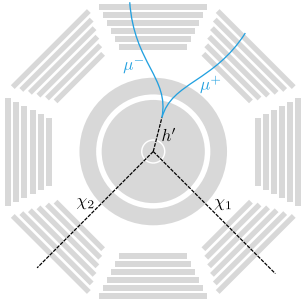
End-to-End Multi-Track Reconstruction using Graph Neural Networks in the CDC

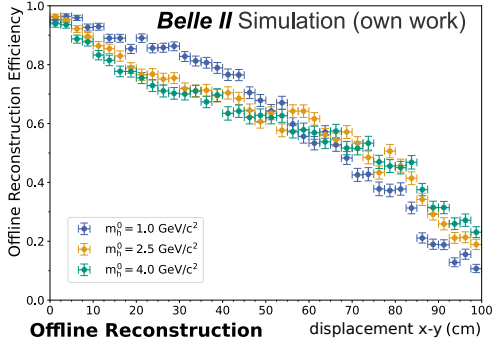
Belle II Germany Meeting 2024

Lea Reuter, Giacomo De Pietro, Torben Ferber, Slavomira Stefkova | 1st October 2024

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)





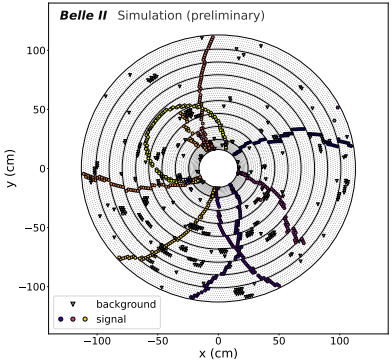
Efficiency decreases depending on displacement (K_S^0, Λ^0, Dark Sector searches)

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

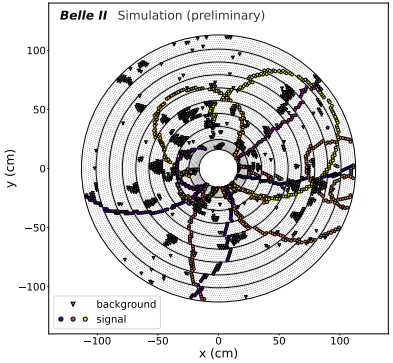
[2] Search for Inelastic Dark Matter produced in association with a Dark Higgs (BELLE2-NOTE-PH-2024-019)

Motivation - High Backgrounds

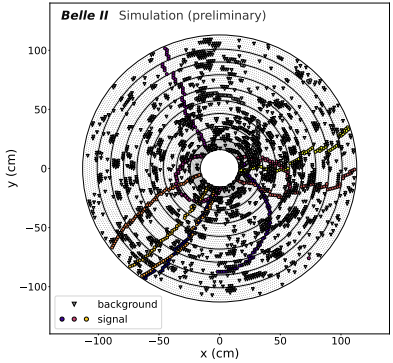
Exp. 1003 backgrounds
360 background CDC hits on average



Exp 26, run 1485 backgrounds
1280 background CDC hits on average



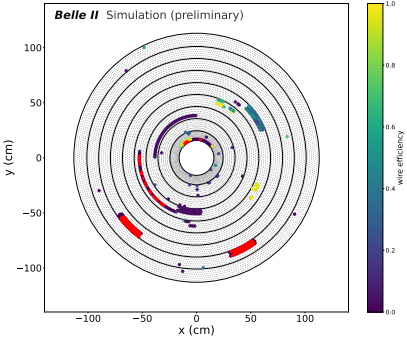
Exp. 0 backgrounds
3000 background CDC hits on average



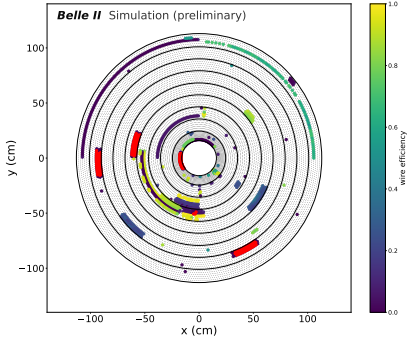
→ Backgrounds getting higher, harder for tracking

Motivation - CDC Wire Inefficiencies[3]

Exp. 22, run 26,
globaltag: `data_reprocessing_prompt`
50 wires off, 368 decreased efficiency,
total of 3% of the CDC wires



Exp 26, run 1485,
globaltag: `data_reprocessing_prompt`
168 wires off, 809 had efficiency,
total of 7% of the CDC wires



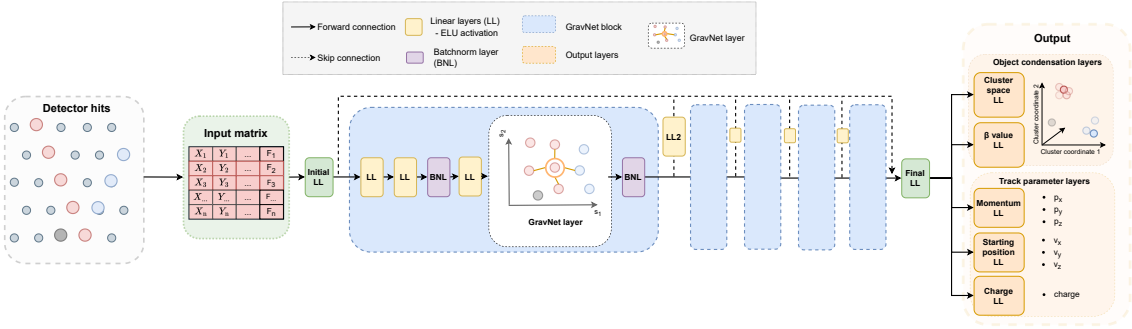
[3]Content of the CDCBadWires payload

Approach

- Find track parameters: momentum, starting position and charge
- Computing resource and time constraint → Graph Neural Networks
- Find unknown number of tracks → Object Condensation (arXiv:2002.03605)

Approach

- Find track parameters: momentum, starting position and charge
- Computing resource and time constraint → Graph Neural Networks
- Find unknown number of tracks → Object Condensation (arXiv:2002.03605)



Inputs:

- x and y wire position
- tdc and adc signal information
- layer, superlayer, and clayer

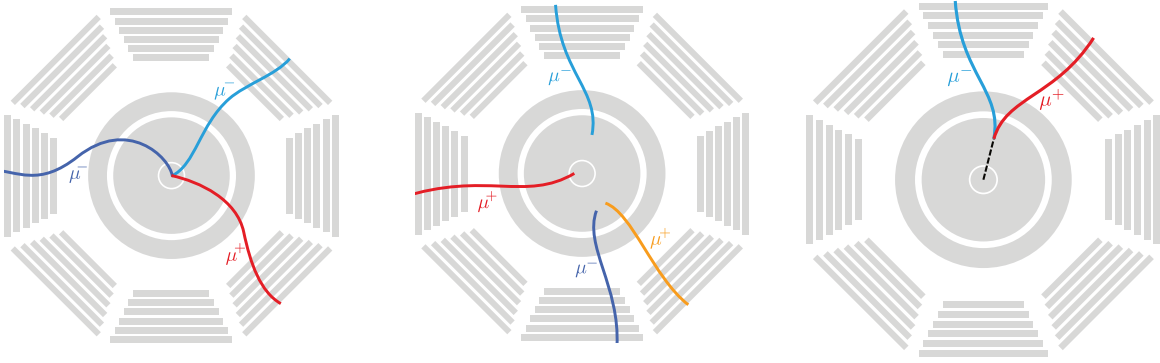
Adjustable Parameters result in a total of 797,812 trainable parameters (3MB weightfiles)

Technical setup

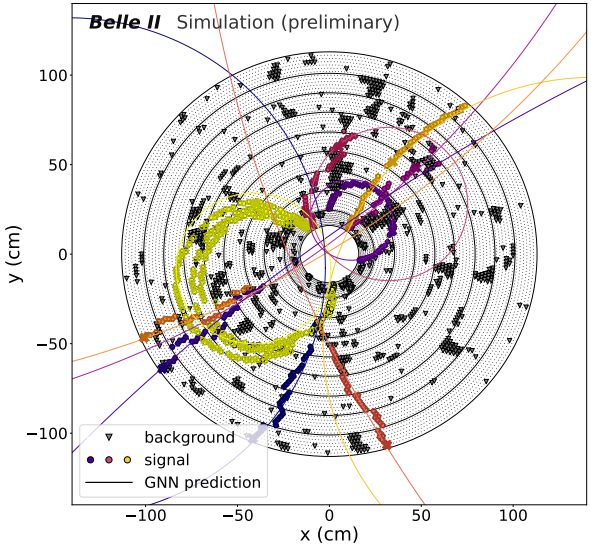
- release : `release-06-00-08`
- run-dependent beam-background files from **exp 26, bucket 36**, runs 1423, 1430, 1438, 1446, 1449, 1485, 1497, 1518, 1519, 1750, 1816, 1824, 1832, 1894, and 1895
  for a total of 1.7 million run-dependent beam-background overlays with similar high number of extra CDC hits
  - globaltag: `mc_production_MC15rd_a_exp26_bucket36`
  - globaltag: `data_reprocessing_prompt`
  - globaltag: `online`
- Comparison for **CDC-only** Tracking!
  Baseline = (`add_cdc_track_finding(with_ca=False)`)

Training Sample Topologies

- Simulate 1 million events with over 4 million tracks with 80/20 split for train/validation
- Train on different sample topologies that cover a large phase space, to not bias the model, **no conservation laws involved here!**
  $\rightarrow$ crucial step to be agnostic about the physics processes



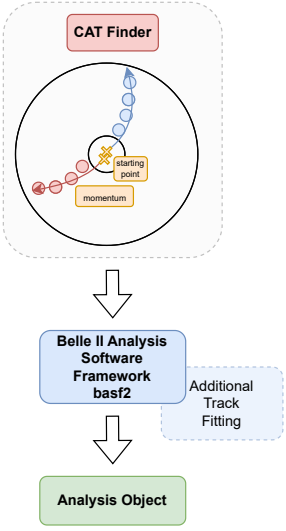
Model Prediction



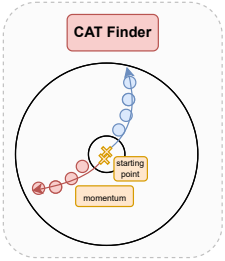
GNN prediction is drawn according to the track parameters predicted by the GNN:

- momentum p_x, p_y, p_z
- starting position v_x, v_y, v_z
- charge

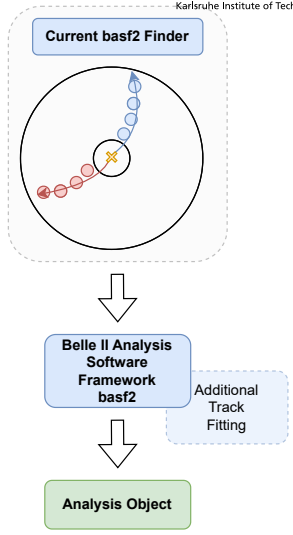
New Track Finder: CDC AI Track (CAT) Finder



New Track Finder: CDC AI Track (CAT) Finder



Fair comparison for track finding algorithms on analysis level for CDC only



Extensive studies and validation

Karlsruhe Institute of Technology

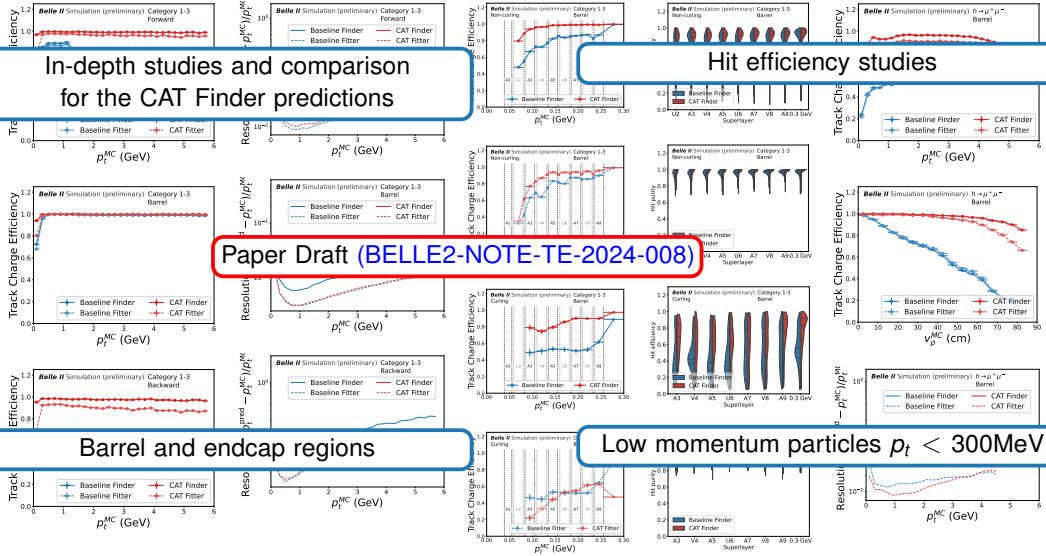
In-depth studies and comparison for the CAT Finder predictions

Hit efficiency studies

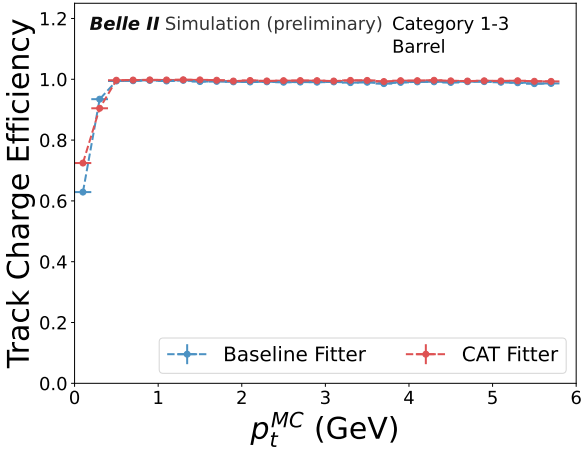
Paper Draft (BELLE2-NOTE-TE-2024-008)

Barrel and endcap regions

Low momentum particles $p_t < 300$MeV



Performance on prompt particles for CDC only



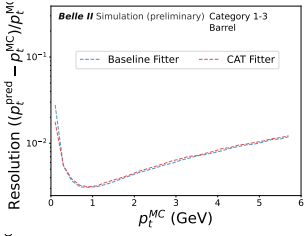
1-12 μ^+ and μ^- per event,
run dependent beam-backgrounds from experiment 26,
barrel acceptance ($35.4° < \theta < 123°$)

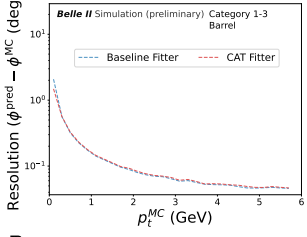
integrated	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$
Baseline	$94.15^{+0.06}_{-0.06}$	$1.83^{+0.04}_{-0.04}$	$0.9^{+0.03}_{-0.03}$
CAT	$95.11^{+0.06}_{-0.06}$	$1.25^{+0.03}_{-0.03}$	$0.34^{+0.02}_{-0.02}$

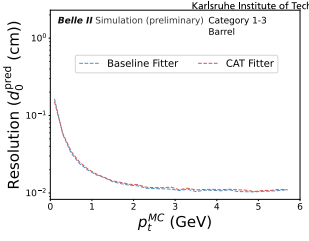
$$\varepsilon_{\text{track}} = \frac{n_{\text{tracks}}(\text{matched to particle and charge})}{n_{\text{simulated}}(\geq 1\text{matched hit})}$$

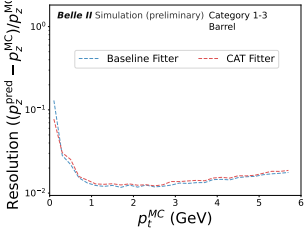
Extensive studies shown in paper draft (BELLE2-NOTE-TE-2024-008)

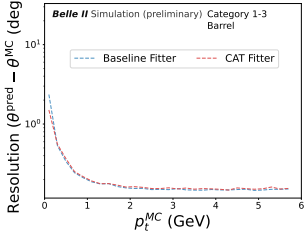
Resolution after Fitting for CDC only

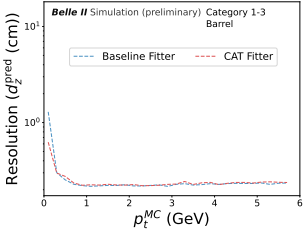






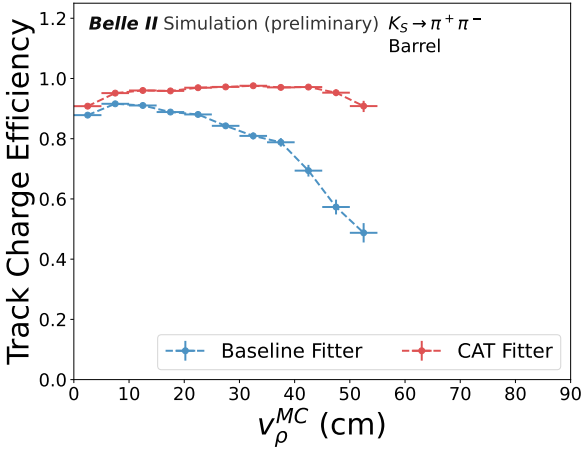






The resolution for the track parameters of the CAT Finder is worse than for the Baseline, but it is sufficient as a seed for the subsequent fitter as shown above.

Performance on displaced particles for CDC only



Particle gun $K_s^0 \to \pi^+\pi^-$, $p_t(K_s^0) = [0.05, 3]$GeV
run dependent beam-backgrounds from experiment 26,
barrel acceptance

integrated	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$
Baseline	$88.1^{+0.1}_{-0.1}$	$5.39^{+0.07}_{-0.07}$	$0.59^{+0.02}_{-0.03}$
CAT	$92.99^{+0.08}_{-0.08}$	$5.13^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$

$$\varepsilon_{\text{track}} = \frac{n_{\text{tracks}}(\text{matched to particle and charge})}{n_{\text{simulated}}(\geq 1\text{matched hit})}$$

Extensive studies shown in paper draft (BELLE2-NOTE-TE-2024-008)

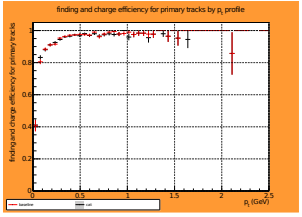
Current Status: Implementation in Full Reconstruction

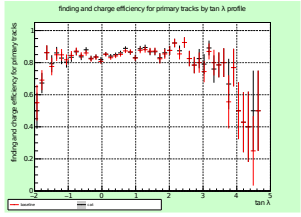
b2validation, `release-08-01-08`, exp. 1003, Full Tracking Validation **including VXD**

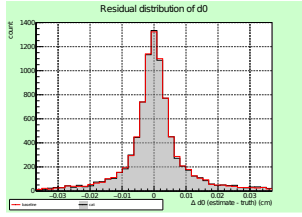
`add_reconstruction(path, useCAT=True/False)`

Extrapolation to the inner tracking detectors work fine, compatible results

CAT Finder vs **Baseline**





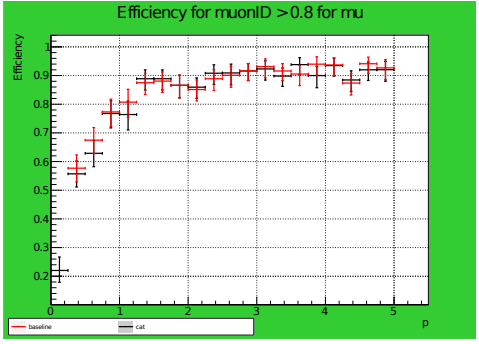


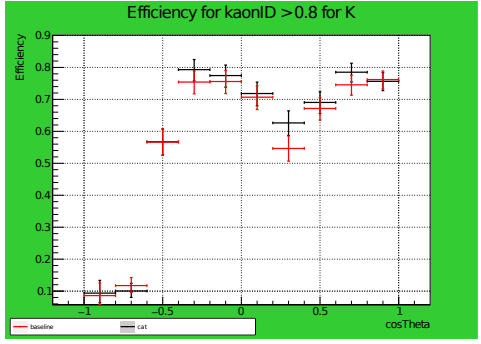
Tag	finding charge efficiency	charge efficiency	charge asymmetry	fake rate	clone rate
Baseline	92.94	98.97	1.47	3.16	3.48
CAT	92.95	99.65	1.30	3.76	5.21

Current Status: Checking PID

b2validation, `release-08-01-08`, exp. 1003,
`add_reconstruction(path, useCAT=True/False)`

Extrapolation to outer detectors (KLM, TOP) works, compatible results
CAT Finder vs **Baseline**

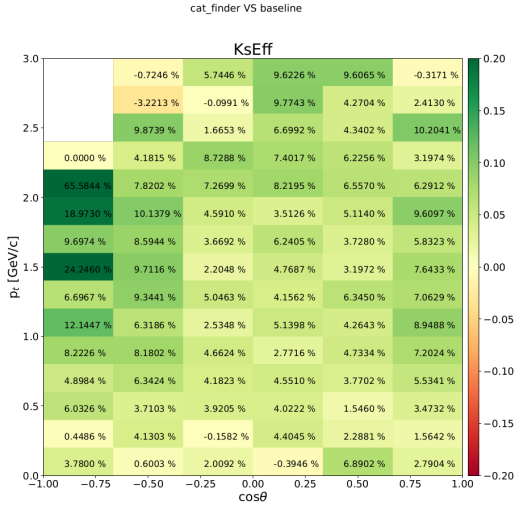




Current Status: Full Tracking Performance for K_s^0

Karlsruhe Institute of Technology

CAT Finder vs **Baseline**



Full Tracking Validation **including VXD**
`add_reconstruction(path, useCAT=True/False)`
`release-08-01-08`, exp. 1003,
Generated samples with Data Production Validation and evaluated with VIBE and tracking validation

Decay $e^+e^- \to c\bar{c}$,
$D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi^+$

Tag	K_S^0 eff.	K_S^0 fake rate	$\pi(K_S^0)$ eff.
Baseline	$72.9^{+0.1}_{-0.1}$	$5.71^{+0.04}_{-0.04}$	$86.7^{+0.1}_{-0.1}$
CAT	$77.8^{+0.1}_{-0.1}$	$6.45^{+0.04}_{-0.04}$	$89.7^{+0.1}_{-0.1}$

Summary and Outlook

Summary

- GNN-based tracking in the CDC is approaching maturity over a wide range of kinematics including prompt tracks, very low transverse momentum tracks, and tracks from displaced vertices

Outlook

- Started implementing the CAT Finder in the full reconstruction and currently validating using b2validation and VIBE on release-08-01-08 and release-09 $\rightarrow$ looks promising
- Focus on higher multiplicity (B-events) and different particles (e^-, p, K^+)
- Extensive validation on collisions data

Paper Feedback

- Paper draft on Belle II Docs (BELLE2-NOTE-TE-2024-008)
- Mailing list: b2n-te-2024-008
- Plan to enter CWR during the B2GM