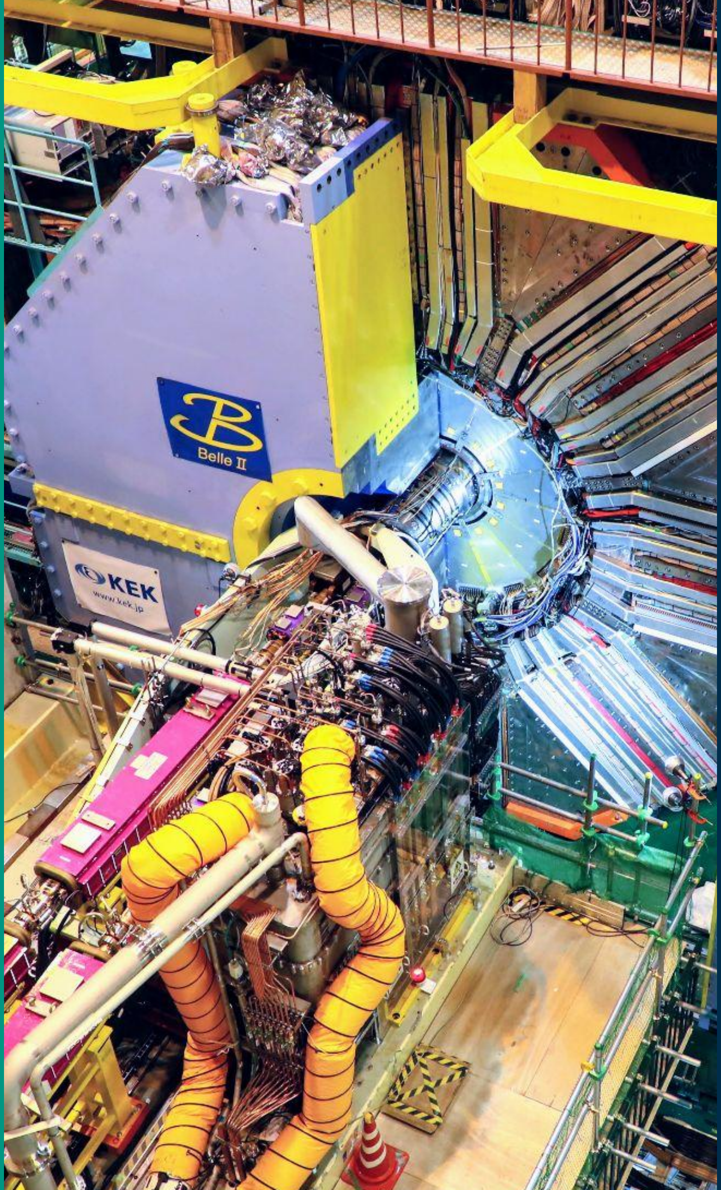
GNN-based Track Finder for SVD and CDC



FSP Meeting 2025 Bonn

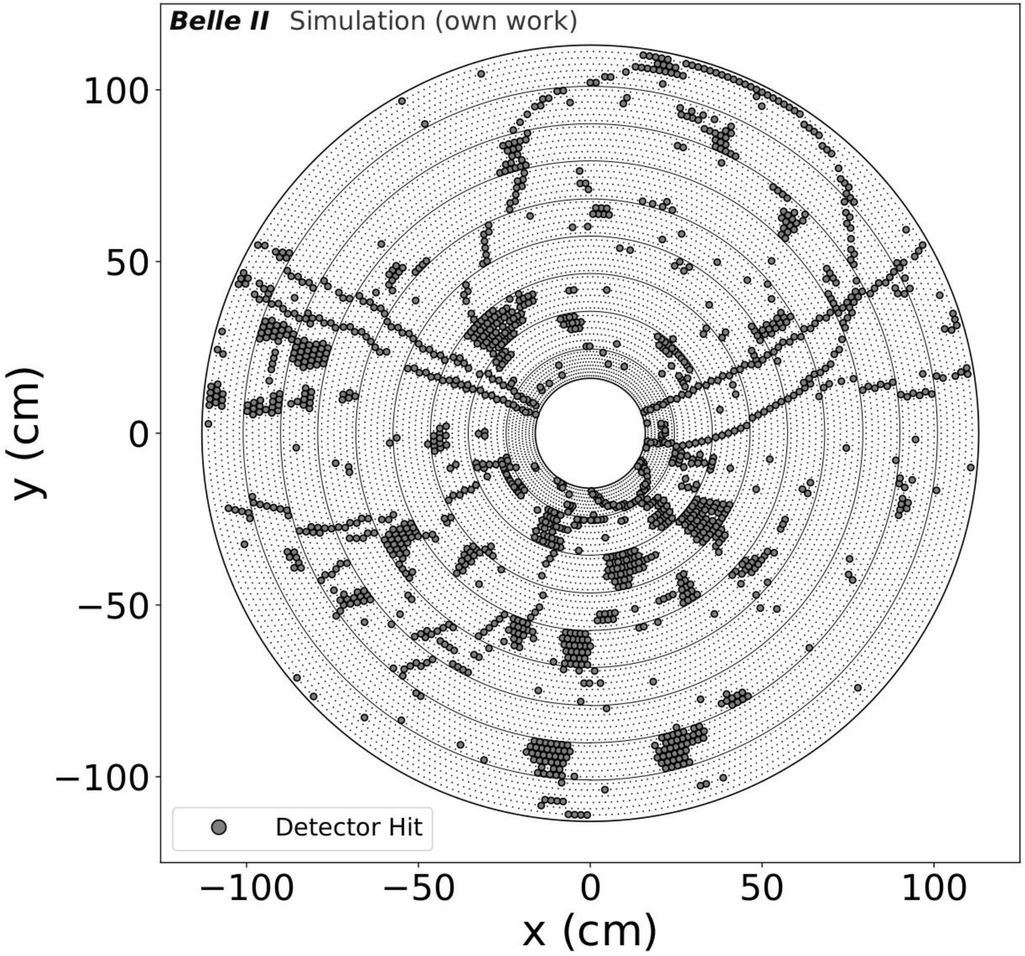
Tristan Brandes, Lea Reuter, Giacomo De Pietro, Torben Ferber

09/09/25

Belle II

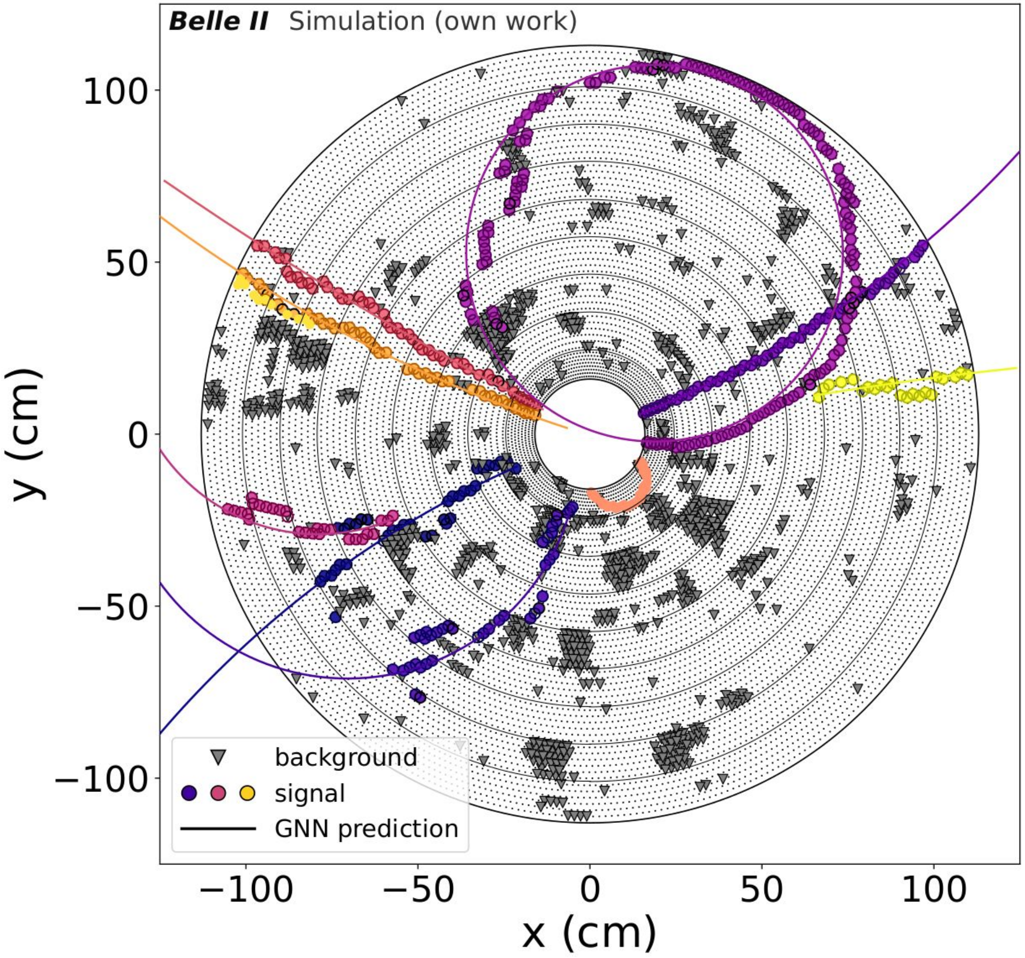
Institute of Experimental Particle Physics

Track Finding in the CDC



- These are all CDC Wire Hits from an exemplary event
- Can you find all the tracks?

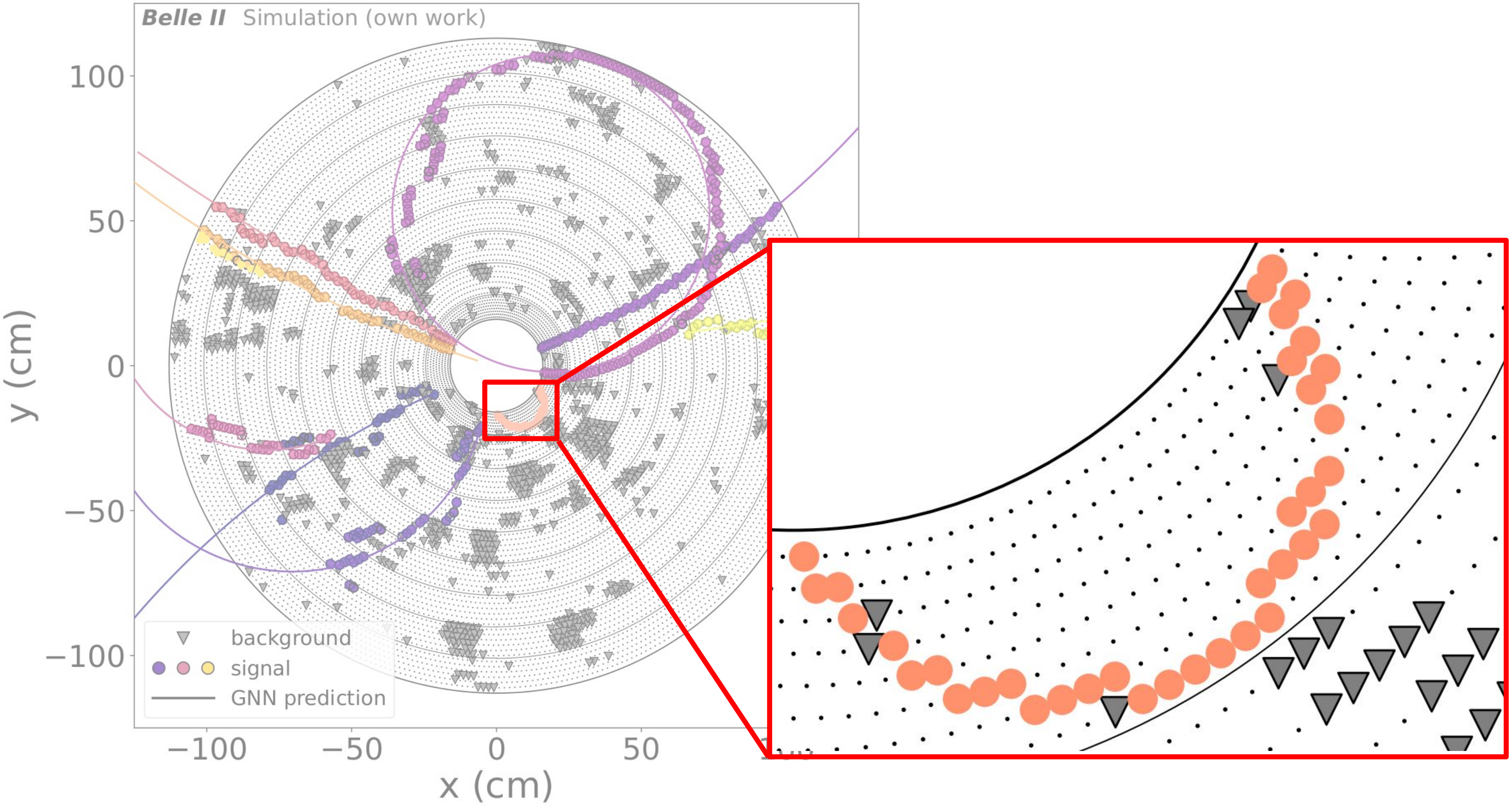
CATFinder



- CATFinder by Lea Reuter et al.
  End-to-End Multi-Track Reconstruction using Graph Neural Networks at Belle II (arXiv:2411.13596)

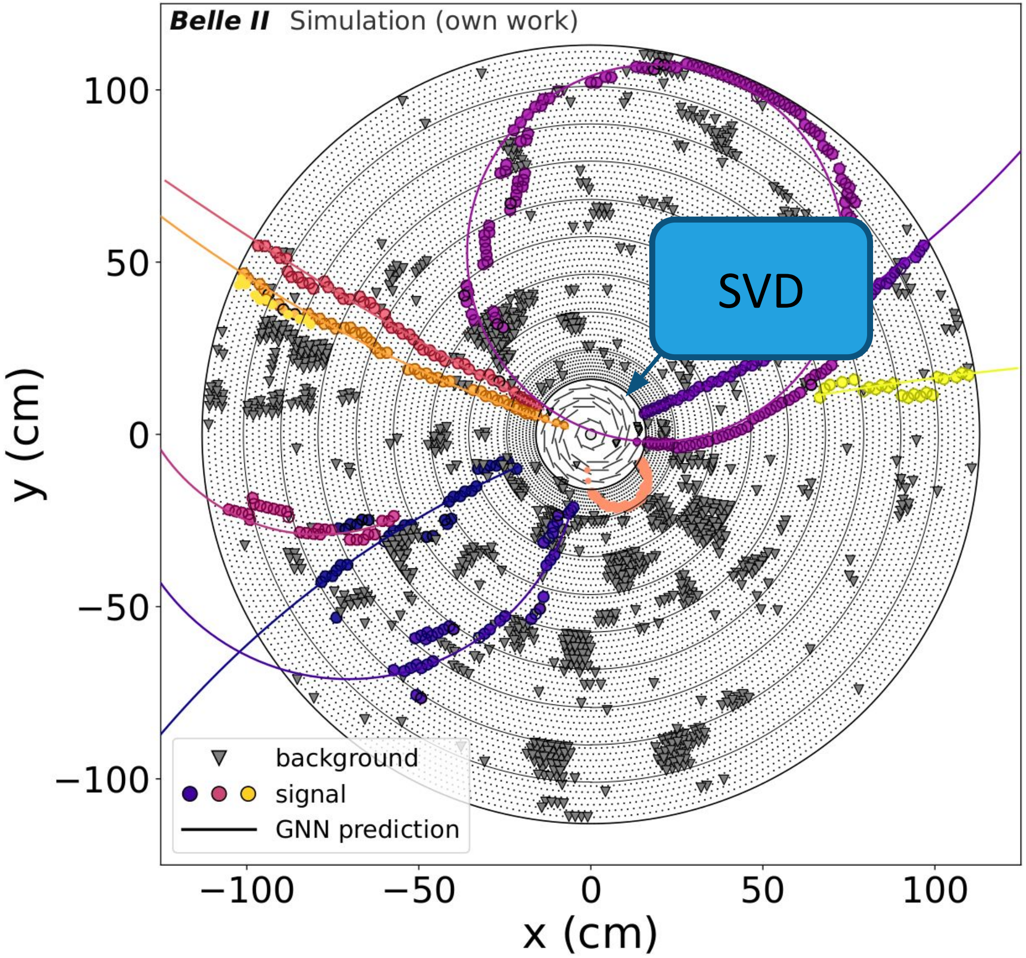
- Successful implementation of Graph Neural Network with Object Condensation loss for Central Drift Chamber
- Provides tracks with parameters given by CATFinder predictions
  - Momentum p_x, p_y and p_z
  - Track starting position v_x, v_y and v_z
  - Charge

CATFinder



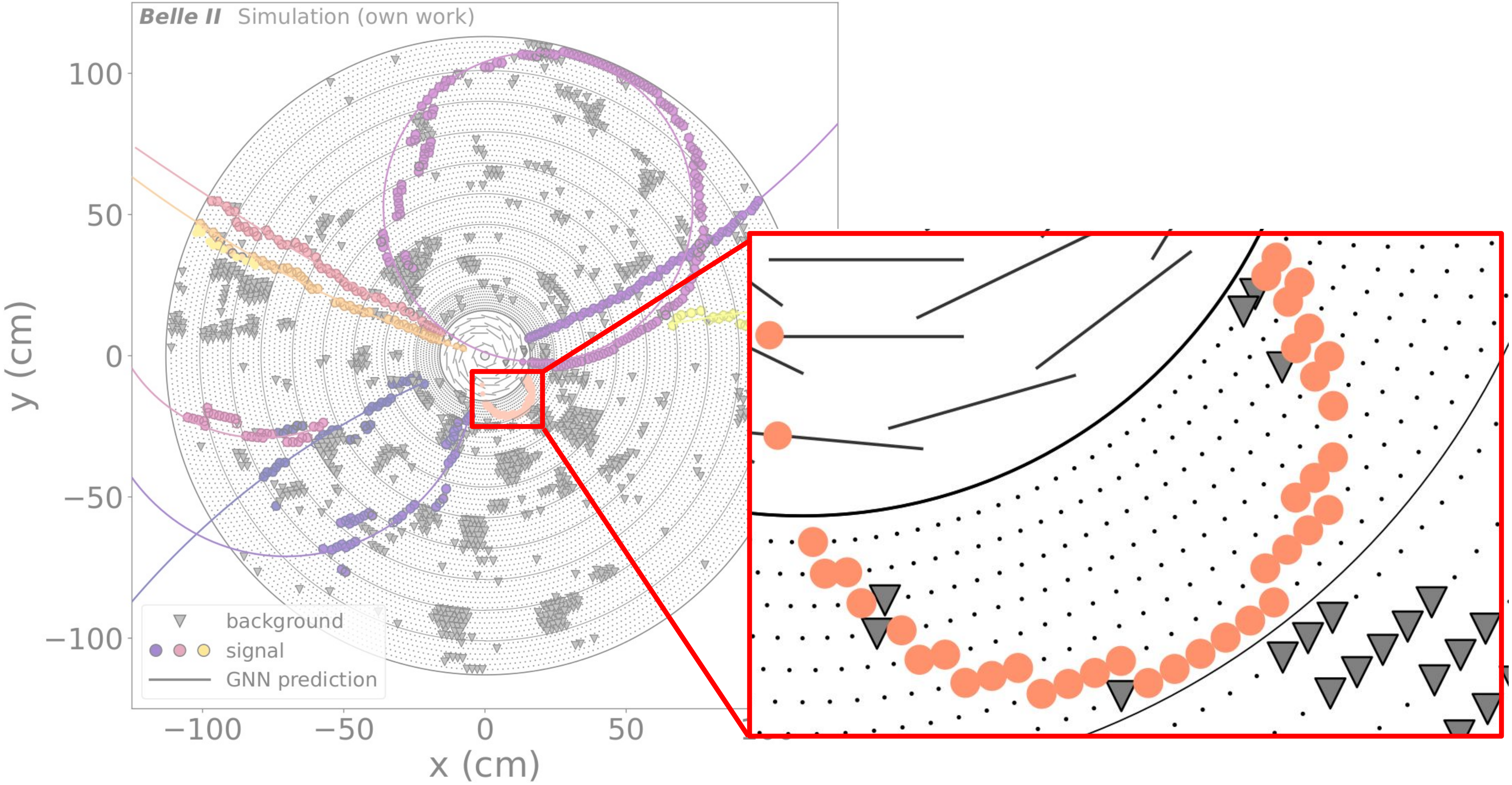
Belle II Simulation (own work)
100
50
0
−50
−100
y (cm)
−100
−50
0
50
x (cm)
background
signal
GNN prediction

Extending CATFinder with SVD Information



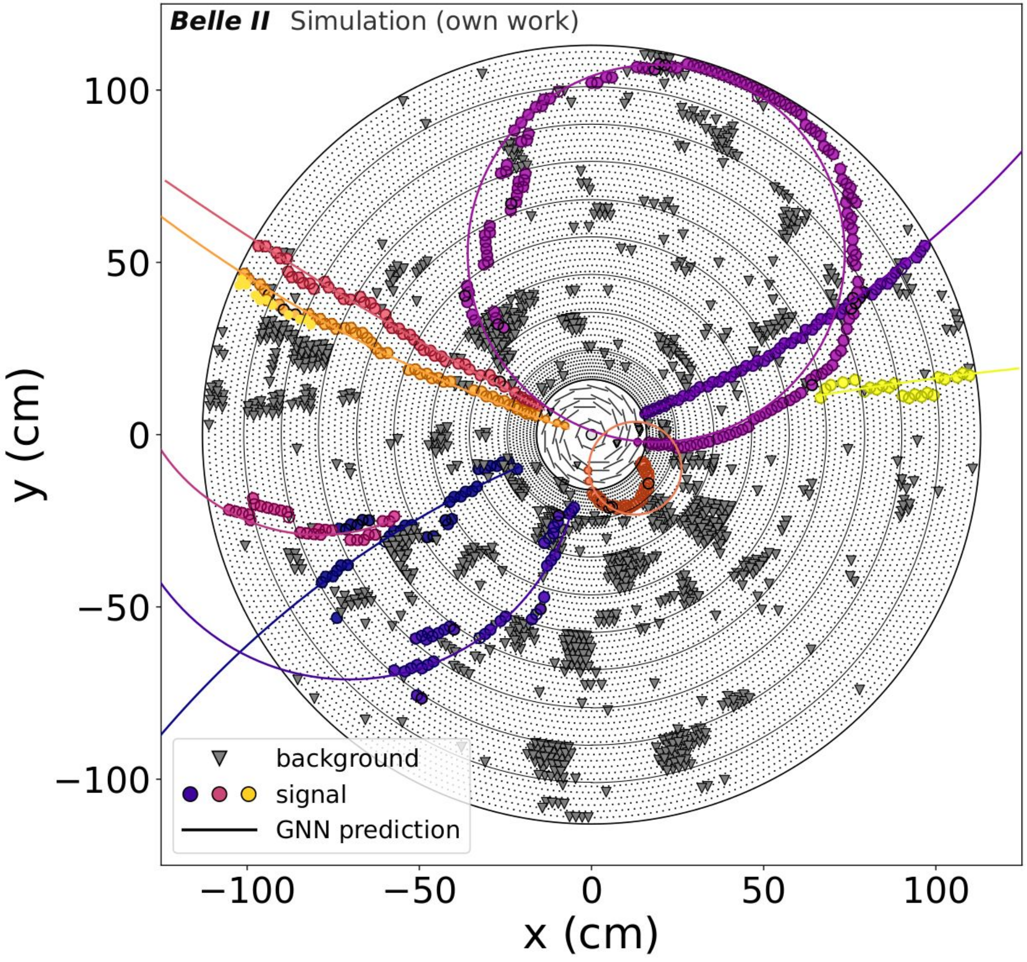
- CATFinder by Lea Reuter et al.
- Successful implementation of Graph Neural Network with Object Condensation loss for Central Drift Chamber
- Provides tracks with parameters given by CATFinder predictions
  - Momentum p_x, p_y and p_z
  - Track starting position v_x, v_y and v_z
  - Charge
- Incorporation of Silicon Vertex Detector as next step

CATFinder



Belle II Simulation (own work)
100
50
0
−50
−100
y (cm)
−100
−50
0
50
x (cm)
background
signal
GNN prediction

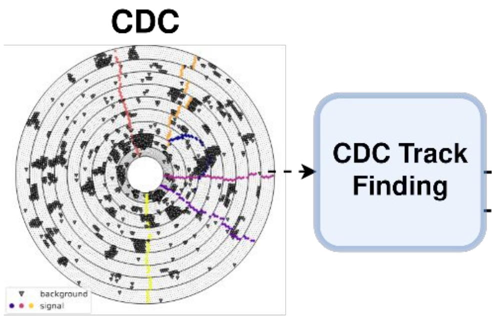
Extending CATFinder with SVD Information



- CATFinder by Lea Reuter et al.
- Successful implementation of Graph Neural Network with Object Condensation loss for Central Drift Chamber
- Provides tracks with parameters given by CATFinder predictions
  - Momentum p_x, p_y and p_z
  - Track starting position v_x, v_y and v_z
  - Charge
- Incorporation of Silicon Vertex Detector as next step

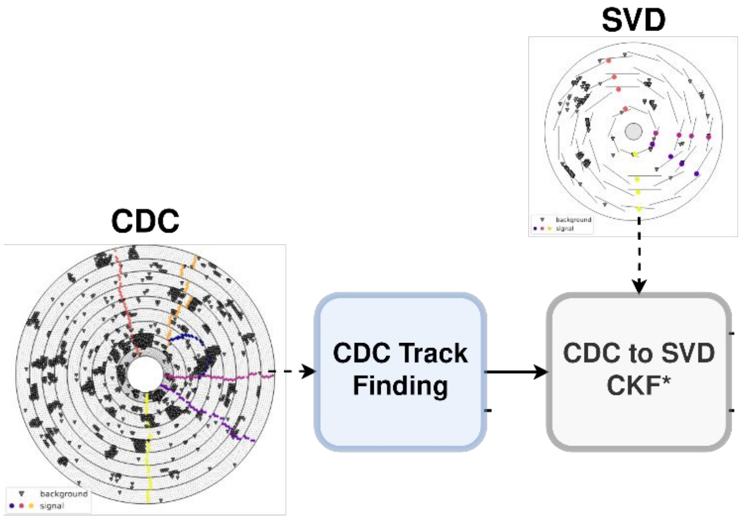
Current Belle II Tracking Chain

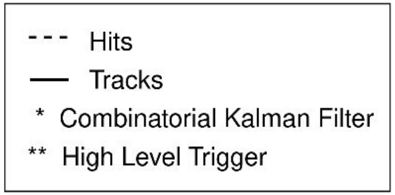
- - - Hits
— Tracks
* Combinatorial Kalman Filter
** High Level Trigger



1. CDC track finding

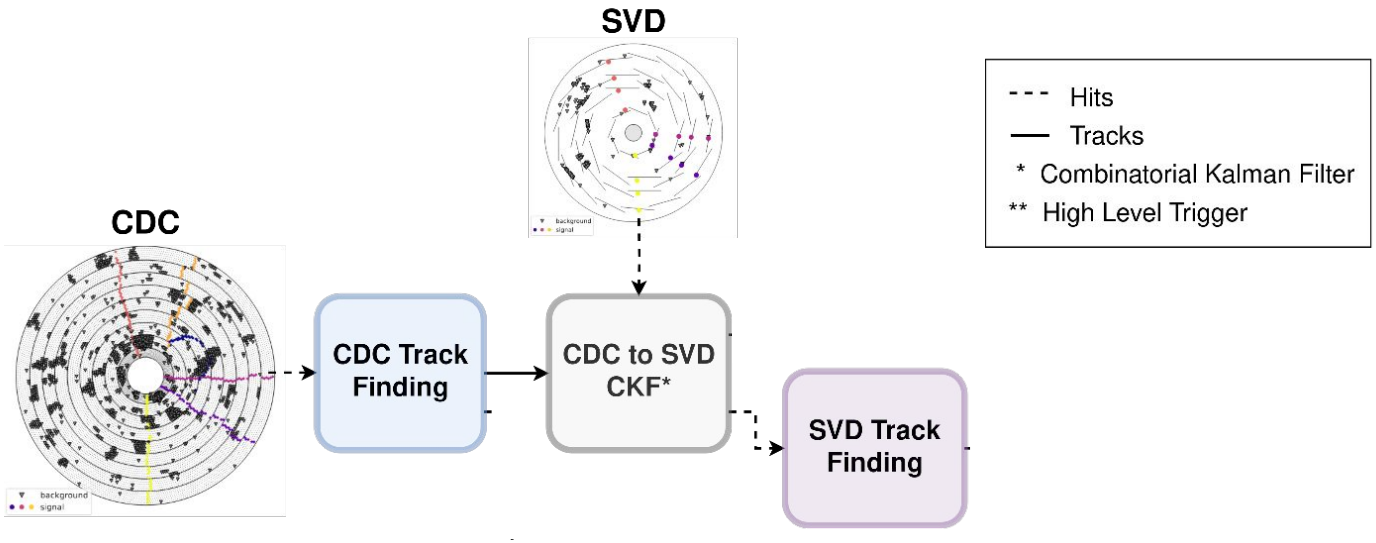
Current Belle II Tracking Chain





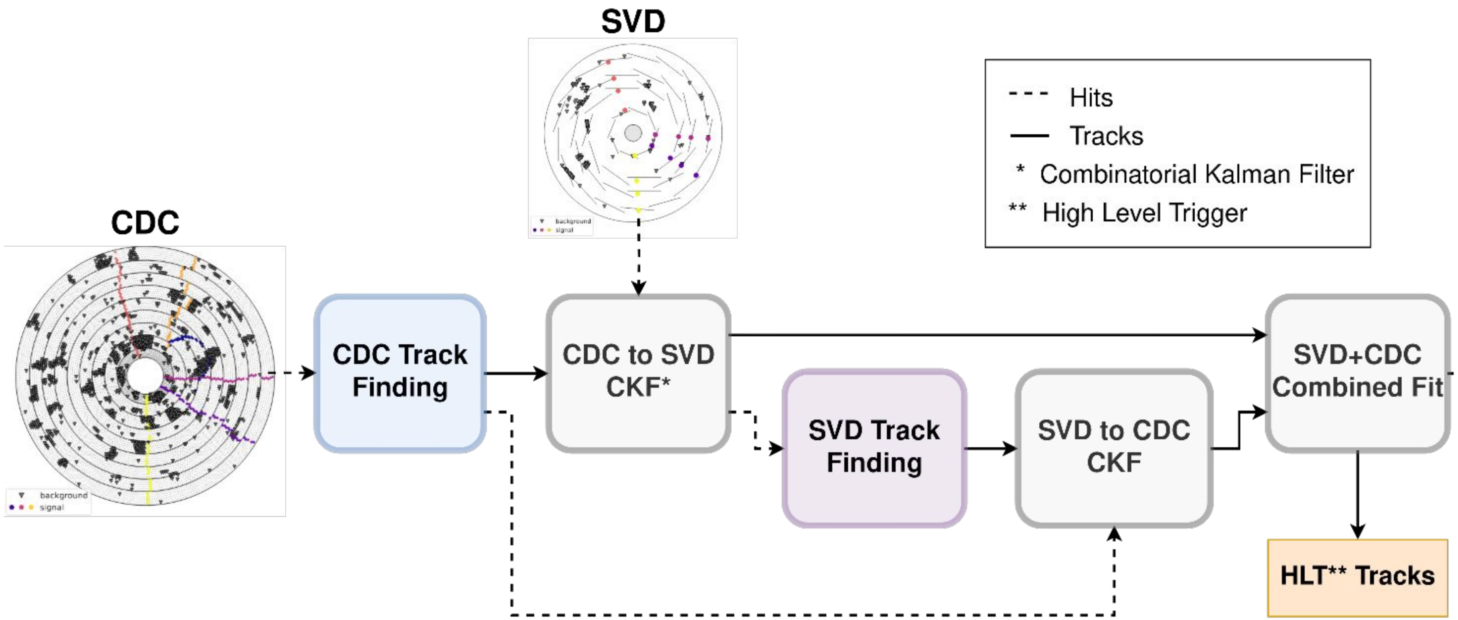
1. CDC track finding
2. Extrapolate found CDC tracks into the SVD and attach matching SVD Hits via CKF (Combinatorial Kalman Filter)

Current Belle II Tracking Chain



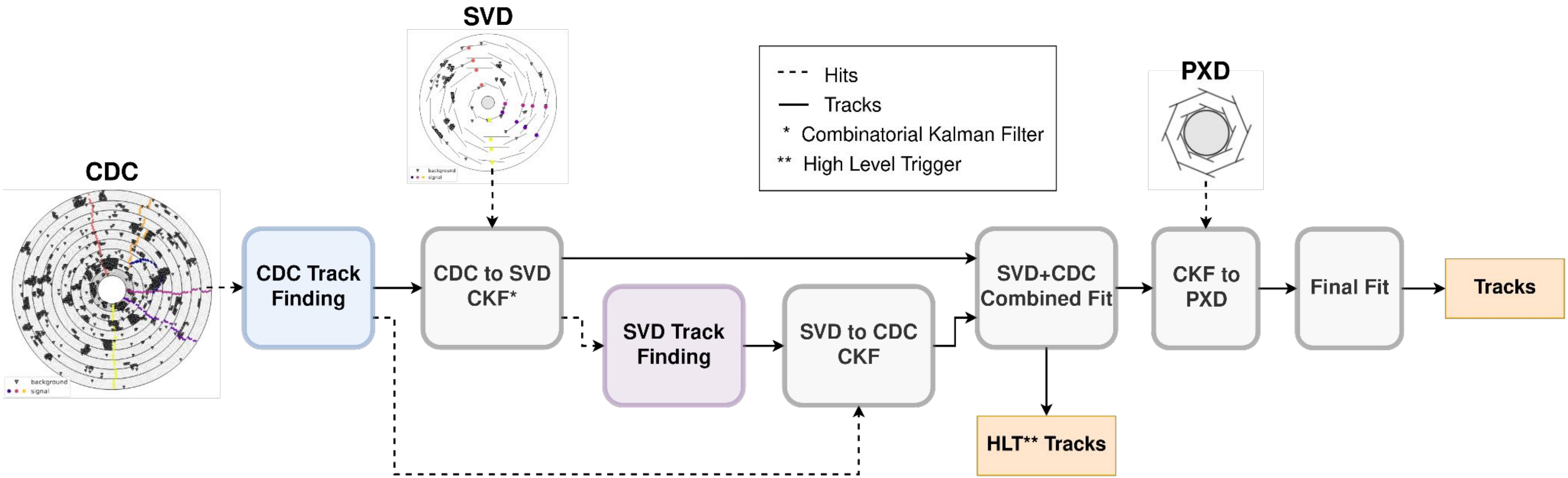
1. CDC track finding
2. Extrapolate found CDC tracks into the SVD and attach matching SVD Hits via CKF (Combinatorial Kalman Filter)
3. SVD standalone track finding on remaining SVD Hits

Current Belle II Tracking Chain



1. CDC track finding
2. Extrapolate found CDC tracks into the SVD and attach matching SVD Hits via CKF (Combinatorial Kalman Filter)
3. SVD standalone track finding on remaining SVD Hits
4. Extrapolate SVD standalone tracks into CDC and attach matching (still unassigned) CDC Hits

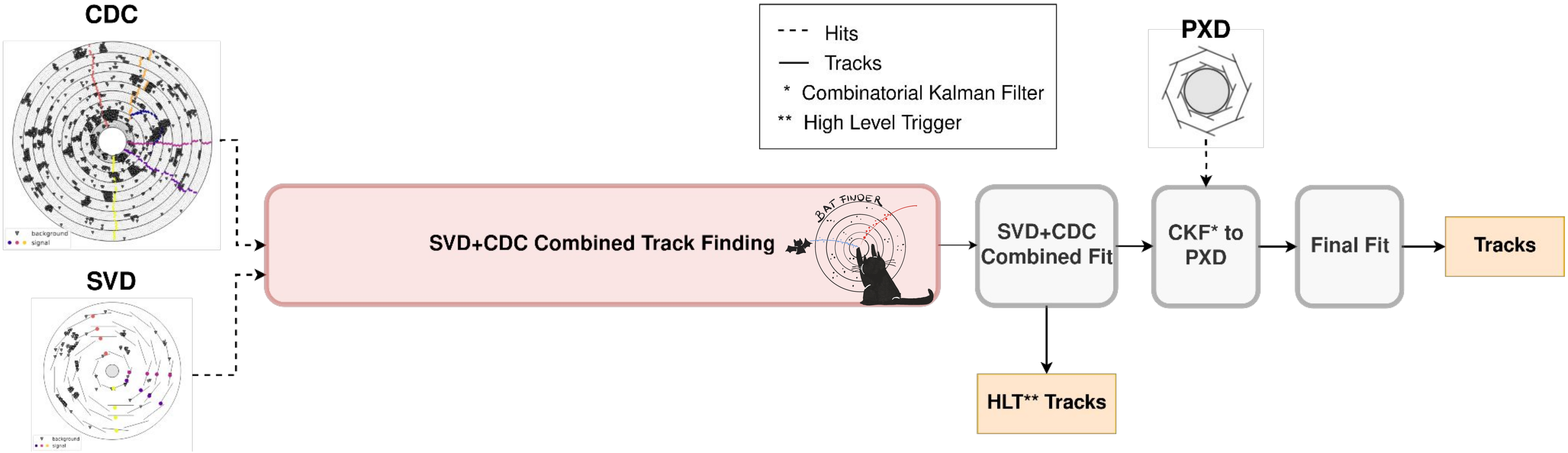
Current Belle II Tracking Chain



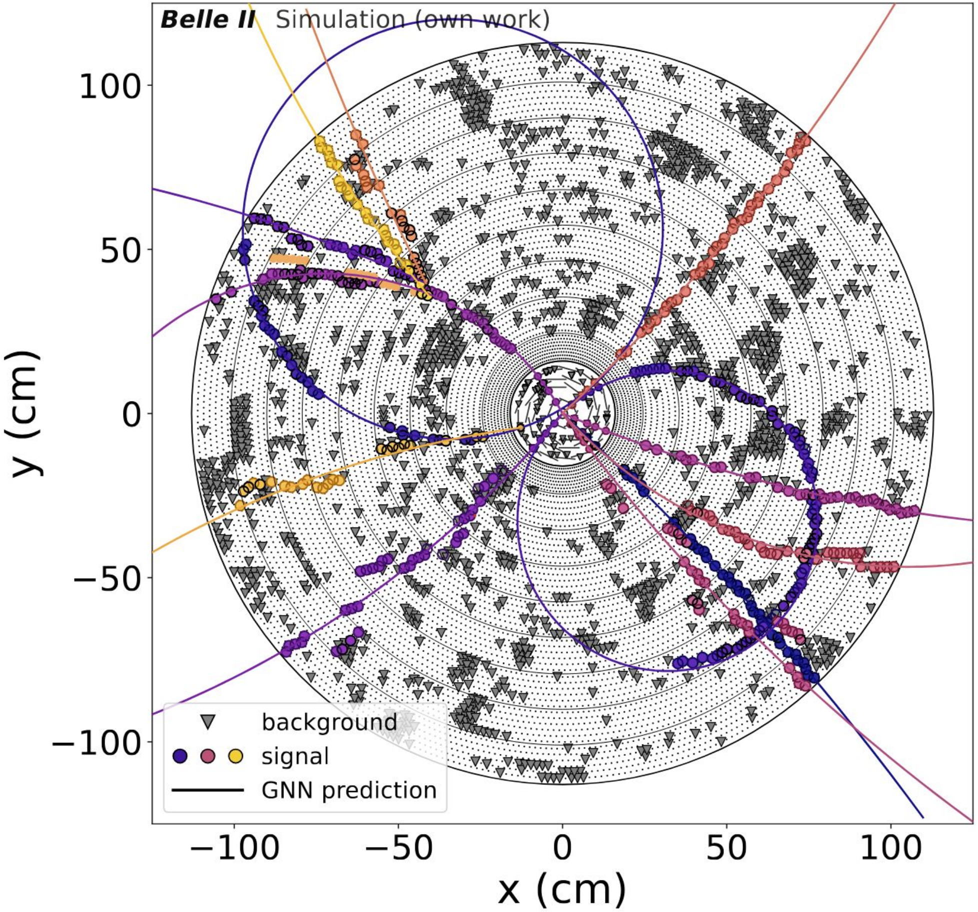
1. CDC track finding
2. Extrapolate found CDC tracks into the SVD and attach matching SVD Hits via CKF (Combinatorial Kalman Filter)
3. SVD standalone track finding on remaining SVD Hits
4. Extrapolate SVD standalone tracks into CDC and attach matching (still unassigned) CDC Hits

Towards a Unified CDC + SVD Track Finder

- We want to provide complete set of track candidates from CDC + SVD in one inference step
  → BATFinder (Belle II AI Track Finder)

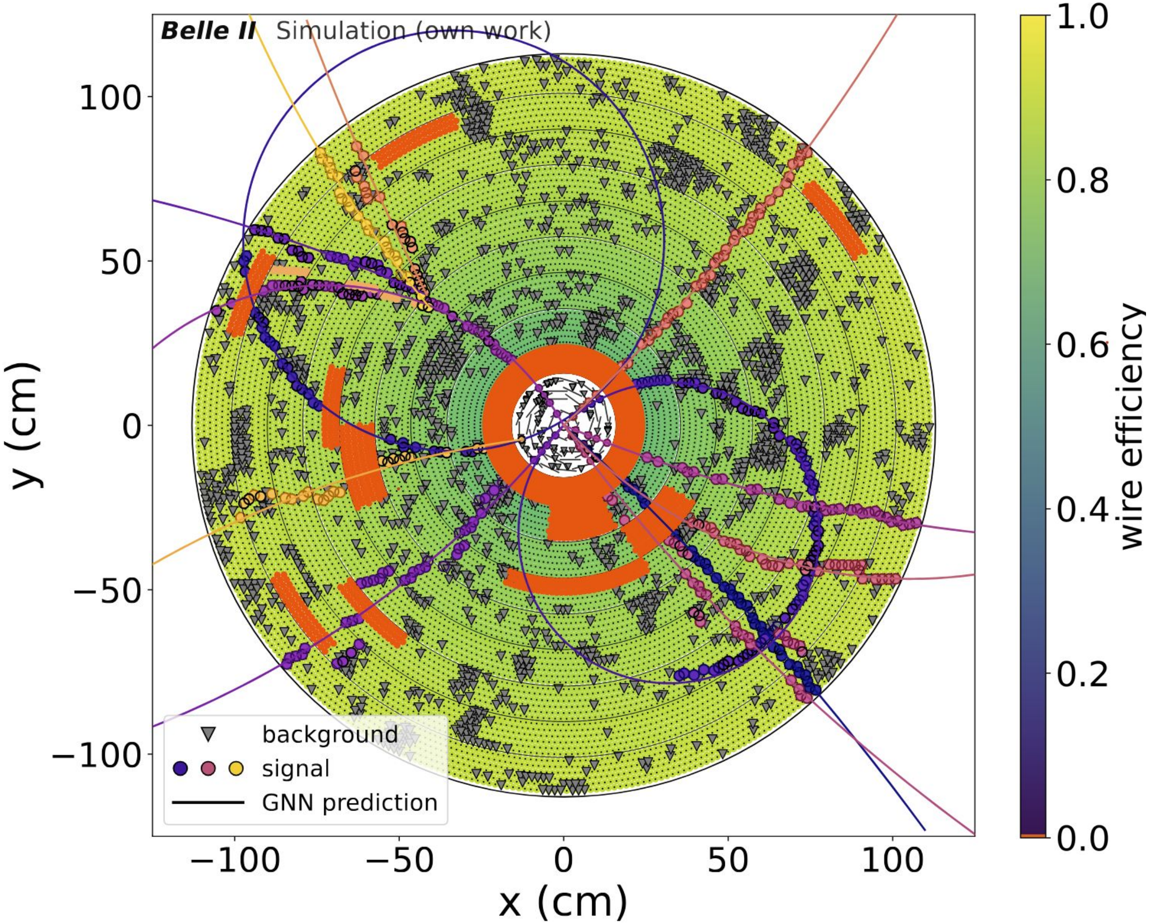


Future CDC Conditions



- Increasing Beam Background Levels

Future CDC Conditions



- Increasing Beam Background Levels
- Wire inefficiencies due to charge buildup at wires

Tracking Approach

- Find unknown number of track candidates per event
- For each track candidate, find
  - Associated hits
  - Estimate the track's momentum, charge, and starting position

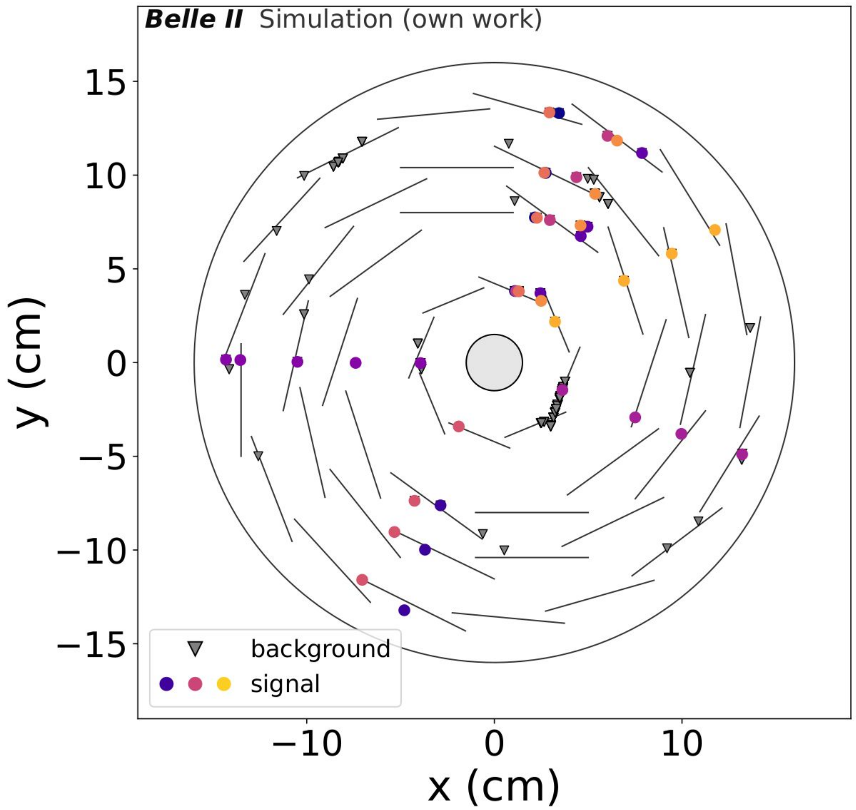
- Model and architecture inspired by CATFinder: Graph Neural Network with Object Condensation
- Graph Neural Network to address
  - Unordered, variable number and spatially irregular number of inputs (Detecor hits)
- Object Condensation Loss to address
  - Arbitrary number of classes (tracks) per event (https://arxiv.org/abs/2002.03605)

- Inputs for the BATFinder are SVD SpacePoints and CDC Hits

Approach: SVD Inputs

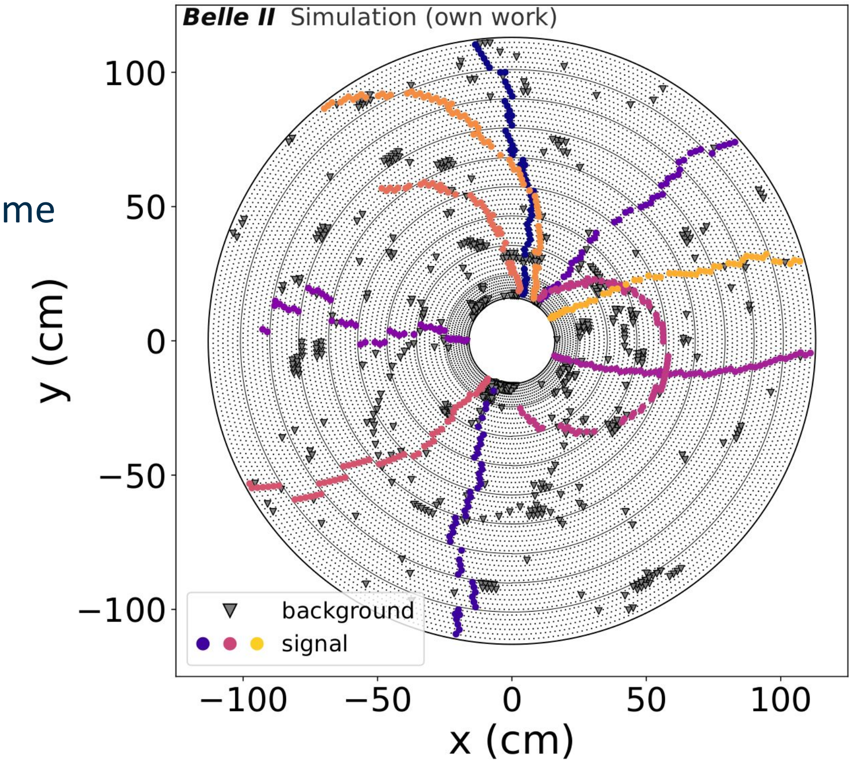
- In SVD reconstruction, SVD SpacePoints are created using SVD Clusters from opposing strip sides

- Per SpacePoint, we provide BATFinder with features from SVD reconstruction
  - 3D coordinates x, y and z
  - Clusters deposited charge
  - Clusters Signal-to-Noise-Ratio (SNR)
  - Cluster TimeU & TimeV



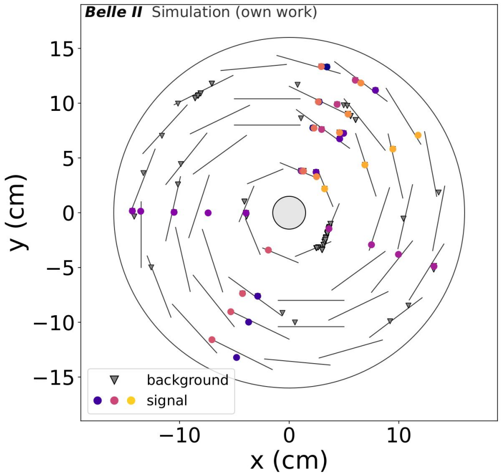
Approach: CDC Inputs

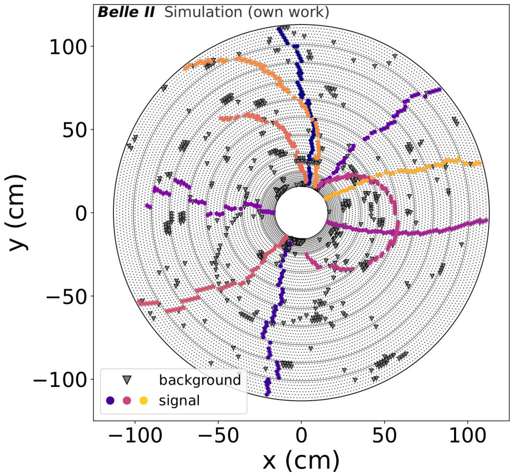
- Per CDC Hit, we provide BATFinder with
  - 2D coordinates **x** and **y** at z = 0
  - **TDC** (Time to digital converter),
    Time when energy deposition exceeds threshold for first time
  - **ADC** (Analog to Digital converter),
    Digitized signal amplitude
  - **TOT** (Time over readout threshold)
  - **Superlayer**, number of superlayer (0 to 8)
  - **Clayer**, global layer number (0 to 55)
  - **Layer**, local layer number in respective superlayer (0 to 7)



BATFinder

- Concept: multimodal GNN → Different types of input data for one network
- Now we combine both inputs from SVD and CDC into multimodal BATFinder GNN
- Main Differences:
  - 3D SpacePoints from SVD vs 2D CDC hits
  - Different number of hits in SVD (on avg. 4) and CDC (~50) per track





Unified SVD and CDC Feature Representation

- Unify SVD and CDC inputs into a common feature vector
- Missing values padded (currently to 0) for compatibility, added is_svd flag

- Example: Feature vectors for SVD Space Point and CDC Hit, zero-padded values in grey

	x	y	z	TimeU	TimeV	SNR	Charge	TDC	ADC	TOT	Superlayer	Clayer	Layer	is_svd
SVD Hit	4	7	6	12	7	20	2000	0	0	0	0	0	0	1

Unified SVD and CDC Feature Representation

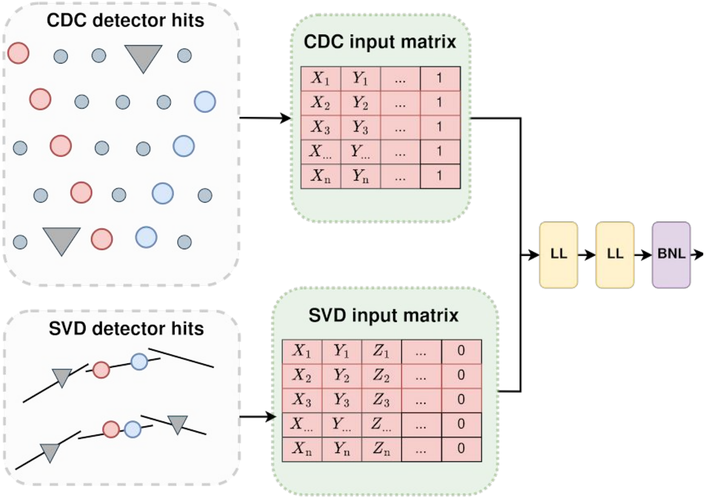
- Unify SVD and CDC inputs into a common feature vector
- Missing values padded (currently to 0) for compatibility, added is_svd flag

- Example: Feature vectors for SVD Space Point and CDC Hit, zero-padded values in grey

	x	y	z	TimeU	TimeV	SNR	Charge	TDC	ADC	TOT	Superlayer	Clayer	Layer	is_svd
SVD Hit	4	7	6	12	7	20	2000	0	0	0	0	0	0	1
CDC Hit	50	45	0	0	0	0	0	4200	700	70	2	20	6	0

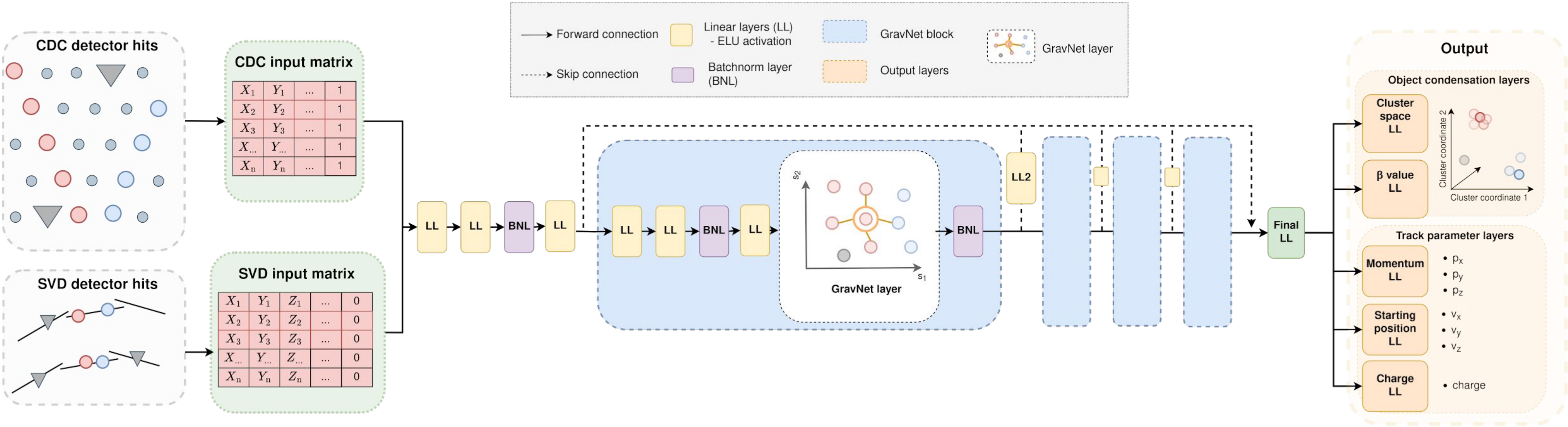
Multimodal GNN Input Representation

- SVD and CDC hits unified via feature padding into a shared input vector
- Additional fully connected layers (LL) map this representation to a latent space before the GNN

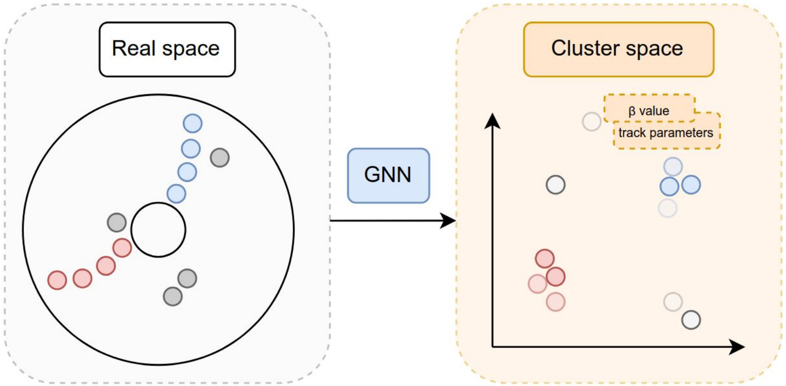


Multimodal GNN Input Representation

- SVD and CDC hits unified via feature padding into a shared input vector
- Additional fully connected layers (LL) map this representation to a latent space before the GNN

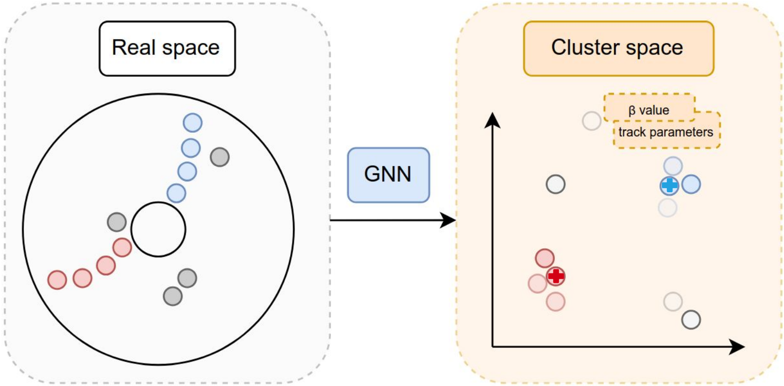


Object Condensation to Track Finding



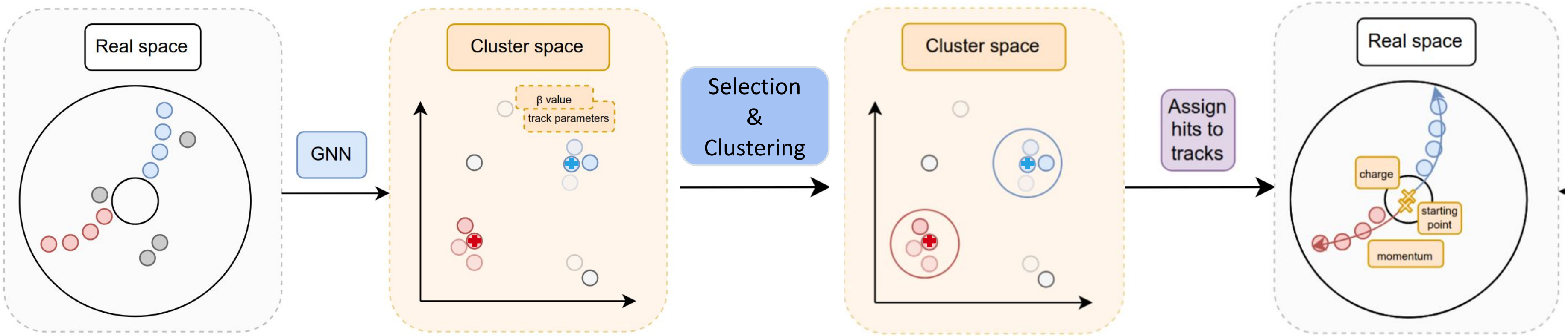
- 1) Hits are translated from real space to latent cluster space, we task model with predicting β-value (How track-like is node) and track parameters
  - Condensation points are selected by employing cut on β-value and requiring separation criteria

Object Condensation to Track Finding



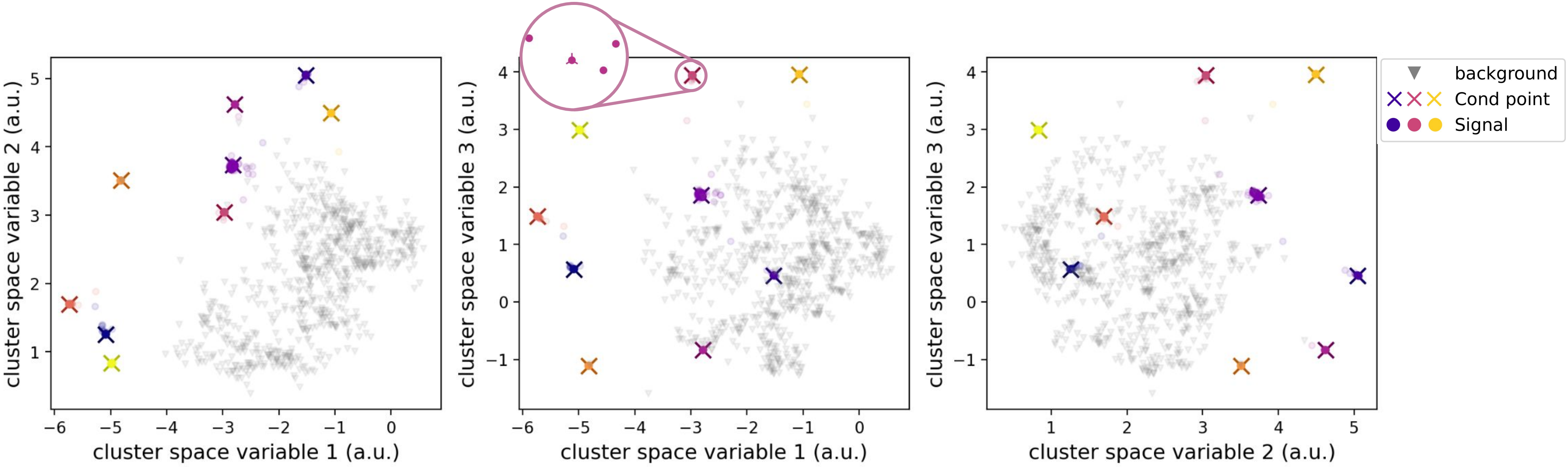
- 1) Hits are translated from real space to latent cluster space, we task model with predicting β-value (How track-like is node) and track parameters
  - Condensation points are selected by employing cut on β-value and requiring separation criteria

Object Condensation to Track Finding



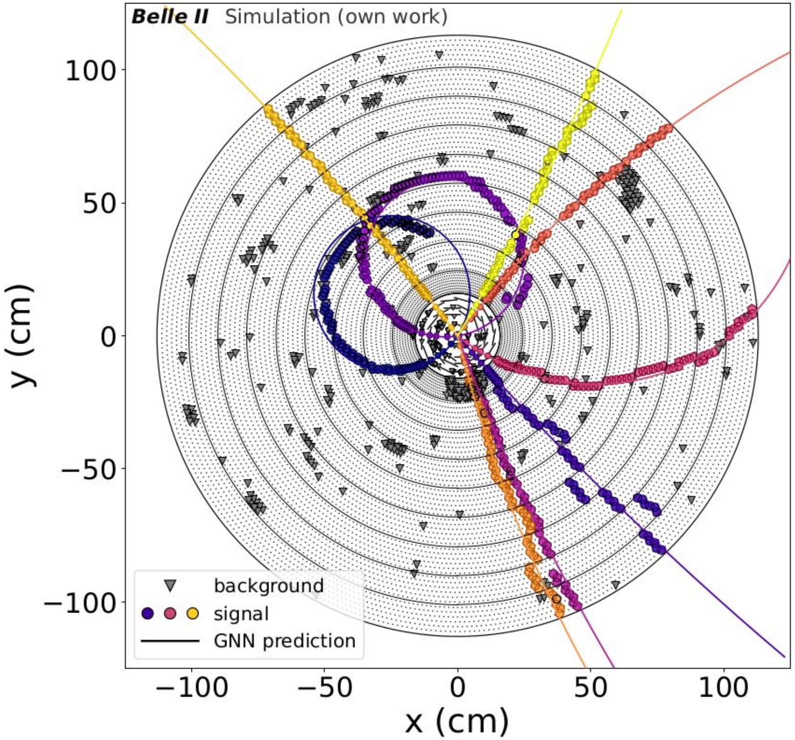
- 1) Hits are translated from real space to latent cluster space, we task model with predicting β-value (How track-like is node) and track parameters
  - Condensation points are selected by employing cut on β-value and requiring separation criteria
- 2) Clustering of nodes based on condensation points, nodes satisfying distance threshold are assigned to condensation point
  - Condensation point and assigned nodes are translated into real space where they form a track

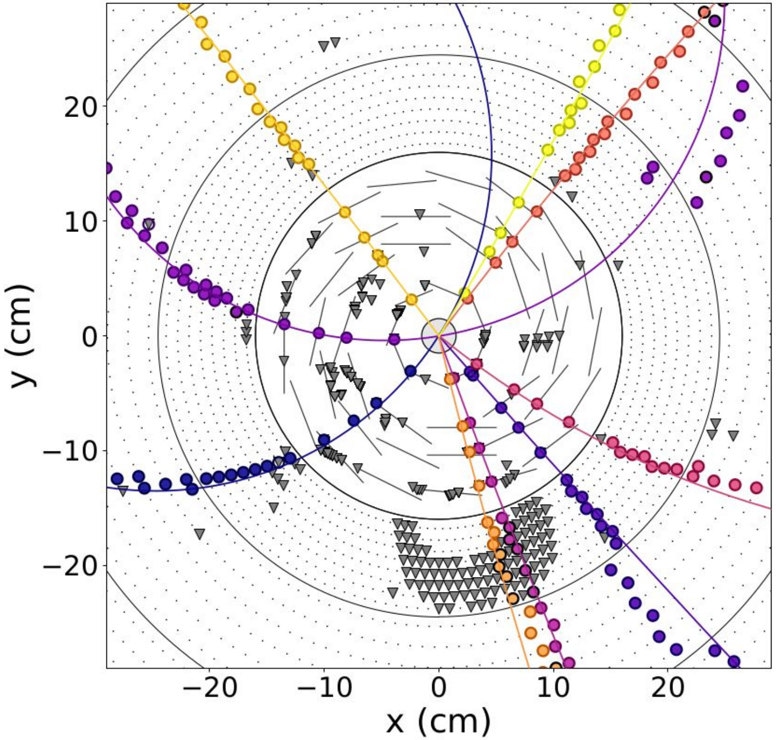
Model Prediction: Condensation Points



- Projections of BATFinder cluster space, each point in cluster space represents SP in SVD or CDC Hit
- Different colors represent the condensation points and their assigned hits
- We identify each condensation point and its assigned hits as a track

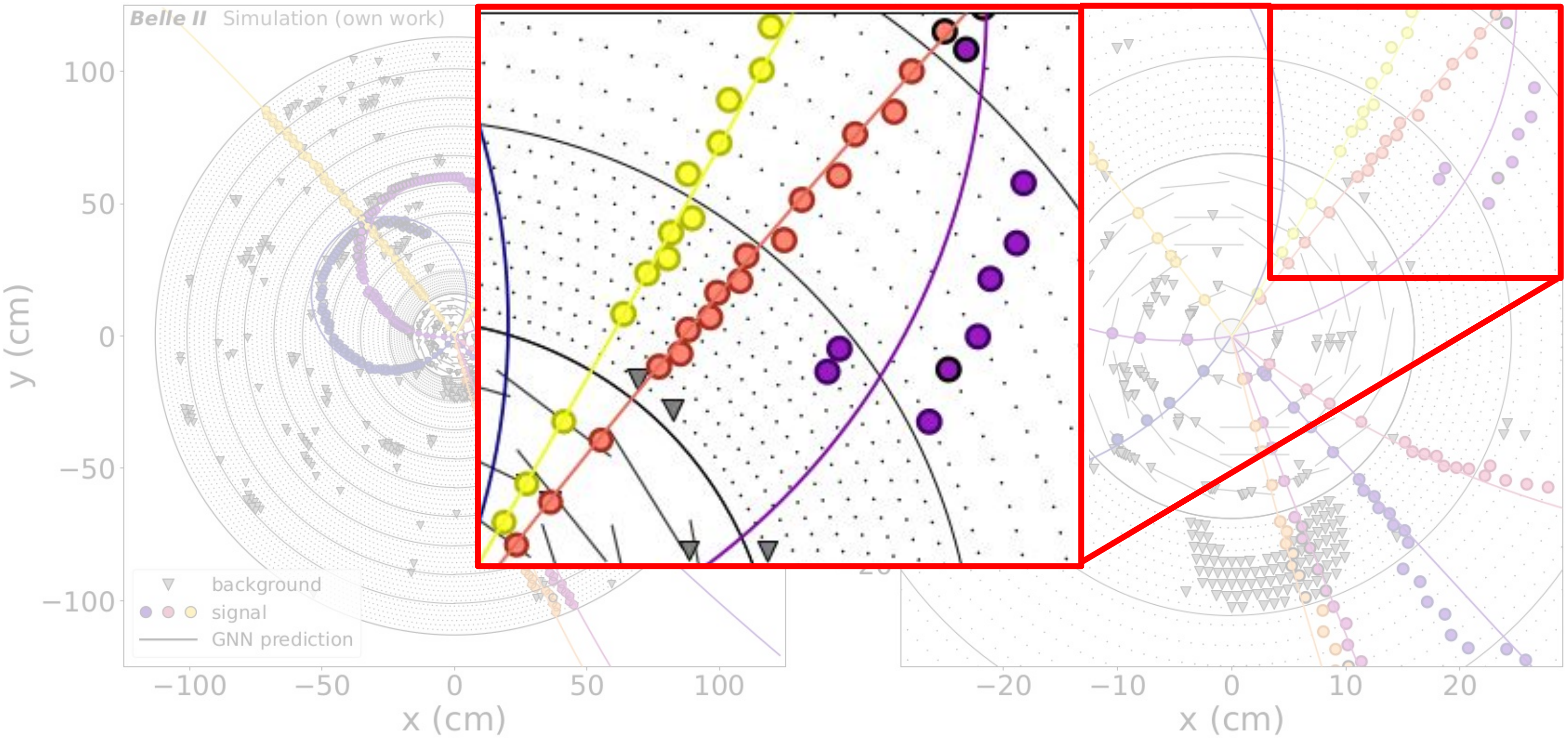
Model Prediction: x-y





- Different colors represent the track candidates and their associated track hits found by BATFinder (Via condensation point selection)
- Drawn Line shows path particle would take with predicted track momentum and starting position
  → should intersect with track hits

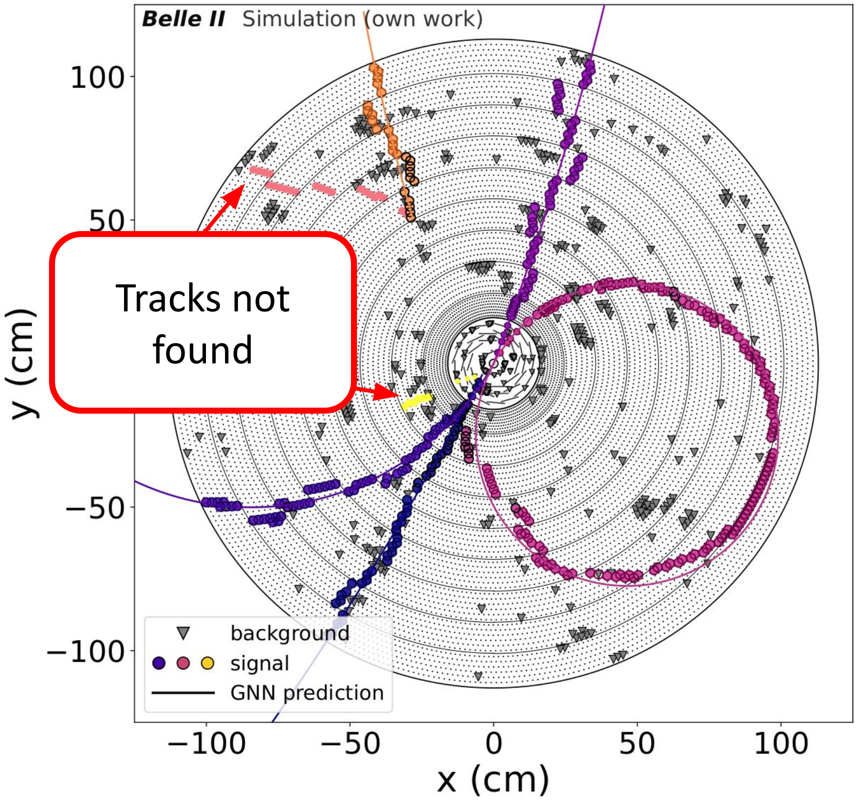
Model Prediction: x-y



- Different colors represent the track candidates and their associated track hits found by BATFinder (Via condensation point selection)
- Drawn Line shows path particle would take with predicted track momentum and starting position
  → should intersect with track hits
- Colored dots represent true track hits, darkened ring means BATFinder found the hit, black ring are not found signal hits

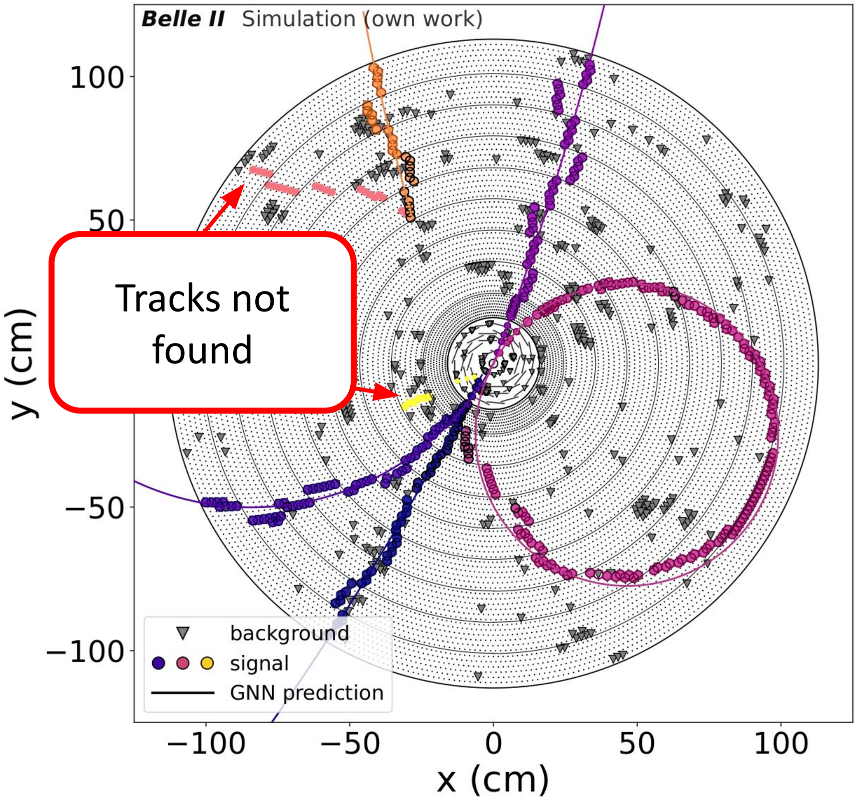
Metrics: Track Efficiency & Fake Rate

$$\text{Track Efficiency} = \frac{\text{true tracks}}{\text{all particles}} = \frac{5}{7}$$

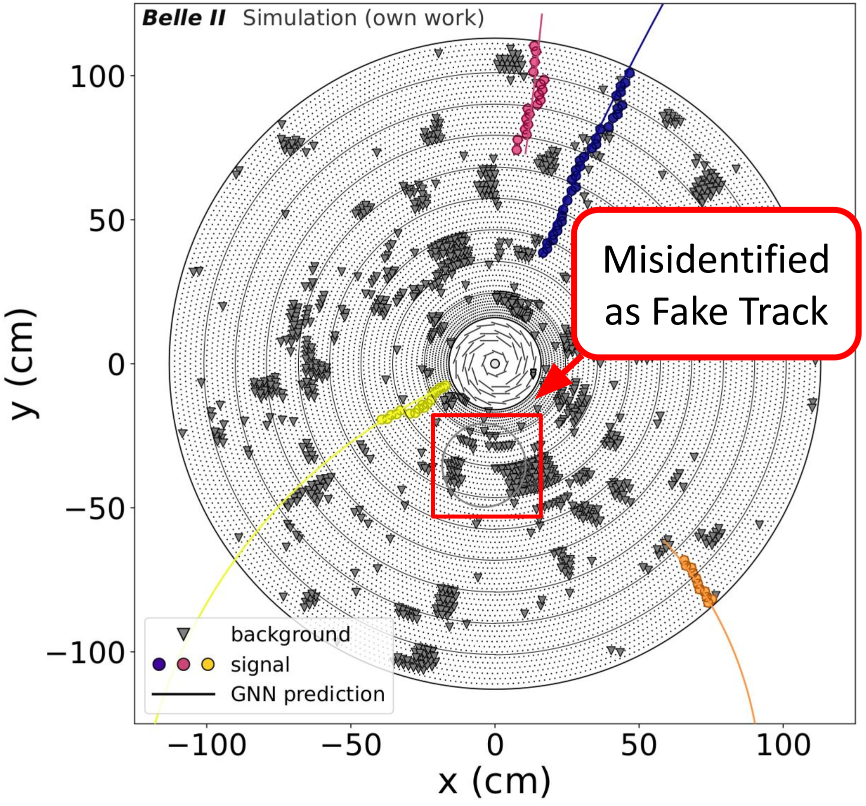


Metrics: Track Efficiency & Fake Rate

$$\text{Track Efficiency} = \frac{\text{true tracks}}{\text{all particles}} = \frac{5}{7}$$



$$\tau_{\text{fake}} = \frac{n_{\text{fake tracks}}}{n_{\text{tracks}}} = \frac{1}{5}$$

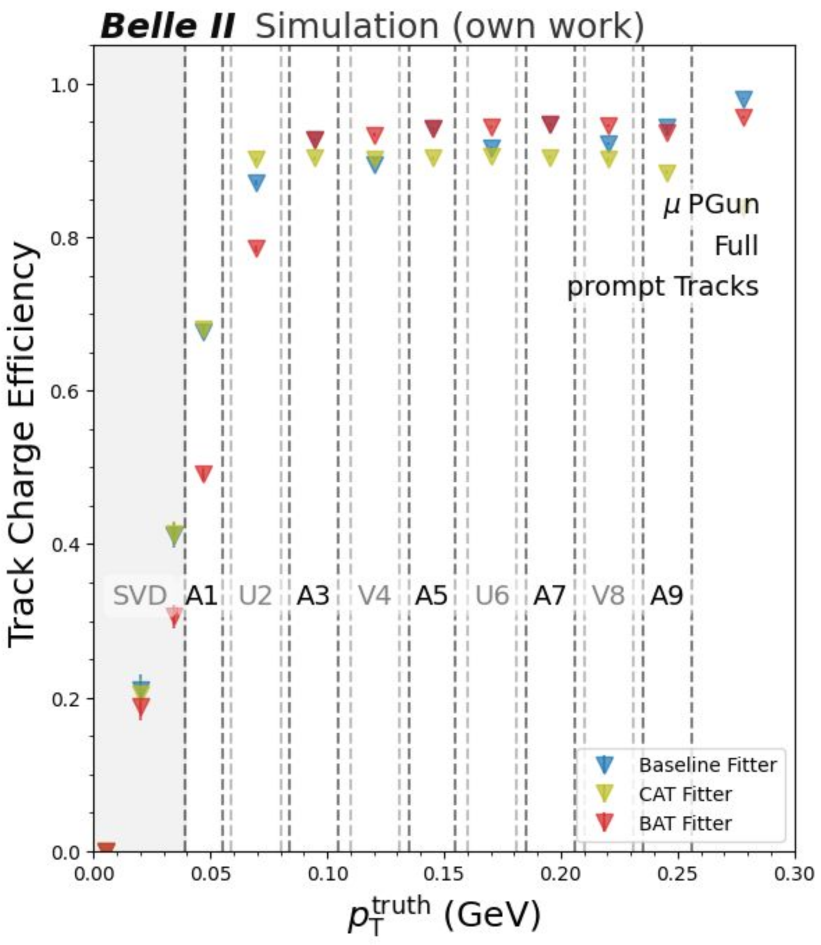


Metrics: Track Efficiency & Fake Rate

- We also require the tracks to have the correct charge prediction
  $\longrightarrow$ track charge efficiency
- High track charge efficiency and a low fake rate are desirable

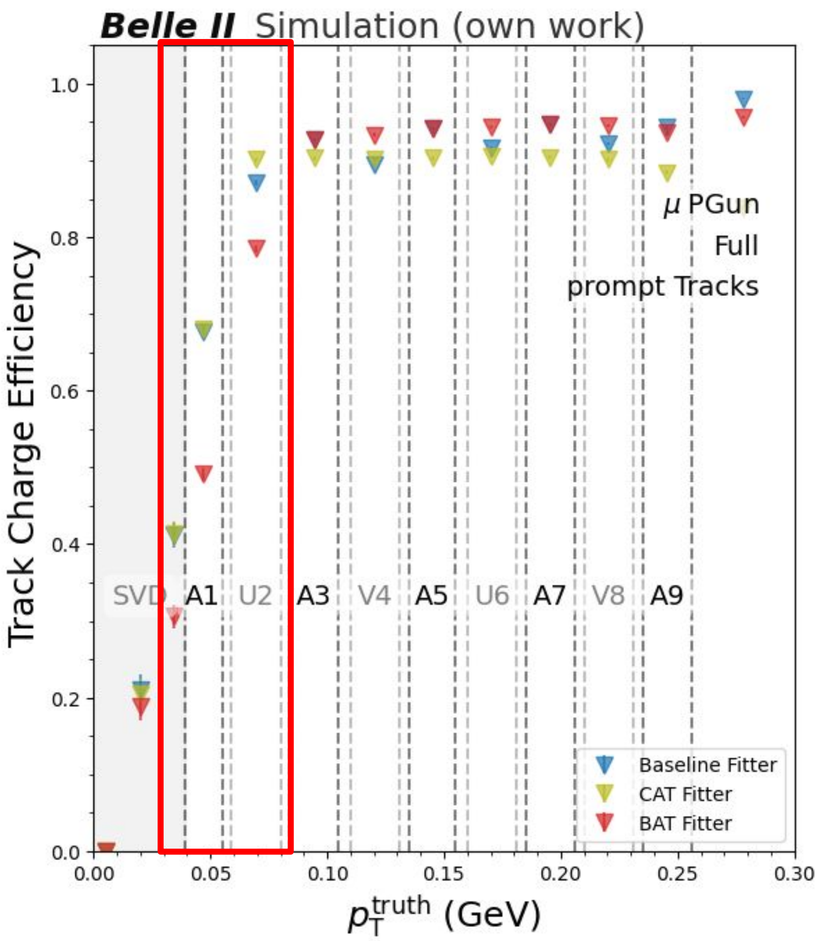
- Found track hits are fitted via the same track fitter that the Baseline employs (`prefilter_track_fit_and_track_creator`), we report the fitted track results

Prompt Track Charge Fitting Efficiency



- Bin names indicate how far a particle with given pt can travel through detector before curling back
- BATFinder (red) shows comparable track charge fitting efficiency performance with Baseline (blue) and CAT (yellow)

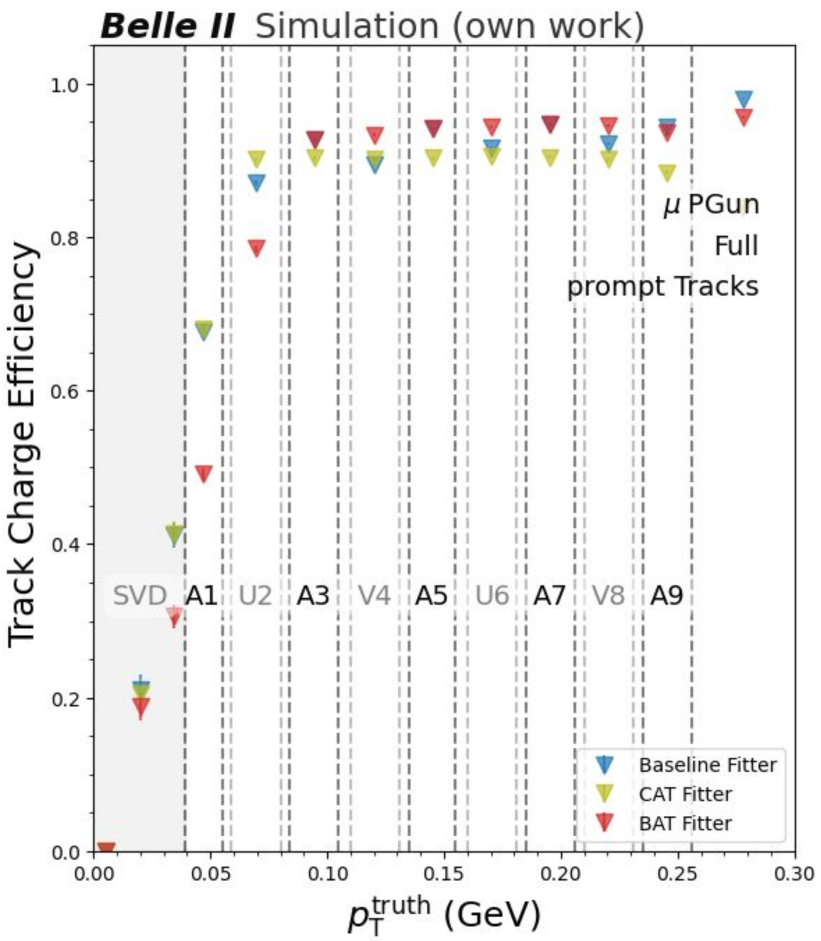
Prompt Track Charge Fitting Efficiency



- Bin names indicate how far a particle with given pt can travel through detector before curling back
- BATFinder (red) shows comparable track charge fitting efficiency performance with Baseline (blue) and CAT (yellow)

- Differences visible in last SVD and A1, U2 bins
  - Loss of 10 percentage points in A1 bin when comparing Baseline (68%) to BATFinder (58%)

Prompt Track Charge Fitting Efficiency

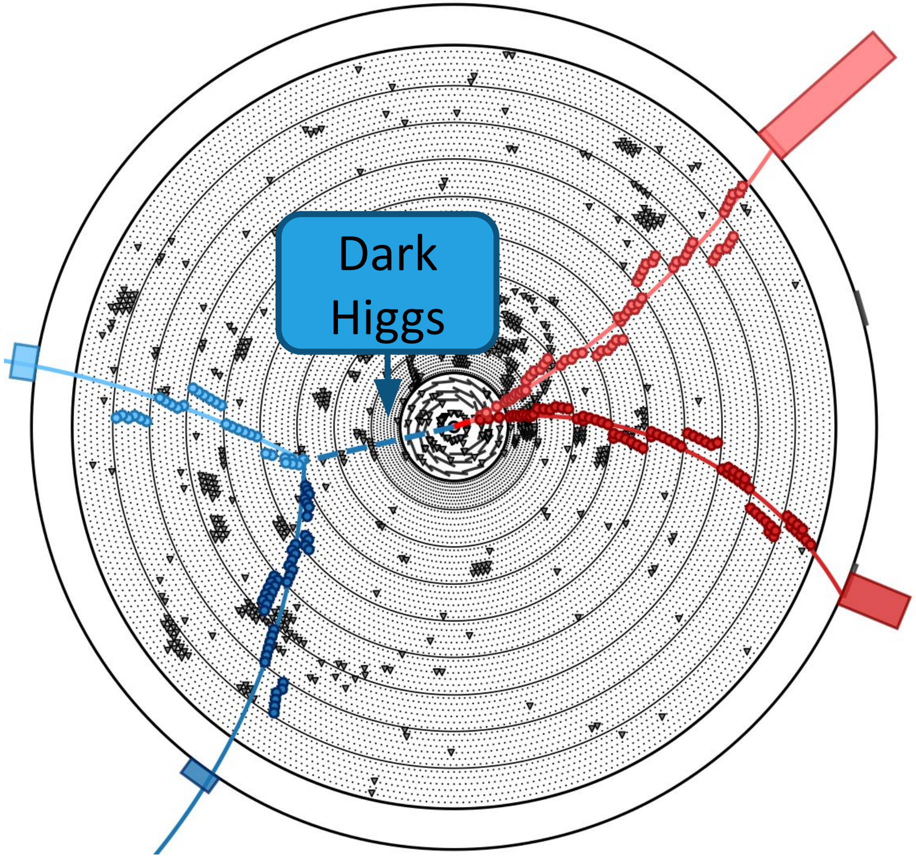


- Bin names indicate how far a particle with given pt can travel through detector before curling back
- BATFinder (red) shows comparable track charge fitting efficiency performance with Baseline (blue) and CAT (yellow)
- Differences visible in last SVD and A1, U2 bins
  - Loss of 10 percentage points in A1 bin when comparing Baseline (68%) to BATFinder (58%)
- BATFinder achieves a fake rate five times lower than Baseline: 0.21% vs 1.05%

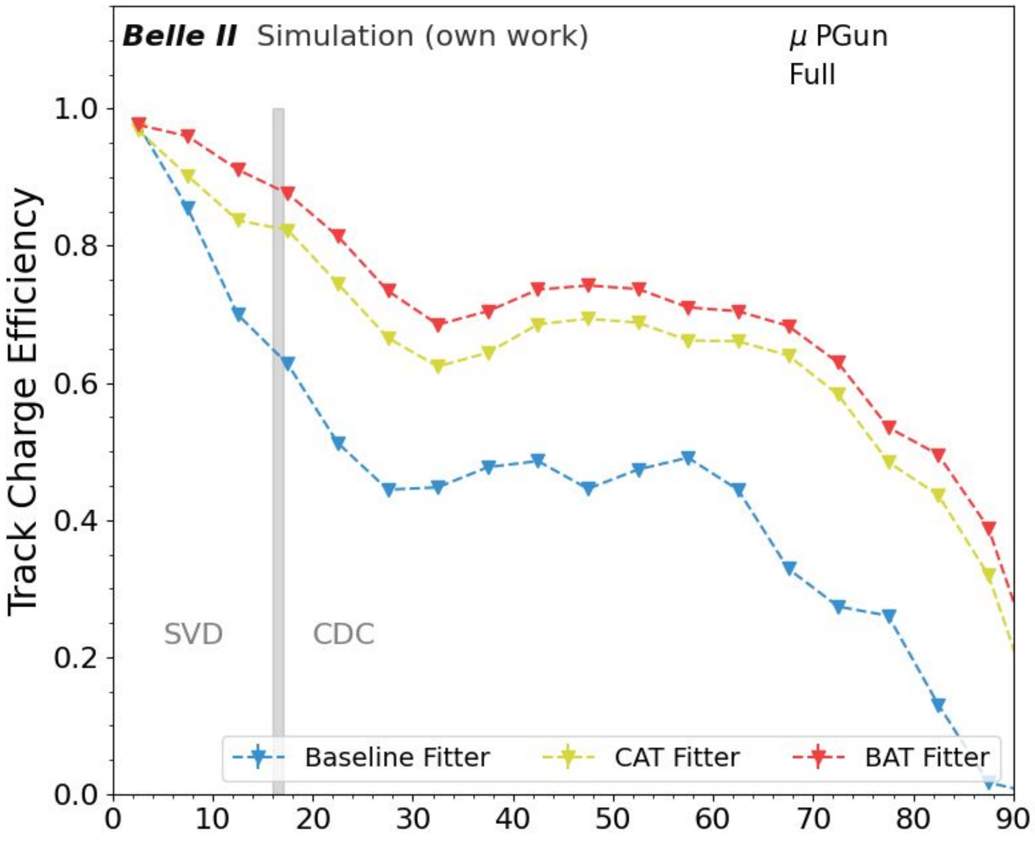
(in %)	t_{fake}	$\varepsilon_{\text{trk,ch}}$
Baseline Fitter	$1.05^{+0.01}_{-0.01}$	$97.82^{+0.01}_{-0.02}$
CAT Fitter	$1.83^{+0.02}_{-0.02}$	$96.92^{+0.02}_{-0.02}$
BAT Fitter	$0.21^{+0.01}_{-0.01}$	$97.70^{+0.02}_{-0.02}$

Performance on Displaced Tracks

Search for a dark Higgs boson produced in association with inelastic dark matter at the Belle II experiment (arXiv:2505.09705)

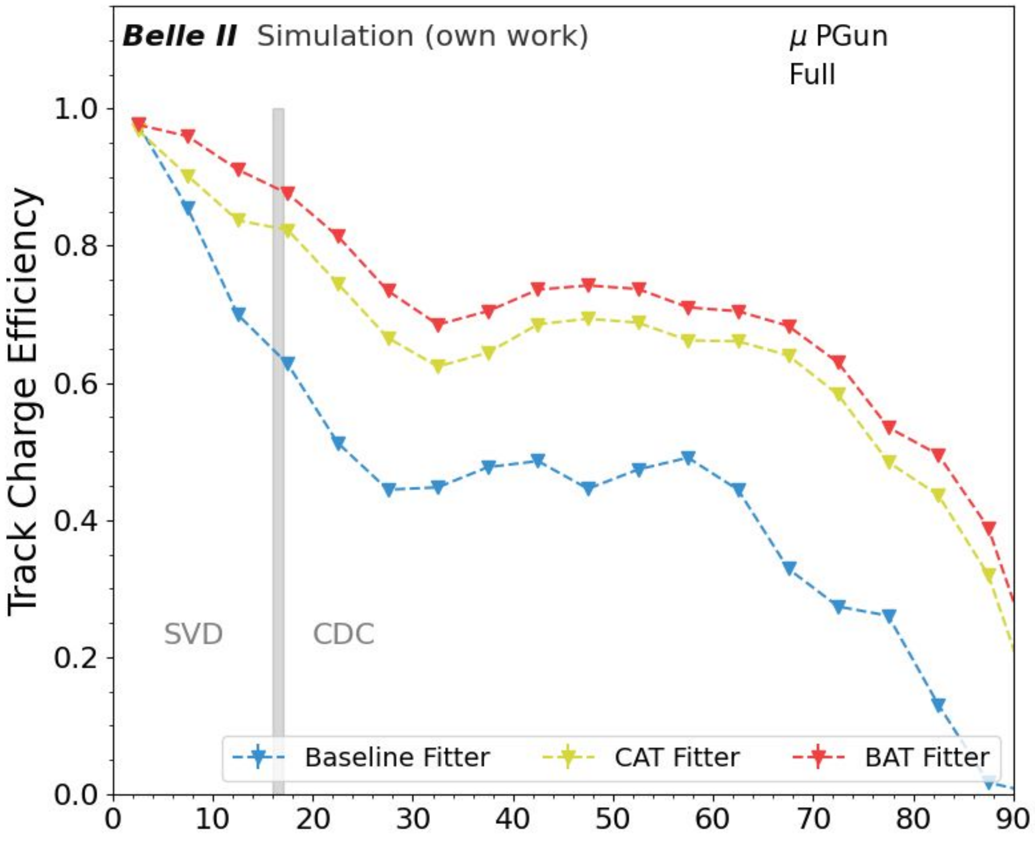


Performance on Displaced Tracks



Distance from IP to track start point

Displaced Track Charge Efficiency

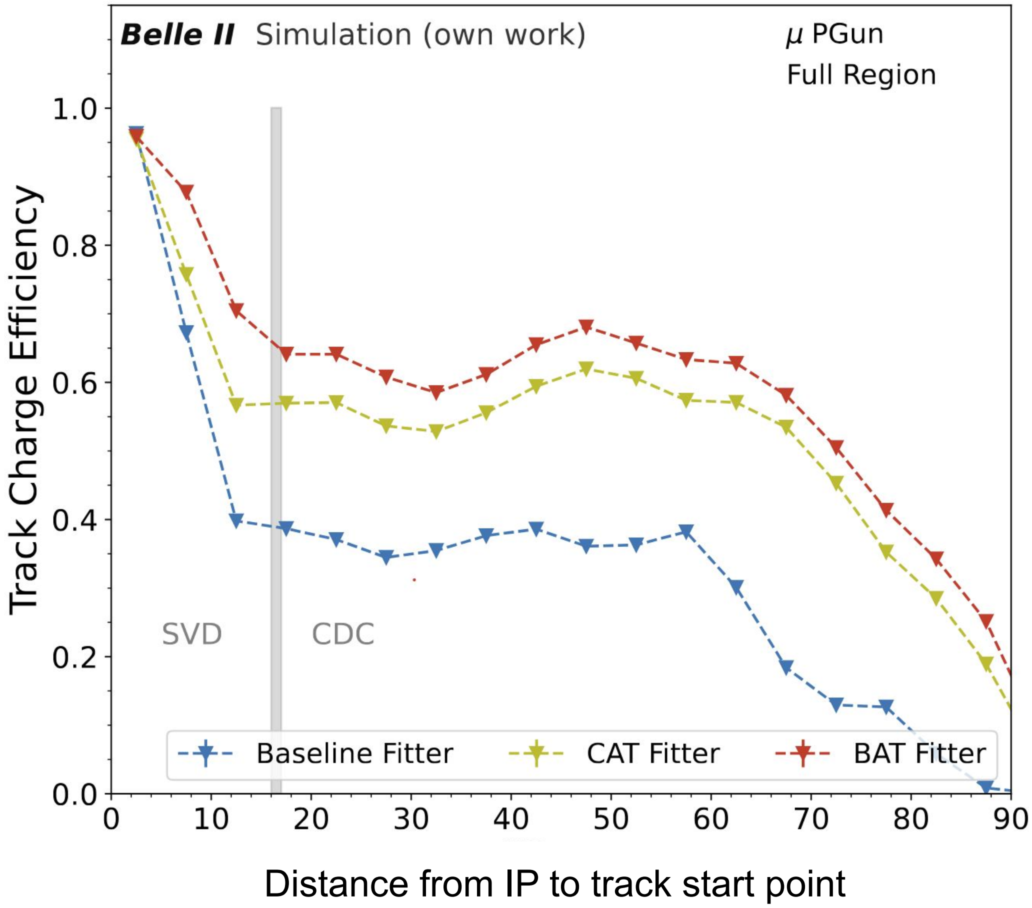


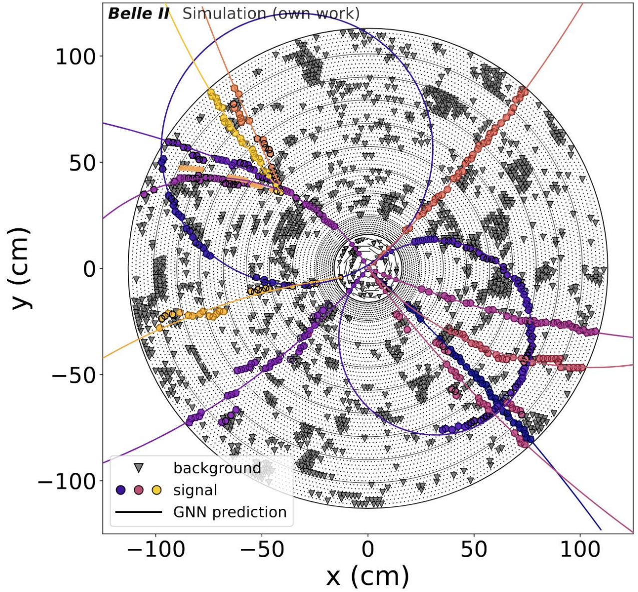
Distance from IP to track start point

- BATFinder achieves the same track charge fitting efficiency improvements for displaced CDC tracks as CATFinder (compared to the Baseline)
- BATFinder achieves a fake rate 20 times lower than Baseline: 0.29% vs 6.44%

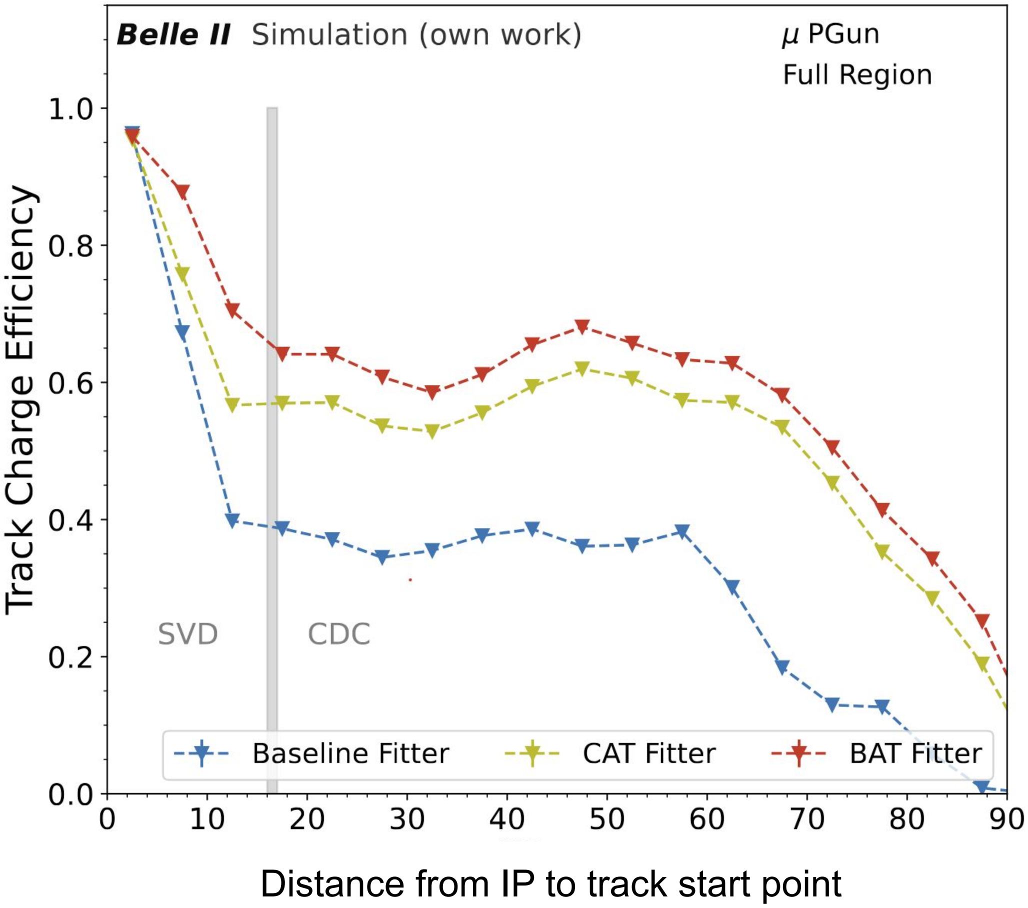
(in %)	t_{fake}	$\varepsilon_{\text{trk,ch}}$
Baseline Fitter	$6.44^{+0.04}_{-0.04}$	$49.73^{+0.06}_{-0.06}$
CAT Fitter	$4.62^{+0.03}_{-0.03}$	$69.35^{+0.05}_{-0.05}$
BAT Fitter	$0.29^{+0.01}_{-0.01}$	$75.13^{+0.05}_{-0.05}$

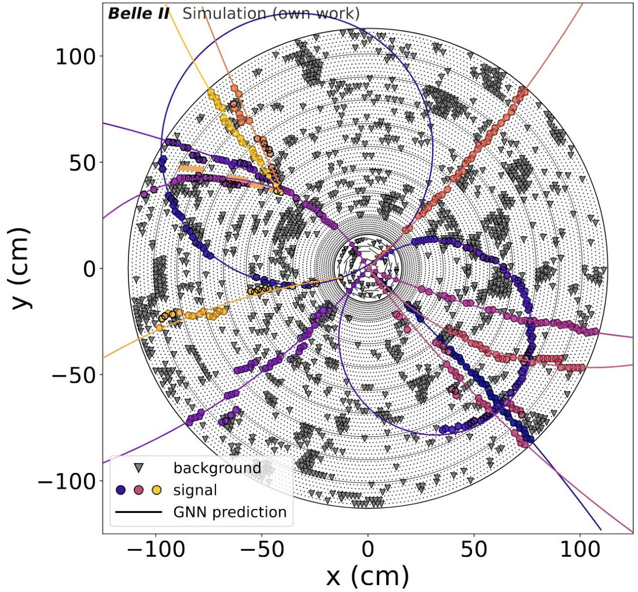
Displaced Track Charge Efficiency: Future Conditions





Displaced Track Charge Efficiency: Future Conditions





(in %)	t_{fake}	$\varepsilon_{\text{trk,ch}}$
Baseline Fitter	$8.28^{+0.10}_{-0.10}$	$35.46^{+0.11}_{-0.11}$
CAT Fitter	$5.49^{+0.07}_{-0.06}$	$54.98^{+0.12}_{-0.12}$
BAT Fitter	$0.65^{+0.03}_{-0.02}$	$62.28^{+0.11}_{-0.11}$

Summary and Outlook

- We have established that BATFinder achieves
  - Comparable performance to Baseline for prompt tracks
  - Surpasses performance of Baseline for displaced tracks (from 49% to 75% track charge efficiency)

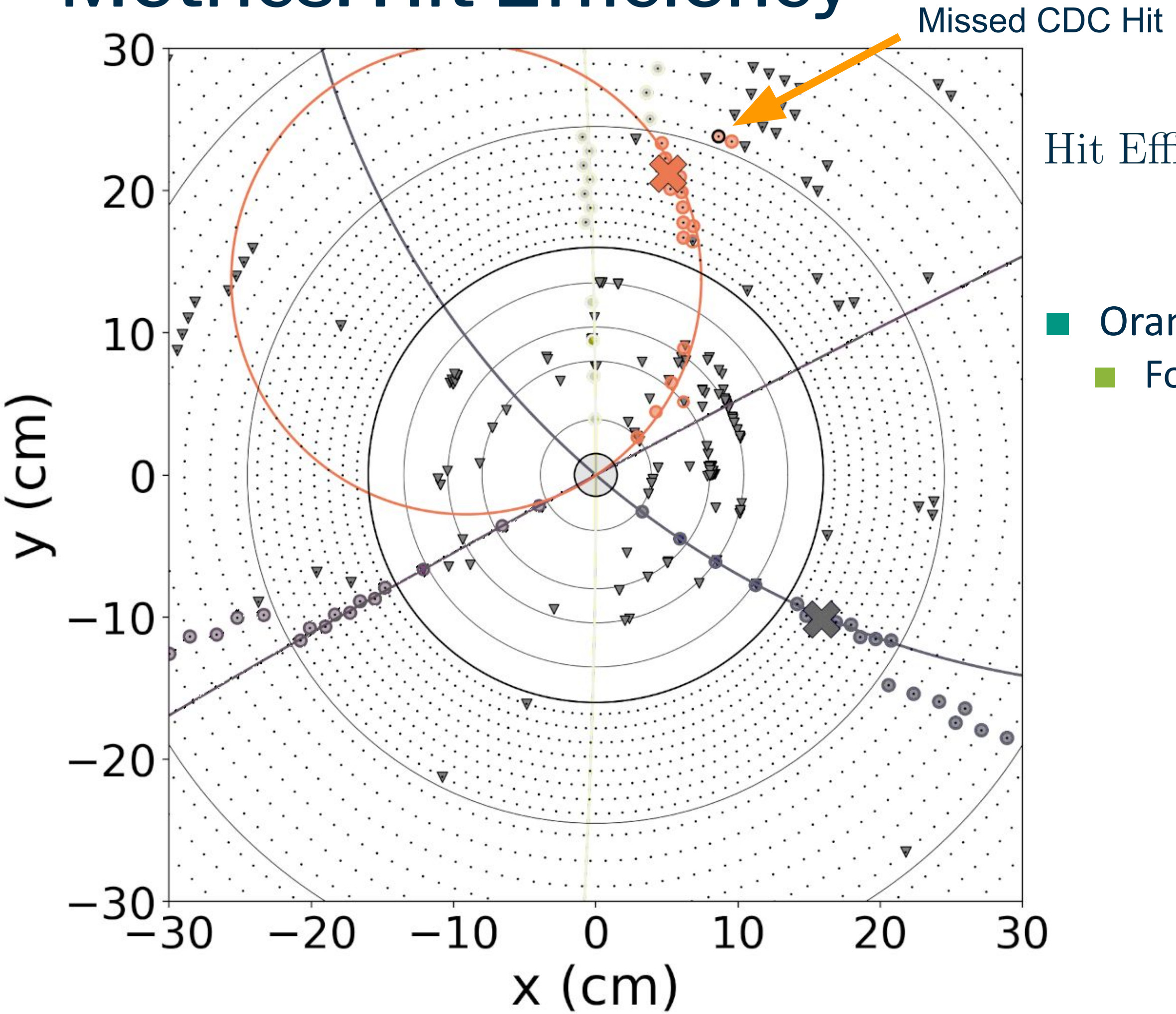
$\rightarrow$ Proof-of-concept for GNN-based object condensation tracking on SVD and CDC

Next steps

- Evaluate on physics samples, currently examining in $B^0\bar{B}^0$
  - Prompt π
  - π from $K_S^0 \rightarrow \pi^+\pi^-$
- Test robustness against future CDC conditions
  - Higher Beam Background levels
  - Masked readout boards and wire inefficiencies
- Investigate Track Fitting procedure

Backup

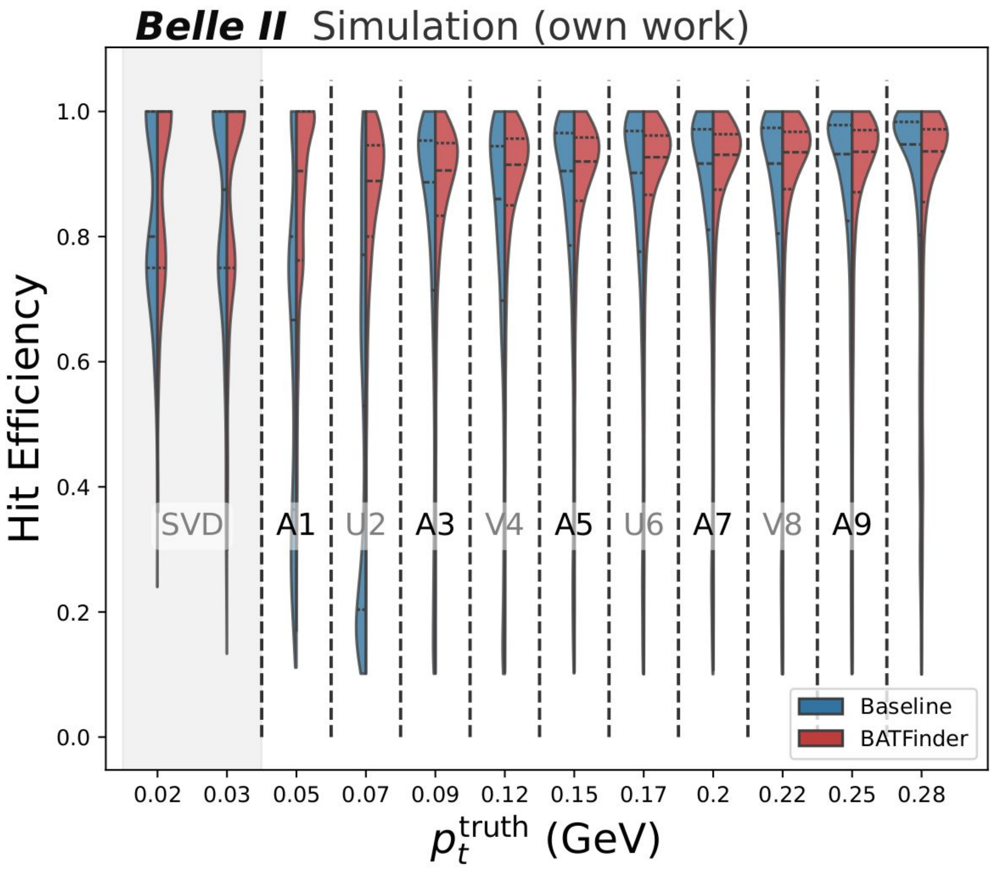
Metrics: Hit Efficiency



$$\text{Hit Efficiency} = \frac{\text{found track hits by particle}}{\text{all hits by particle in SVD and CDC}}$$

- Orange Track
  - Found 15/16 total hits → Hit Efficiency 94%

Hit Efficiency

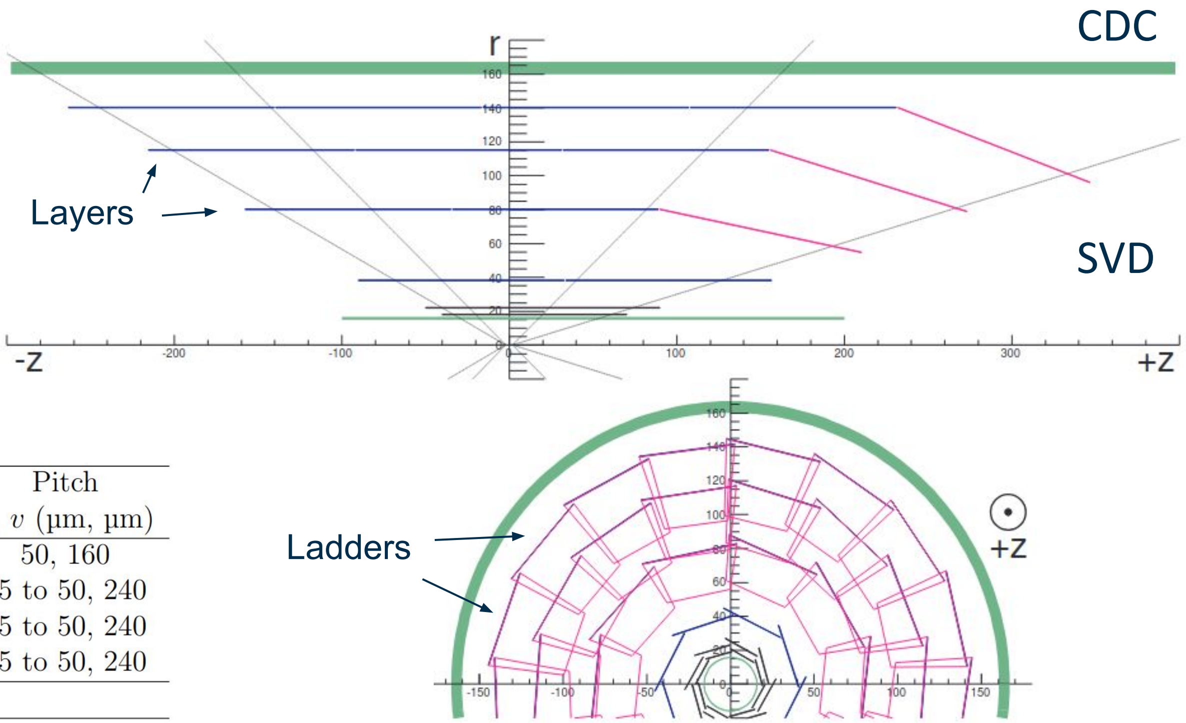


- Comparable tit efficiency for tracks that only reach SVD
- Strong hit efficiency gain for prompt tracks reaching only the first CDC layers (A1, U2)

  - BATFinder: attaches SVD + CDC hits jointly in one inference step
  - Baseline: After initial CDC trackfinding (where hits may get missed), extrapolates SVD only tracks outward to CDC and tries to attach CDCHits belonging to track from SVD → Seems to work suboptimally for first CDC layers

Silicon Vertex Detector

- Four-layer double-sided silicon-strip detector
- Strips on opposite sides perpendicular to one another



- SVD specifications:

Layer	Radius (mm)	Ladders	Sensors	Strips/Sensor u, v	Pitch u, v (µm, µm)
3	39	7	14	768, 768	50, 160
4	80	10	30	768, 512	75 to 50, 240
5	104	12	48	768, 512	75 to 50, 240
6	135	16	80	768, 512	75 to 50, 240
Sum		35	172	132 096, 91 648	

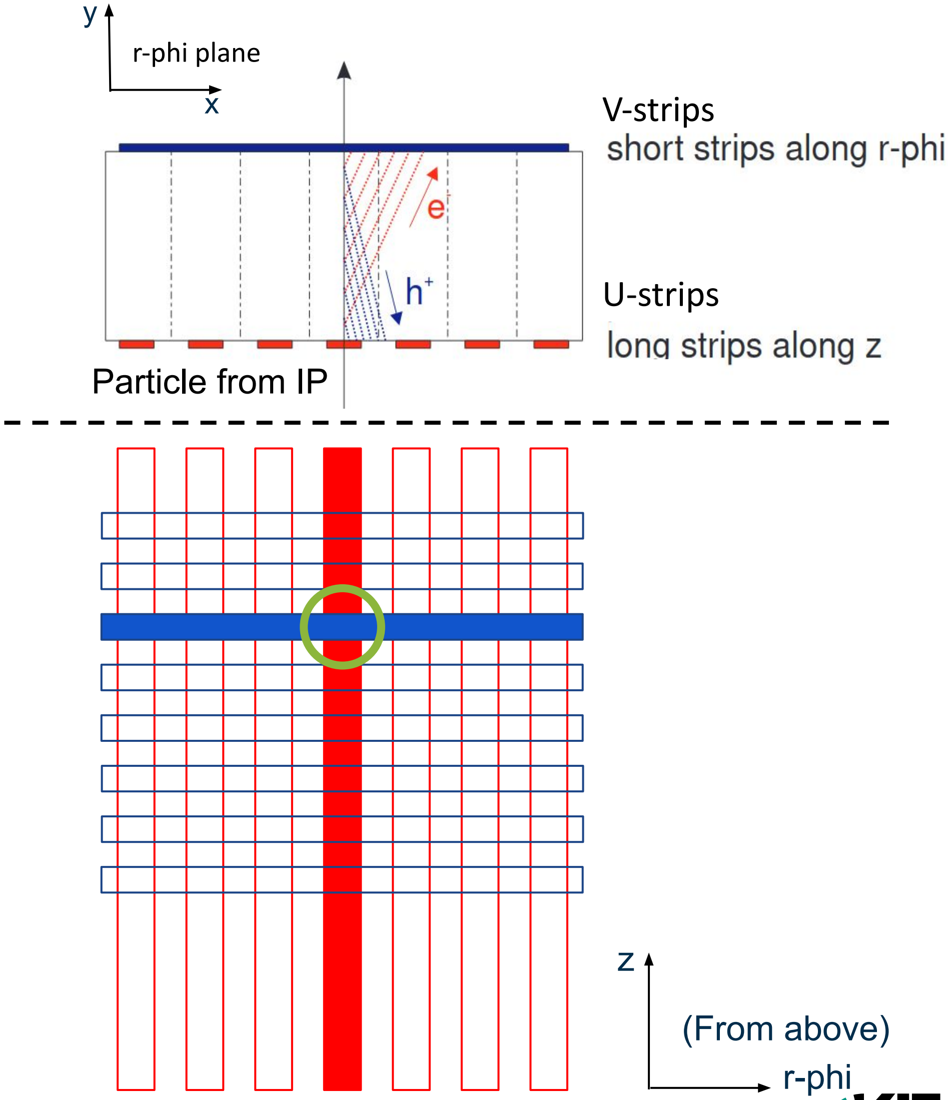
SVD Data Taking

- Charged particle traverses detector, ionizes material
- Electrons / holes drift to V / U side of Ladder

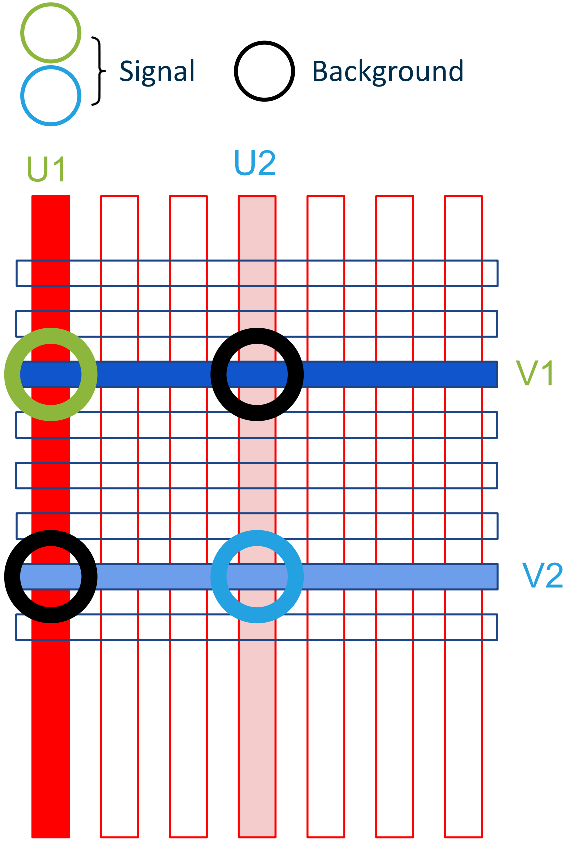
- Cluster neighboring strips on either side based on deposited charge
- Size between 1 and 3 strips

- Combine clusters from both sides to get 3D SpacePoints (SP)

- Example:
  - Two clusters, one U (measures r-phi) and one V (measures z) consisting of one strip form a SpacePoint at intersection



Approach: SpacePoints



- For SP creation, all U clusters are combined with all V clusters
- Example: 1st particle creates clusters U1 and V1, 2nd particle U2 and V2 → Results in 4 SP
- Leads to SP not having definitive labels (e.g. U-cluster hit is from bkg and V-cluster hit is from signal)
- For the training on SP, require that associated clusters to signal SP are related to the same MCParticle, otherwise assign background flag

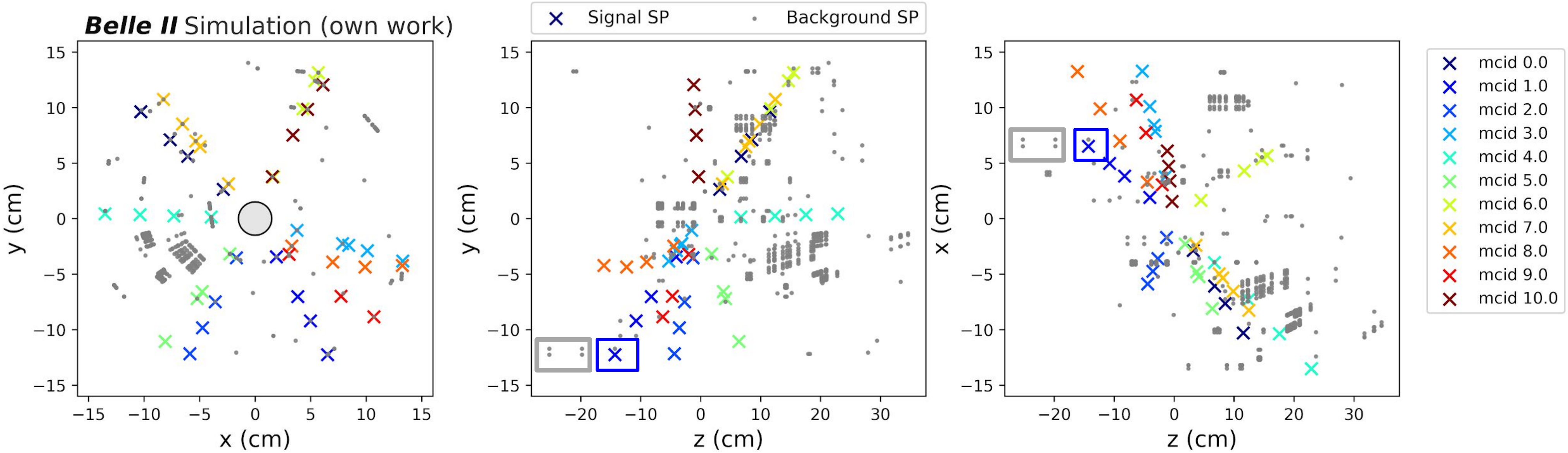
Approach: SpacePoint matching

Signal: MC ID 1

Not Signal: MC ID -9999

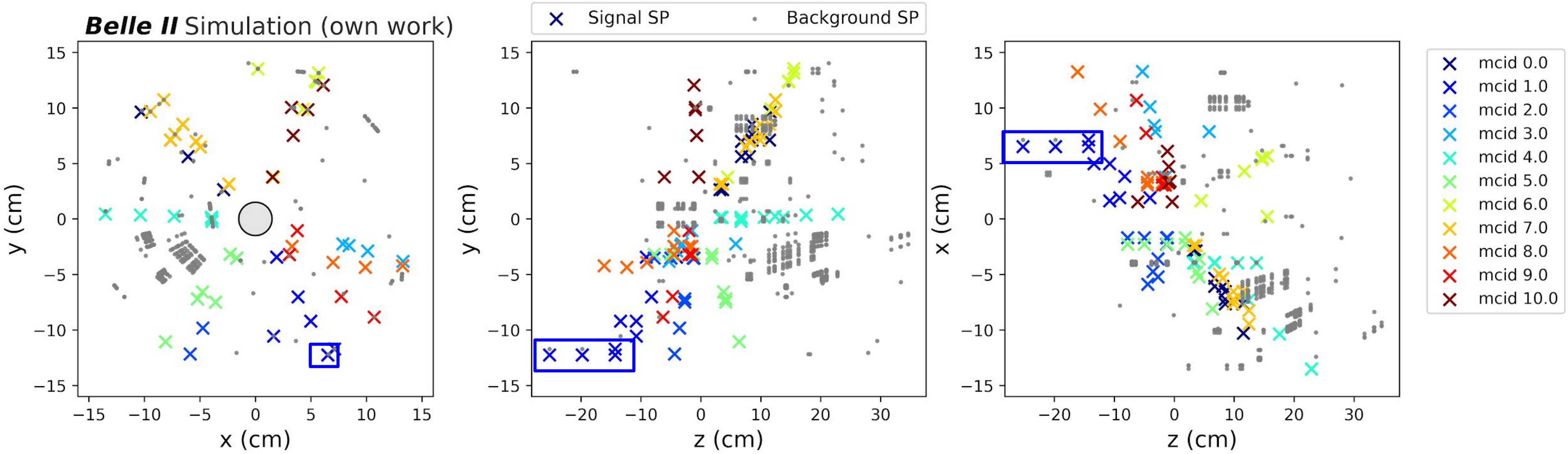
- Solution: We require that associated clusters to signal SP are related to the same MCParticle, otherwise assign background flag

Approach: SpacePoints matched to mc id



- Provides unambiguous truth for training, circled sections are example track belonging to previously ambiguous SP (id 1) which can now be clearly identified

Approach: SpacePoints related to mc id



- Projections of distribution of all SP we obtain from svd reconstruction
- Different colors represent SpacePoints related to MCParticles, Grey is background
- Ambiguous SP eg. same x-y, different z for dark blue MCParticle (id 1 in projections 2 & 3) ⟶ MCParticle appears to hit last layer in three different locations, we cannot use this as truth for SATFinder training

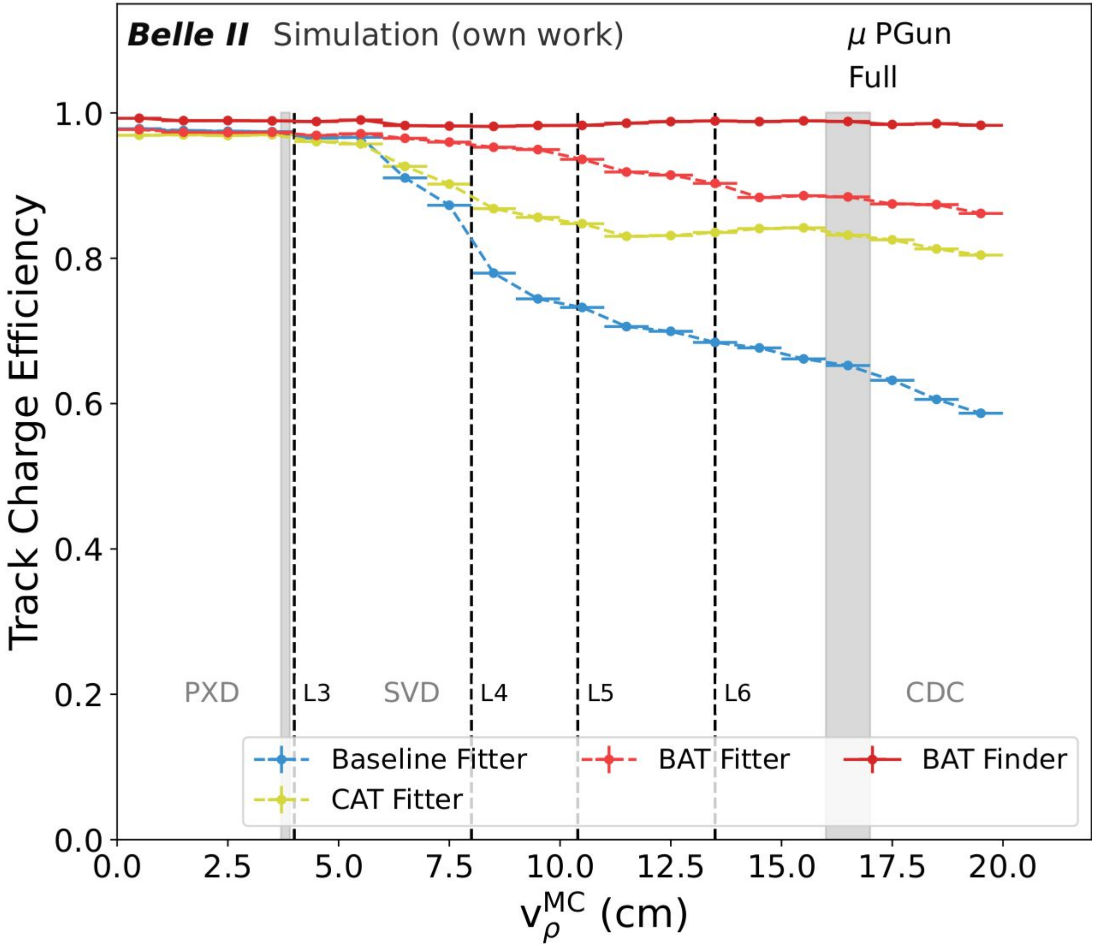
Training Sample Topology

- Sample very similar to CATFinder (Link to published paper)
- I simulated ~250k events in total from category 1-6 and category 8, containing up to 9 tracks

category	name	p_t or p [GeV]	θ_{gen} [°]	r^{3D}_{gen} [cm]	α^{3D}_{gen} [°]
1	prompt fwd	0.05 – 6.0	17.0 – 35.4	0	0
2	prompt brl	0.05 – 6.0	35.4 – 123.04	0	0
3	prompt bwd	0.05 – 6.0	123.04 – 150.0	0	0
4	displaced fwd	0.05 – 6.0	17.0 – 35.4	0 – 100	0
5	displaced brl	0.05 – 6.0	35.4 – 123.04	0 – 100	0
6	displaced bwd	0.05 – 6.0	123.04 – 150.0	0 – 100	0
7	displaced angled	0.05 – 6.0	17.0 – 150.0	0 – 100	0 – 30
8	prompt full	0.05 – 6.0	17.0 – 150.0	0 – 100	0
9	vertex large	0.05 – 6.0	17.0 – 150.0	0 – 100	0 – 90
10	vertex small	0.05 – 6.0	17.0 – 150.0	0 – 100	0 – 25
11	mix 8+10	-	-	-	-

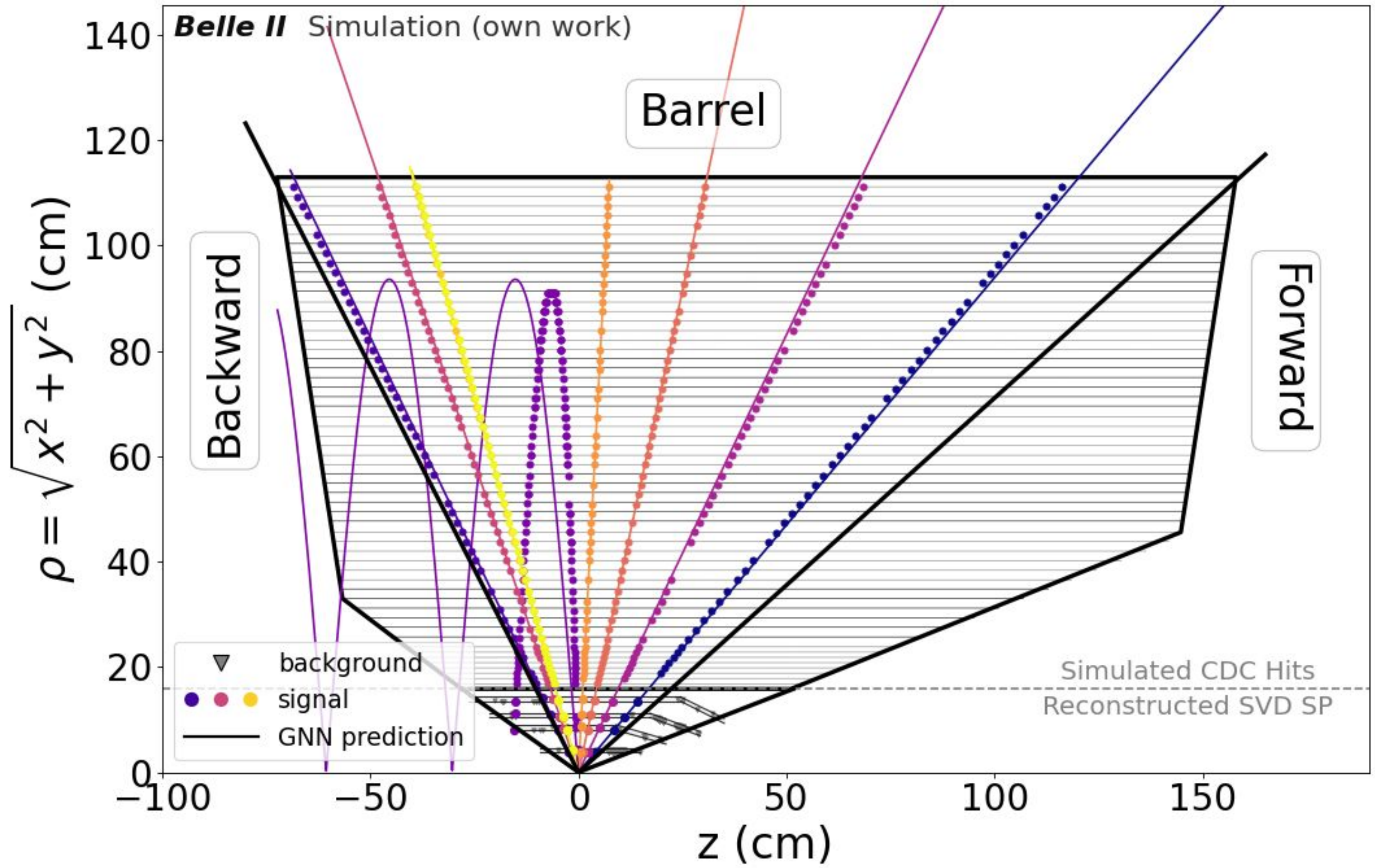
Table taken from CATFinder paper

Displaced within SVD tracks



- BAT Finder achieves >90% track charge fitting efficiency for displaced tracks within the SVD

Display: rho-z



- CDC hits are CDCSimHits
- SVD hits are reconstructed hits