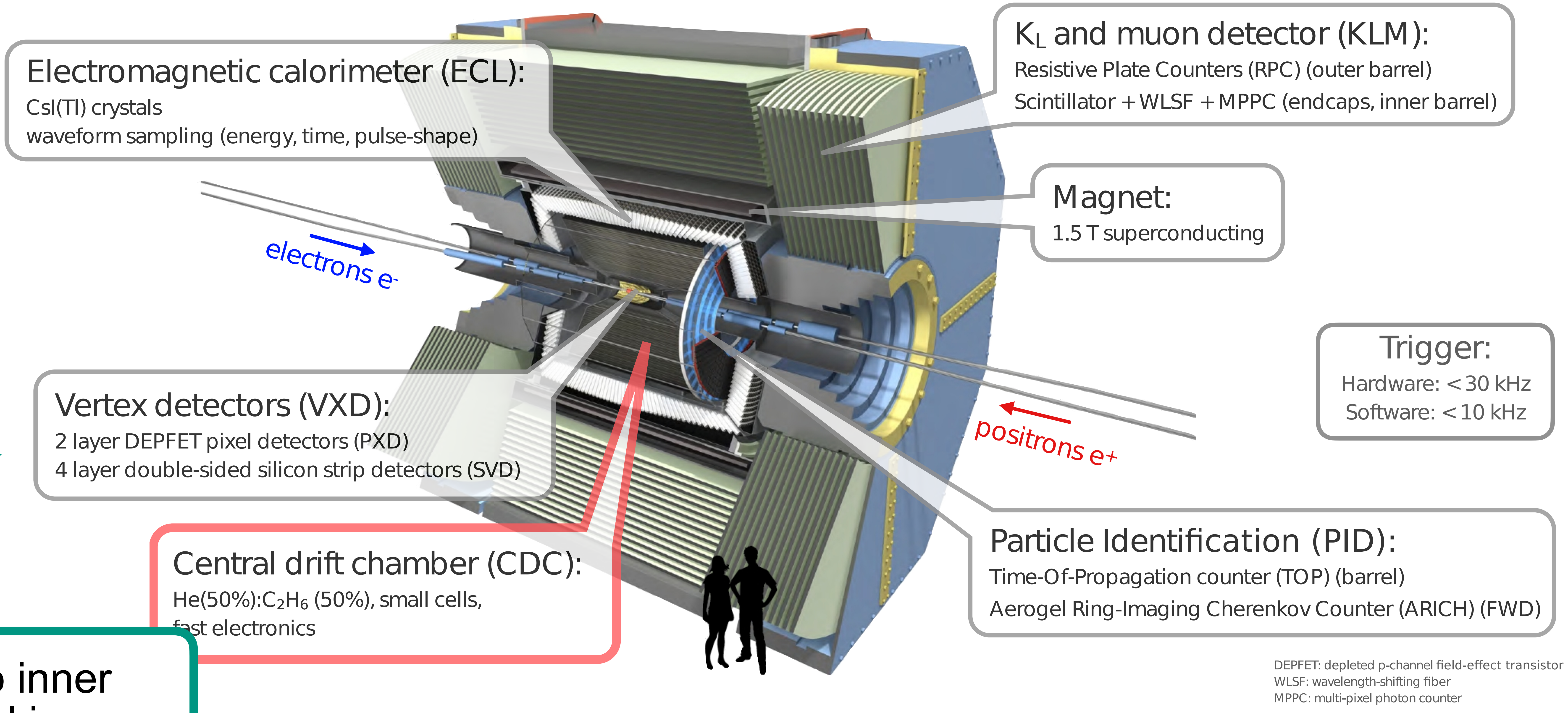


(e) relative p_t resolution.



(f) relative p_z resolution.

Implemented the CAT Finder in the Full Reconstruction algorithm



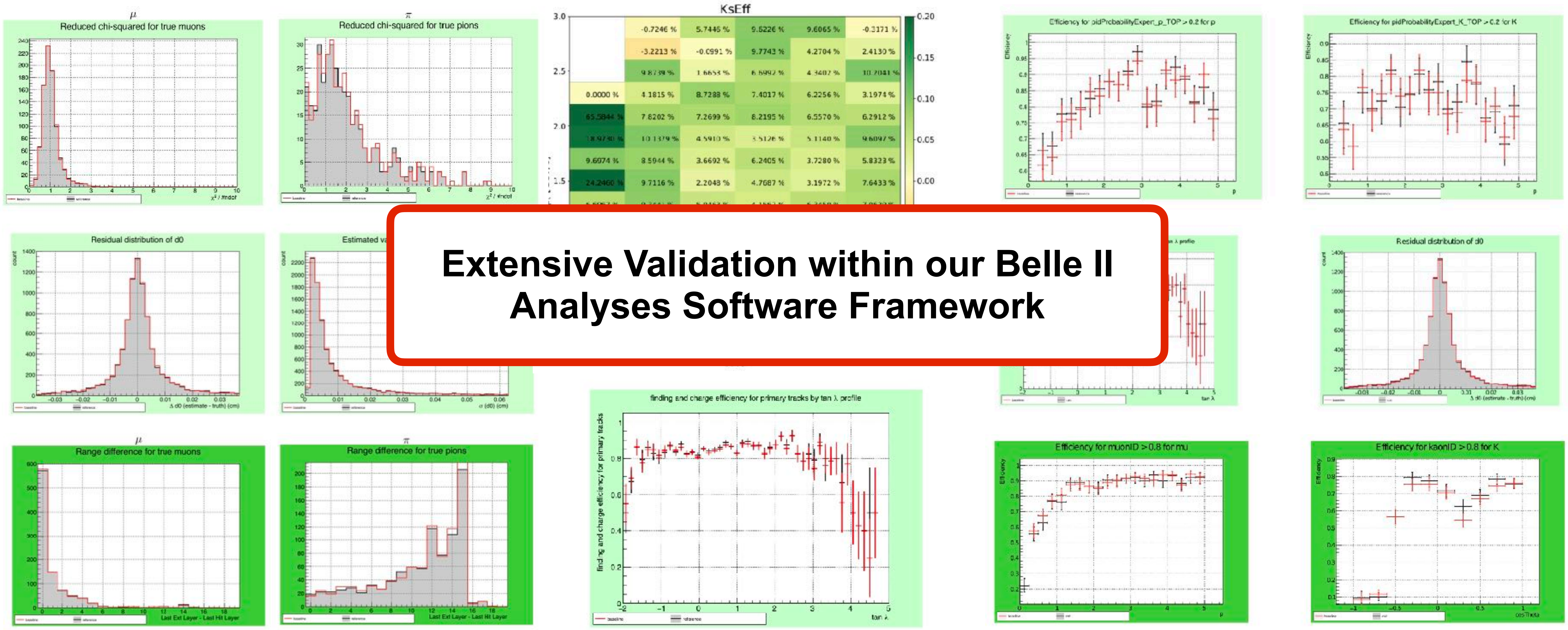
Extrapolation to outer detectors (KLM, TOP) is working

Extrapolation to inner detectors is working (SVD, PXD)

Implemented the CAT Finder in the Full Reconstruction algorithm - Validation

CAT Finder vs Baseline

cat_finder VS baseline



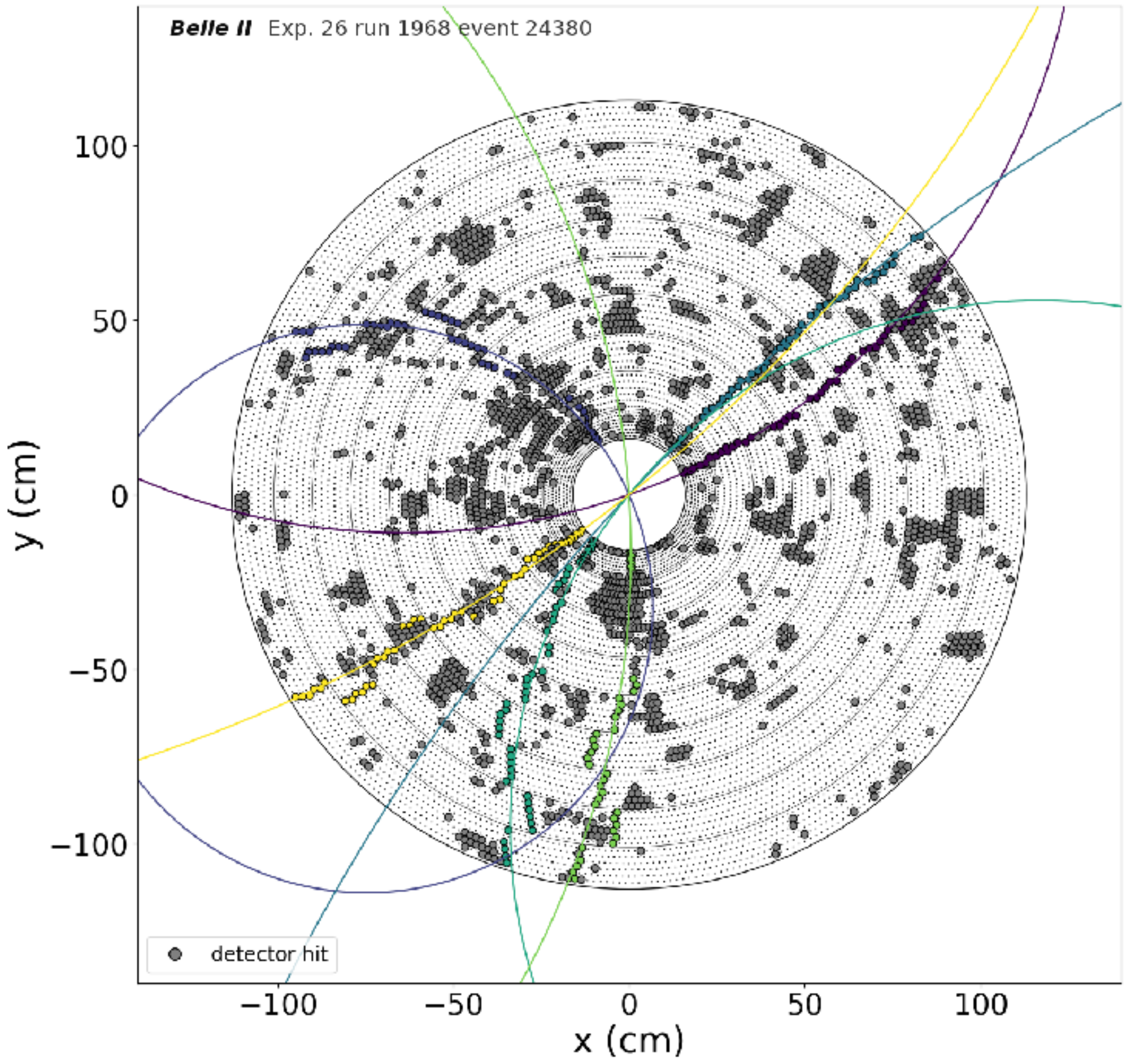
Extensive Validation within our Belle II Analyses Software Framework

Summary

- Implemented end-to-end multi-object track reconstruction algorithm using GNNs in official Belle II software
- Applicable to wide kinematic range and overlapping objects, efficiency largely independent from starting position
- Suppress the varying backgrounds and achieve higher purity than baseline algorithm
- Efficiency outperforms baseline algorithm for displaced tracks and has similar performance on tracks coming from the collision point
- Preprint on arXiv: 2411.13596

Summary

- Implemented end-to-end multi-object track reconstruction algorithm using GNNs in official Belle II software
- Applicable to wide kinematic range and overlapping objects, efficiency largely independent from starting position
- Suppress the varying backgrounds and achieve higher purity than baseline algorithm
- Efficiency outperforms baseline algorithm for displaced tracks and has similar performance on tracks coming from the collision point
- Preprint on [arXiv: 2411.13596](https://arxiv.org/abs/2411.13596)



Outlook

- Implementation of CAT Finder as C++ module (Currently targeting SOFIE from ROOT)
- Tracking comparison on data (K_s^0 peak in high-multiplicity events)
- Checking performance on analysis $B \to KS$