Karlsruhe Institute of Technology

ETP-KA/2025-23

Track Finding with Graph Neural Networks in the Belle II Drift Chamber

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Lea Reuter

Tag der mündlichen Prüfung: 31. Oktober 2025
Referent: Prof. Dr. Torben Ferber
Korreferent: Prof. Dr. Markus Klute

Disclaimer

Algorithm development and data analyses in high-energy physics such as the work presented in this doctoral thesis are a collaborative effort. The SuperKEKB particle accelerator which provides the particle beams essential for all studies at Belle II was built and is operated and maintained by the SuperKEKB accelerator group. The Belle II detector was built and is maintained and operated by the Belle II collaboration. The Belle II collaboration also creates the centrally provided simulated and recorded datasets and maintains the computing infrastructure necessary to process them. The software environment necessary for studies with Belle II data plays an important role and was created and is maintained by the collaboration. I have been a part of the Belle II collaboration since 2021 and performed all studies detailed in this thesis except for the following:

- The $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ and $e^+e^- \to \mu^+\mu^-$ candidate selections on measured data, that are implemented as validation modes in VIBE[*].
- The selections of the high level trigger during data acquisition are implemented in the Belle II Analysis Software Framework.

The algorithm developed in this thesis has been published in [1], together with part of the results in Chapters 3 to 6.
I implemented my algorithm into the Belle II Analysis Software Framework together with Giacomo De Pietro.

This thesis incorporates the use of Artificial Intelligence (AI) tools to help with grammatical or stylistic improvement of text, and program code creation.

Grammarly[†] and DeepL Write[‡] are utilised throughout the thesis for spell and grammar checks, as well as for paraphrasing individual, selected sentences to improve clarity and precision in academic writing. I have approved all suggested changes.

ChatGPT[§] is used to aid the development of C++ and Python code, in particular code restructuring and optimisation that do not constitute the core scientific work of this thesis. I have approved and tested all suggestions to provide robust and reliable results.

---

[*] VIBE: Validation Interface for the Belle II Experiment. See https://vibe.belle2.org/ (Access Date: 2025-09-18)

[†] Grammarly: An AI writing assistant. See https://app.grammarly.com/ (Access Date: 2025-09-18).

[‡] DeepL Write: AI-powered writing companion. See https://www.deepl.com/en/write (Access Date: 2025-09-18)

[§] ChatGPT: A virtual AI assistant based on large language models. See https://openai.com/chatgpt/ (Access Date: 2025-09-18).

Contents

Chapter 1

Introduction

Particle physics experiments such as the Belle II experiment probe the Standard Model of particle physics (SM) and search for new physics that is not explained by the SM. These experiments rely on reconstructing the paths of charged particles through the detector from the ionization trails left in the tracking subdetectors. By combining the reconstructed trajectories with measurements from the other subdetectors, the properties of the particles can be determined. A tracking algorithm that accurately reconstructs all true particle trajectories with high purity is essential for performing any analysis or search at Belle II. The Belle II experiment aims to collect the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 50\,\mathrm{ab}^{-1}$ [2] to achieve a higher sensitivity for very rare processes. To reach this, the SuperKEKB accelerator steadily increases its instantaneous luminosity towards the design value. However, the higher instantaneous luminosity amplifies beam-induced effects and detector noise, which increases the detector occupancy. As a result, the performance of the current track finding algorithm of Belle II decreases with increasing backgrounds [3].

One of the biggest questions in particle physics is understanding the nature of dark matter (DM). Many models predict DM candidates such as weakly interacting massive particles, axions, or sterile neutrinos [4, 5]. They are often accompanied by new mediator particles forming a dark sector. If DM interacts non-gravitationally with the SM, it could be produced directly at Belle II. Almost all searches at Belle II for DM explicitly search for an additional light, MeV-GeV mediator that decays into invisible DM particles or back to SM particles [6–8].

One interesting research direction at Belle II are searches for long-lived neutral mediators like dark photons or dark scalars [9], or inelastic DM [9, 10]. In contrast to B mesons, which are a major part of the Belle II physics program [11] and decay promptly due to their short lifetimes, these long-lived particles can have much longer lifetimes. This allows them to travel measurable distances before decaying. Their signatures include tracks that are displaced from the interaction point. Two searches for these displaced signatures were carried out at Belle II up to today: the search for a dark Higgs boson produced in association with inelastic dark matter [12], and the search for a long-lived spin-0 particle in $b \to s$ transitions [13]. Both searches found that reconstruction efficiency for displaced tracks decreases significantly for Belle II, and improvements would greatly benefit new searches.

The reduced reconstruction efficiency for displaced tracks and the loss of performance with higher backgrounds necessitate the development of a new track algorithm.
In this thesis, I present the development and implementation of a graph neural network (GNN)-based tracking reconstruction algorithm *CAT Finder (CDC AI Track Finder)* for the main tracking detector of Belle II, the CDC. The *CAT Finder* uses object condensation [14] to simultaneously identify the unknown number of tracks in the event and determine their respective parameters. Additionally, the *CAT Finder* assigns the detector hits to the tracks for subsequent track fitting. By integrating this algorithm into the Belle II Analysis Software Framework (basf2), it also provides seeds for track finding and fitting that combine all three tracking detectors.
In Chapter 2, I give an overview of the accelerator, experiment, and the subdetectors, including the CDC. I describe the simulated datasets with backgrounds used for training and evaluation, as well as selections on the measured data to study the *CAT Finder* performance in Chapter 3. Chapter 4 defines the key metrics to evaluate track reconstruction algorithms at Belle II. The full tracking chain, the current CDC track finding algorithm in basf2 (*Baseline Finder*), and my new *CAT Finder* algorithm are described in Chapter 5. In Chapter 6, I compare the *Baseline Finder* and the *CAT Finder* for CDC-only performance. I then extend the evaluation to full detector reconstruction, replacing *Baseline Finder* with *CAT Finder* as CDC track finding algorithm. In Chapter 7, I validate the combined performance of all three tracking detectors and study the impact on all subdetectors that depend on reconstructed trajectories. The evaluation on measured data is presented in Chapter 8. Since one goal of *CAT Finder* is to improve displaced tracks, I perform a sensitivity study for a long-lived spin-0 particle S in $B^+ \to K^+ S$ in Chapter 9.

In Chapter 10, I report on the ongoing work and directions for future improvements. Finally, I summarize and conclude the thesis in Chapter 11.
Natural units are used throughout this thesis for mass, energy, and momentum. Charge conjugated decays are implied in this thesis.

Chapter 2

The Belle II Experiment

I describe the experimental setup in this thesis in the following chapter. I start with giving an overview of the SuperKEKB accelerator [15], followed by the Belle II detector [16]. In the last chapter, I highlight the main tracking detector of Belle II, the CDC, which I will focus on in this thesis.

2.1 The SuperKEKB Accelerator

The SuperKEKB accelerator is an electron-positron collider. An overview of the accelerator is given in Fig. 2.1. While electrons are emitted via a photo cathode, the positrons are produced in collision of an electron beam with a tungsten target, which requires a damping ring to reduce the emittance. Electron and positron bunches are accelerated with a linear accelerator and once they reach their target energy, they are injected into the storage rings. This injection happens continuously during the operation. The SuperKEKB accelerator is an asymmetric collider, where the electron ring operates at $7\,\mathrm{GeV}$ (High-Energy Ring (HER)) and the positron beam at $4\,\mathrm{GeV}$ (Low-Energy Ring (LER)). The collision takes place at the Belle II experiment with the center-of-mass system (cms) collision energy is the $\Upsilon(4S)$ resonance at $\sqrt{s} = 10.58\,\mathrm{GeV}$, so that it enables decays into B-meson pairs. The lab frame is boosted due to the beam asymmetry with a Lorentz boost of $\beta\gamma = 0.28$ in the electron beam (forward) direction. This enables the precise measurements of B- and D-meson decays, where the decay time is expressed in travel length. The main processes including their cross section is discussed in Section 3.4.
The accelerator is designed to achieve an instantaneous luminosity of $\mathcal{L} = 6 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, and an integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 50\,\mathrm{ab}^{-1}$[2]. Achieving this high instantaneous luminosity requires the particle bunch lengths to be small, and extremely small horizontal and vertical beam sizes, which is known as the nano-beam scheme, and shift the bunches using crab cavities to align the bunches at the interaction point [17]. The luminosity is then expressed using the beam currents $I_\pm$ are increased, the vertical beam-beam parameter $\xi_{y\pm}$

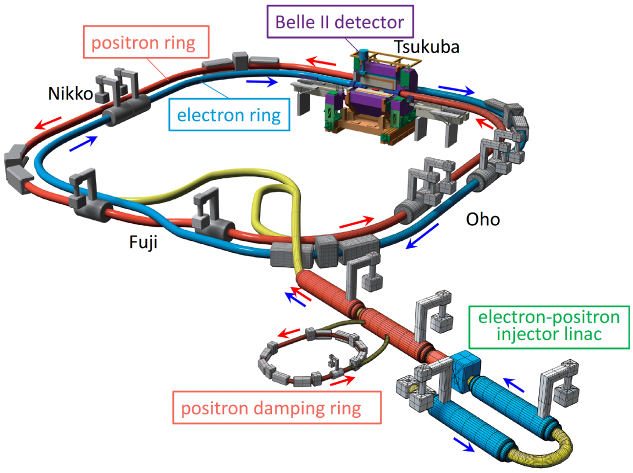


Figure 2.1: Schematic view on the SuperKEKB accelerator complex. Taken from [15].

is increased, and the vertical beta function at the Interaction Point (IP) β^*_γ is decreased[2]:

$$\mathcal{L} \propto \frac{I_\pm \xi_{y\pm}}{\beta^*_\gamma}. \tag{2.1}$$

Increased beam-beam parameter increases the collision probability but also makes the beam more unstable. The parameters above are continuously adjusted to achieve the target luminosity and the highest instantaneous luminosity achieved at the time of this thesis is $\mathcal{L} = 5.1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$[18].

2.2 The Belle II Detector

Particle collisions occur at the IP, with the Belle II detector [16] detecting the particles produced in the collision. Starting from the IP, it consists of a charged particle spectrometer, followed by particle identification systems, the electromagnetic calorimeter, a superconducting solenoid, and the K^0_L and muon detector. A schematic view of the detector is given in Fig. 2.2, including the relevant subdetectors arranged around the IP to achieve a 4π coverage in a cylindrical shape. The detector is asymmetric with a larger detector region towards the forward direction, due to the boosted lab frame in the direction of the electron beam.

The Belle II coordinate system is centered at the IP, with the positive z-axis pointing in the direction of the electron beam. The x-axis is horizontal, pointing outward from the accelerator center, while the y-axis is vertical and points upwards. The polar angle θ is defined from 0 to π, where $\theta = 0$ describes the forward direction, building the longitudinal plane aligned with the detector's solenoid axis. The azimuthal angle ϕ is defined from $-\pi$ to π in the transverse x-y plane orthogonal to the detector's solenoid axis, with positive x

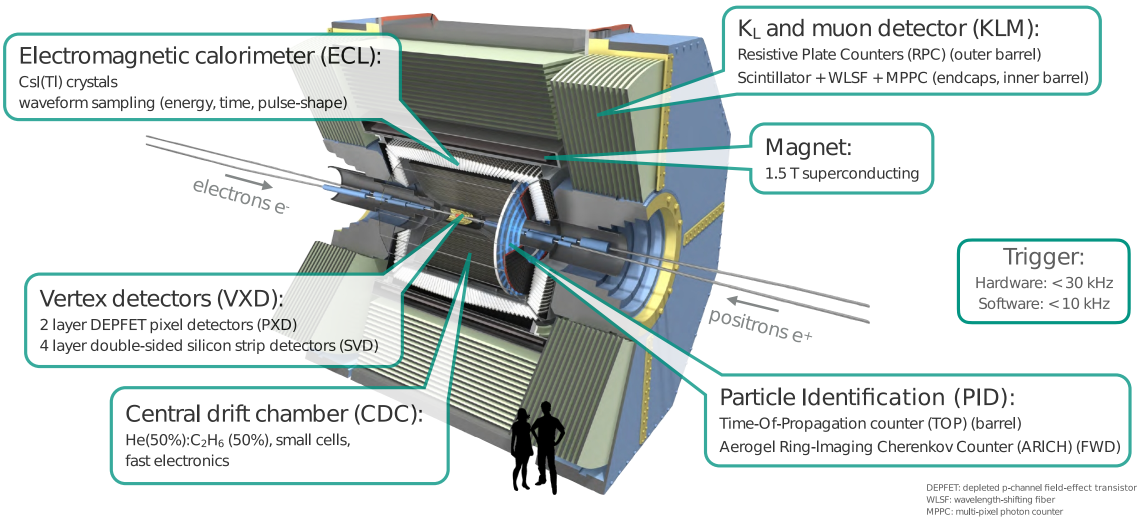


Figure 2.2: Schematic overview of the Belle II detector at the collision point of the electron-positron collider, including the sub-detectors. Taken from [19]

and $y = 0$ being $\phi = 0$.

I give a short explanation of the different subdetectors in the following, with a complete description provided in [16].

2.2.1 Tracking and Vertexing Detectors

The charged particle spectrometers measure the trajectory of the electromagnetically charged particles, referred to as tracks. As these particles traverse through the detector, they ionize the detector material, with the resulting signals recorded. Due to the magnetic field of 1.5 T the particle trajectories are bend, which enables the identification of the electromagnetic charge of particles and their momentum. The Belle II tracking detectors consist of three subdetectors. The first two are the high granularity Pixel Detector (PXD) [20, 21] around the IP, followed by the Silicon Vertex Detector (SVD) [22] detector. Both detectors have a high spatial resolution to resolve the decay vertices close to the IP from among others B-meson, D-meson, or K_s^0 decays. The main tracking detector is the CDC [23, 24], which I describe in detail in Section 2.3 as the main detector of this work.

I describe the remaining two detectors in the following.

Pixel Detector

The PXD detector consists of two silicon sensor layers, located around the beam pipe with a radius of 1 cm around the IP. The first PXD layer is located at 1.4 cm, spanning from -4 cm to 5 cm in the z-direction. The second PXD layer is at 2.2 cm with the z-coordinate ranging from -5 cm to 9 cm. To minimize the material budget, the PXD sensors are made up from Depleted Field Effect Transistors to enable very thin layers. The pixel sensors on the two layers range from $50\,\mu\text{m} \times 55\,\mu\text{m}$ to $50\,\mu\text{m} \times 85\,\mu\text{m}$, arranged in modules of

250×768 sensors, with 16 modules on the first layer and 24 modules on the second layer. Charged particles traversing the sensor produce electron-hole pairs, where the electrons are measured with their three-dimensional coordinates with a spatial resolution of PXD of $12\,\mu$m.

During data taking from 2018-2022 (run I), only the first layer was functional, besides two modules that were installed on the second layer, due to a module on the first layer not working. The second layer was installed during the time between run I and data taking starting from 2024 (run II).

Silicon Vertex Detector

The SVD is constructed from four layers of double-sided silicon strip sensors. The layers are located at the radius (and z-coordinate) of 3.9 cm (-9 cm to 14.7 cm), 8 cm (-16 cm to 21 cm), 11.5 cm (-21 cm to 29 cm), and 14 cm (-26 cm to 35 cm), covering an angular acceptance of $17^\circ < \theta < 150^\circ$. The forward modules of the SVD are slanted in the direction of the beam pipe. When a charged particle traverses the material and ionizes the material, electron-hole pairs are created. While holes travel to the strips parallel to the z-axis, electrons travel to the strips aligned perpendicular at $r - \phi$ in the transverse plane. By combining the information of the measured clusters on both sides of the silicon strip sensors, the three-dimensional information can be determined, enabling a spatial resolution between $18\,\mu$m to $35\,\mu$m.

2.2.2 Particle Identification Detectors

While all subdetectors except the PXD contribute to identifying the different particle types traversing the detector, there are two detectors dedicated to identifying charged pions and kaons.

Time-Of-Propagation Counter

The Time-Of-Propagation counter (TOP) detector [25] covers an angular region of $31^\circ < \theta < 128^\circ$ and is located outside of the CDC in the barrel region of Belle II. The TOP uses Cherenkov radiation to identify particles. When a charged particle traverses one of the 16 quartz bars aligned parallel to the z-axis, it emits Cherenkov photons at a specific angle depending on its velocity. The photons travel through the quartz, with a mirror placed at the forward end, and are read out with a photomultiplier at the backward end. By measuring the time the photons take to propagate through the detector together with the position of the photons at the backward readout, pions and kaons can be identified.

Aerogel Ring-Imaging Cherenkov Detector

The Aerogel Ring-Imaging Cherenkov detector (ARICH) detector [26] is located in the forward direction of the detector and has an acceptance of $14^\circ < \theta < 30^\circ$. As the TOP detector, the particle identification is based on Cherenkov radiation emitted according to the particle velocity. This detector uses proximity-focusing ring-imaging detector technology based on two layers of silica aerogel tiles, where the photons are collected via hybrid avalanche photo detectors.

2.2.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECL) [27] measures the energy of particles. It is a homogeneous calorimeter with scintillating thallium-doped caesium iodide crystals with a depth of approximately 16 radiation lengths. The crystals are ordered in the barrel and the endcaps in forward and backward direction, covering an angular acceptance of $12.4^\circ < \theta < 155.1^\circ$ with two gaps between $31.4^\circ < \theta < 32.2^\circ$ and $128.7^\circ < \theta < 130.7^\circ$. The energy of the particles is measured with the electromagnetic shower they induce in the crystals. The photons that are created within this shower are collected by photodiodes at the end of the crystals, where the collected energies are clustered together building the ECL clusters. These ECL clusters are combined with the tracking information to find out if the shower came from a charged particle leaving a track or neutral particles like photons or π^0. They can also be used to identify particles, as for example electrons deposit nearly their full energy in the ECL, while muons deposit low amounts of energy.

2.2.4 K_L^0 and Muon Detector

The outermost detector after the solenoid is the $\mathrm{K_L^0}$ and muon detector (KLM)[28, 29], detecting muons that only interact minimal with the previous detector material and long-lived neutral kaons K_L^0. Similar to the ECL, it is arranged in three regions with the barrel and two endcaps, leading to a total angular acceptance of the KLM of $18^\circ < \theta < 155^\circ$. The KLM consists of alternating absorber plates and active material detector layers. The absorbers are iron plates, which also serve as magnetic flux return for the solenoid, which provide together with the calorimeters's 0.8 interaction lengths for the K_L^0 to shower hadronically and additional 3.9. These hadronic showers can then be detected by the ECL or KLM.

Due to the high background particle flux, scintillators are used in the endcaps and for the first two barrel detector layers. The later 13 detector layers in the barrel are made up from resistive plate chambers. While K_L^0 are identified by their hadronic shower, clustering the KLM measurements together, muons are identified by comparing KLM measurements with their extrapolated trajectory.

2.2.5 Trigger System

While the bunch crossing rate at SuperKEKB is 250 MHz, the process rate for the $\Upsilon(4S)$ at the target luminosity is at 0.67 kHz[11]. Particle collisions are much less frequent than the bunch crossing, with the most common occurrence being Bhabha scattering $e^+e^- \rightarrow e^+e^-(\gamma)$ with an angle that the electron or positron can be detected within the ECL at a rate of 44.6 kHz. Because there are bandwidth restriction in the Data Acquisition (DAQ), the storable rate cannot exceed 10 kHz. To achieve this reduced rate, the measured data is reconstructed and selections are applied to decide if the measured data should be saved or can be discarded, referred to as trigger systems.

This reconstruction is first done using a hardware-based trigger system, which is called the Level 1 trigger (L1 trigger) in Belle II, in which the rate is reduced to 30 kHz. The L1 trigger uses field-programmable gate array (FPGA)s to combine the information from the CDC, ECL, KLM and TOP to decide if an event should be kept or not.

In a second step, the software-based High Level Trigger (HLT) reconstructs the event using the Belle II reconstruction software on CPUs excluding the PXD. The PXD is excluded due to the placement of the readout electronics outside of the detector acceptance, which decreases the material budget but also makes simultaneous readout of the pixels no longer possible. Instead, the time of readout is integrated which in return increases the occupancy making the PXD clusters unable to be included in the reconstruction. Instead, once the HLT decides to keep an event, the reconstructed track information is used to define regions of interest in the PXD, where the pixels are read out. With the HLT, the rate is reduced to 10 kHz.

These triggers are employed to filter interesting physics events according to the Belle II physics program[11], while rejecting uninteresting events such as Bhabha scattering or beam interactions outside of the collision point.

2.3 The Central Drift Chamber

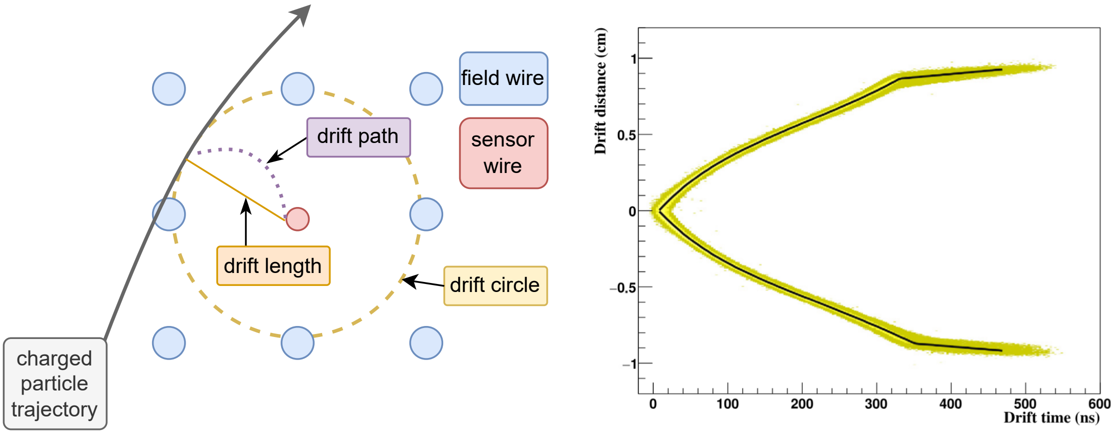
The CDC is the main tracking detector of Belle II, necessary to measure the momentum of particles and provide particle identification for low momentum particles that do not reach the outer detector systems.

Charged particles travel through the 50% helium and 50% ethane gas mixture while ionizing the gas. In comparison to the PXD and SVD silicon based detectors, the CDC has a lower material budget. Through an approximately radial electric field, the electrons and ions are separated and the electrons are accelerated and amplified in the gas towards the sensor anode due to the high wire voltage [30].

The CDC is a 233 cm long cylinder with a radius from 16 cm to 113 cm, that consists of 14,336 gold-plated tungsten sensor wires and 42,240 aluminium field wires. The wires are aligned to form $10 \times 8\text{mm}^2$ and $18 \times 18\text{mm}^2$ large drift cells, with the sensor wire in the middle and the field wires surrounding to span the electric field in which the electrons drift to the sensor, as shown in Fig. 2.3. When a signal is measured, the timing signal is is recorded with a Time-to-Digital Converter (TDC) with a resolution of $\alpha = 0.98\,\text{ns}$. The total timing of the measurement is given with

$$\alpha \cdot \text{TDC} = T_0 - T_{tof} - T_{drift} - T_{prop} \tag{2.2}$$

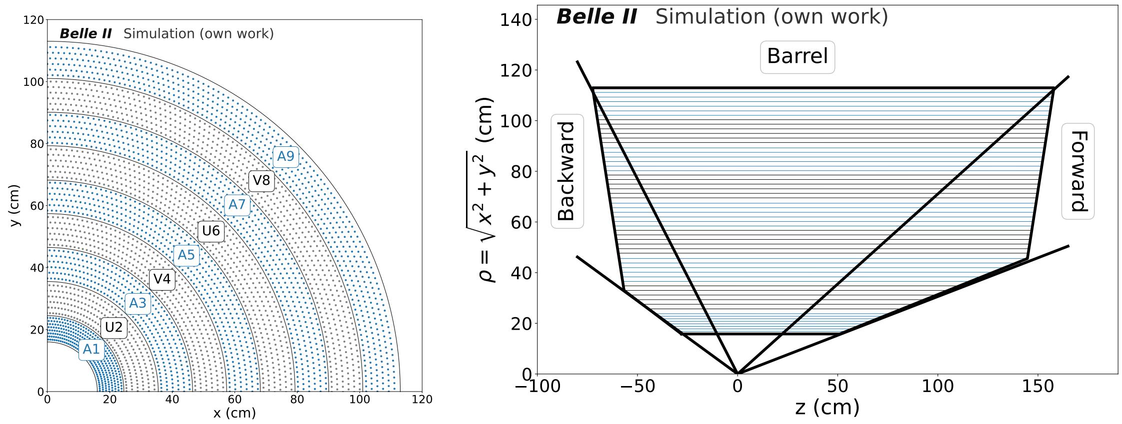
with the collision time T_0, the time-of-flight from the collision point to the drift cell T_{tof}, the drift time of the ionized electrons to the sensor wire T_{drift}, and the time the signal needs to propagate through the wire to the backward readout electronics T_{prop} [24].
The drift length is calculated from the drift time using the time-to-space $x - t$ relation as shown for example in Fig. 2.3. The drift length resolution is about $120\,\mu\text{m}$. As it is unknown where in the drift cell the particle passed through, the measured signal is expressed in the drift circle around the sensor wire, as only the length the ionized electrons travelled is known.
While the time of flight and propagation time is $\mathcal{O}(1)\,\text{ns}$, the drift time of the ionized electrons is much larger and reaches values over 600 ns depending on the readout window. The propagation time of the signal along the wire is orders of magnitude shorter than the electron drift time. As the signal is only read out at the backward end of the wire, the measurement provides no information about the position of the hit along the sensor wire. In addition to the time of the signal, the digitized signal amplitude Analog-to-Digital Converter (ADC) count, which is proportional to the energy deposition of the particle, and the Time over Threshold (TOT) are also provided.



(a) Drift cell schematic.

(b) Drift cell $x - t$ relation.

Figure 2.3: The left figure shows a schematic view of a charged particle (grey) passing a drift cell bound by the field wires (blue), including the drift path of the electrons (purple) and the resulting drift length (orange) with the drift circle (yellow). The right figure, taken from [24], shows the non-linear $x - t$ relation between the drift time and drift length of such a drift cell.



(a) Superlayer arrangement in $x - y$.

(b) Superlayer arrangement in $\rho - z$.

Figure 2.4: CDC geometrical overview in the $x - y$ plane (left) and the $\rho - z$ plane, where $\rho = \sqrt{x^2 + y^2}$ (right). Axial superlayers are colored blue, stereo superlayers colored grey. The superlayers are annotated according to their naming scheme in Table 2.1.

The full arrangement of the sensor wires is given in Fig. 2.4 for the transverse $x - y$ plane and the $\rho - z$ plane, with $\rho = \sqrt{x^2 + y^2}$. The CDC covers the acceptance region of $17^\circ < \theta < 150^\circ$ and the full azimuthal angle in 2π. The wires are arranged in 56 layers, with groups forming the axial and stereo superlayers as seen in Table 2.1. Particles with a polar angle between $35.4^\circ < \theta < 123^\circ$ traverse all layers of the CDC if their momentum is high enough. Particles with a polar angle between $17^\circ < \theta < 35.4^\circ$ leave the CDC early in the forward endcap and particles with a polar angle between $123^\circ < \theta < 150^\circ$ leave the CDC early in the backward endcap. Axial superlayers (A) are aligned in parallel to the z-axis, enabling the two-dimensional trajectory. Stereo superlayers (U,V) on the other hand are tilted in respect to the z-axis according to the stereo angles given in Table 2.1 and visualized in Fig. 2.5. This shift in the wire alignment allows for the measurement of particle trajectories in z-direction. A typical event display for an $\Upsilon(4S) \to B^+B^-$ event in the CDC is shown in Fig. 2.6, where as reference position for the stereo wire x and y

Table 2.1: Configuration of the CDC sensor wires, taken from [16].

Superlayer	N_{layer}	$N_{\text{sensor cells}}$	r (cm)	Stereo angle (mrad)
Axial A1	8	160	16.80 – 23.80	0.0
Stereo U2	6	160	25.70 – 34.80	45.4 – 45.8
Axial A3	6	192	36.52 – 45.57	0.0
Stereo V4	6	224	47.69 – 56.69	-55.3 – -64.3
Axial A5	6	256	58.41 – 67.41	0.0
Stereo U6	6	288	69.53 – 78.53	63.1 – 70.0
Axial A7	6	320	80.25 – 89.25	0.0
Stereo V8	6	352	91.37 – 100.37	-68.5 – -74.0
Axial A9	6	384	102.09 – 111.14	0.0

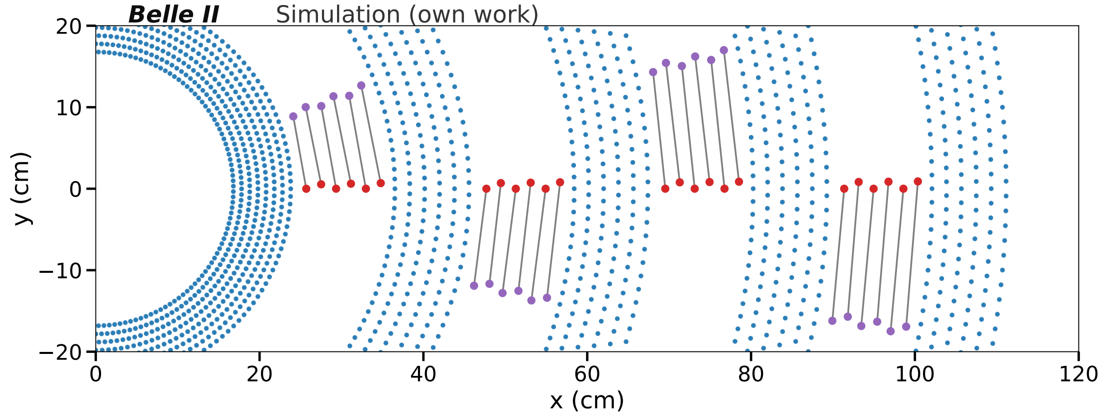
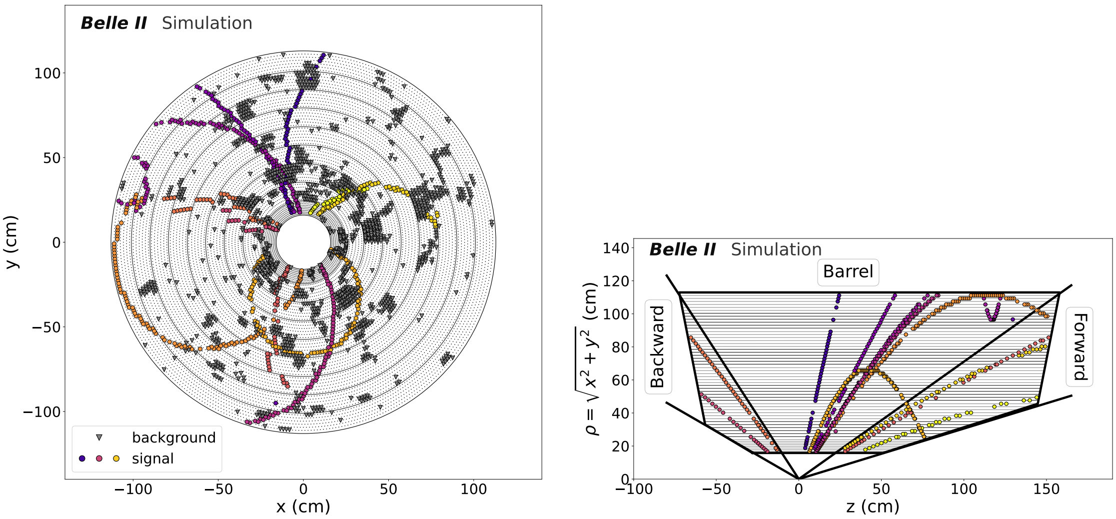


Figure 2.5: This figure shows the $x - y$ projection of the CDC sense wires. The axial wires, which are aligned along the z-axis and therefore have the same projected x and y position over their entire length, are drawn in blue. The stereo wires, which are tilted with respect to the z-axis, are shown by indicating their positions at both ends of the wire: the backward end in red and the forward end in purple. A grey line connects these two positions, illustrating the projected displacement of the stereo wires in the x–y plane as a function of z.



(a) Event display in the x-y plane.

(b) Event display in the z-ρ-plane.

Figure 2.6: Figure and text taken from [1]: Typical event display in the x-y plane (Fig. 2.6a) and the z-ρ plane (Fig. 2.6b) for a simulated $\Upsilon(4S) \rightarrow B^+B^-$ event with *high data beam backgrounds*. In the x-y plane, filled colored circular markers represent signal hits, while filled gray triangular markers represent background hits. These markers correspond to the locations of the sense wires at the z position of the wire center, for wires with recorded ADC signals. In the z-ρ plane, only the signal hits are shown. The three detector regions, forward endcap, barrel, and backward endcap, are also indicated in the z-ρ plane.

position the middle of the wire as shown in Fig. 2.5.

Chapter 3

Dataset

In this chapter, I provide a detailed overview of the datasets utilized throughout this thesis. I begin by describing the simulation in Section 3.1. Next, I discuss the datasets employed for training the novel GNN algorithm developed in this work in Section 3.2. Following the description of the training data, I present the evaluation samples used for both measured data and simulation in Sections 3.3 and 3.4. Finally, I describe the dark sector motivated samples in Section 3.5.

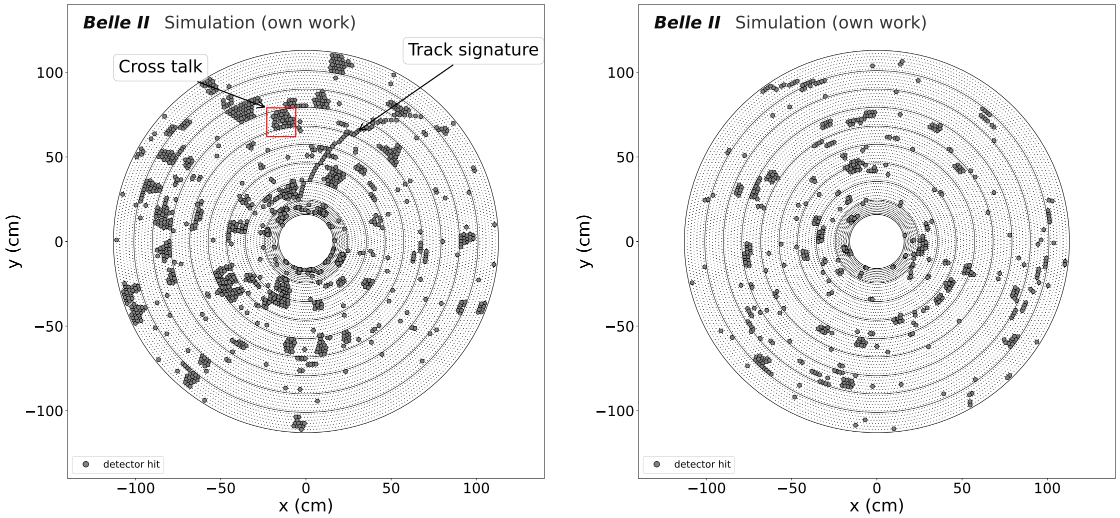
3.1 Simulation

To develop and evaluate a new tracking algorithm, it is necessary to know the underlying truth. For this, I simulate events described in the following Section 3.2 with basf2. Within basf2, the complete detector geometry and the interaction of the generated particles with the detector are simulated using GEANT4. The detector response is also simulated, resulting in digitized detector outputs close to the data taken at Belle II.

3.1.1 Beam Backgrounds

Besides the simulated signal samples, due to the beam interplay there are additional measurements in the CDC not coming from the signal process, that are now described following [31, 32]. Due to the increased beam current and Belle II nano-beam scheme, particles can deviate from their original orbit and can hit the beam pipe or other material close to the detector. This can generate shower particles that can then be measured in the detector, referred to as beam induced backgrounds. While [31] gives a complete overview over all effects, the most relevant effects in the CDC regarding the track finding are the following:

Particle Scattering Two particles in the same beam bunch can coulomb scatter, which causes one of the particles to lose and and the other to gain energy, which results in deviating from the expected beam orbit. This effect is called the Touschek rate and is



(a) Beam background taken from data. (b) Simulated beam background.

Figure 3.1: Example event displays in the x-y-plane for beam-background overlays. Figure 3.1a shows an example pattern for the cross talk in the red box as well as annotate the track signature in one beam background event taken from data from experiment 26 run 1430. Figure 3.1b shows one example event for the simulated beam background.

proportional to the beam current squared and inversely proportional to the number of bunches in the ring and the beam size that is reduced due to the Belle II nano-beam scheme, increasing the number of background hits with higher luminosity achieved during the run period. Besides scattering within itself, the beam can also scatter at gas atoms in the beam pipe, either by coloumb scattering or bremsstrahlung, or other structures. If this interaction occurs close to the Belle II experiment, the shower particles can reach the detector. This effect is proportional to the beam current and the number of gas molecules.

Luminosity backgrounds The collision of beams has a very high cross section for radiative bhabha scattering and two photon processes, which is uninteresting physics wise but increases the occupancy of the detector.

Cross-talk If there is a large charge deposition in one of the readout channels in the CDC, it might happen that this leads to neighbouring channels being triggered too, due to the charge leaking in to the neighbors.

Particle scattering and luminosity-related backgrounds can produce track-like signatures or isolated wire hits, for example when Bremsstrahlung photons convert into electron–positron pairs that are trapped in the magnetic field and detected by only a single wire. An example of a track signature is shown in Fig. 3.1a. Cross talk results in large cluster-like patterns, highlighted for one example in Fig. 3.1a.

The number of background hits is dependent on the beam conditions and the luminosity, as well as the gas condition of the CDC. Fig. 3.2 shows the number of background hits, with

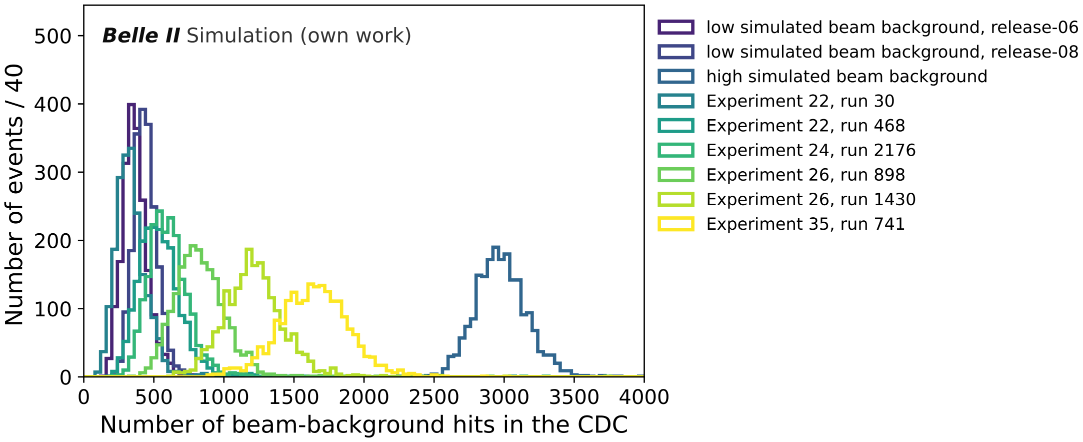


Figure 3.2: Number of beam background CDC Hits for different experiments and runs as well as the simulated backgrounds.

their respective median, standard deviation and associated luminosity values in Table 3.1. The numbers are shown for different experiments and runs of Belle II data taking. Runs represent continuous data-taking periods, where the run number increases whenever data taking is restarted. Belle II groups runs into experiments based on consistent detector conditions and similar L1 trigger and HLT software configurations. The recorded instantaneous luminosity is the peak instantaneous luminosity, which varies during the run due to adjusted beam parameters, but shows a trend where an increase in instantaneous luminosity increases the number of beam-background hits. Especially towards the end of run I, which includes experiments up to 26, the number of beam-background hits increased drastically over 7 months, from 338 to 1260 for the later runs. Additionally, the spread of the distribution also increased, with the standard deviation of the distribution going from 111 to 229 hits. However, the instantaneous luminosity or the beam conditions alone cannot explain the strong increase in beam background hits. Run 2 started in January 2024 and experiment 35, run 714 is from October. Here, although the beam conditions are comparable to experiment 26, run 1430 and the peak instantaneous luminosity is significantly lower, the number of beam-background hits is significantly higher. One assumption for the high increase is either the gas conditions of the CDC or the Malter effect [33, 34], which is explained in Section 6.4.

For the simulation, I overlay randomly triggered events from data that have a very low probability to contain actual collision data from the data-taking period of run I with beam-background distribution of experiment 26, run 1430. As the number of these overlays is limited, I combine multiple runs with an average instantaneous luminosity of about $\mathcal{L}_{\text{beam}} = 3.53 \times 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ and similar median number of beam-background hits, resulting in the high data beam backgrounds with over 2 million overlays available. If not specified differently, the beam-backgrounds used for simulation for any results shown within this thesis are the high data beam backgroundss.

Table 3.1: Experiment and Runs with their time and respective luminosity. The choice of runs is inspired by [3].

Experiment	Run	Date	Peak inst. luminosity (in $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	Beam-background hits Median	Beam-background hits Std. Dev.
22	30	30.11.2021	2.4	338	111
22	468	15.12.2021	3.2	541	166
24	2176	21.04.2022	3.4	604	167
26	898	31.05.2022	3.2	836	223
26	1430	11.06.2022	3.6	1216	229
35	714	27.10.2024	2.1	1645	268
low simulated beam background release-06			1.06	367	91
low simulated beam background release-08			3.0	440	84
high simulated beam background release-08			60.0	3980	185

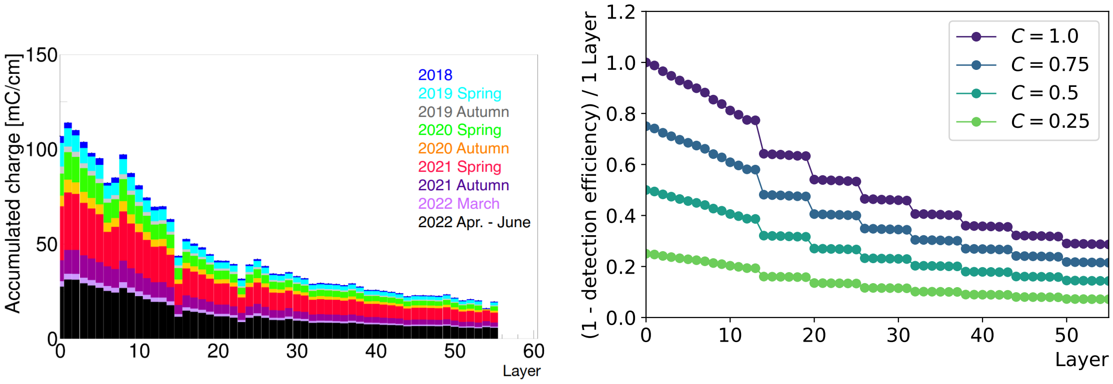
There are also centrally produced simulated beam background events approximating the collider conditions in 2021 [31, 32] low simulated beam backgrounds corresponding to an instantaneous luminosity $\mathcal{L}_{\mathrm{beam}} = 1.06 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ available. The statistics are way higher, but some effects, such as cross talk and the track signatures, are underestimated. This can be seen when comparing the example display of Fig. 3.1a taken from data with the simulated beam-background display of Fig. 3.1b. It is explicitly stated when the low simulated beam backgrounds are used.

3.1.2 Detector Degradation

Due to all the effects mentioned in the previous chapter, the occupancy increases which also increases the current in the CDC. This can lead to a faster chamber ageing as depositions can build-up on the wires and will be explained shortly following [30].

Depositions on the anodes can lead to a decrease of the amplification power. If the build-up is conductive, the wire radius is increasing, reducing the amplification range. If the build-up is insulating, a counter field can build, also reducing the amplification. In the worst case, whiskers that deform the wire can build up, impacting the electric field structure and enabling point discharges, either in a permanent underlying current or random signals.

If insulating depositions are build-up on the cathodes, then the Malter effect can happen [33]. Here, positively charged ions accumulate at the insulation layer of the cathode. This ion layer builds as field between itself and the cathode and can become strong enough to induce field emissions of electrons from the cathode. These electrons then drift towards the anode, being amplified which generates more positive ions, that collect at the insulating layer, which either maintains or increases the field strength. In the end, this leads to a continuous current without any external impact. This is known as the malter current, where due to the constant current signals are measured in the affected wires even in the absence of



(a) Accumulated charge per layer during run I. (b) Approximated detection inefficiency per layer.

Figure 3.3: Figure 3.3a shows the accumulated charge per layer of the CDC as reported in Belle II internal meetings [36] for the data taking period of run I. Figure 3.3b shows the approximated detection inefficiency per layer according to Eq. (3.1) for four different values of C.

ionizing particles going through the gas. Furthermore, real signals measured by a charged particle ionizing the gas can be overshadowed by the malter current, making it unable to differentiate or measure the signal at all. The detector operation can also become very unstable.

The detector degradation is measured with the accumulated charge on each wire, shown for the run I period in Fig. 3.3a, with 150 mC/cm in the innermost layer and around 35 mC/cm in the outermost layer. The ageing of the CDC and its effect on the trackfinding performance is also discussed in detail in the thesis [35] I supervised . I model two effects regarding the detector degradation.

Wire inefficiencies

With the amplification power decreasing or signals being obscured by either a permanent current or sudden spark, this means that the wires get less sensitive to detect less signal. This results in not measuring certain hits in the particle trajectory. As the detector degradation effect is dependent on the particle flux [30], the inner part of the CDC is the most effected, as this region is the most dense one. The accumulated charge on the CDC layers over the run I datataking period is shown in Fig. 3.3a. I use the following formula to approximate the wire efficiency for both anode-aging and malter effects:

$$\varepsilon_{\mathrm{layer,i}} = 1 - C \cdot \frac{N_0}{N_0} \cdot \frac{L_0}{L_{\mathrm{i}}}. \tag{3.1}$$

N_0 and N_{i} refer to the number of wires in the layer 0 and the i-th layer respectively of the CDC, as the charge accumulation decreases the more wires are available. The length of the wires is also relevant for the accumulation, where longer wires are less effected by the charge accumulation, described by the fraction of the length of the 0-th layer L_0 to the i-th layer L_i. Furthermore, the value C is introduced as a tunable parameter to adjust the efficiency, where at a value of $C = 0$ all wires have perfect detection capabilities and with a value of $C = 1$ the innermost layers of the CDC are not able to detect anything. Figure 3.3b shows the inefficiency of the wire with

$$\text{wire inefficiency} = 1 - \text{detection efficiency},$$

matching the shape of the accumulated charge during data taking, which is especially impacted in the first layers.

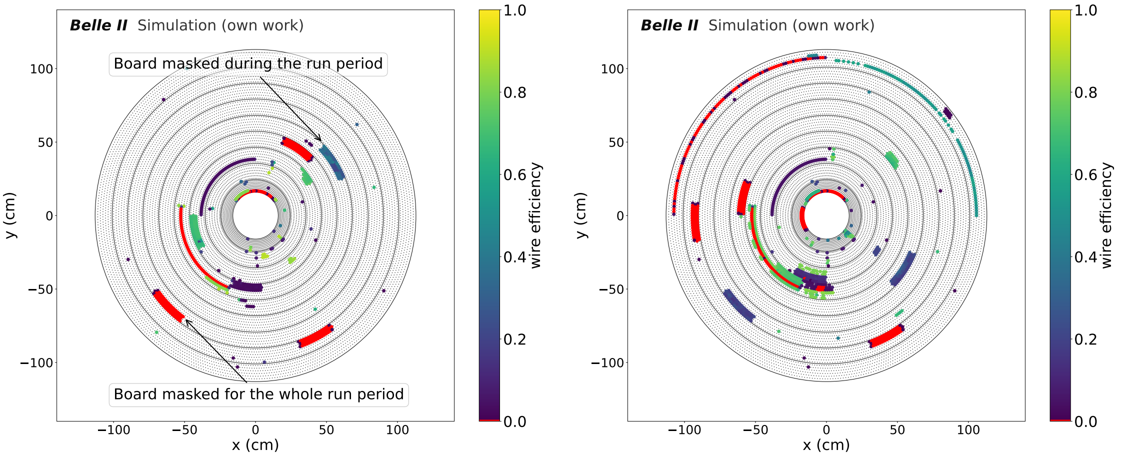
Disabled Boards

During data taking, it can happen that one readout board produces multiple errors. Possible errors can result from radiation on the readout boards as they are located in the detector at the CDC. Another issue may arise from high occupancy due to the beam background conditions, where the buffer might not be enough to handle the large data amounts. Instead of stopping the run and additional time required for bug fixing, its possible to disable the problematic board for the ongoing run, which minimizes the data-taking downtime. As a drawback, the detector coverage is now reduced as nothing is measured for this respective region.

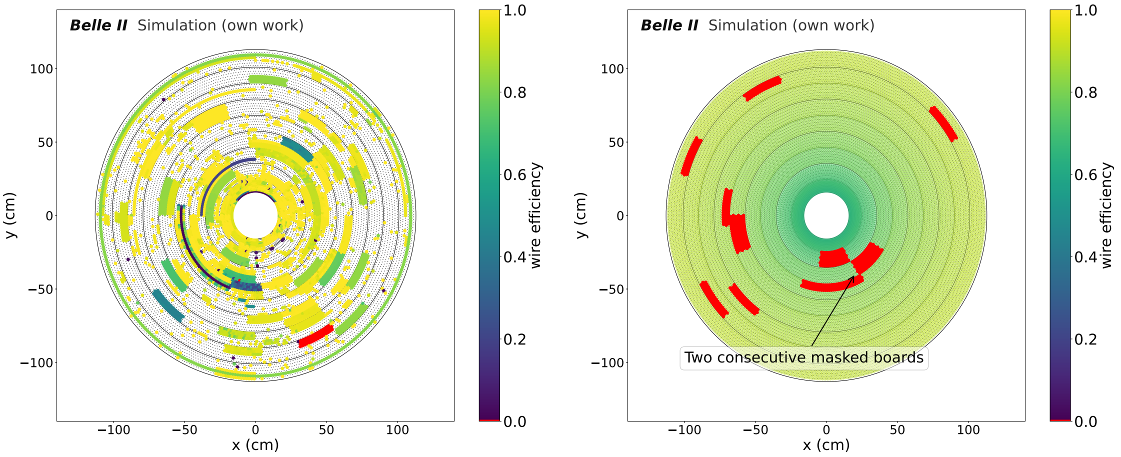
Wire Efficiency Map

Both the wire inefficiencies and the disabled boards are included in the simulation. They are measured in data for a combination of runs in an experiment over a certain time period. Figure 3.4a shows the wire conditions for experiment 22, run 30 from 2021.

A total of 121 wires is completely off, and 430 wires have decreased efficiency. Figure 3.4b shows the wire conditions for the end of run I. In comparison to 2021, now a total of 168 wires are off and 809 have decreased efficiency, increasing the rate of affected CDC wires from a total of 3.8 % to 6.8 %. These specific wire maps are used for the simulation with the data beam backgrounds. For the simulation including the low simulated beam backgrounds, the averaged wire efficiency map over the full run I period is used which is shown in Fig. 3.4c. Additionally, I include one approximation of conditions for the run II using Eq. (3.1) with $C = 0.35$ and a worst-case scenario for the disabled boards, where 2 boards in one superlayer are disabled right after each other, effectively disabling the superlayer in this ϕ-region, for being able to probe the tracking robustness against this scenario. One can see that the inner superlayer of the CDC is impacted the most.



(a) Experiment 22, run 30 (b) Experiment 26, run 1430



(c) run I average (d) Approximated conditions for run II.

Figure 3.4: Wire efficiency maps of the CDC used for simulation for different experiments and run in Fig. 3.4a and Fig. 3.4b and the average wire efficiency map over the full run I data taking period, used in combination of the low simulated beam background samples. Coldred wires have a decreased efficiency of less than one, and wires that are coloured red are completely off. The large regions correspond to disabled boards. If these regions are red, the boards were disabled for the full integrated time period, if they are coloured differently, they were disabled during this time period because an issue was raised.

3.2 Training Datasets

For the development of the new tracking algorithm, the training dataset is the most important part. The dataset used in this thesis is also published in [1], and this section will follow the description closely. The following are the key signatures for the tracking algorithm:

1. Tracks with a low momentum ($p_t > 0.4\,\text{GeV}$) have a large curvature and form circles in the CDC, even with the possibility to curl through the detector multiple times, whereas tracks with a high momentum have a small curvature and leave the detector

Table 3.2: Text and content adapted from [1]: Event samples used for training and validation. See text for details.

category	name	θ_{gen} (°)	r^{3D}_{gen} (cm)	α^{3D}_{gen} (°)
1	prompt fwd	17.0 – 35.4	0	0
2	prompt brl	35.4 – 123.04	0	0
3	prompt bwd	123.04 – 150.0	0	0
4	prompt full	17.0 – 150.0	0 – 100	0
5	displaced fwd	17.0 – 35.4	0 – 100	0
6	displaced brl	35.4 – 123.04	0 – 100	0
7	displaced bwd	123.04 – 150.0	0 – 100	0
8	displaced angled	17.0 – 150.0	0 – 100	0 – 30
9	vertex large	17.0 – 150.0	0 – 100	0 – 90
10	vertex small	17.0 – 150.0	0 – 100	0 – 25
11	mix 4+10	-	-	-
category	name	Generator		
12	B^0B^0	EvtGen		
13	B^-B^+	EvtGen		

in an approximately straight line.

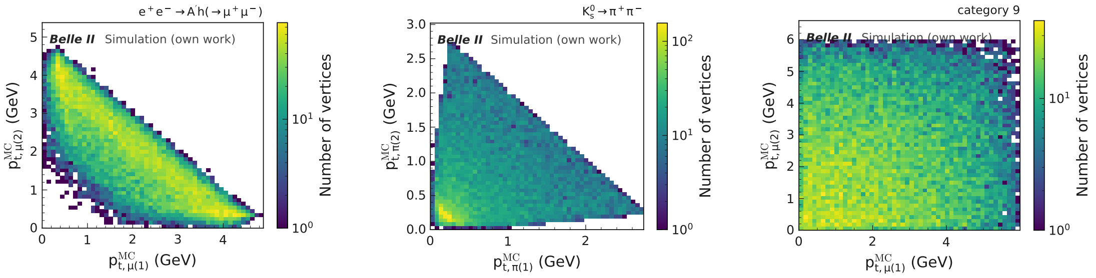
2. Tracks can either exit the detector in the barrel region, traversing the full detector, or leave the detector early in either the forward or backward direction.

3. Displaced vertices can have small opening angles, potentially overlapping in their trajectory, or large opening angles where the tracks are well separated but might go along one CDC layer.

The goal is not to bias the tracking algorithm towards specific physics processes. Therefore, I create a dataset that does not follow conservation laws for charge, energy, momentum, lepton flavour, and more, but instead uses particles drawn randomly from the parameter space defined in Table 3.2. Most of the events in the detector have very prompt tracks that are coming from the IP, or very close to it, so that they decay in the inner tracking detectors. The main target of this work is to develop a tracking algorithm that works better for tracks coming from a displaced vertex than the current one, where the displacement is so large that it decays in the CDC. For the tracks in the categories 1-11, muons are used, with their charge-conjugate state implied. All the generated quantities for these categories are drawn randomly from independent uniform distributions for each charged particle.

3.2.1 Prompt Sample

The first four datasets in Table 3.2 are prompt tracks that come from the IP. I simulate samples in equal parts that go into the forward, barrel, and backward direction to enrich the dataset with shorter track signatures, categories 1-3. To prevent the model to learn detector regions as a signature, I also add one combined category 4 where tracks can go to



(a) $h \to \mu^+\mu^-$ (b) $K^0_S \to \pi^+\pi^-$ (c) Category 9.

Figure 3.5: Transverse momentum of the two particles $p^{\text{MC}}_{t,(1)}$ and $p^{\text{MC}}_{t,(2)}$ in the different vertex samples from $h \to \mu^+\mu^-$ (left), $K^0_S \to \pi^+\pi^-$ (middle) and the category 8 samples described in Table 3.2

the full CDC acceptance region. The momentum for these tracks is sampled uniformly in $p_{\text{t,gen}}$ from 0.05 GeV to 6 GeV. I do not sample momentum higher than $p_{\text{t,gen}}$ =6 GeV, as these tracks are mostly going straight through the detector with a small curvature. For these tracks, the resolution in the end is dominated by the track fitting algorithms described in Section 5.1, and an initial estimate of up to 6 GeV event for higher momentum tracks is sufficient.

3.2.2 Displaced Vertex Sample

Using specific physics processes with displaced vertices, such as dark matter candidates ($h \to \mu^+\mu^-$) [12] or K^0_S mesons [37] (see Section 3.5), introduces a bias in the training, since the model adapts to the underlying physics parameters.

For example, in the case of the $h \to \mu^+\mu^-$ sample, the two displaced tracks from the h vertex in the chosen parameter configuration combined to a transverse momentum of $p_{\text{t}} = 4.5$ GeV. This can be seen in Fig. 3.5a. When training on these events, the model learned that for vertices with displacements greater than 30 cm, the sum of the track momenta should reproduce this value. As a result, it effectively applied a scaling factor to individual tracks whenever the true combined p_{t} differed.

The same effect is observed when using K^0_S events for training (Fig. 3.5b. Furthermore, due to the lifetime of 2.7 cm[37], not enough statistics are available for displaced tracks over 30 cm.

To avoid any bias, because particle mass, momentum, and opening angle are strongly correlated, I instead generate training samples using random opening angles for each track from the vertex.

At first, a starting position is chosen with v_x, v_y, and v_z in 3D with

$$r^{3D}_{\text{gen}} = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{3.2}$$

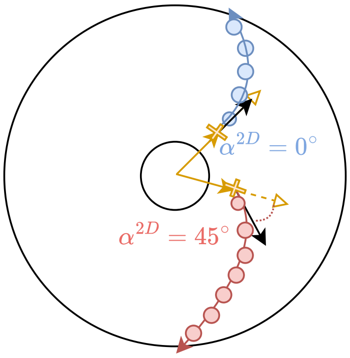


Figure 3.6: Pointing angle for $\alpha^{2D} = 0°$ for the blue particle and $\alpha^{2D} = 45°$ for the red particle. The yellow cross shows the starting position of the particle, where the yellow solid line shows the vector connecting the interaction point and the starting position (the yellow dashed line is the extension), and the black vector shows the particle momentum.

Two particles with opposite charge are generated at this starting position. For each particle, I create a momentum vector, randomly sampled from a uniform distribution in p from 0.05 GeV to 6 GeV. The momentum direction is generated from the rotation angle α_{gen}^{3D} with respect to the vector connecting the origin and r_{gen}^{3D} for a randomly selected perpendicular direction. One example is shown in Fig. 3.6 for the simplified 2D case with α_{gen}^{2D}.
By using this approach, the momentum of both particles is no longer correlated as shown in Fig. 3.5c.

One important track signature is particles flying close together compared to those with a large opening angle. Therefore, I simulate the vertex sample with the large opening angle for category 9, as well as events with small opening angles in category 10.

3.2.3 Transition Sample

I include a transition sample between the prompt tracks and the displaced vertices. This approach improved the training result, as the model was not able to predict tracks from displaced vertices, and in the worst case, the loss also did not converge for this reason. I displace particles for r_{gen}^{3D} in the direction of their momentum, so that the momentum direction and the starting direction align. I also enrich these samples regarding the endcap region, resulting in samples 5-7. Following the vertex sample generation, I also create single, displaced, angled particles with α_{gen}^{3D} in category 8.

3.2.4 Low Momentum Particles

I enrich samples of categories 1-4 with low momentum particles with the momentum p or p_t. To address low momentum track signatures, I randomly sample from a Poisson distribution with $\lambda = 1$. This means that in about 37% of cases the event is left unchanged, in 37% of cases one additional low-momentum track is added, and in around 26% of cases two or more such tracks are added.

In earlier iterations, I included dedicated samples consisting only of low-momentum tracks. However, this led the model to learn a distinction between events with only high- or only low-momentum tracks, rather than capturing a realistic mixture. The reason for this is driven by the beam-background overlays (Fig. 3.1a), which can contain beam-background tracks that are often low-momentum. The probability of obtaining a low-momentum signal track in the 0.05 GeV–6 GeV range is only about 6%, which further reinforced this imbalance. Without careful treatment, the model can confuse these backgrounds with genuine signal tracks. Ideally, truth information for background tracks would be available to resolve this, but including it in simulated beam-backgrounds is computationally expensive, and it is simply not accessible in recorded data, as these would rely on the current track-finding algorithm that struggles with reconstructing tracks that do not originate from the interaction point (Section 5.2).

Since these tracks are background and do not come from the collision, thus not relevant for physics analyses, I explicitly include their signatures in the training data and train the model to classify them as background. Enriching the events with low-momentum particles and creating a mix of topologies proved to be the best approach.

3.2.5 Physics Dataset

While muons are a suitable particle type for training full trajectories through the CDC, as they interact minimally within the Belle II detector material, I also aim to extend the training to different particle species, as they have different material interactions, which lead to different energy losses and trajectories. All particles that do not decay within the Belle II detector are considered Final State Particle (FSP) particles. The ones relevant for tracking are charged particles such as electrons, muons, pions, kaons, protons, and their antiparticles. While heavier particles such as the deuteron are also FSP, these processes are rare and statistically limited. I want to train on events with higher track multiplicities, as for $B\bar{B}$ events in Belle II, an average of 11 charged particles is expected in the CDC. In addition to muons, I generated samples with 1–11 particles consisting of a mixture of electrons, muons, pions, kaons, protons, and their antiparticles, and also increased the average event multiplicity while reducing the momentum range to 0.05–3 GeV. Nevertheless, the model achieved the best performance when trained directly on $B\bar{B}$ events. For interactions of other particle types with the detector material, I give an update on the training samples in Section 10.2, as finishing this was out of scope of this thesis.

Training on $B\bar{B}$ events provides a suitable choice, as these events include realistic momentum and angular distributions. Given the large variety of possible decay channels, it is highly unlikely that the model learns any correlations, particularly as the energy and momentum conservation in the CDC are further influenced by photon contributions. For the training, I use EvtGen [38] to simulate $B^0\bar{B}^0$ and B^-B^+ events.

3.2.6 Training Dataset Composition

For all events in categories 1–8, I generate 1–6 charged particles, drawn from an independent uniform distribution. Examples for the event displays are given in Fig. A.1. For categories 1-4, I use an independent uniform distribution of $0.05 < p_t < 6.0\,\mathrm{GeV}$, with additional low momentum particles as described above, and for categories 5-8 an independent uniform distribution of $0.05 < p < 6.0\,\mathrm{GeV}$. Each sample of categories 1-8 contains 60,000 events.

On average, events in categories 1–8 contain 4.5 particles, corresponding to 276,000 particles per sample. Categories 9 and 10 contain either two, four, or six charged particles, again drawn from an independent uniform distribution, resulting in 120,000 events with a total of 480,000 particles. Category 11 is defined as a mixture of categories 4 and 10. The number of charged particles is drawn from a Poisson distribution with $\lambda = 1.5$ for the full momentum range $0.05 < p_t < 6.0\,\mathrm{GeV}$, and enriched with additional low momentum tracks with $0.05 < p_t < 0.4\,\mathrm{GeV}$ also drawn from a Poisson distribution with $\lambda = 1.5$. The number of vertices is also drawn from a Poisson distribution with $\lambda = 1.5$, resulting in two oppositely charged particles per vertex.

This yields 240,000 events in each of the major category groups: prompt (categories 1–4), displaced (categories 5-8), and vertex (categories 9–10). The category 11 sample contains 300,000 events.

I define two training datasets used throughout this thesis:

***CAT Finder* dataset** This dataset consists of 1,180,000 simulated events from categories 1–11. About 2% of the events are discarded because they either do not contain any particle with a sufficient number of signal hits in the CDC for training. The maximum number of particles in this training set is 15.

***CAT B Finder* dataset** In addition to the *CAT Finder* dataset, this dataset includes category 12 and 13 samples with 120,000 $B^0\bar{B}^0$ and 120,000 B^-B^+ events without requiring a minimum number of particles per event. This introduces an additional category group of high-multiplicity multi-particle events, bringing the total to 1,400,000 simulated events.

For both datasets, I use 80% for training and 20% for validation. As part of the simulation, I generate one dataset overlayed with low simulated beam backgrounds and high data beam backgrounds.

3.3 Technical Evaluation Samples

The evaluation on the training dataset categories is done on a statistically independent dataset from categories 1-9, containing between 90,000 to 150,000 events with over 1 million evaluated events.

I also add neutral kaon $K_S^0 \to \pi^+\pi^-$ events containing one K_S^0 each. The K_S^0 decay vertex is calculated from the nominal K_S^0 lifetime [37] with a uniformly generated transverse momentum of $p_t(K_S^0) = [0.05 - 3]$ GeV. The average transverse decay distance is $v_\rho = 8.24\,\text{cm}$.

All evaluation events use the high data beam backgrounds overlays, except for the inference under varying beam background conditions, where I test the pre-trained model robustness performance on both events with low simulated beam backgrounds and high data beam backgrounds.

3.4 Belle II Physics Evaluation Samples

It is important to evaluate the performance of a tracking algorithm on relevant processes for the Belle II physics program[11]. The cross section of Belle II for various processes at the cms collision energy $\sqrt{s} = 10.58\,\text{GeV}$ is given in Table 3.3. The processes are generated using EvtGen[38], PYTHIA8[39], KKMC[40], Tauola[41] and BabaYaga@NLO[42]. For studies in this thesis, I generate an integrated luminosity of $\int L dt = 2\text{fb}^{-1}$ of $\Upsilon(4S)$, continuum, and tau pairs according to the cross section of Table 3.3 for all algorithms that I evaluate. I use both high data beam backgrounds and low simulated beam backgrounds for the simulation. For the low multiplicity events, due to the large cross section of the Bhabha scattering, I do not generate either of these final states according to the expected cross section, nor do I include additional low multiplicity physics events. Instead I directly simulate 100,000 $e^+e^-(\gamma)$ and $\mu^+\mu^-(\gamma)$ events including initial state radiation and bremsstrahlung.

The following sections go into detail about event selections for multiple analyses performed

Table 3.3: Total cross sections of various physics processes at $\sqrt{s} = 10.58\,\text{GeV}$ at Belle II including continuum, $\Upsilon(4S)$ decays, τ pair, and low multiplicity events including Bhabha scattering and $e^+e^- \to \mu^-\mu^+(\gamma)$ taken from [11], and the generator used for production.

Physics process	**Cross section (nb)**	Generator
	$\Upsilon(4S)$	
$B^0\bar{B}^0$	0.510	EvtGen + PYTHIA8
B^+B^-	0.540	EvtGen + PYTHIA8
	Continuum	
$u\bar{u}(\gamma)$	1.605	KKMC + PYTHIA8
$d\bar{d}(\gamma)$	0.401	KKMC + PYTHIA8
$s\bar{s}(\gamma)$	0.383	KKMC + PYTHIA8
$c\bar{c}(\gamma)$	1.329	KKMC + PYTHIA8
	Tau pairs	
$\tau^+\tau^-(\gamma)$	0.919	KKMC + Tauola
	Low multiplicity events	
$e^+e^-(\gamma)$	295.800	BabaYaga@NLO
$\mu^+\mu^-(\gamma)$	1.148	KKMC

on measured data in the later part of the thesis. Analyses on simulation for low multiplicity events are inaccessible as they would require a large computational effort to simulate sufficient numbers of background samples, and are therefore not included in the thesis.

3.4.1 $e^+e^- \to \mu^+\mu^-(\gamma)$

For the evaluation on simulation, I use KKMC to generate radiative muon pairs $e^+e^- \to \mu^+\mu^-(\gamma)$.

To evaluate di-muon events on measured data, the selections in Table 3.4 are applied on the HLT to sort events during data taking. There are two different HLT selections for di-muon events, one for events where both muons are reconstructed as matched clusters in the ECL, and one for events where only one muon is reconstructed as a matched cluster in the ECL.
The first one, where both muons have matched clusters, requires two tracks in the event with opposite charge, and a back-to-back opening angle between 170° to 190° in both the azimuthal angle ϕ and the polar angle θ. As muons are minimum ionizing particles, the energy of each matched cluster is required to be less than 0.5 GeV, and the total energy from track and photon clusters in the ECL to be less than 2 GeV. Furthermore, the momentum of both tracks is required to be larger than 0.5 GeV.
The second selection, where only one muon is reconstructed with a matched ECL cluster, requires at least two tracks with opposite charge, and also an opening angle between 170° to 190° in both the azimuthal angle ϕ and the polar angle θ between two opposite charged tracks. The matched ECL cluster is required to have an energy less than 1 GeV, and the reconstructed mass of both tracks to be between 8 GeV and 12 GeV (with $\sqrt{s} = 10.58$ GeV). There can be more than two tracks in the second selection.
The events have to pass either one of the two selections to be classified as a di-muon event and are tailored towards $e^+e^- \to \mu^+\mu^-$, where no radiative photon is emitted with high energy.

For the case where an additional photon is radiated, a different HLT selection is applied, given in Table 3.5. The event should have fewer than four tracks. For two tracks in the event, they have to be in the CDC acceptance, with at least one CDC hit, and a matched ECL cluster with less than 0.5 GeV energy. At least one track has to be matched to a KLM cluster, if the second one is not, then the ECL cluster energy is required to be less than 0.25 GeV, and the momentum of the track to be less than 2 GeV. The highest track momentum in the event is required to be larger than 1 GeV, and the lowest track momentum to be less than 3 GeV. The opening angle between the two tracks has to be larger than 90°, and the recoil momentum $p_{\text{recoil}} = |p_{\text{in}} - p_1 - p_2|$, with the initial beam four-momentum p_{in} and the momentum of the two tracks p, of the event larger than 0.1 GeV. Further selections are given in the respective chapter.

Table 3.4: Event HLT selection criteria for $e^+e^- \to \mu^+\mu^-$.

HLT selection	Selection criteria
Matched clusters	two tracks with opposite charge within the CDC acceptance
	matched KLM cluster(s) or ($E_{\text{cluster}} < 0.25\,\text{GeV}$ and $p < 2.0\,\text{GeV}$)
	$\|\|\phi_1^* - \phi_2^*\| - 180^\circ\| < 10^\circ$
	$\|\|\theta_1^* + \theta_2^*\| - 180^\circ\| < 10^\circ$
	each matched clusters with $E_{\text{cluster}} < 0.5\,\text{GeV}$
	total $E^{\text{ECL}} < 2\,\text{GeV}$
	both tracks $p > 0.5\,\text{GeV}$
Missed cluster	at least two tracks with two opposite charge
	$\|\|\phi_1^* - \phi_2^*\| - 180^\circ\| < 10^\circ$
	$\|\|\theta_1^* + \theta_2^*\| - 180^\circ\| < 10^\circ$
	at least one with ECL cluster matched and $E_{\text{cluster}} \leq 1\,\text{GeV}$
	reconstructed mass of both tracks between 8 GeV and 12 GeV.

Table 3.5: Event HLT selection criteria for $e^+e^- \to \mu^+\mu^-\gamma$.

HLT selection	Selection criteria
Additional photon	Number of tracks < 4
	two opposite charged tracks, each $\|d_0^{IP}\| < 2.0\,\text{cm}$ and $\|z_0^{IP}\| < 4.0\,\text{cm}$
	one track $E_{\text{cluster}} < 0.5\,\text{GeV}$, the other $E_{\text{cluster}} \leq 0.25\,\text{GeV}$
	In CDC acceptance with at least one CDC hit
	at least one matched KLM cluster
	matched KLM cluster(s) or ($E_{\text{cluster}} < 0.25\,\text{GeV}$ and $p < 2.0\,\text{GeV}$)
	Highest $p > 1\,\text{GeV}$
	Lowest $p < 3\,\text{GeV}$
	$\|\phi_1 - \phi_2\| \geq 90^\circ$
	Recoil momentum $p_{\text{recoil}} > 0.1\,\text{GeV}$

3.4.2 $e^+e^- \to e^+e^-(\gamma)$

To select Bhabha scattering events ($e^+e^- \to e^+e^-(\gamma)$), several complementary criteria are applied, targeting both track-based and calorimeter-based signatures. The selection logic is summarized in Table 3.6 and detailed below. The event needs to satisfy one of these categories to pass the HLT selection.

For the first category, events are required to have at least two tracks with opposite charge. The three-dimensional opening angle between the two tracks, α^{3D}, must satisfy $\alpha^{\text{3D}} > 165^\circ$, ensuring a back-to-back topology in the center-of-mass frame. Each track must fulfill either a momentum requirement $p/\sqrt{s} > 0.175$, where p is the track momentum, or an energy-over-momentum requirement $E_{\text{cluster}}/p > 0.8$, with E_{cluster} being the energy of the associated ECL cluster. At least one track must be matched to an ECL cluster. The total reconstructed energy in the ECL, defined as the sum of energies of all clusters matched to tracks as well as the remaining unmatched clusters in the event, must exceed 4 GeV.

Alternatively, events can be selected based on calorimeter information alone. Two ECL

Table 3.6: Event HLT selection criteria for Bhabha scattering.

HLT selection	**Selection criteria**
Track reconstruction	at least two tracks with opposite charge two-track opening angle $\alpha^{*3D} > 165^\circ$ $p/(\sqrt{s}) > 0.175$ or $E_{\text{cluster}}/p > 0.8$ and at least one matched track Total reconstructed energy of all clusters > 4 GeV
ECL reconstruction	$165^\circ < \lvert\phi^*_{\text{cluster},1} - \phi^*_{\text{cluster},2}\rvert < 178.5^\circ$ $178^\circ < (\theta^*_{\text{cluster},1} + \theta^*_{\text{cluster},2}) < 182^\circ$ both $E^*_{\text{cluster}}/\sqrt{s} > 0.4$ at least one $E^*_{\text{cluster}}/\sqrt{s} > 0.45$
Radiative Bhabha	two opposite charged tracks in the event, both at least 1 CDC hit both $0.7 < E_{\text{cluster}}/p < 1.3$ CDC ionization loss within 30% of electron expectation

clusters are required, with the azimuthal angle difference $|\phi^*_{\text{cluster},1} - \phi^*_{\text{cluster},2}|$ between 165° and 178.5°, and the sum of their polar angles ($\theta^*\text{cluster},1 + \theta^*_{\text{cluster},2}$) between 178° and 182°, making them back-to-back. Both clusters must have a normalized energy $E_{\text{cluster}}/\sqrt{s} > 0.4$, and at least one must exceed 0.45.

For the last category, two oppositely charged tracks are required, each with at least one CDC hit. Both tracks must have an energy-over-momentum ratio $0.7 < E_{\text{cluster}}/p < 1.3$, and the CDC ionization loss (dE/dx, see Section 7.3.1) must be within 30% of the expectation for electrons, ensuring electron identification and suppressing hadronic backgrounds. Further selections are given in the respective chapter.

3.4.3 J/ψ and K^0_s Decays in High Multiplicity Events

The high multiplicity HLT selection requires at least three tracks in the event with a transverse momentum $p^*_t > 0.2\,\text{GeV}$ and $|d^{\text{IP}}| < 2$ and $|z^{\text{IP}}| < 4$ and the event should not be classified as Bhabha scattering according two the track reconstruction selection in Table 3.6. I evaluate both the J/ψ and K^0_s candidates in high-multiplicity events using measured data. The event selections are described in the respective chapter.

3.5 Dark Sector Evaluation Samples

Another interesting field to study is the search for new physics, especially DM, as described in Chapter 1. For this, the focus in this work is on signatures of long-lived particles decaying into displaced tracks. Two searches performed at Belle II probe such signatures. The first one is the search for a dark Higgs boson produced in association with inelastic dark matter [12]. The second is the search for a long-lived scalar particle in flavour-changing neutral current $b \to s$ transitions [13].

Both searches highlight that the Belle II reconstruction is optimized for short lifetimes, necessary to measure charge-parity violation in B-meson oscillation. However, the reconstruction efficiency decreases for vertices displaced further than the SVD but still inside the tracking volume. As I aim to improve the reconstruction for displaced tracks, I include the samples of these two searches in my evaluation.

3.5.1 Dark Higgs Boson Produced in Association with Inelastic Dark Matter

A complete overview of the search performed in Belle II, including the theoretical foundations, is given in [3]. The underlying model of this search is a two-mediatior model. It includes the dark photon [6] A' that mixes kinetically with the SM photon, and is assumed massive in the Belle II search. As this mixes with the SM photon, it can decay into SM particles or the DM candidate. The DM candidate in this model is a stable Majorana fermion χ_1 [9, 10], that can be excited to χ_2 via the absorption of a massive dark photon, called inelastic DM. The mechanism to give mass to the dark photon is done via a dark Higgs h mechanism similar to the SM. The dark Higgs can then decay in either DM, or via mixing with the SM Higgs boson into SM particles.
The samples used in this thesis are dark Higgs

$$e^+e^- \to A'h(\to \mu^+\mu^-)$$

events, generated using MadGraph5@NLO [43] with one dark Higgs $h \to \mu^+\mu^-$, and a fully invisible decay of a light dark photon A'. The four momentum of the h and A' can be calculated via cms energy with the dark Higgs masses $m_h = [0.5, 2.0, 4.0]$ GeV. The dark Higgs decay vertex $h \to \mu^+\mu^-$ is pointing back to the interaction point. Here, pointing refers to the alignment of the momentum vector of the h aligns with the direction of the vector from the IP to the decay vertex of the h . The dark Higgs decay vertex position is drawn randomly from a uniform r^{3D}_{gen}-distribution up to 100 cm to ensure enough statistics of the parameter space of very displaced vertices.
I generate one additional sample near the expected Belle II sensitivity given in [12] with $m_h = 1.5\,\text{GeV}$ and the kinetic mixing angle θ with $\sin(\theta) = 10^{-4}$ and a lifetime of $c\tau = 21.5\,\text{cm}$.

Additionally, I include samples following the full decay chain in [12], given by

$$e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1),$$

with the masses $m_h = [0.5, 1.0, 2.5]\,\text{GeV}$, $m_{\chi_1} = 2.5\,\text{GeV}$, and an virtual A' $m_{A'} = 3m_{\chi_1}$. Here, two displaced vertices are present, one from $h \to \mu^+\mu^-$ pointing back to the interaction point, and one from the three-body χ_2 decay, where the reconstructable vertex from the visible SM particles e^+e^- is non-pointing. Non-pointing refers to the case where the momentum vector and the vertex vector of the χ_2 do not align. Both decay positions are

distributed uniformly in r^{3D}_{gen} up to 100 cm to cover large displacements.
These samples provide both pointing and non-pointing displaced vertices, making them ideal to test the reconstruction efficiency of my algorithm in a low multiplicity track environment.

3.5.2 Long-lived Spin-0 Particle in $b \to s$ Transition

A complete overview over the $B \to KS$ search performed in Belle II including the theoretical foundations is given in [44]. Here, a light scalar S is produced in $b \to s$ transitions. The $b \to s$ transitions within the SM are suppressed [45], which makes $B \to KX$ decays rare and an opportunity to search for small deviations with respect to the SM. The S exhibits a displaced vertex signature in combination with the kaon, resulting in a clean experimental environment with low background. I consider decays for the light scalar for $S \to \mu^+\mu^-/\pi^+\pi^-/K^+K^-$. In contrast to the inelastic dark matter case from the previous section, which involves only four charged particles in the final state, B decays at the $\Upsilon(4S)$ resonance contain on average about six charged particles in addition to the three particles forming the signal B-meson. This provides a high-multiplicity track environment to test the performance of my track reconstruction algorithm for displaced tracks under more complex event conditions.
The signal sample is generated with EvtGen[38], using the decay models VSS for $\Upsilon(4S) \to B^+B^-$ of a vector particle to two scalar particles, and PHSP (two-body phase space) for both the $B^+ \to K^+S$ and $S \to x^+x^-$ generic decays. The S is generated first with a lifetime of zero and then displaced according to $c\tau$. The lifetimes for the signal samples are simulated for a reduced range in comparison to [13] with

$$c\tau = [0.01, 0.1, 1.0, 0.5, 5.0, 10.0, 50.0, 100.0, 10^3, 10^4]\ \text{cm}.$$

I will follow the Belle II convention using m for the reference mass and M for the measured mass throughout this thesis. The masses simulated for the signal events, where the final states are kinematically accessible, are

$$m_S = [0.22, 0.4, 0.9, 1.4, 1.9, 2.4, 2.9, 3.4, 3.9, 4.4]\ \text{GeV}$$

for the $S \to \mu^+\mu^-$,

$$m_S = [0.4, 0.9, 1.4, 1.9, 2.4, 2.9, 3.4, 3.9, 4.4]\ \text{GeV}$$

for $S \to \pi^+\pi^-$ and

$$m_S = [1.0, 1.4, 1.9]\ \text{GeV}$$

for the $S \to K^+K^-$ final state. There are only three mass points for the $S \to K^+K^-$ due to computational limitations.. These within the mass range relevant for the model described in [13]. The event selection is given in the respective chapter.

Chapter 4

Metrics

The metrics to evaluate the performance of a new tracking algorithm are defined in [1] and repeated closely following [1] here. The first step is to find out if a reconstructed track can be matched to the simulated truth. Given a simulated particle with a number of hits matched to this particle, the hit efficiency ε_{hit} is defined as the number of hits matched to this particle and included in the track divided by the total number of hits matched to this track:

$$\varepsilon_{\text{hit}} = \frac{n_{\text{hits}}(\text{matched and} \in \text{track})}{n_{\text{hits}}(\text{matched})}. \tag{4.1}$$

When all hits matched to the simulated particle are included in the track, this results in perfect hit efficiency of 1.0. The next step is the hit purity $\mathfrak{p}_{\text{hit}}$ per track. This is the number of hits matched to the particle and included in the track, divided by the total number of hits included in the track:

$$\mathfrak{p}_{\text{hit}} = \frac{n_{\text{hits}}(\text{matched and} \in \text{track})}{n_{\text{hits}}(\in \text{track})}. \tag{4.2}$$

The hit purity of a track is 1.0 if all hits included in the track are matched to the same particle.

To relate a track to a simulated particle, a minimum hit efficiency of $\varepsilon_{\text{hit}} > 0.05$, hit purity of $\mathfrak{p}_{\text{hit}} > 0.66$ and a minimum 7 hits to be included in the track.

The reason for the low hit efficiency is due to tracks that curl within the tracking volume, thus leaving many hits. The hit purity requirement ensures that at least 66% of the hits need to be associated with a single unique particle. The hit efficiency and hit purity requirement are calculated using only CDC hits for the CDC only evaluation, and all hits of the tracking detectors for the full reconstruction evaluation.

If multiple tracks are related to a particle, the track with the highest purity is chosen to be matched. If two tracks have the same hit purity, the highest hit efficiency is chosen. The other tracks are defined as clone tracks. If a track does not achieve the hit efficiency and hit purity requirements, it is defined as a fake track.

The track efficiency ε_{trk} can now be defined using the matched tracks to the number of simulated particles that leave at least one hit in the tracking detector:

$$\varepsilon_{\text{trk}} = \frac{n_{\text{trks}}(\text{matched to part.})}{n_{\text{simulated}}(\geq 1\text{matched hit})}. \tag{4.3}$$

The track charge efficiency $\varepsilon_{\text{trk,ch}}$ is the most important performance metric, where, in addition to the track being matched to the simulated particle, it also needs to have the correct charge:

$$\varepsilon_{\text{trk,ch}} = \frac{n_{\text{trks}}(\text{matched to part., corr. charge})}{n_{\text{simulated}}(\geq 1\text{matched hit})}. \tag{4.4}$$

The track purity $\mathfrak{p}_{\text{trk}}$ is defined as the number of matched tracks to the total number of tracks:

$$\mathfrak{p}_{\text{trk}} = \frac{n_{\text{trks}}(\text{matched to part.})}{n_{\text{trks}}}. \tag{4.5}$$

The clone rate is the ratio of the clone tracks to all tracks that are related to a simulated particle

$$\mathfrak{r}_{\text{clone}} = \frac{n_{\text{clone trks}}}{n_{\text{tracks}}(\text{related to part.})}, \tag{4.6}$$

and the fake rate the ratio of the fake tracks to the total number of tracks

$$\mathfrak{r}_{\text{fake}} = \frac{n_{\text{fake trks}}}{n_{\text{trks}}}. \tag{4.7}$$

The wrong charge rate is defined by the matched track that has a wrong charge prediction dived by the number of matched tracks.

$$\mathfrak{r}_{\text{wrong ch.}} = \frac{n_{\text{trks}}(\text{matched to part., wrong ch.})}{n_{\text{trks}}(\text{matched to part.})}. \tag{4.8}$$

As tracks can be found, but fail the fitting step, the metrics above are distinguished according to this difference:

Track finding performance metrics The track finding efficiency, track charge finding efficiency, track finding clone rate, track finding fake rate and wrong finding charge rate are calculated with the track objects after the track finding

Track fitting performance metrics The track fitting efficiency, track charge fitting efficiency, track fitting clone rate, track fitting fake rate and wrong fitting charge rate are calculated with the track objects after the track finding and track fitting.

These parameters are defined as the performance metrics for the CDC only evaluation. The full reconstruction efficiency of different FSP (f) is defined as

$$\text{Efficiency}(f) = \frac{n_{\text{trks}}(\text{matched to part. } f\text{, corr. charge})}{n_{\text{simulated } f}(\geq 1\text{matched hit})}. \tag{4.9}$$

with track objects after the full reconstruction.
The track momentum resolution is evaluated using the normalized residuals of the transverse and longitudinal momentum components, p_t and p_z, by comparing the reconstructed and simulated values for matched tracks:

$$\eta(p_{t,z}) = \frac{p_{t,z\,\mathrm{rec}} - p_{t,z\,\mathrm{simulated}}}{p_{t,z\,\mathrm{simulated}}}. \tag{4.10}$$

For an unbiased reconstruction, the distribution of $\eta(p_{t,z})$ is expected to peak at zero. The resolution $r(p_{t,z})$ is defined as the 68% coverage of the normalized residual distribution,

$$r(p_{t,z}) = P_{68\%}\left(\left|\eta(p_{t,z}) - P_{50\%}(\eta(p_{t,z}))\right|\right), \tag{4.11}$$

where P_q denotes the qth quantile of the $\eta(p_{t,z})$ distribution, with $P_{50\%}$ corresponding to the median [46]. In the case of a normal distribution, this definition is equivalent to the standard deviation.

Chapter 5

Track Reconstruction Algorithms

In this chapter, I discuss the track reconstruction algorithms. Conventional track finding approaches typically start from seeds placed either at the inner or outer boundary of the tracking detector volume and then iteratively build the particle trajectories, for example using a Combinatorial Kalman Filter (CKF)[47]. This strategy cannot be applied to the CDC. Track finding cannot start from the outside, since a large fraction of tracks never reach the outermost layers. Moreover, the inner and outer axial layers are parallel to the z-axis and therefore provide no z-information, only by combining them with the stereo layers can a full three-dimensional trajectory be reconstructed. For these reasons, a different approach is used at Belle II.

The reconstruction is split into two stages. In the first stage, hits originating from the same particle are grouped together, including an initial estimate of the track kinematics. In the second stage, a dedicated track fitting algorithm refines this estimate by performing a fit. I begin by describing the track fitting algorithm in Section 5.1, since its requirements strongly influence the design of the track finding algorithms. I then present the *Baseline Finder*, currently used for track finding in the CDC at Belle II and implemented in basf2, in Section 5.2. In Section 5.3, I introduce the new GNN-based track finding algorithm, which is published in [1]. Plots, text, and tables closely reference the paper. Finally, I provide a summary of the full track reconstruction chain, including all three tracking detectors, in Section 5.4.

5.1 Track Fitting

The CDC track finding algorithm needs to provide the following for the subsequent track fitting algorithm:

- an initial estimate of the track parameters, which include the track starting point, the momentum and the charge,
- a set of ordered hits from the starting point,

- an initial estimate of the covariance matrix of the track parameters.

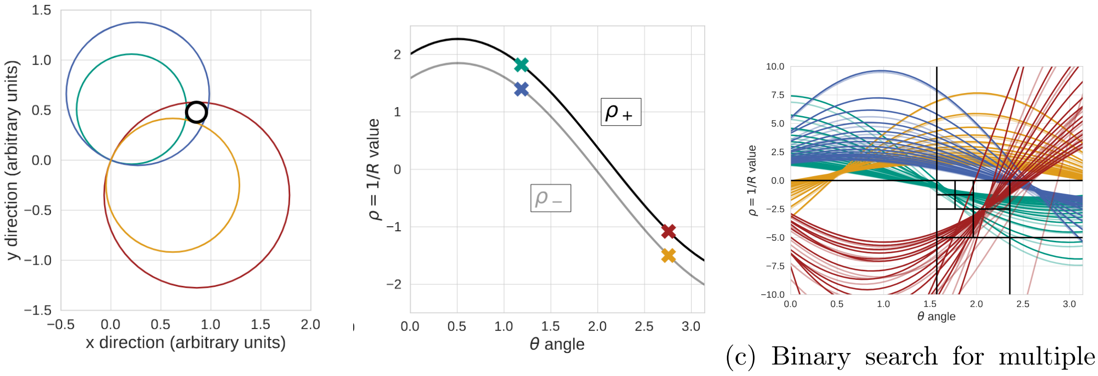
The track fitting is performed using Kalman Filter algorithms implemented in GENFIT2 [48–51]. This algorithm recursively predicts the next state of a particle, including the position, momentum, and the covariance according to the seed. The Runge-Kutta-Nyström method [52] is used to handle material effects such as energy loss or multiple scattering, as well as non-uniform magnetic fields to extrapolate the particle's trajectory. By comparing this prediction of the particle movement with the actual measurement, a new prediction is obtained for all three parameters, which each hit improving the track parameters. This approach is very dependent on the ordering. As it might be that the track finding algorithms assign wrong hits to the track that come from other particles or beam-background, this decreases the track fit and can result in the fit failing. To prevent this, in addition a Deterministic Annealing Filter (DAF) is used to add weights to the hits in the trajectory. This downweights hits far away from the fitted trajectory, with hits being removed from the trajectory. If too many hits are rejected by the DAF, the track fit can fail. There are three particle mass hypotheses used in the Belle II track fitting step used to calculate the energy loss and material effect: pions, kaons, and protons. The latter two are added as these improve the resolution for kaons and protons. As I focus on muons and pions in Chapter 6, I only use the pion mass hypothesis for the fitting for the CDC-only results in Chapter 6. For the chapters from Chapter 7 on, the kaon and proton mass hypotheses are used in addition.

5.2 Baseline Track Finding

The current Belle II tracking algorithm is described in detail in [46]. The track finding operates in two steps, a global Legendre transformation [53] and a local cellular automaton [54]. Both of these approaches use pre-filtered hits, using a cut based approach based on TDC, ADC and TOT (see Section 2.3).
The global Legendre approach operates assuming that the particle trajectory originates close to the IP, which makes it particularly suited for prompt tracks. This algorithm operates only on the hits from the five axial superlayers of the CDC. The method is conceptually related to the Hough transform, where circles in a plane are mapped into lines in a parameter space. In the case of charged particles in a magnetic field, neglecting the energy loss, the trajectories are described as circles in the x–y plane. To simplify their description, a conformal transformation is applied to the hit coordinates

$$u = \frac{2x}{x^2 + y^2}, \quad v = \frac{2y}{x^2 + y^2}, \tag{5.1}$$

which maps circles passing through the origin into straight lines in the (u, v) plane, requiring they pass through the IP at (x, y) =(0,0). Unlike point-like hits from the silicon detectors, either directly from PXD clusters or the combination of the silicon strips, the measurement



(a) Drift circle with four potential trajectories in $x-y$.

(b) Legendre transformation of the drift circle in $\rho-\theta$.

(c) Binary search for multiple drift circles from four different particles in an event.

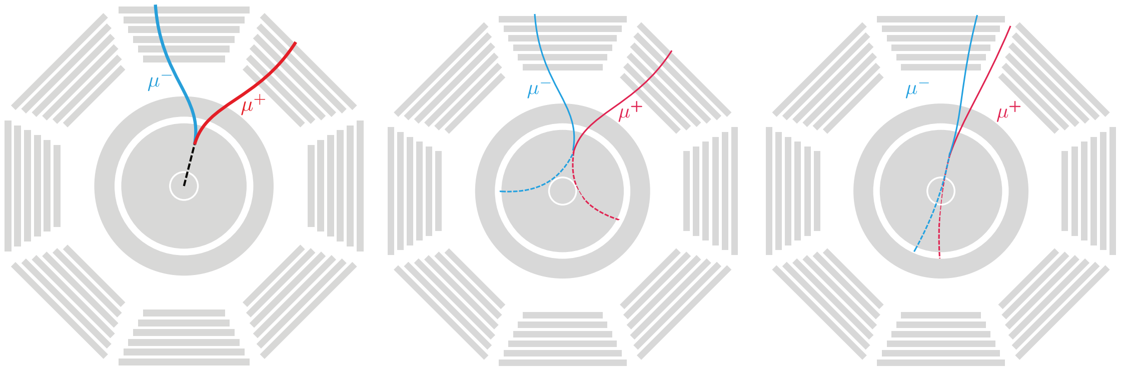
Figure 5.1: The left plot shows a drift circle in black, and two potential trajectories for the left-hit hypothesis in red and green, and for the right-hit hypothesis in orange and blue. All four trajectories are going through the IP at (0,0). The middle plot shows the Legendre transformation of the drift circle under the left hit hypothesis for ρ_+ and the right hit hypothesis ρ_-. The right plot visualizes multiple drift circles coming from four different particle according to their colour. All figures are taken from [55].

of the CDC corresponds to the drift circle with radius d around the wire position (x, y) (Section 2.3). To account for this, the Legendre transformation is used to describe the two tangents to each drift circle, corresponding to the left-hit (+) and right-hit (-) hypotheses. An example for four different, tangent trajectories for a single drift circle are shown in Fig. 5.1a. The transformation into the Legendre parameter space (ρ, θ) is given by

$$\rho_{\pm}(\theta) = \frac{2}{x^2 + y^2 - d^2} \left(x \cos(\theta) + y \sin(\theta) \pm d\right), \qquad \theta \in [0, \pi] \tag{5.2}$$

shown in Fig. 5.1b for the example drift circle. If several drift circles originate from the same trajectory, their $\rho_{\pm}(\theta)$ curves intersect at a common point in parameter space. After transforming all drift circles that passed the background filter in the event, a binary search in the in (ρ, θ) space is performed to find all sets of tracks as shown in Fig. 5.1c The main advantage of the Legendre algorithm is its robustness against missing hits along the trajectory, which may occur due to reduced wire efficiency or masked boards (see Section 3.1.2). However, since it relies on the assumption that tracks pass close to the IP, it is not suitable for particles originating from displaced vertices, as illustrated in Fig. 5.2a. In such cases, the extrapolated trajectories do not cross the IP as seen in Fig. 5.2b, and the Legendre transformation cannot describe or reconstruct them. This effect is most pronounced for particles with low transverse momentum and large opening angles. For high-momentum particles with small opening angles, the trajectory may by chance pass through the IP, in which case the algorithm can still find the track (Fig. 5.2c).

To reconstruct the full three-dimensional trajectory, hits from the stereo superlayers must



(a) Neutral particle decaying displaced in two charged muons. (b) Extrapolated trajectories for low momentum particles. (c) Extrapolated trajectories for high momentum particles.

Figure 5.2: The left figure shows a displaced vertex decaying in two opposite-charged muons traveling outwards of the detector according to their solid lines. The middle figure adds the extrapolated trajectory with a dashed line for each particle, showing that the trajectory misses the interaction point. On the right, the opposite-charged muons have a higher momentum and a smaller opening angle, which results in their extrapolated trajectories passing by the IP.

be incorporated in a second step. Since stereo wires are tilted with respect to the z-axis, their measured (x, y) position depends on the assumed z coordinate and cannot be assigned uniquely. While stereo hits are added in a second step to allow the z-determination, it may happen that parts of the trajectory are missing due to the tight constraints of the Legendre transformation. To address this, a local Cellular Automaton algorithm is used to identify short track segments within each superlayer by linking neighbouring hits. These segments are then combined with the Legendre track candidates using a Boosted Decision Tree (BDT). During this process, the stereo hit x- and y-reference position assuming $z = 0$ (see Fig. 2.5) is updated, to the value where the (x, y) position of the stereo hit aligns with the two dimensional trajectory from axial hits.

Finally, the Legendre algorithm fits the axial hits to a circle in the transverse plane, providing an initial estimate of the curvature and thus the transverse momentum. A subsequent linear fit in the ρ–z plane using the stereo hits yields the estimate of the longitudinal momentum.

5.3 Graph Neural Network based Track Finding

Machine learning (ML) is widely used in high energy physics (HEP), and has been shown to improve performance in multiple areas. An up-to-date overview is provided in the living review [56]. The task of recognizing an unknown number of objects, in this case the tracks from the detector hits, is called object detection [57] in computer vision. Object detection is typically performed using convolutional neural networks (CNNs) [58] or transformer-based methods [59], which rely on Euclidean, grid-like inputs such as images. Due to the non-uniform and irregular structure of the CDC, especially for the axial and stereo superlayers, CNNs are not suitable for track reconstruction in Belle II. Graph representations are better suited for such data, with nodes representing objects and edges representing their relationships. GNNs operate directly on graphs and have been applied in multiple domains like materials science and chemistry [60] and medicine [61]. A comprehensive overview of GNN methods and applications is provided in [62].

In HEP, GNNs are used to model irregular detector geometries and have successfully learnt latent representations [63, 64], which will be the basis of the algorithm presented in this thesis. They have been applied to combine information from subdetectors for particle reconstruction [65], electromagnetic calorimeters [66, 67], and decay chain reconstruction [68]. Significant efforts on track reconstruction using GNNs [69–72] have been driven by the TrackML challenge [73, 74]. However, the use case is different as for Belle II, as these algorithms target the very high-multiplicity environments of the Large Hadron Collider (LHC) experiments. In these experiments, there are significantly more tracks compared to Belle II, a higher fraction of sensor hits belonging to the tracks, and simpler track kinematics, as the tracks are coming from the interaction point with high transverse momentum. Studies on track reconstruction algorithms that are detector independent typically focus on idealized detector simulations and tracks with simple kinematics [71, 75], that are not displaced. For gaseous detectors, GNNs have been explored for edge classification (PANDA [76], BES III [77]) and hit classification (COMET [78]), although none address complex event topologies. Additionally, these approaches require multiple additional steps to extract the tracks with their respective parameters and currently none provide end-to-end ML solutions that are proven to be less resource demanding while achieving the same performance [79–81]. Drift detectors as the CDC differ significantly from pixel and strip detectors used at the LHC and High Luminosity LHC (HL-LHC), which provide 3D spatial information. In contrast, the CDC uses indirect measurements of the drift time and wire position (see Section 2.3), making track reconstruction more complex due to the lack of spatial information as described earlier in Section 5.2. Beyond GNNs, other ML methods, such as large language models [82] and transformers [83], are also explored for tracking tasks.

In addition to handling irregular input, GNNs offer practical advantages. The occupancy of the CDC varies significantly with up to 15% per event, depending on beam background conditions (Section 3.1.1) and the number of signal tracks.
As the track finding algorithm needs to run on the CPU-only HLT, it is required to have less than 2 GB memory consumption and $\mathcal{O}(100)$ ms processing time per event. GNNs can naturally process variable-sized inputs without the need for any sorting or padding, i.e., the addition of zero-values to standardize input dimensions. This reduces computational overhead and makes GNN more memory efficient than transformer models and, in return,

a suitable choice for track reconstruction in Belle II. I will go into detail about the input features and the targets for the GNN in this work in Section 5.3.1, and follow with the detailed explanation of the chosen GNN architecture and training loss, the postprocessing, and the hyperparameter optimization afterwards.

5.3.1 Input Features and Prediction Targets

The most relevant input feature is the position of the CDC hits. Because the CDC is read out only from the backward end of the wires, the z-position of a hit along the wire cannot be determined (see Section 2.3). The stereo wires (Fig. 2.5) cover a large range in the x–y projection, so a fixed reference position along each wire has to be chosen for use as input to the GNN model. I consider two reference positions:

$z = 0$: the position of the x- and y-coordinates of the wires at $z = 0$.

middle of the wire: x- and y-coordinates at the middle of the wire as the reference position.

The difference on the corresponding x- and y-coordinates is shown in Fig. 5.3 for particles with different θ angles ranging from $20\,^\circ$ to $135\,^\circ$.

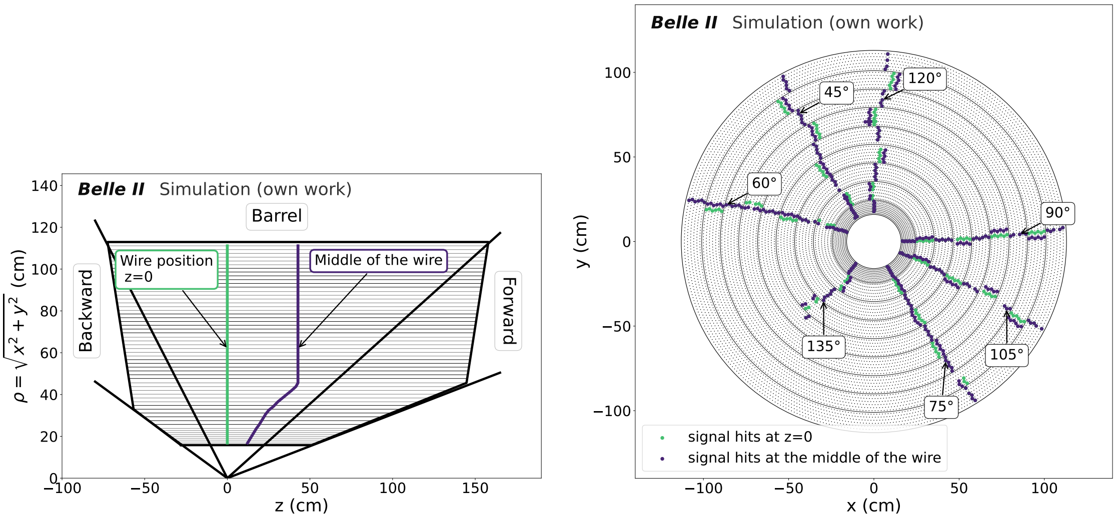
The middle of the wire position has the benefit that as the collisions' cms is boosted towards the forward direction, which results in a smaller shift between the axial and stereo layers for the majority of the tracks. This can be seen as the stereo layers are aligned for $\theta \approx 60\,^\circ$ for the middle of the wire in Fig. 5.3b.

The $z = 0$ position simplifies the learning task. The model does not have to learn the z-dependence of the wire map shift for the middle of the wire as shown in Fig. 5.3a, but instead can more easily translate from the stereo layers to the momentum in the z-direction and the starting position of tracks in the z-direction. This was shown in [84], where the track charge fitting efficiency increased on the technical samples (Section 3.2 from 91%, when using the middle-of-the-wire reference position, to 92% when using the $z = 0$ reference position.

Since the technical samples do not include the forward boost, the difference is expected to be less pronounced in physics samples. A detailed evaluation of measured data will therefore be required in future studies.

While polar coordinates are a natural choice for representing the wire geometry, angle-based representations introduce discontinuities at 360°. This abrupt transition from 360° to 0° made it difficult for the model to learn. Using Cartesian coordinates for the input features was crucial for model's performance.

To account for the loss of radial information when using Cartesian coordinates for the input features, additional wire descriptions are provided. Specifically, each wire has as input its layer number (ranging from 1 to 56), its superlayer (from 1 to 9), and the local layer



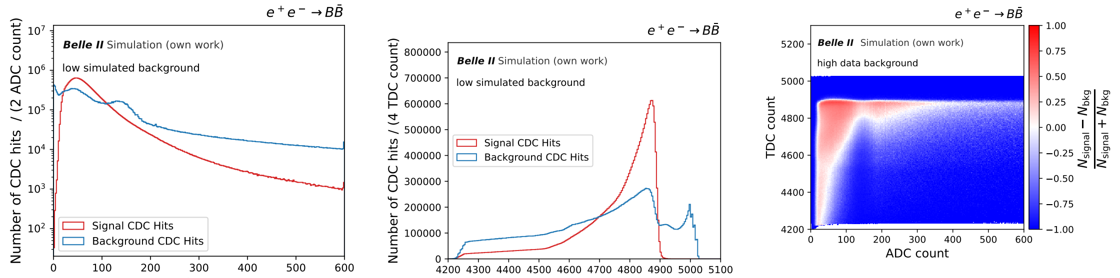
(a) Reference position in $\rho - z$ view.

(b) Reference position in $x - y$ view.

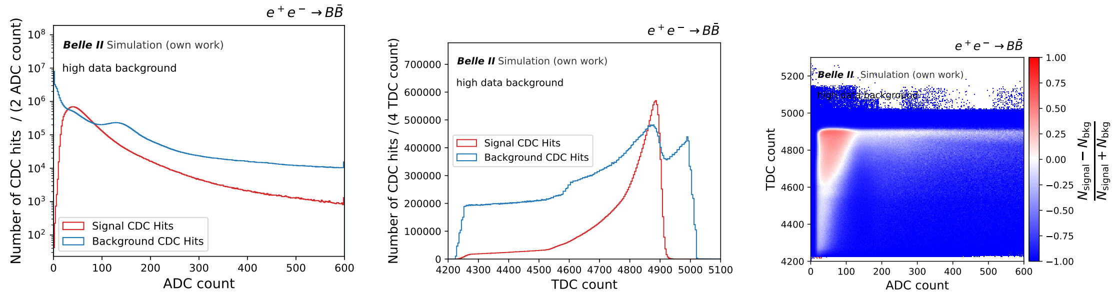
Figure 5.3: Reference z position for the approximated wire position shown in $\rho - z$ view in Fig. 5.3a and $x - y$ view for different θ angles in Fig. 5.3b for $z = 0$ in green and z in the middle of the wire in blue.

number within each superlayer (see Section 2.3). This structured representation captures the radial position of the hit without introducing the issues associated with continuous geometric parameters. Additionally, the digitized measurements of the wire signal are used as input with the ADC and TDC, as described in Section 2.3. The signal and background distributions are shown in Fig. 5.4, for both low simulated beam backgrounds and high data beam backgrounds. As the simulation does not correctly capture cross-talk, not only the yield but also the shapes differ between the simulated and measured background distributions. While separating the signal and background distribution is not possible for only ADC or TDC, the combination provides separation power as seen in Fig. 5.4c and Fig. 5.4f. Furthermore, the distributions for different particle types for both ADC and TDC are shown in Fig. 5.5. While the TDC distribution does not differ too much, as the time-of-flight is $\mathcal{O}(100)$ less than the drift-time, the ADC differed between the particle types. Especially K^+ and p deposit more energy in the drift cells ([30]), which leads to higher ADC counts. While I considered the drift-time and drift-length, these values are very dependent on the calibration (discussed in detail in Section 10.1), and using the raw digitized measurements proved to be more efficient (investigated again in [84]).

A dedicated study about the input features to use during the training of the algorithm is reported in the bachelor thesis I supervised in [84]. Although TOT was considered as an input feature in [84], I encountered issues with this feature on measured data. I will discuss the details in Section 10.1. Next, I define the prediction targets. Particle trajectories within Belle II are parametrized using the point of closest approach (POCA)[85], including d_0 as distance from the IP to the POCA in the transverse plane and z_0 for the z-distance. Furthermore, their momentum is defined with the inverse of the curvature radius and the

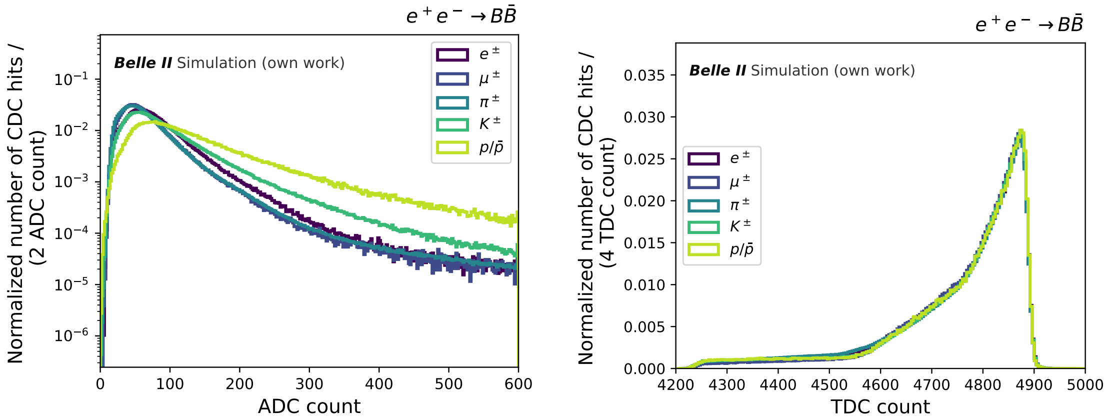


(a) low simulated beam backgrounds ADC (b) low simulated beam backgrounds TDC (c) low simulated beam backgrounds TDC over ADC asymmetry



(d) high data beam backgrounds ADC (e) high data beam backgrounds TDC (f) high data beam backgrounds TDC over ADC asymmetry

Figure 5.4: Input features of the simulated samples with low simulated beam backgrounds (top) and high data beam backgrounds (bottom) in $B\bar{B}$ events. The left shows the ADC counts, the middle the TDC counts, where the background distribution is in blue and signal distribution in red. The asymmetry between the signal hits N_{signal} and background hits N_{bkg} is defined as $\frac{N_{\text{signal}}-N_{\text{bkg}}}{N_{\text{signal}}+N_{\text{bkg}}}$ for TDC over ADC is shown on the right, with blue having more background hits and red having more signal hits in the bin.



(a) ADC (b) TDC

Figure 5.5: Normalized distribution of the ADC (left) and TDC (right) count distribution from different signal particles in $B\bar{B}$ events.

angle of the transverse momentum at the POCA. The z-direction is defined by the angle

of the momentum at POCA and the transverse plane.
Using the helix radius of particle tracks as a training target proved ineffective: for high transverse momentum ($p_t > 6\,\text{GeV}$) the radius diverges, while for low-p_t tracks it becomes very small, which introduces numerical instability in the prediction targets. This wide dynamic range made it impossible to normalize the target values, and therefore impossible for the model to learn. Instead, I train the model on the momentum vector p_x^{MC}, p_y^{MC} and p_z^{MC}. Even though these outputs range from 0 to 6 GeV, scaling them did not improve the model performance, therefore I keep them unscaled for convenience. Cartesian coordinates are used as prediction targets for the starting position vector described with v_x^{MC}, v_y^{MC}, and v_z^{MC}. Cartesian coordinates are chosen for the same reason here as for the input features. In addition, the model has to predict if a particle has a positive charge or a negative charge. The final set of input features and prediction targets used for the model is given in Table 5.1. All input features are scaled to ensure they are suitable for model training. The x- and y-positions are originally given in centimeters and are converted to meters to bring their values into a consistent range of approximately -1.1 to 1.1. All other features are normalized to the [0,1] range by first shifting the distribution by its minimum value, and then dividing by the resulting respective maximum values. To limit the influence of outliers and focus on the region relevant for the signal hits, as shown in Fig. 5.4a and Fig. 5.4d, the ADC value is clipped at 600, where the signal hits are mostly between 25 and 300.

Table 5.1: Final set of input features and prediction targets, including their ranges before and after scaling.

Feature	Original range	Scaled range
Input features		
x at wire midpoint (cm)	$[-111, 111]$	$[-1.11, 1.11]$
y at wire midpoint (cm)	$[-111, 111]$	$[-1.11, 1.11]$
Global layer	[0, 56]	[0, 1]
Superlayer	[0, 8]	[0, 1]
Local layer	[0, 7]	[0, 1]
ADC (counts)	$[0, 600]^*$	[0, 1]
TDC (counts)	[5100, 4200]	[0, 1]
Prediction targets		
p_x^{MC} (GeV)	$[0, 6]$	-
p_y^{MC} (GeV)	$[0, 6]$	-
p_z^{MC} (GeV)	$[0, 6]$	-
v_x^{MC} (m)	$[-111, 111]$	$[-1.11, 1.11]$
v_y^{MC} (m)	$[-111, 111]$	$[-1.11, 1.11]$
v_z^{MC} (m)	$[-73, 157]$	$[-0.73, 1.57]$
Charge$^{\text{MC}}$ (sign)	$\{-1, +1\}$	$\{0, +1\}$

5.3.2 Graph Neural Network Model

There are three major tasks the algorithm has to achieve:

1. identify the unknown number of tracks in the event;
2. predict the three-momentum, starting point, and the charge of each track;
3. assign the associated hits to each track.

Due to not knowing the number of tracks beforehand and the need to handle overlapping tracks, I use an object condensation loss [14], which predicts the number of objects (tracks), and associated features (track parameters) in a GNN-based one-shot approach.

GravNet

GNNs operate on graph-structured data, where nodes represent detector hits and edges define relationships between them, allowing information to be exchanged during the message passing step. This graph construction is a crucial preprocessing step, but constructing a graph during the transition between axial and stereo superlayers is challenging. This is due to a significant offset and a change in sign in the spatial gap between particles moving in the forward versus backward directions, particularly noticeable between 45° and 120°, as shown in Fig. 5.3b. Additionally, this offset varies depending on the reference z-position of the input features.

GravNet [64] addresses these issues by allowing the model to dynamically learn meaningful connections between nodes, removing the need for manually designing complex and geometry-dependent graphs. An overview of GravNet is illustrated in Fig. 5.6.

In the first step, the inputs to this layer are transformed into a learnable representation space S, and learned feature space F_{LR}.

In the representation space S, the n_S-dimensional Euclidean distance between each node j and all other nodes is calculated, and edges are built for node j with the k nearest neighbours, visualized in step two.

In the third step, the learned features f_{LR} of all the connected nodes are exponentially weighted by their Euclidean distance and aggregated by summation.

These aggregated node features are concatenated with the initial node features and the learned features in the final step four.

Architecture

The architecture of my model is shown in Fig. 5.7, which is implemented in `PyTorch Geometric` [86].

The model input is every wire hit as a graph node, with the input features defined in Section 5.3.1 The first layer is a Batch Norm Layer (BNL) [87], followed by $N = 4$ GravNet Blocks. Each GravNet Block starts with global average pooling [88], where the

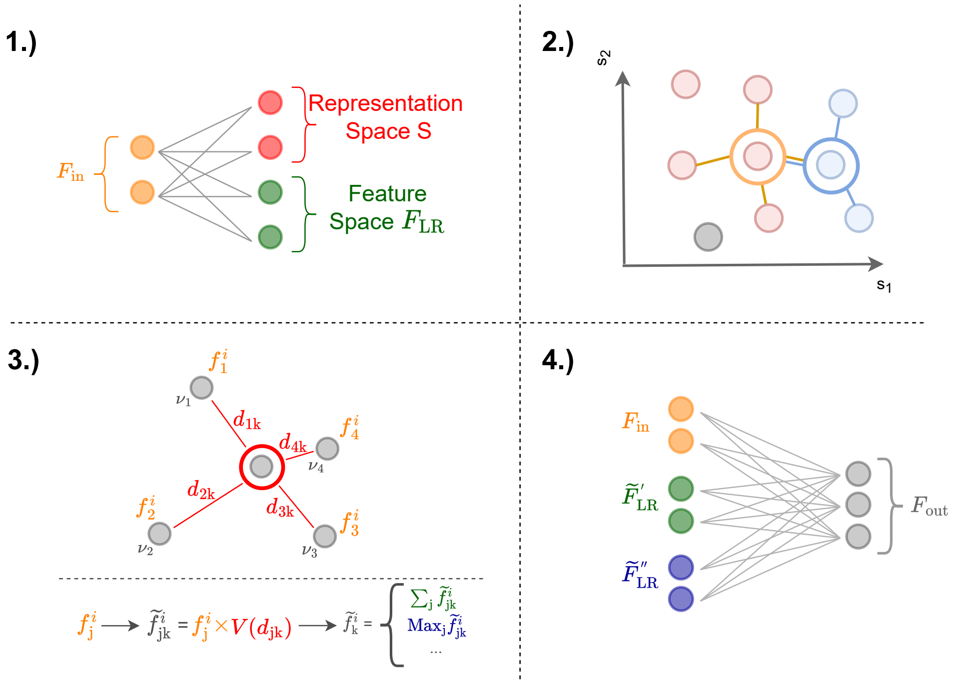


Figure 5.6: This is an overview over the GravNet layer, taken from [66]. See the text for details.

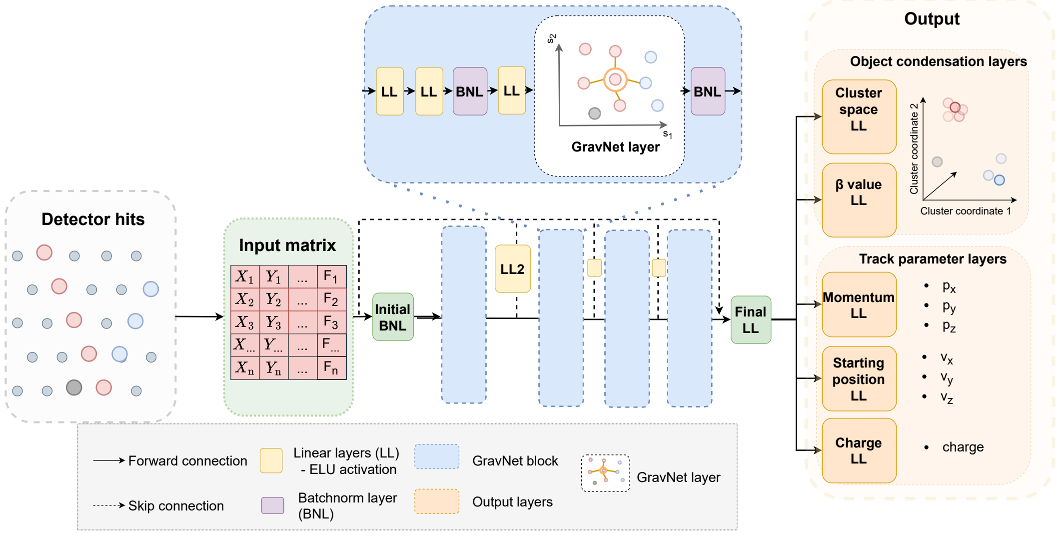


Figure 5.7: Figure and text taken from [1]: illustration of the GNN architecture.

mean value for each feature per event is calculated. Each mean feature is then appended to the original node features, so that a global information exchange of the event graph is possible, thus improving the model learning. After the pooling layer, the network includes two consecutive Linear Layers (LLs), followed by a BNL and another LL. All LLs use the exponential linear unit (ELU) activation function [89]. The feature transformation follows a GravNet layer, responsible for the graph building and message passing between the nodes. The output of the GravNet layer is propagated through another BNL and then forwarded to the next GravNet block. Skip connections [90] prevent vanishing gradients in the earlier layers of the model. With these skip connections, the output of each GravNet block is passed through an extra LL (LL2) directly to the final linear layer.

There are five output layers of the model, addressing both the track assignment and the parameter prediction task.

Object Condensation

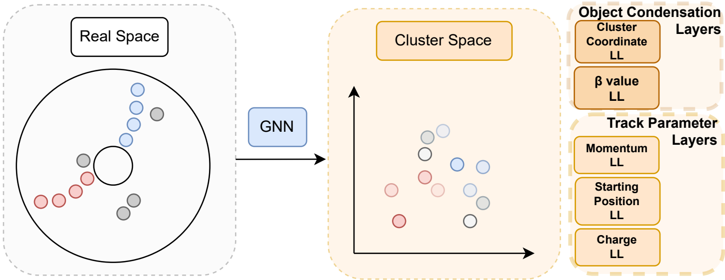
In computer vision, object detection often uses anchor boxes and bounding boxes to identify and separate objects, resolving overlaps by analyzing bounding box intersections. These methods rely on distinct object centers, making them less suitable for tracking, where tracks can overlap spatially and lack well-defined centers. They are also sensitive to object size and density [79], which vary widely in Belle II, with tracks ranging from a few to over a thousand measured CDC hits per signal particle. Traditional seed-based or multi-step clustering approaches can lose information due to successive thresholds and segmentation stages. Object condensation [14] was proposed as a one-shot approach for calorimeter clustering in sparse, overlapping environments without requiring fixed cluster counts or centers.

I visualize the steps of object condensation in Fig. 5.8 for two particles with their associated hits with additional background hits. If a CDC hit results from the energy deposition of a signal particle, the corresponding node is assigned a unique integer object ID per particle, starting from 1. In contrast to the simplified schematic in Fig. 5.8, for the training only signal particles with at least 7 matched hits in the event are labelled signal. If the CDC hit is associated with any signal particle that has less than 7 matched hits, the node is labelled as background.

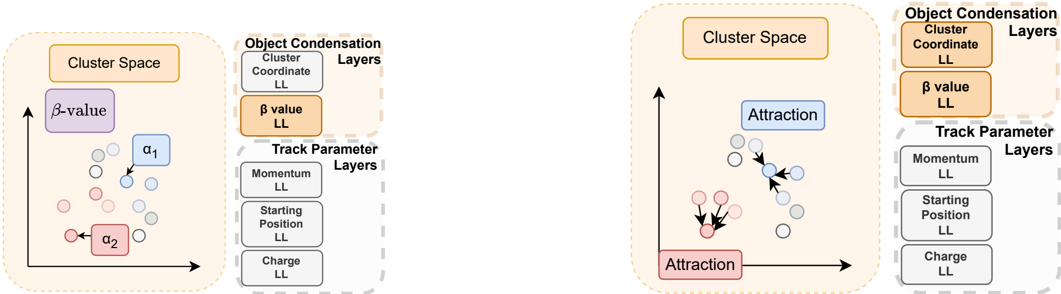
The concept of object condensation is to group nodes belonging to the same object by pulling them together in a learned clustering space. The goal is to condense the information of each object onto a single representative node, while simultaneously repelling nodes that belong to different objects. This is achieved by introducing a learned potential in the clustering space.

The following notation is adapted from the original notation in [14] to fit the use case of this work. Each node is assigned a β-value. This β-value is given by one output LL with a single neuron, followed by a sigmoid activation function to constrain the value between 0 and 1. Utilizing this output, a weight is calculated for each node. The weight q_i per node i is calculated using the β-value, that is clipped at a maximum of $1-\epsilon$, with $\epsilon = 10^{-7}$ to avoid infinity and $q_{\text{min}} > 0$, to avoid a value of 0:

$$q_i = \operatorname{arctanh}(\beta_i) + q_{\text{min}}. \tag{5.3}$$



(a) Translation from real space with the GNN.

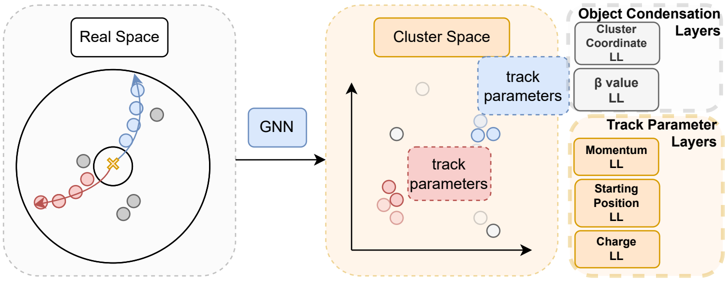


(b) For each object K, select the point with the highest β-value α.

(c) Attracting nodes from the same object using the attractive loss $L_{\text{attractive}}$ (Eq. (5.6)).

(d) Repel nodes from the same object using the repulsive loss $L_{\text{repulsive}}$ (Eq. (5.7)).

(e) Combined object condensation loss term including $L_{\text{attractive}}$, $L_{\text{repulsive}}$,L_{β}, and $L_{\text{suppress noise}}$ (Eq. (5.13)).



(f) Track parameter prediction on the track parameter layers.

Figure 5.8: Visualized object condensation loss for the two particles with associated red or blue hits, as well as grey background hits: (a) GNN translation, (b) α selection, (c) attractive potential, (d) repulsive potential, (e) combined object condensation loss terms visualized, and (f) track parameter prediction.

Using these weights, the potential affecting each node can be calculated. Instead of the computationally expensive task of calculating it between each node to every other node, a simplification is done where the potential is approximated by only taking the node with the highest weight q_α for each object, shown in Fig. 5.8b. These approximated weights q_α can now be used to attract nodes from the same object k together, as shown in Fig. 5.8c, with the attractive potential

$$V_{\text{attractive},k}(x) = ||x - x_\alpha||^2 q_{\alpha k}. \tag{5.4}$$

The term $||\cdot||$ describes the n-dimensional Euclidean distance according to the cluster space dimension CS. To repulse nodes from different objects than k as seen in Fig. 5.8d, the repulsive potential

$$V_{\text{repulsive},k}(x) = \max(0, 1 - ||x - x_\alpha||) q_{\alpha k} \tag{5.5}$$

is applied. As soon as the distance between the nodes with the highest weight is larger than one, this potential goes to zero, thus no longer affecting the loss. The distance between two objects is therefore bound by this value 1. With the introduction of the matrix element M_{jk} that is 1 if the node j belongs to object k and 0 else, the attraction potential loss for all nodes N and all objects K is defined as

$$L_{\text{attractive}} = \frac{1}{N} \sum_{j=1}^{N} q_j \sum_{k=1}^{K} \left(M_{jk} V_{\text{attractive},k}(x_j) \right) \tag{5.6}$$

and the repulsive potential loss as

$$L_{\text{repulsive}} = \frac{1}{N} \sum_{j=1}^{N} q_j \sum_{k=1}^{K} \left((1 - M_{jk}) V_{\text{repulsive},k} \right). \tag{5.7}$$

Using this potential also ignores the issue of class balance. Track objects with a low number of hits (< 10), as they either go into the endcap directions or are very displaced, are weighted the same as low-momentum tracks curling in the detector and leaving well over 200 hits. One minimum of L_V can be achieved by every node having the value of $\beta = \epsilon$. To counteract the model to only assign this value to every node, the L_β loss

$$L_\beta = \frac{1}{K} \sum_{k=1}^{K} (1 - \beta_{\alpha k}) \tag{5.8}$$

is added, requiring that the node with the highest weight is assigned a β value of 1 and else penalized. This also combines with the desired behaviour of condensing the information of each object down to one node where the β value is at maximum. There is also the need to suppress the background nodes N_B, for this the additional loss term $L_{\text{suppress noise}}$ is added

$$L_{\text{suppress noise}} = s_B \frac{1}{N_B} \sum_{i=1}^{N_B} \beta_i, \tag{5.9}$$

thus encouraging the model to assign low β values that are close to 0 to the background nodes. I set $s_B = 1$, since the noise suppression term is already normalized by the number of background nodes. This allows the model to focus more on the other loss components, without disproportionately emphasizing noise suppression. The interplay of the losses for the object condensation layers is shown in Fig. 5.8e.

Lastly, the model also has to predict the parameters of each object. For this, the three track parameter LLs shown in Fig. 5.7 are used to predict the momentum, starting position, and charge. Because the goal at the end of the training is to condense the information on only one point with the highest weight, it is also only necessary to predict the correct parameters on the nodes with the highest β-value. Therefore, all parameter prediction loss terms $L_{\text{parameter}}$ calculated from the prediction p and truth t taken from the matched particle are scaled with the weight $\hat{q}_j$, resulting in the weighted loss term

$$L'_{\text{parameter}} = \frac{\sum_{j=1}^{N} \sum_{k=1}^{K} \hat{q}_j \cdot M_{jk} \cdot L_{\text{parameter}}(t_k, p_k)}{\sum_{j=1}^{N} \sum_{k=1}^{K} \hat{q}_j M_{jk} + \epsilon}. \tag{5.10}$$

The weight $\hat{q}_j$ is calculated according Eq. (5.3) but using a different value for $\hat{q}_{\text{min}}$. For the momentum prediction and the starting position prediction, I use one linear layer each with three output nodes without an activation function. The output nodes predict p_x, p_y, and p_z in GeV for the momentum. For the starting position, I use the Cartesian coordinates described in section Section 5.3.1. For the momentum and starting position loss term L_m and L_v, I calculate the mean absolute error between the predicted value p and the true value t using

$$L_{p/v} = |t_{p/v} - p_{p/v}|. \tag{5.11}$$

The charge prediction is achieved by one LL using one output node with a sigmoid activation function. Then, a binary cross entropy loss is calculated resulting in the charge loss

$$L_{\text{charge}} = -t_{\text{charge}} \ln(p_{\text{charge})} + (1 - t_{\text{charge}}) \ln(1 - p_{\text{charge}}). \tag{5.12}$$

For the total loss, I sum over all individual loss terms

$$L = L_{\text{attractive}} + L_{\text{repulsive}} + L_\beta + L_{\text{suppress noise}} + L'_{\text{p}} + L'_{\text{v}} + L'_{\text{charge}}. \tag{5.13}$$

The values of q_{min} and $\hat{q}_{\text{min}}$ can be used to prioritize the prediction of the object properties with $\hat{q}_{\text{min}}$ or focus on object segmentation with q_{min} using a high value. As the parameter prediction is not as relevant as the object segmentation in my case, I chose a high value for $q_{\min} = 1$ and a low value for $\hat{q}_{\min} = 0.1$. By selecting a low value for $\hat{q}_{\min}$, the model emphasizes learning parameter predictions on the nodes with the highest β values, since the relative differences between β dominate over the constant offset $\hat{q}_{\min}$. Increasing $\hat{q}_{\min}$ would diminish this effect, causing the model to treat all nodes more equally in the track parameter prediction. For object condensation, I prefer to treat all nodes more equally to ensure clear differentiation between objects and to cluster all nodes belonging to the same object closely together. Therefore, I chose a larger offset for this purpose. Similarly to increasing $q_{\min}$ to weight the condensation losses even more than the parameter prediction losses, a scaling factor can be applied to the individual loss terms. Preliminary studies showed that this did not improve the object condensation performance, but the tuning of these hyperparameters in Eq. (5.13) could be revisited in future studies.

5.3.3 Graph Neural Network Post-Processing

The process I use to retrieve track information from the model's inference step is outlined in Fig. 5.9. It consists of six stages carried out in sequence:

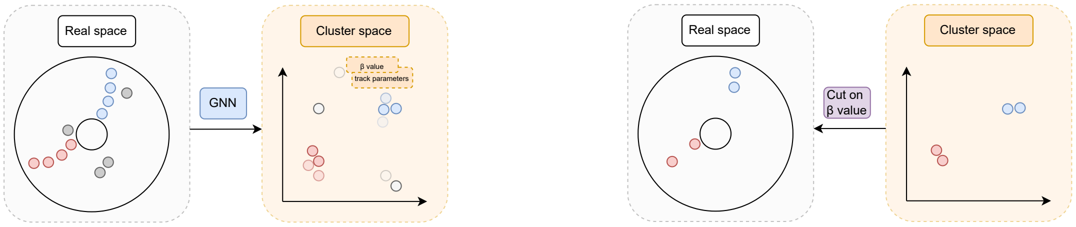
1. I apply the trained model to the event to obtain predictions at the node level as shown in Fig. 5.9a. Each node has the output of all 5 output layers described in the previous section, so a position in the learned clustering space, a β-value indicating its condensation relevance, and values for all seven track parameters.

2. To begin identifying potential tracks, I apply a threshold t_β to the predicted β-values. Nodes with β-value above this threshold are considered as candidates for condensation points, as shown in Fig. 5.9b.

3. I then refine this set of candidates by enforcing spatial separation in the cluster space. Starting with the node that has the highest β-value, I compute n-dimensional Euclidean distances r to all other candidates in the cluster space. Using a threshold t_d, I discard any candidates within $r < t_d$ of the current one. This step is repeated iteratively until all remaining candidates are sufficiently spaced apart. The surviving nodes are the final condensation points, each representing an individual track (see Fig. 5.9c).

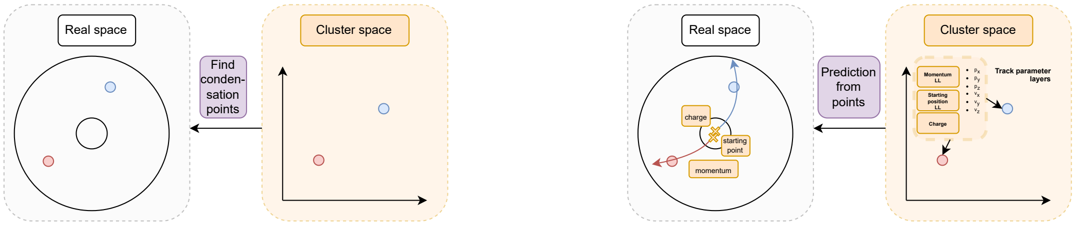
4. Once the condensation points are identified, the parameters for each track, such as momentum, position, and charge, are obtained directly from the model's predictions at the corresponding condensation points (see Fig. 5.9d).

5. To assign hits to tracks, I compute the distance from each condensation point to every node in the latent space. Nodes within a radius $r < t_h$ of a condensation point are assigned to that track, as illustrated in Fig. 5.9e. If $t_h > t_d/2$, a hit could be associated with multiple condensation points. To avoid assigning any hit to more than one track, as this is currently not handled by the subsequent fitting stages, I



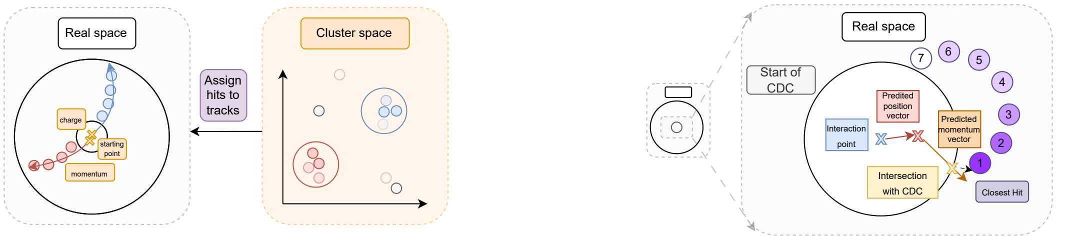
(a) Translation from real space to latent cluster space.

(b) Condensation point candidate selection based on β threshold.



(c) Condensation point selection based on isolation criteria.

(d) Parameter extraction for condensation points.



(e) Clusterisation based on condensation points.

(f) Hit ordering in real space, zoom to start of the CDC.

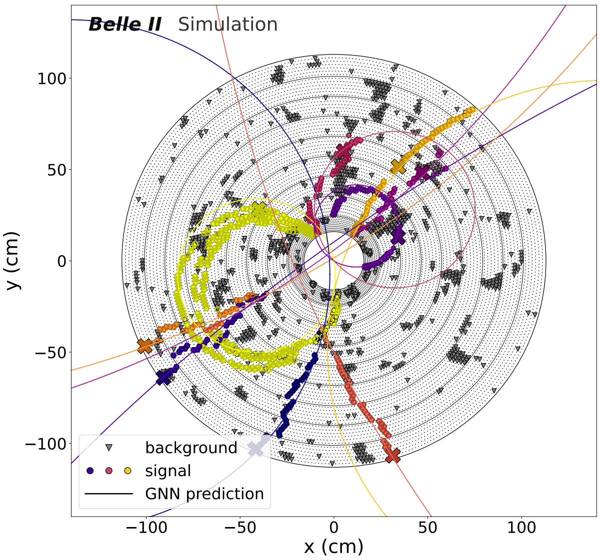
Figure 5.9: Figure and text taken from [1]: Track finding using object condensation: (a) Latent space, (b) condensation point candidate selection based on β threshold, (c) condensation point selection based on isolation, (d) parameter extraction, (e) clustering, and (f) hit ordering in real space.

assign it only to the closest condensation point. Each condensation point with its set of associated nodes constitutes a candidate track. However, I require at least seven hits per track to proceed, and tracks with fewer hits are excluded. Alternative clustering algorithms were tried in the supervised thesis [91], but it was found that the method described here is the fastest and most performant. I will go into detail about the alternative clustering algorithms in Section 10.3.

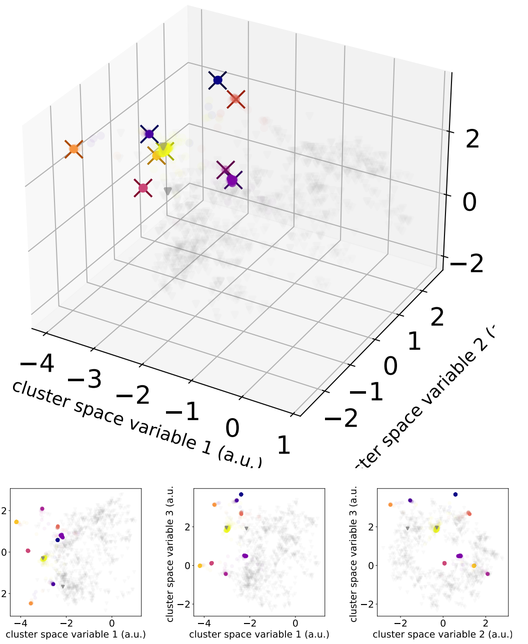
6. Once the hits are assigned, I sort them to prepare for track fitting. The ordering starts from the hit closest to the predicted entry point of the CDC. This is either the predicted starting position if it lies within the CDC, or the intersection of the predicted direction with the inner CDC surface, starting at the predicted position and extending in the direction of the predicted momentum as shown in Fig. 5.9f. I calculate Euclidean distances in the $x - y$ plane and iteratively select the nearest hit

> to construct the track sequence. While this method performs well for most cases, it fails for low momentum tracks that curl back and re-enter the CDC multiple times. One approach to address this is to apply a Hough transformation (see Section 5.2) to isolate only the first arm of the loop, as explored in the bachelor's thesis [92]. However, this was found to degrade performance in high-multiplicity events. An alternative solution is to train the model exclusively on the first arm of a curling track, ensuring that both hit ordering and subsequent fitting are more likely to succeed. I trained models both on the full and only the first arm and report results for curlers in Section 6.1.1.

An example event with corresponding learned latent space representation is shown in Fig. 5.10. Unlike the *Baseline Finder*, the *CAT Finder* does not offer a predefined covariance matrix for initializing track fits. As a workaround, I initialize all covariance matrix elements to 0.1. While this affects the time required for the track fit to converge, I have not observed any significant impact on track finding efficiency with different initialization. Both this initialization and the ordering of hits for low-momentum curling tracks are aspects that



(a) Example event display in the x–y plane. Filled colored circular markers show signal hits, filled gray triangular markers show background hits. Markers with colored outlines are found by the GNN to belong to the same track object. The GNN predictions (colored lines) are drawn using the predicted starting point and three-momentum for the predicted particle charge, and the corresponding condensation point is marked by a colored cross.



(b) Top: 3D cluster space with condensation points marked by crosses. Bottom: 2D projections. Colors match Fig. 5.10a.

Figure 5.10: Figure and text taken from [1]:: Event display and (Fig. 5.10b) cluster space representation of one example event from category 11 (Table 3.2) for high data beam backgrounds.

could be more appropriately handled during the track fitting step rather than the track finding stage. Tuning the track fitting procedure should be addressed in future work and will be discussed in Section 10.6.

5.3.4 Hyperparameter Optimization

For the *CAT Finder*, both model and track finding hyperparameters (t_β, t_d, t_h) are optimized in two stages, beginning with the model hyperparameters. This ordering is chosen because optimizing the model hyperparameters requires retraining, which is computationally expensive, whereas the track-finding hyperparameters can be tuned after training at significantly lower cost. For the *CAT Finder*, the hyperparameters should be optimized for the track finding efficiency, purity, and parameter resolution. However, performing full reconstruction after each training step for every model is not computationally possible within the scope of my thesis. Even evaluating the track finding step alone is not feasible, as this still increases the time after each step.

Model Hyperparameter Optimization

For the most time-efficient approach, the model's validation loss is used to approximate the model's performance. The loss is used as the target metric to minimize in the hyperparameter optimization. The search is conducted on a reduced, independent training dataset of 62,000 events (6% of the original training dataset), maintaining the same sample composition as the main training set (see Table 3.2) for a maximum of 200 epochs. The considered hyperparameters and their ranges are listed in Table 5.2. To save computing time, a Bayesian hyperband optimization [93] is used. This efficiently explores the hyperparameter space by using information from previous trials to focus on the most promising regions rather than searching randomly. Furthermore, runs are stopped early if the target metric is significantly worse than in previous trainings. This is implemented using the `Weights and Biases` [94] sweep method, with the number of runs limited to 300 due to computational constraints.[*].

The most influential hyperparameters, as shown in Fig. 5.11, are the number of nearest neighbours in GravNet (k), the width of the LL, and the number of GravNet blocks (N), all of which are strongly anti-correlated with respect to the loss. This anti-correlation means that increasing their values generally decreases the loss and improves model performance, favouring larger models. The remaining four hyperparameters show little correlation and have a lesser impact. The final selected values, ranked by importance, are summarized in Table 5.2. The optimal model consists of 797,812 trainable parameters.

While this optimization provides good initial starting parameters for the model training, it should be repeated only after finalizing all other aspects of the track finding pipeline, including fitting optimization and final dataset preparation, due to the significant compu-

---

[*] The 300 runs are equivalent to approximately 300 days of computing time

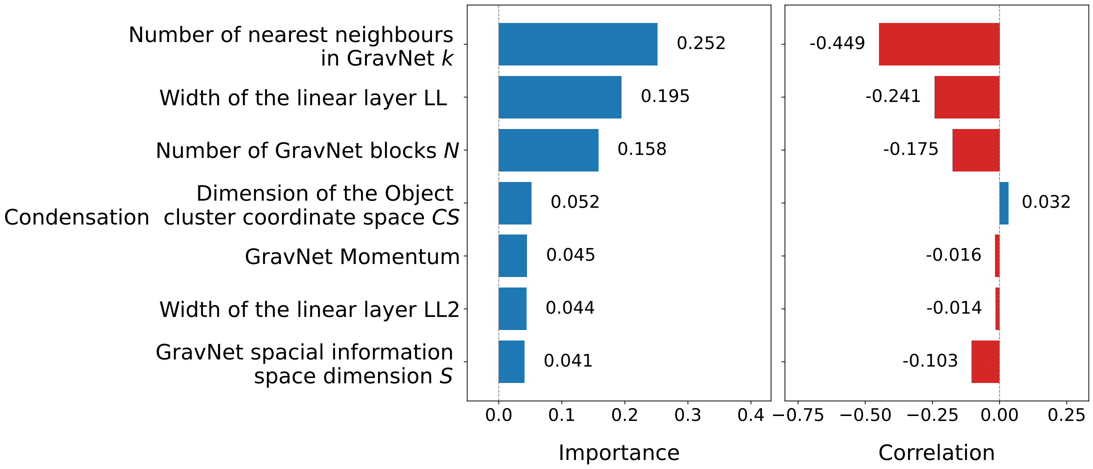


Figure 5.11: Parameter importance and correlation regarding the loss of the hyperparameter search using by `Weights and Biases` [94].

tational cost. Thus, this hyperparameter search is not repeated even for different dataset compositions discussed in Section 3.2. During final training, I reduce the learning rate from 0.001 by a factor of two each time the loss does not improve for 30 epochs, using a batch size of 1024.

To accelerate training, I apply a two-phase strategy. In the first phase, I train the model on the low simulated beam background dataset to learn track signatures, focusing on condensing the hits from the track objects and track parameter prediction. This phase is the most time-consuming part that typically requires around 500 epochs to converge. In the second phase, I fine-tune the model on a dataset with a higher beam background, using a learning rate reduced by a factor of 10 compared to the first phase.

This fine-tuning step focuses on background suppression and improves overall performance

Table 5.2: Table and text taken from [1]: Overview of the GNN model hyperparameters, including the explored ranges and the final values obtained after optimization. The parameters are listed in order of importance, as illustrated in Fig. 5.11. The optimization is based on the event categories given in Table 3.2, using an independent dataset corresponding to 6% of the full training sample. This dataset is split into 80% for training and 20% for validation. The final hyperparameters are selected based on the lowest validation loss.

Hyperparameter	Examined range	Result
Number of nearest neighbours in GravNet k	2-100	54
Width of the linear layer LL	32-128	126
Number of GravNet blocks N	2-7	4
Dimension of the Object Condensation - cluster coordinate space CS	2–5	3
GravNet Momentum	0.1-0.8	0.77
GravNet spacial information space dimension S	3-6	4
Width of the linear layer LL2	16-64	16

on high data beam backgrounds. Training on the high data beam backgrounds takes over three times longer per epoch due to higher hit occupancy, as low simulated beam backgrounds consists of on average 440 beam-background hits in comparison to high data beam backgrounds with 1216 average beam background hits on average. This two-step approach significantly reduces total training time, as the model converges after only 50 additional epochs in the second step. This is substantially more time efficient than training exclusively on high data beam backgrounds samples from the start.

Track Finding Hyperparameter Optimization

I optimize the three track finding hyperparameters t_β, t_d, and t_h targeting the typical events expected at Belle II (see Section 3.4) using prompt samples from category 2 (see Table 3.2) and displaced $K_S^0 \to \pi^+\pi^-$ events, where the K_S^0 momentum is directed into the CDC barrel region (see Section 3.2 for details).

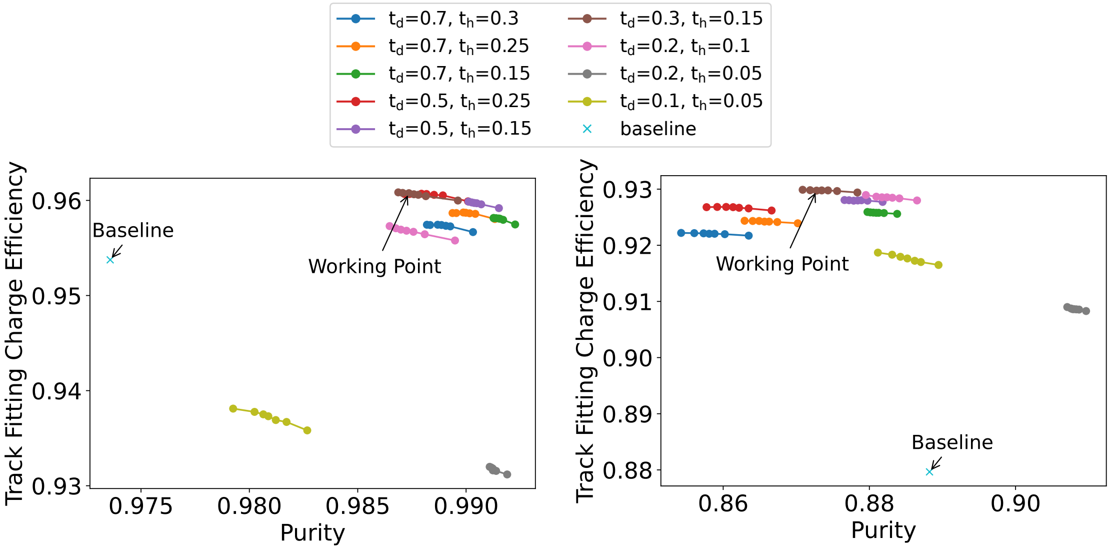
To evaluate performance, I compute the track finding and fitting efficiency ε_{trk} and purity $\mathfrak{p}_{\text{trk}}$ for multiple values of $t_\beta \in \{0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95\}$, generating ROC curves to illustrate the trade-off between efficiency and purity. This is done for several combinations of the condensation point distance $t_d \in \{0.1, 0.2, 0.3, 0.5, 0.7\}$ and the hit radius $t_h \in \{0.05, 0.1, 0.15, 0.25, 0.3\}$. The resulting performance, based on the combined track finding and fitting charge efficiency, is shown in Fig. 5.12. The final working point $\mathfrak{w} = (t_\beta, t_d, t_h)$ is selected to ensure that, first

$$\varepsilon_{\text{trk}}(\mathfrak{w}_i) + \mathfrak{p}_{\text{trk}}(\mathfrak{w}_i) \geq \varepsilon_{\text{trk}}(\text{baseline}) + \mathfrak{p}_{\text{trk}}(\text{baseline}), \tag{5.14}$$

and then

$$\max_i \left(\varepsilon_{\text{trk}}(\mathfrak{w}_{\mathrm{i}})_{\text{category 2}} + \varepsilon_{\text{trk}}(\mathfrak{w}_{\mathrm{i}})_{K_S^0 \to \pi^+\pi^-} \right), \tag{5.15}$$

where $\varepsilon_{\text{trk}}(\mathfrak{w}_{\mathrm{i}})_x$ is the track finding and fitting efficiency on the category 2 or the $K_S^0 \to \pi^+\pi^-$ sample. This results in the optimal values $t_\beta = 0.3$, $t_d = 0.3$ and $t_h = 0.15$. I repeated this study on the tracking fitting charge efficiency for $B\bar{B}$ events in Fig. B.1, yielding similar results as the category 2 sample. Increasing t_h, nearly as large as the t_d, did show promising results on $B^0\bar{B}^0$ simulated events (see [91]), but this did not translate to data, and the reconstruction efficiency dropped for larger t_h.



(a) Category 2.

(b) $K_S^0 \to \pi^+\pi^-$ (barrel).

Figure 5.12: Figure and text taken from [1]: Combined track finding and fitting charge efficiency as function of purity for the *CAT Finder*, and the respective value for the *Baseline Finder* for (a) category 2 and (b) $K_S^0 \to \pi^+\pi^-$ for *high data beam backgrounds*. See text for details.

5.4 Full Detector Reconstruction

The tracking chain [46, 85] for the full reconstruction, which incorporates all three tracking detectors, begins with the hits in the CDC. The full tracking chain implemented in basf2 is shown in Figure 5.13. For the *Baseline Finder*, a cut-based background filter is applied using the ADC, TDC, and TOT features described in Section 5.2, as well as a cross-talk identification algorithm based on the CDC front-end readout chips (see Section 3.1.1). In contrast, this background filtering step is not applied for the *CAT Finder*, as the relevant features are already used directly as input to the *CAT Finder* model, allowing for a more effective and data-driven optimization compared to the cut-based approach. This data-driven approach enables the machine learning model to learn patterns in the CDC hits distributions, including subtle correlations between features like ADC, TDC, and TOT, potentially leading to better background discrimination. The tracks found by the CDC track finding algorithms *Baseline Finder* (in blue) and *CAT Finder* (in red) are used as a starting point for the remaining tracking detectors.

After the CDC tracks are found, the CDC hits assigned to the track are marked and cannot be used again during any of the later steps. The CDC tracks are then extrapolated to the SVD with a CKF. During this step, SVD clusters matching the trajectory of the CDC track are attached to the CDC track, resulting in CDC+ SVD tracks shown in Fig. 5.13. The SVD clusters attached to the CDC+ SVD tracks are also marked and cannot be used in the subsequent steps.

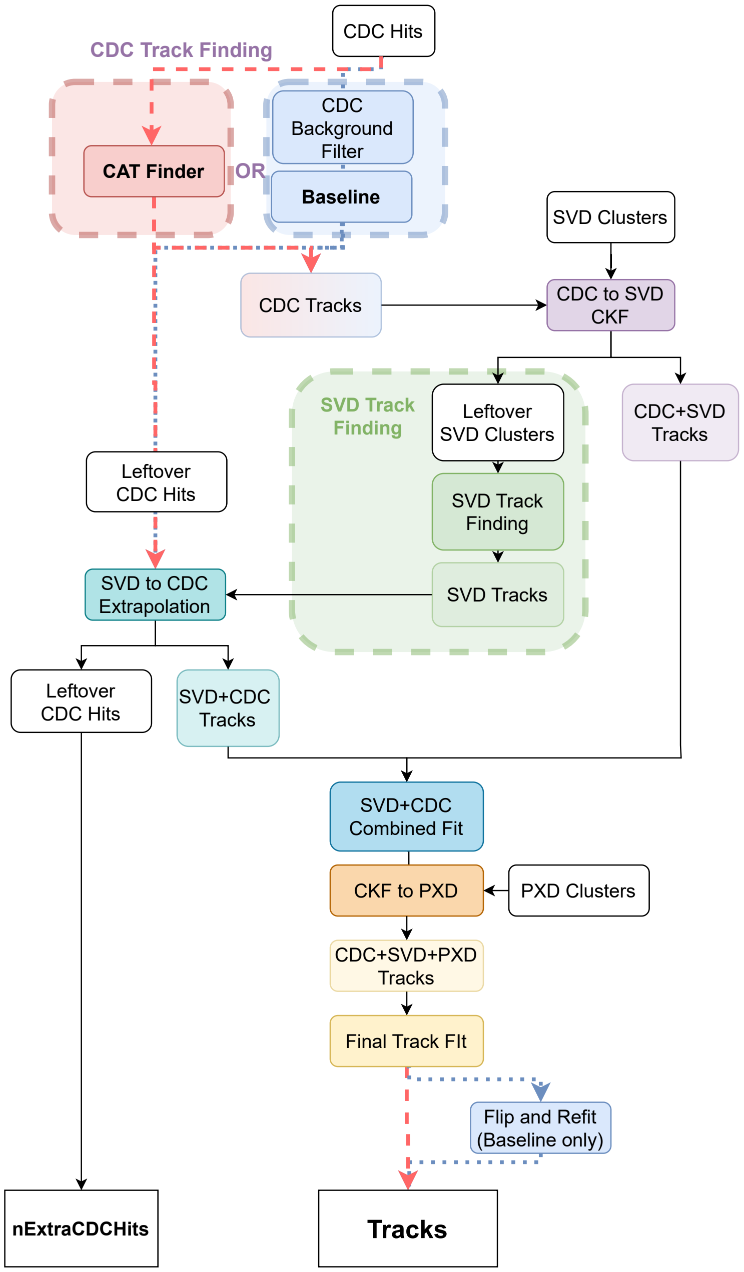


Figure 5.13: An illustration of the Belle II tracking pipeline.

All remaining clusters are used in the SVD standalone track finding utilizing an advanced filter concept called Sector Map [95] and a cellular automaton [54], described in more detail in [46].
This is especially relevant for low-momentum particles that do not or barely reach the CDC, leaving no to less than 6 hits in the CDC. The tracks found here are then extrapolated to the CDC and CDC hits that match the trajectory of the SVD tracks are attached. This results in SVD+ CDC tracks, which have their origin from the SVD track finding as

shown in Fig. 5.13.

Both the SVD+ CDC and CDC+ SVD tracks are fitted in a combined fit. The PXD occupancy is too high to directly include the PXD clusters in a combined PXD+ SVD stand-alone track finding (Section 2.2). Additionally, since the PXD is excluded from the HLT reconstruction, using it in full reconstruction track finding could cause significant discrepancies between online data taking and offline reconstruction. Instead, the combined SVD+ CDC and CDC+ SVD tracks are extrapolated to the PXD using a CKF to attach the PXD clusters to improve the resolution of the impact parameters. Then, a final track fit is performed with clusters and hits from all three tracking detectors.

For the *Baseline Finder*, low transverse momentum tracks are selected using an multivariate analysis (MVA) and on these tracks an additional track fit is performed with the opposite charge, thus flipping the track. Both the original track fit and the flipped track are compared, and an additional MVA decides which track to keep. This is not necessary for the *CAT Finder*, and discussed in detail in Section 6.1.1 Furthermore, all CDC hits that were not attached to any track are counted, used by analysts to gauge the level of beam backgrounds (see Section 3.1.1).
After all three tracking detectors of Belle II are combined to return the tracks, these tracks are now extrapolated to the particle identification (PID) detectors, the ECL, and the KLM.
For the *CAT Finder*, I implemented the code used in this thesis as a basf2 `Python` module, fully integrated into the basf2 framework. A `C++` version of my *CAT Finder* implementation is currently under development (see [96]) and will be included in the upcoming release of basf2, which enables its use in addition to the existing algorithm.

Chapter 6

Track Finding Comparison in the CDC

A large part of my results shown in this chapter are published in [1]. The plots, text, and tables closely reference the paper.
This chapter focuses on the track reconstruction in the CDC. I show a comparison of the *CAT Finder* and *Baseline Finder* with and without the track fitting described in Section 5.1. First, I discuss the track finding and fitting charge efficiency for prompt particles, followed by the comparison for displaced particles in Section 6.1. Next, I compare the momentum resolution between the algorithms in Section 6.2, and report the position reconstruction of the *CAT Finder* in Section 6.3. Finally, I evaluate the *CAT Finder* robustness against varying beam background conditions and detector ageing in Section 6.4.

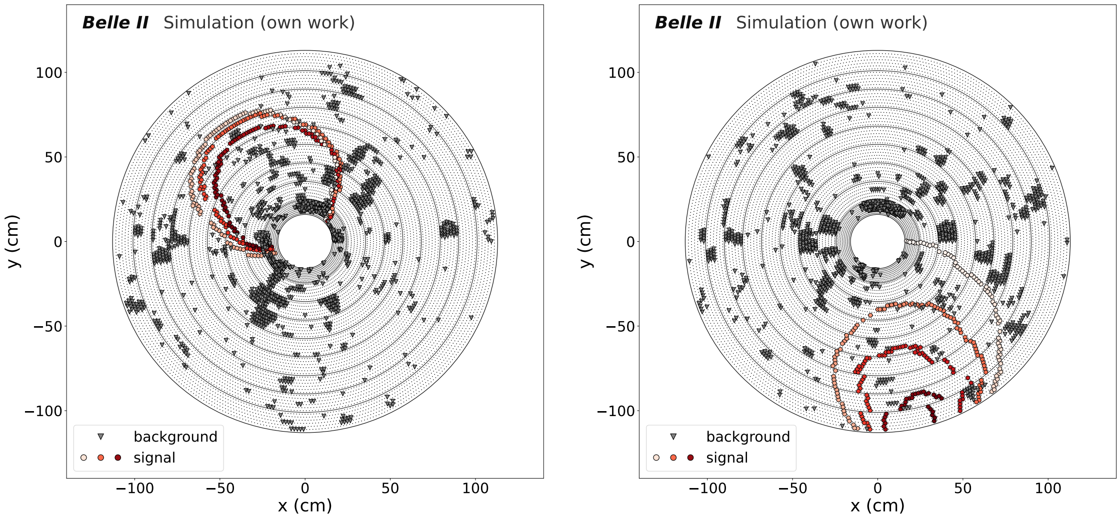
6.1 Track Finding and Track Fitting Efficiency

6.1.1 Prompt Tracks

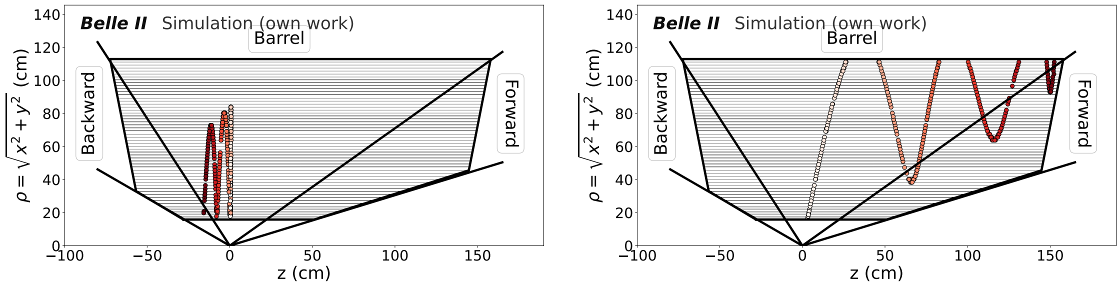
The technical prompt sample from category 1-3 (see Table 3.2) is ideal for an in-depth evaluation of difficult prompt track signatures. There are three difficult track signatures for the prompt track finding:

- Particles that leave the detector in endcap regions do not traverse the full detector, thus leaving a low number of hits. I report the efficiencies differentiated between the endcap regions and the barrel region for all algorithms.

- Particles with a transverse momentum of $p_t < 0.255$ GeV do not have enough momentum to reach the end of the CDC and instead start doing a loop inside the CDC. If their polar angle is within the CDC geometry $80^\circ \lesssim \theta \lesssim 95^\circ$, they can re-enter the CDC without significant energy loss in the vertex detectors (*inner curlers*). An example of an inner curler is shown in Fig. 6.1a. Figure 6.1c shows the $\rho - z$ view. The inner curler stops if it reaches the electronics or readouts outside of



(a) Inner Curler. (b) Outer Curler.



(c) Inner Curler. (d) Outer Curler.

Figure 6.1: Event displays in the x-y-plane (top) and the $\rho - z$ plane (bottom) for (Fig. 6.1a, Fig. 6.1c) *inner curler* (left) and (Fig. 6.1b, Fig. 6.1d) *outer curler* (right) for *high data beam background*. Filled, coloured circular markers indicate signal hits, where the colour represents the time of the first interaction in the drift cell as given by the GEANT4 simulation. The simulated time represents the time of the first interaction of the particle in the drift cell as given by the GEANT4 simulation. The colours start with white for the earliest hit and transition to darker red as time progresses, illustrating the particle trajectory. Filled gray triangular markers in the $x - y$ view show background hits. Only signal hits are shown for the $\rho - z$ view.

the tracking detector, or has a hadronic interaction between the SVD and CDC, highlighting the dependence on detector geometry.

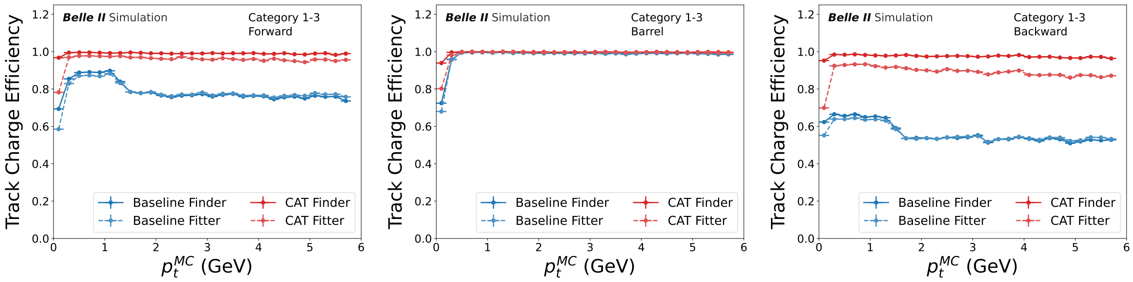
- Minimum-ionizing muons and pions with transverse momentum in the range 0.255 GeV $\leq p_t \lesssim$ 0.3 GeV can cross the entire CDC and then pass through the outer detectors TOP and ECL with minimal energy loss. If their polar angle lies within the CDC barrel region, they may re-enter the CDC (*outer curlers*). Heavier particles and electrons typically lose too much energy in the ECL to do so. Figure 6.1b presents an example of an outer curler. Compared to the inner curler, the energy loss per loop outside the tracking volume is significantly higher, resulting in a decreasing loop radius with each turn.

As already discussed in Section 5.3.3, neither the current hit ordering nor the track fitting algorithms correctly handle the curlers. Thus, I split the discussion of the samples from categories 1-3 in tracks that are non-curling and tracks that curl. This split is only done for the technical samples and for no other samples. Curling tracks are identified if two consecutive hits have a distance larger than 20 cm or if there are more than 32 signal hits in the first CDC superlayer A1. No events are discarded, the particles are sorted into the respective categories accordingly. Due to the low-momentum enrichment, many tracks populate the inner CDC, making this sample the most fitting to evaluate the prompt performance.

Non-curling tracks

The track finding charge efficiencies and the track fitting charge efficiencies for the three detector regions forward endcap, barrel, and backward endcap are shown in Fig. 6.2. The performance metrics are summarized in Table 6.1. The first challenging track signature involves particles leaving few hits as they exit the CDC early. Despite the high beam background density in the inner superlayers and no pre-filtering, the *CAT Finder* achieves higher hit purity and significantly better track efficiency in the endcap compared to the *Baseline Finder*. A similar trend is observed for the combined finding and fitting charge efficiency. While the *CAT Finder* has a higher initial fake rate than the *Baseline Finder*, it drops after fitting—indicating some *CAT Finder* tracks cannot be fitted. The clone rate remains low and comparable for both in the endcap.

In the barrel, the *CAT Finder* shows higher finding efficiency than the baseline but with increased fake and clone rates. Fig. 6.2b shows efficiency drops for $p_t \lesssim 0.3\,\text{GeV}$ in both



(a) Forward endcap. (b) Barrel. (c) Backward endcap.

Figure 6.2: Figure and text taken from [1]: Track finding charge (empty markers, connected by solid lines to guide the eye) and combined track finding and fitting charge efficiency (filled markers, connected by dashed lines to guide the eye) for the prompt evaluation samples (category 1-3, see Table 3.2, *high data beam backgrounds*) with curler tracks removed, as function of simulated transverse momentum p_t^{MC} for the *Baseline Finder* (blue) and the *CAT Finder* (red) in the (a) forward endcap, (b) barrel, and (c) backward endcap. The vertical error bars that show the statistical uncertainty are smaller than the marker size. The horizontal error bars indicate the bin width. The uncertainties of the different track finding algorithms are correlated since they use the same simulated events.

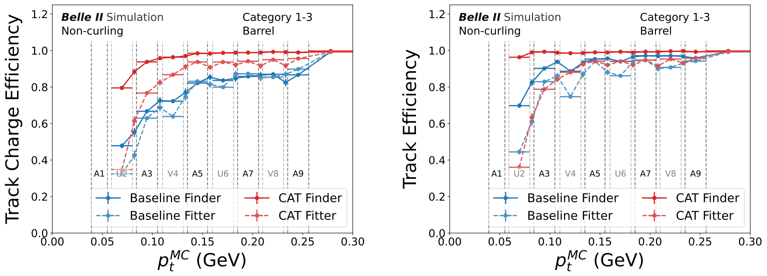
Table 6.1: Figure and table taken from [1]: The performance metrics for the prompt evaluation samples (category 1-3, *high data beam backgrounds*, see Table 3.2 and Section 3.2 for details) for non-curling tracks for *CAT Finder* and *Baseline Finder* in different detector regions. Uncertainties below <0.01% are not shown in the table.

(in %)	$\varepsilon_{\mathrm{trk}}$	$\mathfrak{r}_{\mathrm{fake}}$	$\mathfrak{r}_{\mathrm{clone}}$	$\varepsilon_{\mathrm{trk,ch}}$	$\mathfrak{r}_{\mathrm{wrong\ ch.}}$
			forward endcap		
Baseline Finder	$80.1^{+0.1}_{-0.1}$	$0.55^{+0.02}_{-0.02}$	0.01	$78.4^{+0.1}_{-0.1}$	$2.06^{+0.04}_{-0.04}$
CAT Finder	$98.94^{+0.03}_{-0.03}$	$1.62^{+0.03}_{-0.03}$	$0.21^{+0.01}_{-0.01}$	$98.89^{+0.03}_{-0.03}$	$0.06^{+0.01}_{-0.01}$
Baseline Fitter	$78.1^{+0.1}_{-0.1}$	$0.49^{+0.02}_{-0.02}$	0.01	$77.1^{+0.1}_{-0.1}$	$1.37^{+0.04}_{-0.04}$
CAT Fitter	$95.93^{+0.06}_{-0.05}$	$0.31^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$94.29^{+0.06}_{-0.06}$	$1.71^{+0.04}_{-0.04}$
			barrel		
Baseline Finder	$97.97^{+0.04}_{-0.04}$	$2.31^{+0.04}_{-0.04}$	$0.05^{+0.01}_{-0.01}$	$95.92^{+0.06}_{-0.06}$	$2.09^{+0.04}_{-0.04}$
CAT Finder	$99.61^{+0.02}_{-0.02}$	$3.34^{+0.05}_{-0.05}$	$0.59^{+0.02}_{-0.02}$	$99.16^{+0.03}_{-0.03}$	$0.46^{+0.02}_{-0.02}$
Baseline Fitter	$96.88^{+0.05}_{-0.05}$	$1.83^{+0.04}_{-0.04}$	0.03	$95.5^{+0.06}_{-0.06}$	$1.42^{+0.03}_{-0.03}$
CAT Fitter	$97.6^{+0.04}_{-0.04}$	$1.26^{+0.03}_{-0.03}$	$0.16^{+0.01}_{-0.01}$	$97.39^{+0.05}_{-0.04}$	$0.22^{+0.01}_{-0.01}$
			backward endcap		
Baseline Finder	$60.5^{+0.1}_{-0.1}$	$1.08^{+0.04}_{-0.04}$	$0.03^{+0.01}_{-0.01}$	$58.0^{+0.1}_{-0.1}$	$4.08^{+0.07}_{-0.07}$
CAT Finder	$97.64^{+0.04}_{-0.04}$	$1.2^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$97.42^{+0.04}_{-0.04}$	$0.22^{+0.01}_{-0.01}$
Baseline Fitter	$58.8^{+0.1}_{-0.1}$	$0.92^{+0.03}_{-0.04}$	$0.02_{-0.01}$	$56.8^{+0.1}_{-0.1}$	$3.28^{+0.06}_{-0.06}$
CAT Fitter	$92.43^{+0.07}_{-0.07}$	$0.69^{+0.02}_{-0.02}$	$0.03_{-0.01}$	$87.67^{+0.09}_{-0.09}$	$5.16^{+0.06}_{-0.06}$

algorithms, detailed in Fig. 6.3.

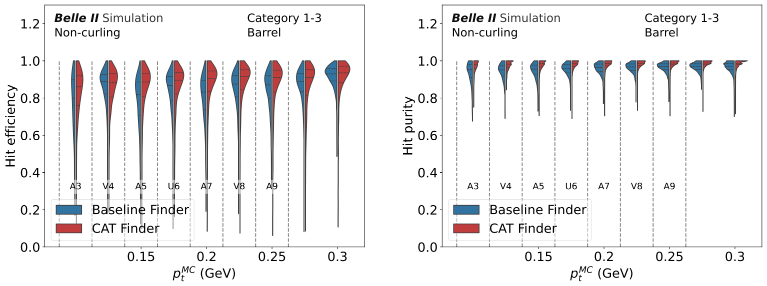
If the transverse momentum is below 0.039 GeV, the transverse momentum is not enough to reach the CDC. If the transverse momentum is between 0.039 GeV< p_t < 0.055 GeV, it only manages to reach and deposit hits in the first superlayer of the CDC. The labels A1–A9 indicate these superlayer boundaries Section 2.3 reached by prompt particles based on their transverse momentum.

The *CAT Finder* significantly outperforms the *Baseline Finder* in track charge efficiency for non-curling particles that do not reach the end of the CDC, as shown in Fig. 6.3a. This holds for both finding and combined finding–fitting efficiency, despite some *CAT Finder* tracks not being fitted. For axial superlayers (A), both algorithms show similar combined efficiencies, while for stereo layers (U, V), the *Baseline Finder* performance drops. At low momenta, the *Baseline Finder* struggles to determine the track's starting direction for one-loop tracks (see Fig. 6.4), leading to a higher wrong charge rate of 1.42% compared to the 0.22% of the *CAT Finder*. The *CAT Finder* predicts the correct charge more often, improving performance. In the 0.255, GeV $\lesssim p_t \lesssim$ 0.3, GeV range, both perform similarly. For matched tracks found by both algorithms (intersecting sample), the *CAT Finder* achieves higher hit efficiency and purity (Figs. 6.3c and 6.3d). In the additional sample (tracks found only by one algorithm), the *CAT Finder* still shows higher hit efficiency and purity, while the number of additional *Baseline Finder* tracks is very small. This indicates that the *CAT Finder* finds more complex, high-quality tracks, though not all can be fitted. Improving their fit requires further tuning beyond the scope of this work.



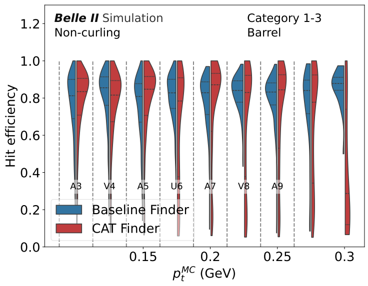
(a) Track finding and fitting charge efficiency.

(b) Track finding and fitting efficiency.

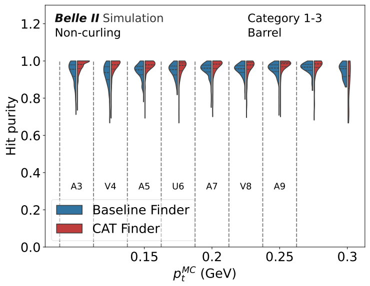


(c) Hit efficiency for intersecting sample.

(d) Hit purity for intersecting samples.

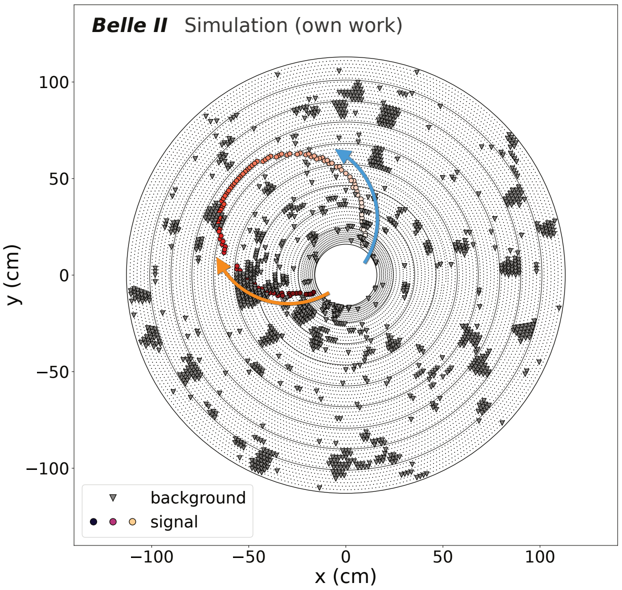


(e) Hit efficiency for additional samples.

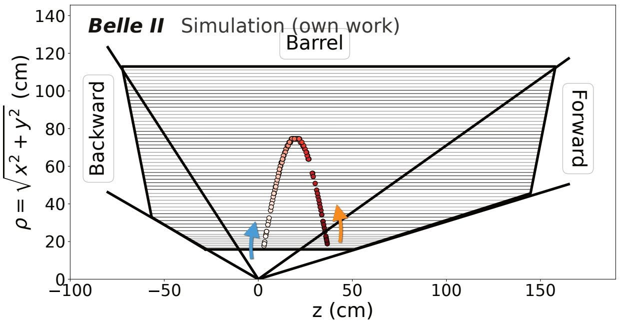


(f) Hit purity for additional samples.

Figure 6.3: Figure and text taken from [1]: The top row shows the low momentum track finding (empty markers, connected by lines to guide the eye) and combined track finding and fitting charge efficiency (filled markers) (b) and the track finding and combined track finding and fitting charge efficiency (a) for the prompt evaluation samples (category 1-3, *high data beam backgrounds*, see Table 3.2) for non-curling tracks. The middle row shows hit efficiency and hit purity for tracks found by both *CAT Finder* and *Baseline Finder* (intersecting sample) (c and d) and the bottom row for the additional found tracks (e and f). The dashed horizontal dark (light) gray lines show the axial (stereo) superlayer boundaries how far the prompt track reaches with the given transverse momentum. (paper figure)



(a) x–y view.

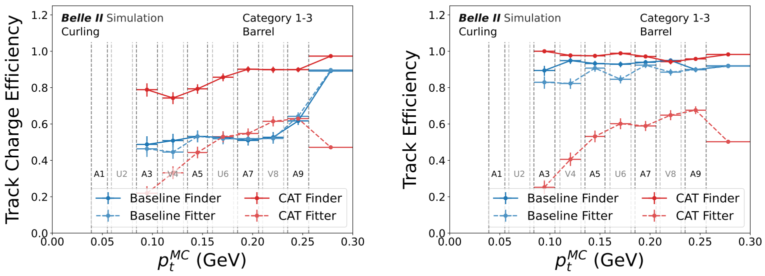


(b) ρ–z view.

Figure 6.4: Event display of a *single-loop curler* under *high data beam background*, shown in the $x - y$ view (top, Fig. 6.4a) and the $\rho - z$ view (bottom, Fig. 6.4b). Filled, colored circular markers represent *signal hits*, with color encoding the global time of the first interaction in each drift cell as provided by the GEANT4 simulation. The color scale progresses from light pink (early hits) to red and dark red (later hits), visualizing the particle trajectory. In the $x - y$ view, filled gray triangular markers indicate *background hits* (see Fig. 2.6 for details). Only signal hits are shown in the ρ–z view. The blue arrow indicates the particle trajectory direction in the $x - y$ and $\rho - z$ view, while the orange arrow indicates the reversed direction.

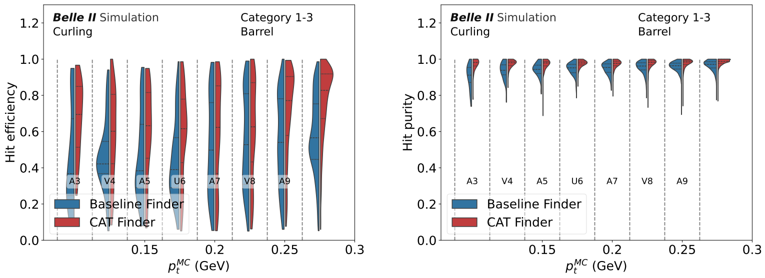
Curling tracks

Figure 6.5 repeats the analysis from Fig. 6.3 for curling tracks. The *CAT Finder* shows significantly better track finding charge efficiency (Fig. 6.5a) than the *Baseline Finder*. However, track fitting efficiency is lower, as current fitting algorithms and the *CAT Finder* hit ordering (Section 5.3.3) cannot handle curling signatures. Fig. 6.6 shows the *CAT Finder* prediction for the two previous example curlers, including the curling signature. The



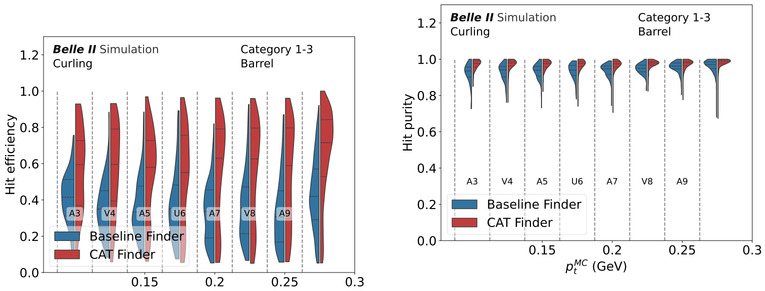
(a) Track finding and fitting charge Efficiency.

(b) Track finding and fitting efficiency.



(c) Hit efficiency for intersecting sample.

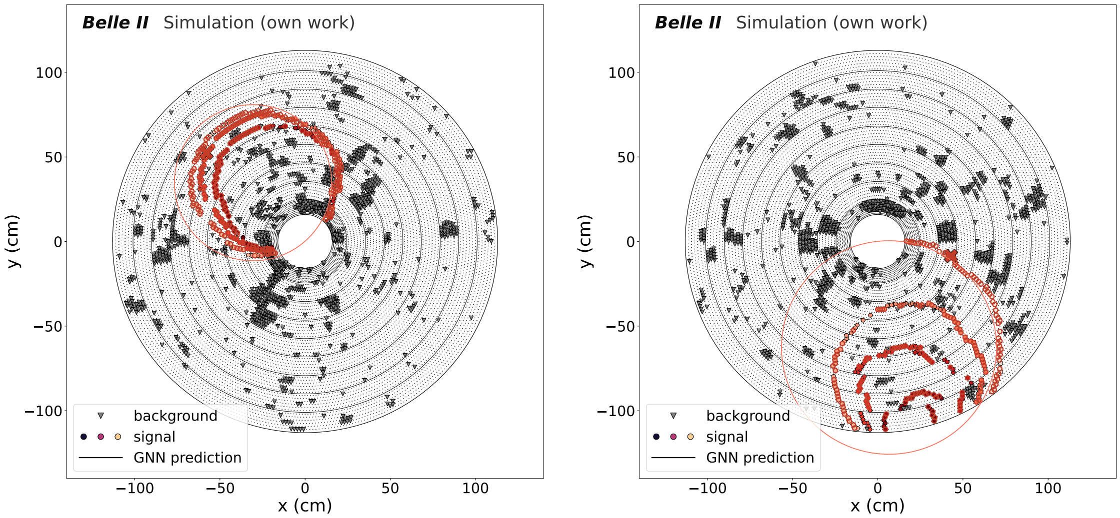
(d) Hit purity for intersecting samples.



(e) Hit efficiency for additional samples.

(f) Hit purity for additional samples.

Figure 6.5: Figure and text taken from [1]: Low momentum track finding (empty markers, connected by lines to guide the eye) and combined track finding and fitting charge efficiency (filled markers) (b) and the track finding and combined track finding and fitting charge efficiency (a) for curling tracks with *high data beam backgrounds*. The middle row shows hit efficiency and hit purity for tracks found by both *CAT Finder* and *Baseline Finder* (intersecting sample) (c and d) and the bottom row for the additional found tracks (e and f). The dashed horizontal dark (light) gray lines show the axial (stereo) superlayer boundaries how far the prompt track reaches with the given transverse momentum.



(a) Inner Curler.

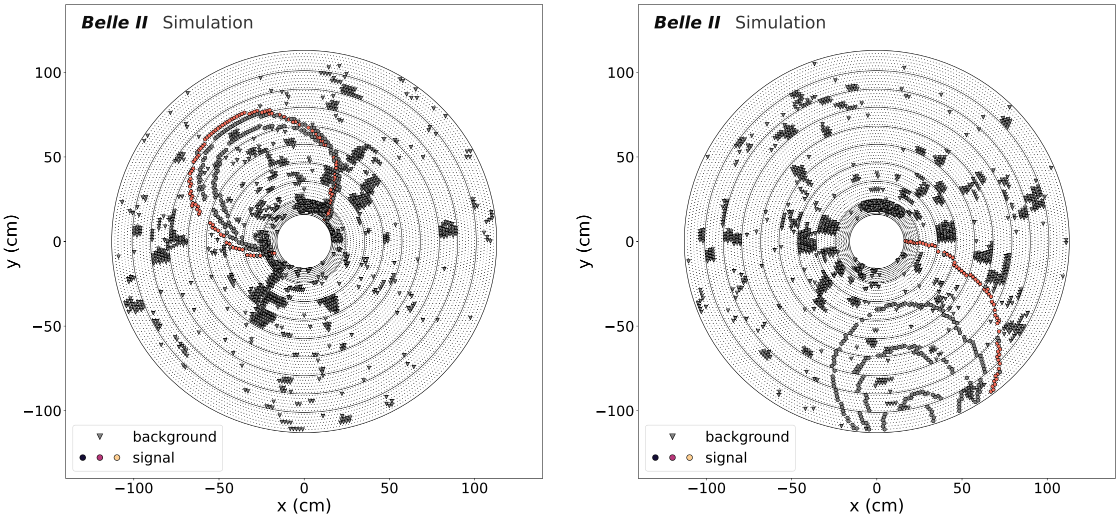
(b) Outer Curler.

Figure 6.6: Text adapted from [1]: Event displays in the x-y-plane for (Fig. 6.6a) *inner curler* (left) and (Fig. 6.6b) *outer curler* (right) for *high data beam background.* Filled, colored circular markers indicate signal hits. Filled gray triangular markers in the $x - y$ view show background hits (see Fig. 2.6 for details). Markers with colored outlines are found by the GNN to belong to the same track object. The GNN predictions (colored lines) are drawn using the predicted starting point and three momentum for the predicted particle charge, and the corresponding condensation point is marked by a colored cross.

Baseline Finder assigns only one loop per track, resulting in low hit efficiency for the intersecting sample (Fig. 6.5c), but its tracks can be fitted reliably. For particles not reaching the end of the CDC, fitting efficiency is similar between both algorithms, but it drops in the last bin—corresponding to *outer curlers.* The *CAT Finder* achieves very high hit efficiency and purity, even in the additional sample. This improved hit efficiency would improve the estimate for the beam-background conditions according to the unmatched number of CDC hits (Section 5.4), reduce clones, and help identify complete tracks. But to propagate this to the end of the reconstruction chain, significant modifications to the fitting algorithms are needed. Current Belle II fitting is not optimized for curling tracks, resulting in low post-fit efficiency. A potential solution involves identifying the outermost first loop for accurate momentum estimation and passing this to the fitter while marking the additional hits as taken to avoid clones. A Hough transformation approach (see Section 5.2) was explored in a thesis I supervised [92]. Tracks were first filtered as curlers and classified as

- *outer curlers* with large polar angle θ and significant energy loss in the TOP and ECL, requiring looser boundaries;
- *inner curlers* with smaller θ and less energy loss, needing tighter constraints.

While promising in concept, the method did not yield a viable solution. To account for the



(a) Inner Curler.

(b) Outer Curler.

Figure 6.7: Text adapted from [1]: Event displays in the x-y-plane for (Fig. 6.7a) *inner curler* (left) and (Fig. 6.7b) *outer curler* (right) for *high data beam background*. Filled, circular markers indicate signal hits while filled, and only colored circular markers are used as training targets for the model. Filled gray triangular markers in the $x-y$ view show background hits (see Fig. 2.6 for details).

fit requirement, I train this new model on only the first loop before the TOP detector as the training signal Fig. 6.7. This increases the chance of capturing a fittable helix within current fitting constraints, the resulting *CAT B Finder* (see Section 3.2) is evaluated in Section C.2. The overall finding and fitting charge efficiency for low momentum particles improves in comparison to *CAT Finder*. The performance will be addressed in the upcoming Section 6.1.1.

Prompt tracks in $\mu^-\mu^+(\gamma)$

In addition to the previously described samples, I evaluated the track finding algorithms on simulated $\mu^-\mu^+(\gamma)$ events. As one of the main calibration samples at Belle II, these events aim for a track fitting charge efficiency of 100% in the barrel. Unlike the category 1–3 evaluation samples with up to 12 particles in an event, $\mu^-\mu^+(\gamma)$ events almost always contain two isolated, prompt, high-momentum particles.

I observed that the *CAT Finder* achieves significantly higher finding charge efficiency in both endcaps, though with a slightly higher fake rate than the *Baseline Finder*. After fitting, the combined charge efficiency of the *CAT Finder* is comparable to the *Baseline Finder*, but with a noticeably lower fake rate. In the barrel, both algorithms perform well, with the *CAT Finder* reaching 99.48% and the *Baseline Finder* 99.07%, and the *CAT Finder* maintaining a lower fake rate. In the endcaps, I measured a combined charge efficiency of 94.7% for the *CAT Finder*, clearly outperforming the *Baseline Finder* at 67.7%. The *CAT B Finder* achieves similar performance as the *CAT Finder* for tracks going in the barrel, but decreased performance for the endcaps with 91.8% I provide detailed plots and numerical results in Section C.3.

Prompt Tracks from $\pi^\pm$ in $B\bar{B}$

The most relevant event types for the Belle II flavour-physics program are $B^0\bar{B}^0$ and B^-B^+ events. In this section, I report the combined track finding and fitting charge efficiencies for prompt pions, results for other particle types ($e^\pm$, $\mu^\pm$, $K^\pm$, $p/\bar{p}$) are presented in Section C.4.

I show the tracking finding charge efficiency for $\pi^\pm$ as a function of p_t^{MC} and $\cos\theta^{\mathrm{MC}}$ for the *Baseline Finder* (Fig. 6.8a), *CAT Finder* (Fig. 6.8b), and *CAT B Finder* (Fig. 6.8c). I observe that both *CAT Finder* and *CAT B Finder* improve efficiency in the endcaps and for *inner* and *outer curlers*, with *CAT B Finder* showing the best performance. Figure C.4 shows efficiency vs. p_t^{MC} over $N_{\mathrm{seen}}^{\mathrm{MC}}$, the number of visible particles. As $N_{\mathrm{seen}}^{\mathrm{MC}}$ increases, efficiency drops across all models due to increased overlap and occupancy from low-momentum tracks.

The same comparisons are repeated for fitted tracks in Fig. 6.9 and Fig. C.5. Here, the benefit of the new *CAT B Finder* becomes clear: unlike *CAT Finder*, it successfully fits *inner curlers* and retains the better endcap performance, with similar results for *outer curlers* compared to the *Baseline Finder*.

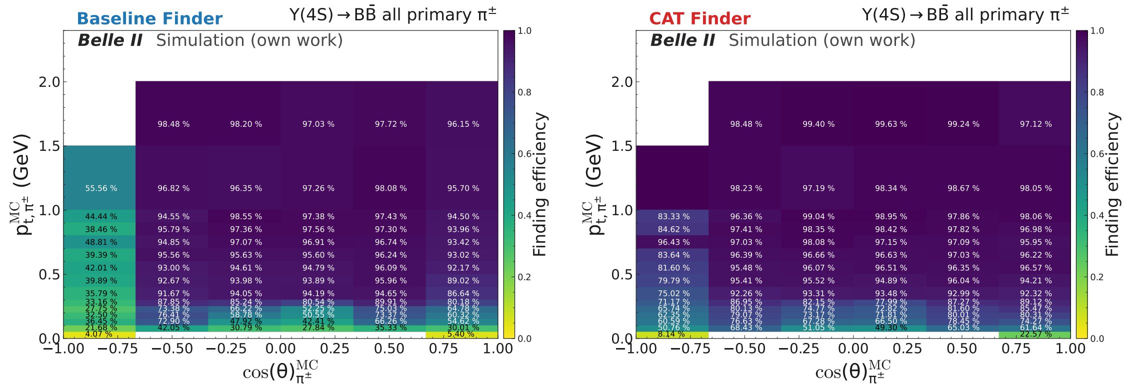
Table 6.2 summarizes the performance for prompt primary pions in $B\bar{B}$ events. In the barrel, I observe a fitting charge efficiency of 81.8% for the *Baseline Finder*, 77.8% for the *CAT Finder*, and 87.% for the *CAT B Finder*, outperforming the *Baseline Finder* in the most relevant figure of merit.

Figure 6.10 presents hit efficiency over $N_{\mathrm{seen}}^{\mathrm{MC}}$ with distributions (Fig. 6.10a) and medians for comparison (Fig. 6.10b). Hit efficiency drops with higher multiplicity for all models, but more steeply for the *CAT Finder*. The higher the multiplicity, the lower the momentum of the particles. The *CAT B Finder* recovers some performance but remains slightly below the *Baseline Finder*. Hit purity (Fig. 6.10d) is higher for both *CAT Finder* and *CAT B Finder*, reflecting a trade-off made during hyperparameter tuning (Section 5.3.4), which favoured purity over efficiency. The results for the other FSP are given in Section C.4. Hit efficiency is highest for muons and comparable between *Baseline Finder* and *CAT B Finder*. Other particle types show lower hit efficiencies, which I attribute to the training data, as only muon particle tracks were used for *CAT Finder*, and over 50% of muon particle tracks for *CAT B Finder*. Training on a mix of different particle types improves general performance but requires careful dataset design, which I will discuss in detail in Section 10.2.

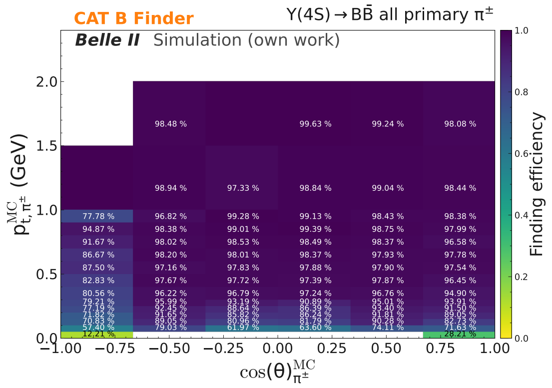
Despite differences in hit efficiency, fitting efficiency remains consistent across particle types in the new model.

Table 6.2: The performance metrics for prompt, primary $\pi^{\pm}$ in $B\bar{B}$ samples for *CAT Finder*, *CAT B Finder* and *Baseline Finder* in different detector regions. Uncertainties below <0.01% are not shown in the table.

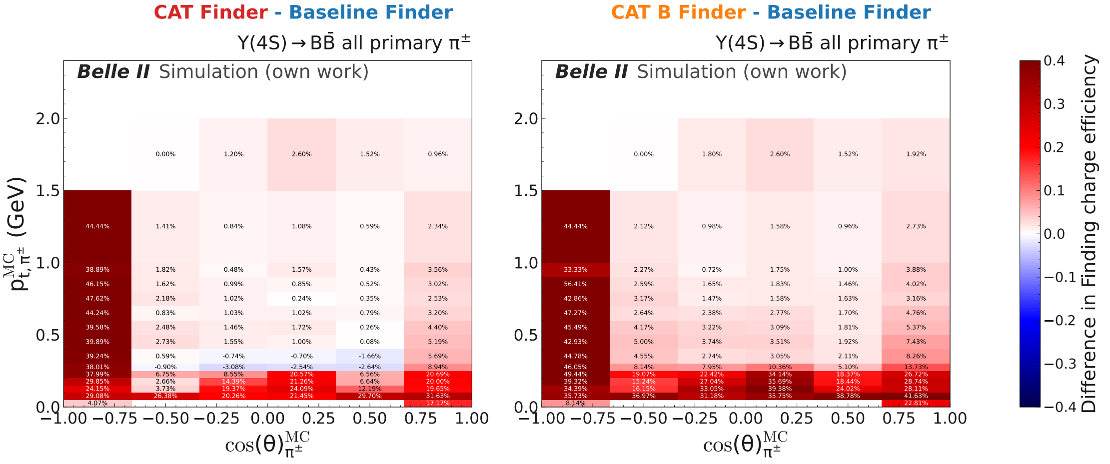
(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
			FWD		
Baseline Finder	$60.5^{+0.5}_{-0.5}$	$2.0^{+0.2}_{-0.2}$	$0.04^{+0.02}_{-0.03}$	$60.4^{+0.5}_{-0.5}$	$0.09^{+0.04}_{-0.05}$
CAT Finder	$85.0^{+0.4}_{-0.4}$	$6.3^{+0.2}_{-0.2}$	$1.5^{+0.1}_{-0.1}$	$84.3^{+0.4}_{-0.4}$	$0.8^{+0.1}_{-0.1}$
CAT B Finder	$90.9^{+0.3}_{-0.3}$	$3.9^{+0.2}_{-0.2}$	$0.85^{+0.08}_{-0.08}$	$90.5^{+0.3}_{-0.3}$	$0.43^{+0.07}_{-0.08}$
Baseline Fitter	$55.2^{+0.5}_{-0.5}$	$2.1^{+0.2}_{-0.2}$	$0.04^{+0.02}_{-0.03}$	$53.9^{+0.5}_{-0.5}$	$2.4^{+0.2}_{-0.2}$
CAT Fitter	$68.0^{+0.5}_{-0.5}$	$2.5^{+0.2}_{-0.2}$	$0.38^{+0.06}_{-0.07}$	$66.1^{+0.5}_{-0.5}$	$2.7^{+0.2}_{-0.2}$
CAT B Fitter	$79.5^{+0.4}_{-0.4}$	$1.8^{+0.1}_{-0.1}$	$0.18^{+0.04}_{-0.04}$	$78.1^{+0.4}_{-0.4}$	$1.8^{+0.2}_{-0.2}$
			BRL		
Baseline Finder	$86.7^{+0.1}_{-0.1}$	$2.58^{+0.06}_{-0.06}$	$0.48^{+0.03}_{-0.03}$	$83.0^{+0.2}_{-0.2}$	$4.23^{+0.09}_{-0.09}$
CAT Finder	$89.9^{+0.1}_{-0.1}$	$5.75^{+0.08}_{-0.08}$	$2.3^{+0.05}_{-0.05}$	$87.9^{+0.1}_{-0.1}$	$2.21^{+0.07}_{-0.07}$
CAT B Finder	$94.77^{+0.1}_{-0.1}$	$4.02^{+0.06}_{-0.06}$	$3.29^{+0.06}_{-0.06}$	$93.4^{+0.1}_{-0.1}$	$1.46^{+0.05}_{-0.05}$
Baseline Fitter	$84.2^{+0.2}_{-0.2}$	$1.79^{+0.05}_{-0.05}$	$0.3^{+0.02}_{-0.02}$	$81.8^{+0.2}_{-0.2}$	$2.83^{+0.08}_{-0.08}$
CAT Fitter	$78.7^{+0.2}_{-0.2}$	$1.99^{+0.05}_{-0.05}$	$0.69^{+0.03}_{-0.03}$	$77.8^{+0.2}_{-0.2}$	$1.23^{+0.05}_{-0.05}$
CAT B Fitter	$87.9^{+0.1}_{-0.1}$	$1.21^{+0.04}_{-0.04}$	$0.81^{+0.03}_{-0.03}$	$87.2^{+0.1}_{-0.1}$	$0.79^{+0.04}_{-0.04}$
			BWD		
Baseline Finder	$51.3^{+0.6}_{-0.6}$	$4.0^{+0.3}_{-0.3}$	$0.04^{+0.02}_{-0.04}$	$51.1^{+0.6}_{-0.6}$	$0.32^{+0.09}_{-0.11}$
CAT Finder	$77.7^{+0.5}_{-0.5}$	$3.6^{+0.2}_{-0.2}$	$0.99^{+0.1}_{-0.11}$	$76.6^{+0.5}_{-0.5}$	$1.4^{+0.2}_{-0.2}$
CAT B Finder	$86.1^{+0.4}_{-0.4}$	$2.2^{+0.2}_{-0.2}$	$0.83^{+0.09}_{-0.09}$	$85.7^{+0.4}_{-0.4}$	$0.48^{+0.09}_{-0.1}$
Baseline Fitter	$47.8^{+0.6}_{-0.6}$	$3.3^{+0.3}_{-0.3}$	$0.04^{+0.02}_{-0.04}$	$46.9^{+0.6}_{-0.6}$	$1.9^{+0.2}_{-0.3}$
CAT Fitter	$60.3^{+0.6}_{-0.6}$	$3.5^{+0.2}_{-0.2}$	$0.33^{+0.07}_{-0.08}$	$58.2^{+0.6}_{-0.6}$	$3.5^{+0.3}_{-0.3}$
CAT B Fitter	$71.2^{+0.6}_{-0.5}$	$2.4^{+0.2}_{-0.2}$	$0.13^{+0.04}_{-0.05}$	$69.2^{+0.6}_{-0.6}$	$2.9^{+0.2}_{-0.2}$



(a) *Baseline Finder*. (b) *CAT Finder*.

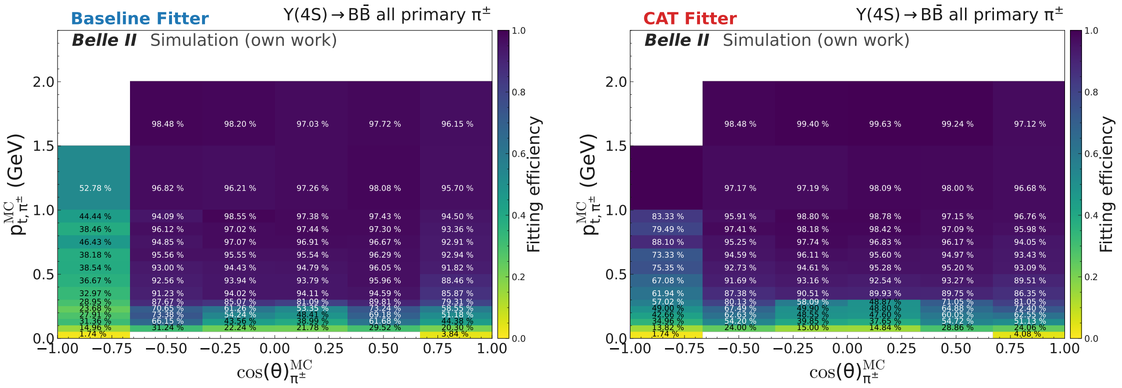


(c) *CAT B Finder*.

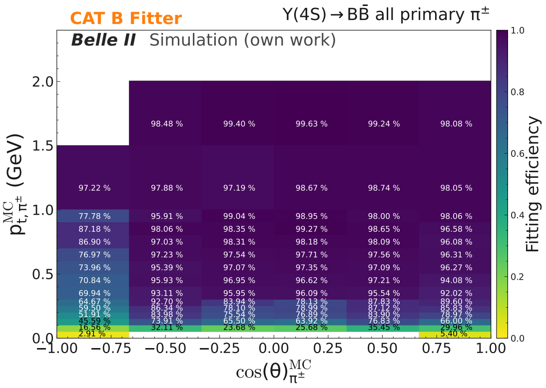


(d) Difference between *CAT Finder* (left)/*CAT B Finder* (right) and *Baseline Finder*.

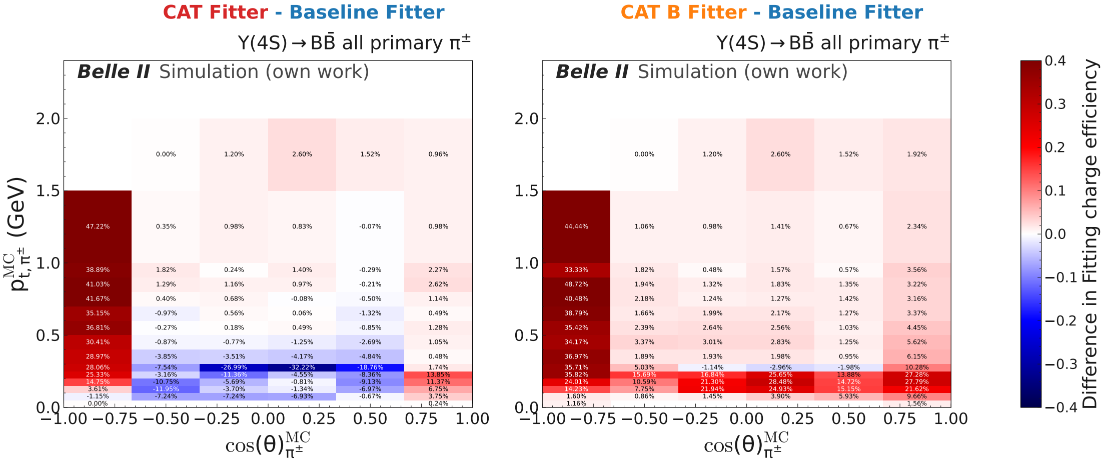
Figure 6.8: The track finding charge efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. 6.8a), *CAT Finder* (Fig. 6.8b), and *CAT B Finder* (Fig. 6.8c). The difference in track finding charge efficiency between the two new tracking algorithms to the *Baseline Finder* is given in Fig. 6.8d.



(a) *Baseline Finder*. (b) *CAT Finder*.

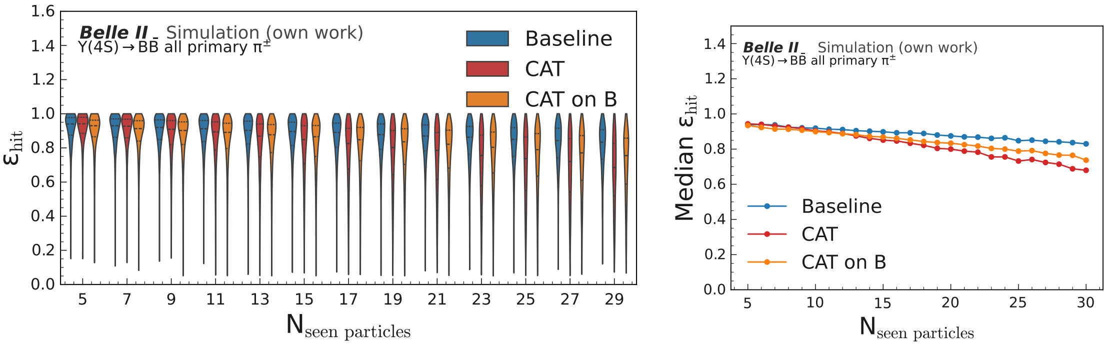


(c) *CAT B Finder*.



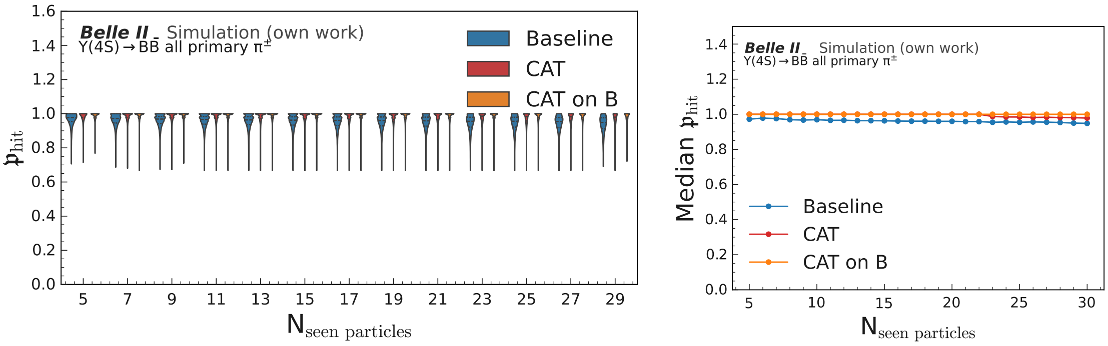
(d) Difference between *CAT Finder* (left)/*CAT B Finder* (right) and *Baseline Finder*.

Figure 6.9: The track fitting charge efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. 6.9a), *CAT Finder* (Fig. 6.9b), and *CAT B Finder* (Fig. 6.9c). The difference in track fitting charge efficiency between the two new tracking algorithms to the *Baseline Finder* is given in Fig. 6.9d. The same plot, but for the track finding charge efficiency is given in Fig. 6.8.



(a) Hit efficiency ε_{hit} distributions.

(b) Median hit efficiency ε_{hit}.



(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

Figure 6.10: Hit efficiency (top row) and hit purity (bottom row) over the number of particles seen in the CDC for all primary matched pions in the intersecting sample. The left column shows the distributions for the *Baseline Finder* in blue, *CAT Finder* in red, and *CAT B Finder* in orange, while the right column shows the median for all three algorithms.

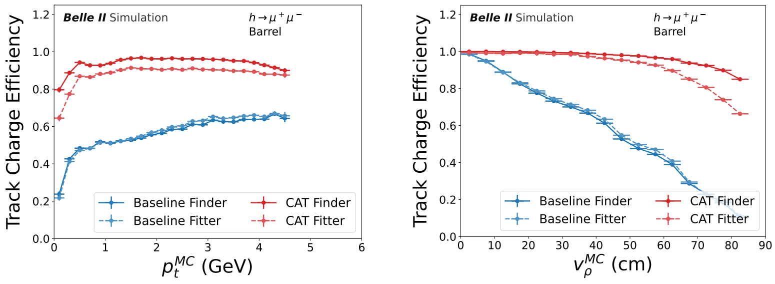
6.1.2 Displaced Tracks

For displaced tracks, two additional difficult track signatures are added on top of the prompt track signatures:

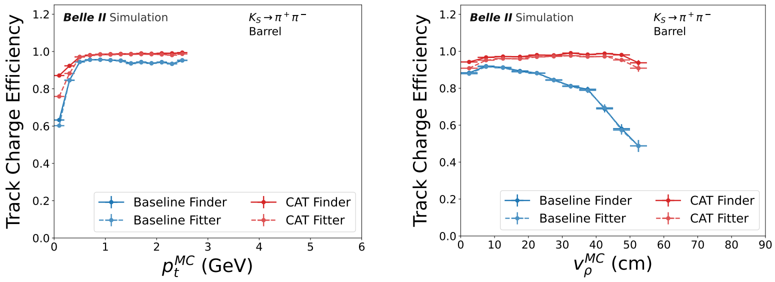
- Particles with small opening angles overlapping close together and with large opening angles, possibly going backwards in the detector.
- Particles where the vertex does not point back to the interaction point.

Pointing Vertices

I evaluated the track finding charge efficiency for pointing displaced vertices using simulated $h \to \mu^+\mu^-$ events, where the dark Higgs decays uniformly along its flight path into two opposite charged muons. I considered three mass hypotheses: $m_h = 0.5$ GeV (small opening angles), $m_h = 2.0$, GeV, and $m_h = 4.0$ GeV (large opening angles). Large masses lead to well-separated muon tracks within the same superlayer, which might even go from outside back to the interaction point, while small masses result in overlapping tracks. Due to the uniform lifetime distribution along the dark Higgs direction, these samples contain many highly displaced tracks.
Additionally, I studied events with single $K_S^0 \to \pi^+\pi^-$ decays, where the decay distance follows an exponential distribution.
Track finding and fitting charge efficiencies for the *Baseline Finder* and *CAT Finder* are shown in Fig. 6.11; finding and fitting efficiencies are in Section C.5. Integrated performance metrics over the full samples are summarized in Table 6.3 and Table 6.4.
Across all p_t, displacements, and detector regions, I observed that the *CAT Finder* consistently achieves higher efficiencies both before and after fitting. It also has the lowest fake (2.5%) and clone rates, reaching 85.4% combined efficiency per track over the full acceptance. In contrast, the *Baseline Finder* achieves 52.2% efficiency with a higher fake rate of 4.1%.
I further tested the algorithms on a signal sample near the expected Belle II sensitivity,[3, 10, 12], with $m_h = 1.5$ GeV, $\sin(\theta) = 10^{-4}$, and $c\tau = 21.5\,\text{cm}$. Here, the *CAT Finder* finds and fits both tracks in 87.2% of events, compared to 44.9% for the *Baseline Finder*. Fake rates are 2.5% (*CAT Finder*) and 3.3% (*Baseline Finder*). Restricting to the barrel, I measured 90.0% efficiency for the *CAT Finder* and 52.2% for the *Baseline Finder*, with fake rates of 2.1% and 3.0% of the tracks, respectively.
Unlike $h \to \mu^+\mu^-$ decays, $K_S^0 \to \pi^+\pi^-$ decays occur closer to the IP and at lower momentum. Despite these differences, the trend remains consistent: the *CAT Finder* significantly improves track finding and fitting efficiency, while maintaining similar or lower fake and clone rates compared to the *Baseline Finder*.



(a) $h \to \mu^+\mu^-$, p_t^{MC}. (b) $h \to \mu^+\mu^-$, v_ρ^{MC}.



(c) $K_S^0 \to \pi^+\pi^-$, p_t^{MC}. (d) $K_S^0 \to \pi^+\pi^-$, v_ρ^{MC}.

Figure 6.11: Figure and text taken from [1]: Track finding (empty markers) and combined track finding and fitting efficiency (filled markers) for (top) displaced tracks in $h \to \mu^+\mu^-$ events and in (bottom) $K_S^0 \to \pi^+\pi^-$ events with *high data beam backgrounds*, as function of (left) the true simulated transverse momentum p_t^{MC}, and (right) the true simulated displacement v_ρ^{MC} in the $x - y$ plane.

Table 6.3: Figure and text taken from [1]: The performance metrics per track for $h \to \mu^+\mu^-$ (m_h = [0.5,2.0,4.0] GeV) samples with *high data beam backgrounds* decaying uniformly along its flight direction into two charged particles (see Section 3.2 for details) for different track finding algorithms in different detector regions.

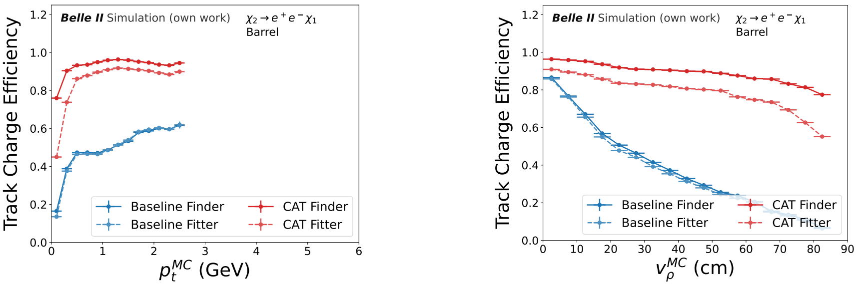
(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
			forward endcap		
Baseline Finder	$36.2^{+0.4}_{-0.4}$	$15.1^{+0.8}_{-0.8}$	$0.3^{+0.1}_{-0.1}$	$33.8^{+0.3}_{-0.4}$	$6.5^{+0.3}_{-0.3}$
CAT Finder	$88.1^{+0.2}_{-0.2}$	$15.8^{+0.4}_{-0.4}$	$0.64^{+0.09}_{-0.10}$	$83.9^{+0.3}_{-0.3}$	$4.8^{+0.2}_{-0.2}$
Baseline Fitter	$35.5^{+0.4}_{-0.4}$	$17.5^{+0.7}_{-0.7}$	$0.22^{+0.1}_{-0.13}$	$34.3^{+0.4}_{-0.4}$	$3.4^{+0.2}_{-0.2}$
CAT Fitter	$79.7^{+0.3}_{-0.3}$	$7.4^{+0.3}_{-0.3}$	$0.1^{+0.03}_{-0.04}$	$75.4^{+0.3}_{-0.3}$	$5.4^{+0.2}_{-0.2}$
			barrel		
Baseline Finder	$59.5^{+0.1}_{-0.1}$	$4.94^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.03}$	$56.4^{+0.1}_{-0.1}$	$5.13^{+0.08}_{-0.08}$
CAT Finder	$96.89^{+0.05}_{-0.05}$	$5.12^{+0.06}_{-0.06}$	$1.56^{+0.03}_{-0.03}$	$94.94^{+0.06}_{-0.06}$	$2.01^{+0.04}_{-0.04}$
Baseline Fitter	$58.8^{+0.1}_{-0.1}$	$3.57^{+0.06}_{-0.06}$	$0.33^{+0.02}_{-0.02}$	$57.4^{+0.1}_{-0.1}$	$2.36^{+0.05}_{-0.05}$
CAT Fitter	$92.75^{+0.07}_{-0.07}$	$2.12^{+0.04}_{-0.04}$	$0.54^{+0.02}_{-0.02}$	$89.21^{+0.08}_{-0.08}$	$3.81^{+0.05}_{-0.05}$
			backward endcap		
Baseline Finder	$17.2^{+0.4}_{-0.4}$	$4.4^{+0.2}_{-0.2}$	$0.32^{+0.06}_{-0.07}$	$15.1^{+0.3}_{-0.3}$	$12.1^{+0.7}_{-0.8}$
CAT Finder	$71.0^{+0.4}_{-0.4}$	$14.8^{+0.3}_{-0.3}$	$0.74^{+0.07}_{-0.07}$	$64.6^{+0.5}_{-0.5}$	$9.1^{+0.3}_{-0.3}$
Baseline Fitter	$16.5^{+0.4}_{-0.4}$	$4.8^{+0.3}_{-0.3}$	$0.23^{+0.05}_{-0.07}$	$15.6^{+0.3}_{-0.3}$	$5.8^{+0.5}_{-0.6}$
CAT Fitter	$58.0^{+0.5}_{-0.5}$	$3.1^{+0.1}_{-0.1}$	$0.08^{+0.02}_{-0.03}$	$53.3^{+0.5}_{-0.5}$	$8.2^{+0.3}_{-0.4}$

Table 6.4: Figure and text taken from [1]:The performance metrics per displaced pion track in $K_S^0 \to \pi^+\pi^-$ samples with *high data beam backgrounds* with a uniformly generated transverse momentum of $p_t(K_S^0) = [0.05 - 3]$ GeV. The average transverse decay distance is $v_\rho = 8.24\,\text{cm}$ (see Section 3.2 for details) for different track finding algorithms in different detector regions.

(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
forward endcap					
Baseline Finder	$63.2^{+0.2}_{-0.2}$	$3.5^{+0.1}_{-0.1}$	$0.13^{+0.02}_{-0.03}$	$62.5^{+0.2}_{-0.2}$	$1.2^{+0.06}_{-0.06}$
CAT Finder	$93.2^{+0.1}_{-0.1}$	$6.8^{+0.1}_{-0.1}$	$0.26^{+0.02}_{-0.02}$	$92.7^{+0.1}_{-0.1}$	$0.45^{+0.03}_{-0.03}$
Baseline Fitter	$61.9^{+0.2}_{-0.2}$	$4.1^{+0.1}_{-0.1}$	$0.1^{+0.02}_{-0.02}$	$61.2^{+0.2}_{-0.2}$	$1.23^{+0.06}_{-0.06}$
CAT Fitter	$88.6^{+0.1}_{-0.1}$	$3.01^{+0.08}_{-0.08}$	$0.08^{+0.01}_{-0.01}$	$86.6^{+0.2}_{-0.2}$	$2.23^{+0.07}_{-0.07}$
barrel					
Baseline Finder	$91.25^{+0.09}_{-0.09}$	$6.89^{+0.08}_{-0.08}$	$0.83^{+0.03}_{-0.03}$	$88.5^{+0.1}_{-0.1}$	$3.0^{+0.06}_{-0.06}$
CAT Finder	$96.15^{+0.06}_{-0.06}$	$11.52^{+0.09}_{-0.09}$	$1.99^{+0.04}_{-0.04}$	$95.56^{+0.07}_{-0.07}$	$0.61^{+0.03}_{-0.03}$
Baseline Fitter	$90.05^{+0.1}_{-0.1}$	$5.39^{+0.07}_{-0.07}$	$0.59^{+0.02}_{-0.03}$	$88.1^{+0.1}_{-0.1}$	$2.16^{+0.05}_{-0.05}$
CAT Fitter	$93.43^{+0.08}_{-0.08}$	$5.13^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$92.99^{+0.08}_{-0.08}$	$0.46^{+0.02}_{-0.02}$
backward endcap					
Baseline Finder	$44.0^{+0.2}_{-0.2}$	$2.51^{+0.08}_{-0.08}$	$0.1^{+0.02}_{-0.02}$	$42.7^{+0.2}_{-0.2}$	$3.0^{+0.1}_{-0.1}$
CAT Finder	$90.1^{+0.1}_{-0.1}$	$9.6^{+0.1}_{-0.1}$	$0.42^{+0.03}_{-0.03}$	$89.4^{+0.1}_{-0.1}$	$0.74^{+0.04}_{-0.04}$
Baseline Fitter	$42.6^{+0.2}_{-0.2}$	$2.24^{+0.08}_{-0.08}$	$0.07^{+0.01}_{-0.02}$	$41.2^{+0.2}_{-0.2}$	$3.2^{+0.1}_{-0.1}$
CAT Fitter	$83.2^{+0.2}_{-0.2}$	$2.35^{+0.07}_{-0.07}$	$0.12^{+0.01}_{-0.02}$	$79.3^{+0.2}_{-0.2}$	$4.7^{+0.1}_{-0.1}$

Non-pointing Vertices

I evaluated the track finding charge efficiency for non-pointing displaced vertices using simulated $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ events, where the χ_2 decays uniformly along its flight path into two charged particles and an invisible χ_1 (see Section 3.5). The results for the electrons are shown in Fig. 6.12 for the barrel region. In comparison to the $h \to \mu^+\mu^-$ samples, a large part of the tracks cannot be fitted, especially for electrons that are in the low momentum region. The total integrated efficiency per track is 43.7% for the *Baseline Finder* with a fake rate of 2.76%, whereas the efficiency for the *CAT Finder* is nearly doubled with 81.9% while keeping a lower fake rate of 1.54%. All track metrics are given in Table 6.5 split in detector regions. In total, the *CAT Finder* performance is far superior for both pointing and non-pointing tracks, even for difficult particle types like electrons, that produce bremsstrahlung.



(a) $\chi_2(\to e^+e^-\chi_1)$, p_t^{MC}.

(b) $\chi_2(\to e^+e^-\chi_1)$, v_ρ^{MC}.

Figure 6.12: Text adapted from [1]: Track finding (empty markers) and combined track finding and fitting efficiency (filled markers) for (top) displaced electron tracks in $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ events with *high data beam backgrounds*, as function of (left) the true simulated transverse momentum p_t^{MC}, and (right) the true simulated displacement v_ρ^{MC} in the $x - y$ plane.

Table 6.5: Text adapted from [1]: The performance metrics per track for $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ (m_h = [0.5, 1.0, 2.5] GeV and m_{χ_1} =2.5 GeV) samples with *high data beam backgrounds* for the χ_2 decaying uniformly along its flight direction into two charged particles and an invisible χ_1 (see Section 3.2 for details) for different track finding algorithms in different detector regions.

(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
			forward endcap		
Baseline Finder	$20.5^{+0.4}_{-0.4}$	$10.1^{+0.6}_{-0.6}$	$0.2^{+0.08}_{-0.12}$	$19.9^{+0.4}_{-0.4}$	$3.2^{+0.3}_{-0.4}$
CAT Finder	$72.5^{+0.4}_{-0.4}$	$7.2^{+0.2}_{-0.2}$	$0.66^{+0.08}_{-0.08}$	$68.2^{+0.4}_{-0.4}$	$5.9^{+0.2}_{-0.3}$
Baseline Fitter	$19.3^{+0.4}_{-0.4}$	$10.8^{+0.5}_{-0.6}$	$0.21^{+0.09}_{-0.13}$	$18.5^{+0.3}_{-0.3}$	$4.2^{+0.4}_{-0.4}$
CAT Fitter	$57.3^{+0.4}_{-0.4}$	$4.3^{+0.2}_{-0.2}$	$0.11^{+0.03}_{-0.04}$	$55.0^{+0.4}_{-0.4}$	$4.1^{+0.2}_{-0.2}$
			barrel		
Baseline Finder	$42.8^{+0.2}_{-0.2}$	$3.66^{+0.06}_{-0.06}$	$0.67^{+0.03}_{-0.03}$	$39.7^{+0.2}_{-0.2}$	$7.1^{+0.1}_{-0.1}$
CAT Finder	$94.38^{+0.08}_{-0.08}$	$3.33^{+0.04}_{-0.04}$	$2.58^{+0.04}_{-0.04}$	$88.7^{+0.1}_{-0.1}$	$5.98^{+0.08}_{-0.08}$
Baseline Fitter	$40.6^{+0.2}_{-0.2}$	$2.41^{+0.05}_{-0.05}$	$0.29^{+0.02}_{-0.02}$	$38.5^{+0.2}_{-0.2}$	$5.2^{+0.1}_{-0.1}$
CAT Fitter	$81.3^{+0.1}_{-0.1}$	$1.33^{+0.03}_{-0.03}$	$0.53^{+0.02}_{-0.02}$	$77.9^{+0.1}_{-0.1}$	$4.18^{+0.07}_{-0.07}$
			backward endcap		
Baseline Finder	$12.6^{+0.4}_{-0.4}$	$3.7^{+0.2}_{-0.2}$	$0.13^{+0.04}_{-0.05}$	$11.7^{+0.4}_{-0.4}$	$7.1^{+0.8}_{-0.9}$
CAT Finder	$63.2^{+0.6}_{-0.6}$	$8.1^{+0.2}_{-0.2}$	$1.1^{+0.07}_{-0.07}$	$56.5^{+0.6}_{-0.6}$	$10.6^{+0.4}_{-0.5}$
Baseline Fitter	$11.5^{+0.4}_{-0.4}$	$3.4^{+0.2}_{-0.2}$	$0.06^{+0.03}_{-0.04}$	$10.7^{+0.4}_{-0.4}$	$6.8^{+0.8}_{-0.9}$
CAT Fitter	$42.3^{+0.6}_{-0.6}$	$1.9^{+0.1}_{-0.1}$	$0.21^{+0.03}_{-0.04}$	$39.6^{+0.6}_{-0.6}$	$6.5^{+0.4}_{-0.4}$

6.2 Track Momentum Resolution

The *CAT Finder* provides estimates of the three-momentum for each condensation point used within this work as starting values for the subsequent track fitting algorithm GENFIT2. The resolution after the fitting step are the result of the trackfit described in Section 5.1. As for high transverse momentum, the tracks appear like straight lines. Detailed information on the drift lengths need to be combined to achieve a good momentum approximation. For this reason, the training dataset momentum range stopped at $p_t = 6\,\text{GeV}$. Even though this biases the model, as it learned that the prediction goes up to around $p_t = 6\,\text{GeV}$, this value is sufficient as a seed for the track fit for higher energetic particles. As the resolution of the *CAT Finder* is biased (see Section C.6 for details), the resolution is not shown for track momenta above $p_t = 4\,\text{GeV}$ for the *CAT Finder* before fitting.

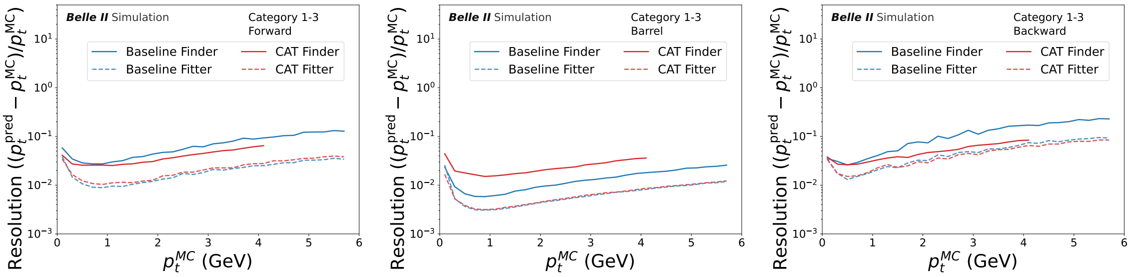
6.2.1 Prompt Tracks

I analyzed the momentum resolution of matched prompt tracks (categories 1–3 in Table 3.2), focusing on non-curling tracks found by both the *CAT Finder* and *Baseline Finder* algorithms (Section 6.1.1). The transverse $\eta(p_t)$ and longitudinal $\eta(p_z)$ resolutions are shown in Fig. 6.13.

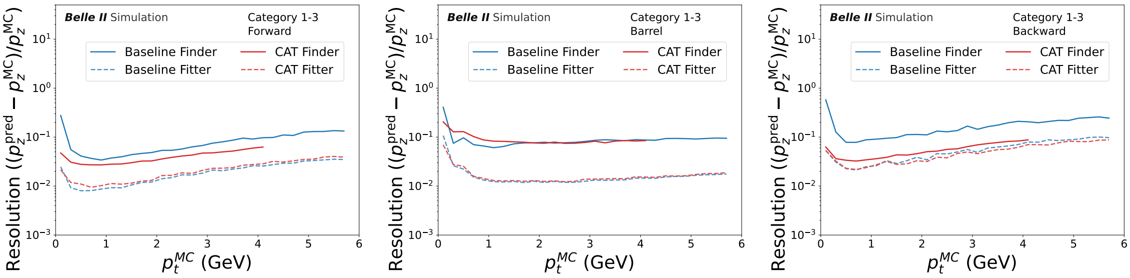
For tracks before fitting, the *CAT Finder* yields a consistent $\eta(p_t)$ of a few percent across all detector regions. In contrast, the *Baseline Finder* achieves better resolution in the barrel, reaching below 1%, but its performance drops in the endcaps due to lower hit efficiency and purity. The longitudinal resolution $\eta(p_z)$ shows similar behavior in the barrel for both algorithms, while *CAT Finder* again performs better in the endcaps.

After fitting with GENFIT2, both *CAT Finder* and *Baseline Finder* show nearly identical resolutions for $\eta(p_t)$ and $\eta(p_z)$ across all regions, which is expected due to the comparable hit efficiencies at higher p_t values.

The helix parameters (see Section 5.3.1) are compared in Section C.7

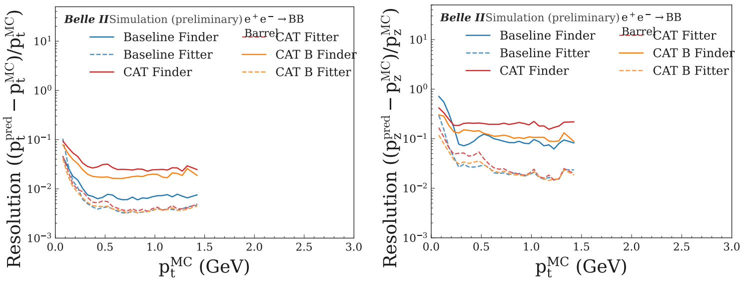


(a) Forward endcap. (b) Barrel. (c) Backward endcap.



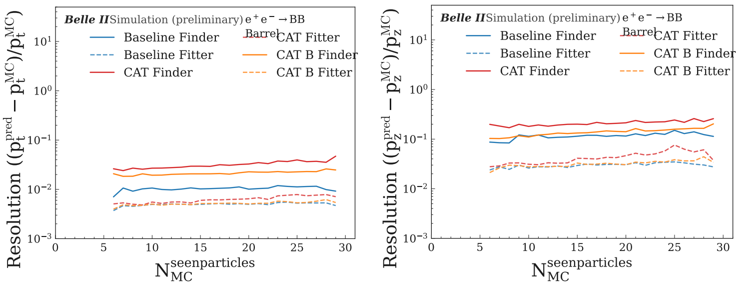
(d) Forward endcap. (e) Barrel. (f) Backward endcap.

Figure 6.13: Figure and text taken from [1]: Relative (top) transverse and (bottom) longitudinal momentum resolution as function of simulated transverse momentum p_t^{MC} for the intersecting prompt evaluation sample (category 1-3, *high data beam backgrounds*, see Table 3.2) in the (left) forward endcap, (center) barrel, and (right) backward endcap for tracks found by both (red) *CAT Finder* and (blue) *Baseline Finder*. For the *CAT Finder* the resolution is shown only for p_t <4 GeV, see Section C.6 for details.



(a) $\eta(p_t)$ for p_t^{MC}.

(b) $\eta(p_z)$ for p_t^{MC}.



(c) $\eta(p_t)$ for $N_{\text{seen}}^{\text{MC}}$.

(d) $\eta(p_z)$ for $N_{\text{seen}}^{\text{MC}}$.

Figure 6.14: Text adapted from [1]: Relative (left) transverse and (right) longitudinal momentum resolution as function of simulated transverse momentum p_t^{MC} (top) and number of seen particles $N_{\text{seen}}^{\text{MC}}$ (bottom) for the intersecting prompt pions from $B\bar{B}$, *high data beam backgrounds*, see Table 3.2) in the barrel, and (right) backward endcap for tracks found by (red) *CAT Finder*, (orange) *CAT B Finder*, and (blue) *Baseline Finder*. For the *CAT Finder* the resolution is shown only for p_t <4 GeV, see Section C.6 for details.

In high-multiplicity $B\bar{B}$ events, the resolution before fitting is also larger for both *CAT Finder* and *CAT B Finder* compared to *Baseline Finder*, shown in Fig. 6.14. Curling tracks are not removed here. However, *CAT B Finder* improves over *CAT Finder*. After fitting, *CAT Finder* performs worse at low p_t. This is correlated with $N_{\text{seen}}^{\text{MC}}$, and can be seen in Fig. 6.14c with resolution decreasing as $N_{\text{seen}}^{\text{MC}}$ increases. The *CAT Finder* shows the steepest drop in hit efficiency, as illustrated in Fig. 6.10b, which directly correlates with its poorer momentum resolution. The *CAT B Finder*, however, matches the *Baseline Finder* in resolution, even though the hit efficiency is slightly decreased compared to *Baseline Finder*.

6.2.2 Displaced Tracks

I evaluate the momentum resolution for matched displaced tracks using events from dark Higgs and single $K_S^0 \to \pi^+\pi^-$ decays, as described in Section 3.3 and Section 3.5. I include curling tracks and focus on intersecting and additional samples in the barrel region.

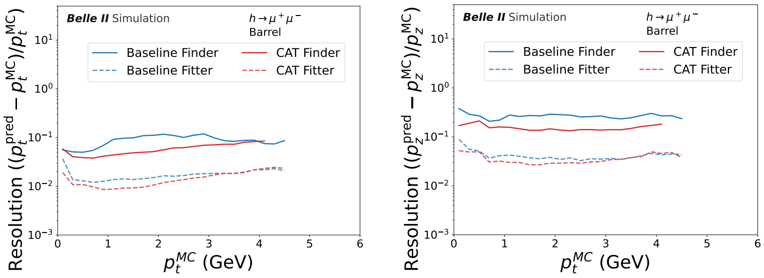
The transverse and longitudinal momentum resolutions, $\eta(p_t)$ and $\eta(p_z)$, for the intersecting sample of both *CAT Finder* and the *Baseline Finder* are shown in Fig. 6.15. Extended results for the additional sample for *CAT Finder* in Section C.8 and the comparison for $K_S^0 \to \pi^+\pi^-$ is in Section C.9. The *Baseline Finder* sample for additional displaced tracks is too small for reliable statistics.

Before track fitting, *CAT Finder* achieves better $\eta(p_t)$ and $\eta(p_z)$ resolutions across most p_t and displacement values, except at very small displacements, where *Baseline Finder* performs better.

For the $K_S^0 \to \pi^+\pi^-$ sample, which has smaller displacements on average, both *CAT Finder* and *Baseline Finder* show improved resolution compared to the $h \to \mu^+\mu^-$ sample.

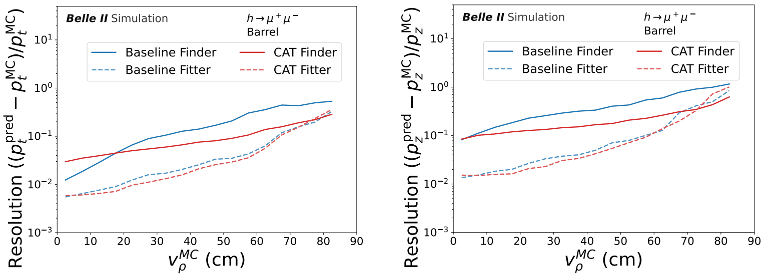
After fitting with GENFIT2, both algorithms yield comparable momentum resolutions.

For electrons from $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$, I observe better resolution for $\eta(p_t)$ and $\eta(p_z)$ after the track finding, except in the very small displacement case in Fig. 6.16. This trend holds even after fitting with GENFIT2 and correlates with the higher hit efficiency achieved by the *CAT Finder* (see Section C.10).



(a) $\eta(p_t)$ as function of p_t in the barrel.

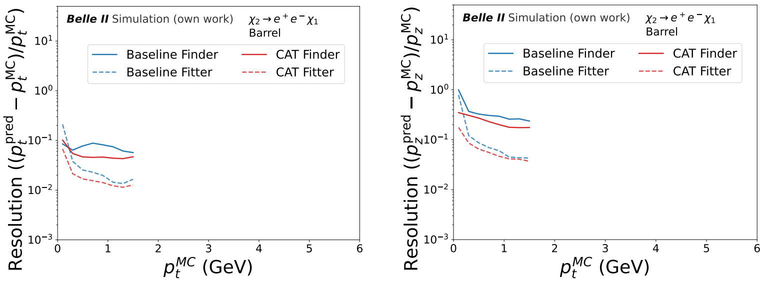
(b) $\eta(p_z)$ as function of p_t in the barrel.



(c) $\eta(p_t)$ as function of displacement v_ρ in the barrel.

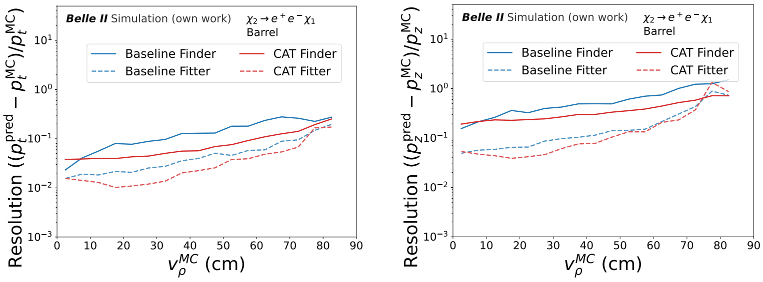
(d) $\eta(p_z)$ as function of v_ρ in the barrel.

Figure 6.15: Figure and text taken from [1]: Relative resolution of (first column) transverse and (second column) longitudinal momentum as function of simulated transverse momentum p_t^{MC} (top row) and simulated displacement v_ρ^{MC} (bottom row) for displaced tracks from $h \to \mu^+\mu^-$ decays with *high data beam backgrounds* in the barrel for tracks found by both (red) *CAT Finder* and (blue) *Baseline Finder* for the intersecting sample. For the *CAT Finder* the resolution is shown only for p_t <4 GeV, see Section C.6 for details.



(a) $\eta(p_t)$ as function of p_t in the barrel.

(b) $\eta(p_z)$ as function of p_t in the barrel.



(c) $\eta(p_t)$ as function of displacement v_ρ in the barrel.

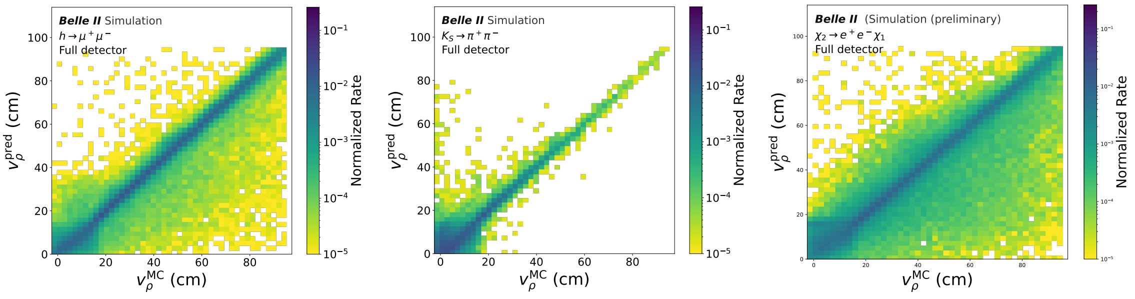
(d) $\eta(p_z)$ as function of v_ρ in the barrel.

Figure 6.16: Text adapted from [1]: Relative resolution of (first column) transverse and (second column) longitudinal momentum as function of simulated transverse momentum p_t^{MC} (top row) and simulated displacement v_ρ^{MC} (bottom row) for displaced electron tracks from $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ decays with *high data beam backgrounds* in the barrel for tracks found by both (red) *CAT Finder* and (blue) *Baseline Finder* for the intersecting sample. For the *CAT Finder* the resolution is shown only for p_t <4 GeV, see Section C.6 for details.

6.3 Position Reconstruction

In this section, I focus on evaluating the position predictions from the *CAT Finder*, which directly uses the GNN output. Neither the *Baseline Finder* nor any post-fitting methods provide comparable position information, as their track models are defined purely by a helix without a specific start point.

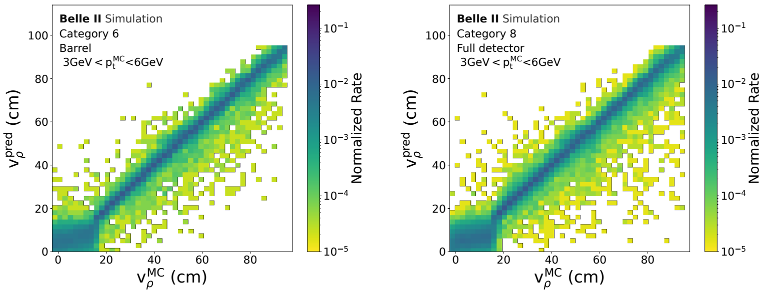
In Fig. 6.17, I show the *CAT Finder* 's position predictions for truth-matched displaced tracks across various samples. For displaced tracks from pointing vertex samples like $h \to \mu^+\mu^-$ and $K_S^0 \to \pi^+\pi^-$, the *CAT Finder* yields unbiased and well-resolved predictions, even in the inner detector region lacking nearby CDC hits (Fig. 6.17a below 16 cm, Fig. 6.17b). For non-pointing decays such as the two-electrons in $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ sample, the resolution broadens overall, but remains accurate overall. This suggests the *CAT Finder* doesn't simply extrapolate from a single trajectory, it likely uses context from nearby tracks to infer the shared decay vertex.



(a) $h \rightarrow \mu^+\mu^-$.

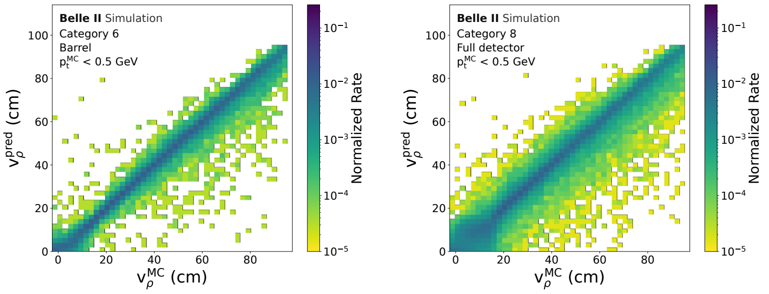
(b) $K_S^0 \rightarrow \pi^+\pi^-$.

(c) $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$.



(d) Displaced tracks (category 6), 3 GeV $< p_t <$ 6 GeV.

(e) Displaced angled tracks (category 8), 3 GeV $< p_t <$ 6 GeV.



(f) Displaced tracks (category 6), low $p_t <$ 0.5 GeV

(g) Displaced angled tracks (category 8), low $p_t <$ 0.5 GeV

Figure 6.17: Figure and text adapted from [1]: Two-dimensional histograms showing the correlation between the reconstructed position of the *CAT Finder* model output v_ρ^{pred} in the $x-y$ plane and the simulated position v_ρ^{MC} for (a) $h \rightarrow \mu^+\mu^-$, (b) $K_S^0 \rightarrow \pi^+\pi^-$, (c) $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$, (d) displaced tracks with high transverse momentum, (e) displaced angled tracks with high transverse momentum, (f) displaced tracks with low transverse momentum, and (g) displaced angled tracks with low transverse momentum, each with *high data beam backgrounds*.

For individual tracks, there is a rather complex and non-trivial geometrical behaviour learned by the model. When examining low p_t tracks within the CDC, I observe that the GNN tends to predict starting points along the helix where the inferred, negative momentum vector points back toward the interaction point. As a large part of the training events are the transition samples (category 5-7) between the prompt tracks and the vertex samples, the training is biased towards tracks being extrapolated towards the interaction

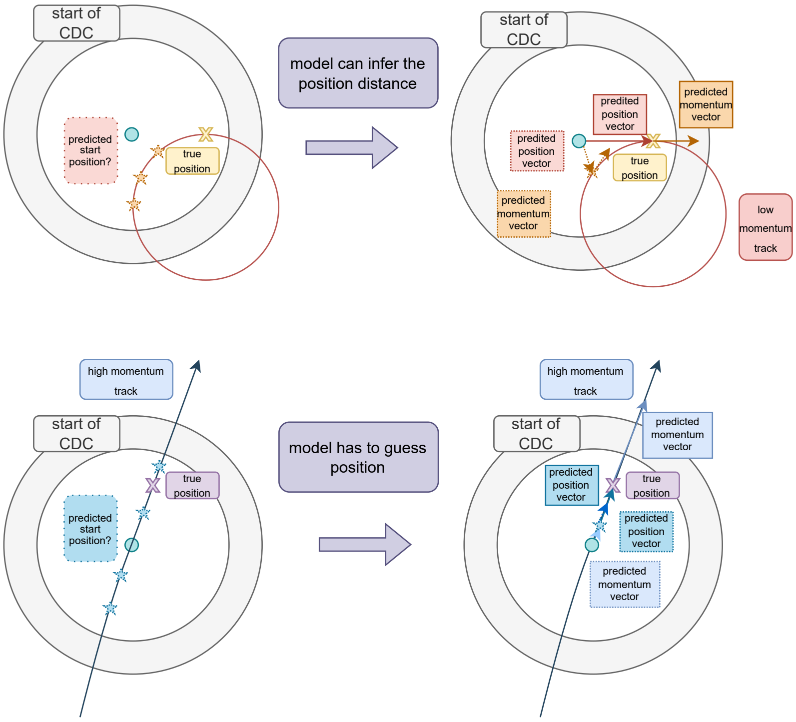


Figure 6.18: Figure and text taken from [1]: Illustration how the GNN learns to predict the starting position for particles with low transverse momentum even without the presence of another nearby track from a common vertex.

point, which the model learned. As a result, the *CAT Finder* infers reliable start positions even for tracks with $v_{\rho}^{MC} \lesssim 16, \text{cm}$ (Fig. 6.17f).

For higher p_t tracks, this inference breaks down. The near-linear trajectory leads to ambiguity in momentum direction, and the predictive power vanishes in the inner region (Fig. 6.17d). To confirm this, I evaluated tracks from category 8 samples containing only individual non-pointing particles. As expected, the *CAT Finder* shows no meaningful prediction in the inner detector (Fig. 6.17e, Fig. 6.17g).

6.4 Robustness to Variable Detector Conditions

6.4.1 Beam-background Conditions

As outlined in Section 3.1.1, beam background conditions fluctuate due to changes in accelerator settings. In particular, changes in the beams' current, which happen during the beam tuning in a run, can lead to substantial variations within a run. It is therefore essential for reconstruction algorithms to remain robust under these dynamic conditions, given by the varying CDC occupancy.

The reconstruction algorithms must be robust under these changes. The current model training strategy I employ consists of pre-training the model on the low simulated backgrounds and fine-tuning it on the high data backgrounds. For analyzing the performance of the model, I evaluate the model trained on low simulated backgrounds with a median occupancy of 440 beam background hits (see Table 3.1). This model is tested on both the low simulated background samples (Table 6.6) and the high data background samples with a median occupancy of 1216 hits (Table 6.7) for the samples of categories 1–3. In addition to the hit occupancy, the wire efficiency maps also differ between the two evaluation samples. They are shown in Fig. 3.4c for the low simulated background and in Fig. 3.4b for the high simulated background. The model trained on high data background samples is reported in Table 6.1. Comparing the track fitting charge efficiency between the low simulated backgrounds and the high data backgrounds for the *Baseline Finder*, I observe a significant drop in the endcaps. This is also the case for the *CAT Finder*, but to a much smaller extent. When applying the pre-trained model directly to high background data, the performance is slightly lower than that of the model trained on high background. Specifically, the track fitting efficiency drops by about 1 percentage point in the barrel and 3–4 points in the endcaps. However, the fake rate is slightly improved, suggesting that hyperparameter tuning might suffice to compensate for efficiency loss.

Overall, even without specific retraining, the *CAT Finder* continues to outperform the *Baseline Finder* across all detector regions, demonstrating strong robustness to beam background fluctuations.

Table 6.6: Table and text taken from [1]: performance metrics for the prompt evaluation samples (category 1-3, see Table 3.2 and Section 3.2 for details) for non-curling tracks for *CAT Finder* and *Baseline Finder* in different detector regions for a model trained and evaluated on *low simulated beam-background.* Uncertainties below <0.01% are not shown in the table.

(in %)	$\varepsilon_{\mathrm{trk}}$	$\mathfrak{r}_{\mathrm{fake}}$	$\mathfrak{r}_{\mathrm{clone}}$	$\varepsilon_{\mathrm{trk,ch}}$	$\mathfrak{r}_{\mathrm{wrong\ ch.}}$
			forward endcap		
Baseline Finder	$87.05^{+0.09}_{-0.09}$	$0.83^{+0.03}_{-0.03}$	0.01	$84.97^{+0.1}_{-0.1}$	$2.39^{+0.05}_{-0.05}$
CAT Finder	$99.26^{+0.02}_{-0.02}$	$1.02^{+0.03}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$99.22^{+0.02}_{-0.02}$	$0.03_{-0.01}$
Baseline Fitter	$85.18^{+0.1}_{-0.1}$	$0.78^{+0.03}_{-0.03}$	0.01	$84.3^{+0.1}_{-0.1}$	$0.99^{+0.03}_{-0.03}$
CAT Fitter	$97.12^{+0.05}_{-0.05}$	$0.32^{+0.02}_{-0.02}$	$0.04^{+0.01}_{-0.01}$	$96.42^{+0.05}_{-0.05}$	$0.72^{+0.02}_{-0.02}$
			barrel		
Baseline Finder	$98.71^{+0.03}_{-0.03}$	$2.06^{+0.04}_{-0.04}$	0.03	$96.73^{+0.05}_{-0.05}$	$2.0^{+0.04}_{-0.04}$
CAT Finder	$99.72^{+0.01}_{-0.01}$	$2.15^{+0.04}_{-0.04}$	$0.42^{+0.02}_{-0.02}$	$99.4^{+0.02}_{-0.02}$	$0.33^{+0.02}_{-0.02}$
Baseline Fitter	$97.68^{+0.04}_{-0.04}$	$1.75^{+0.04}_{-0.04}$	0.01	$96.27^{+0.05}_{-0.05}$	$1.44^{+0.03}_{-0.03}$
CAT Fitter	$98.13^{+0.04}_{-0.04}$	$0.97^{+0.03}_{-0.03}$	$0.13^{+0.01}_{-0.01}$	$97.97^{+0.04}_{-0.04}$	$0.17^{+0.01}_{-0.01}$
			backward endcap		
Baseline Finder	$69.5^{+0.1}_{-0.1}$	$0.72^{+0.03}_{-0.03}$	0.02	$66.2^{+0.1}_{-0.1}$	$4.66^{+0.07}_{-0.07}$
CAT Finder	$98.54^{+0.03}_{-0.03}$	$0.75^{+0.02}_{-0.02}$	$0.11^{+0.01}_{-0.01}$	$98.43^{+0.03}_{-0.03}$	$0.11^{+0.01}_{-0.01}$
Baseline Fitter	$67.8^{+0.1}_{-0.1}$	$0.63^{+0.03}_{-0.03}$	0.02	$65.8^{+0.1}_{-0.1}$	$2.98^{+0.06}_{-0.06}$
CAT Fitter	$95.12^{+0.06}_{-0.06}$	$0.3^{+0.02}_{-0.02}$	$0.03_{-0.01}$	$91.66^{+0.08}_{-0.08}$	$3.63^{+0.05}_{-0.05}$

Table 6.7: Table and text taken from [1]: The performance metrics for the evaluation samples for different track finding algorithms in different detector regions evaluated on *high data beam background*, but trained on *low simulated beam background* for non-curling tracks. Uncertainties below <0.01% are not shown in the table.

(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
			forward endcap		
Baseline Finder	$80.1^{+0.1}_{-0.1}$	$0.55^{+0.02}_{-0.02}$	0.01	$78.4^{+0.1}_{-0.1}$	$2.06^{+0.04}_{-0.04}$
CAT Finder	$97.73^{+0.04}_{-0.04}$	$1.46^{+0.03}_{-0.03}$	$0.59^{+0.02}_{-0.02}$	$97.18^{+0.05}_{-0.05}$	$0.56^{+0.02}_{-0.02}$
Baseline Fitter	$78.1^{+0.1}_{-0.1}$	$0.49^{+0.02}_{-0.02}$	0.01	$77.1^{+0.1}_{-0.1}$	$1.37^{+0.04}_{-0.04}$
CAT Fitter	$93.89^{+0.07}_{-0.07}$	$0.22^{+0.01}_{-0.01}$	$0.19^{+0.01}_{-0.01}$	$91.93^{+0.08}_{-0.08}$	$2.09^{+0.04}_{-0.04}$
			barrel		
Baseline Finder	$97.97^{+0.04}_{-0.04}$	$2.31^{+0.04}_{-0.04}$	$0.05^{+0.01}_{-0.01}$	$95.92^{+0.06}_{-0.06}$	$2.09^{+0.04}_{-0.04}$
CAT Finder	$99.5^{+0.02}_{-0.02}$	$2.25^{+0.04}_{-0.04}$	$1.94^{+0.04}_{-0.04}$	$98.69^{+0.03}_{-0.03}$	$0.82^{+0.03}_{-0.03}$
Baseline Fitter	$96.88^{+0.05}_{-0.05}$	$1.83^{+0.04}_{-0.04}$	0.03	$95.5^{+0.06}_{-0.06}$	$1.42^{+0.03}_{-0.03}$
CAT Fitter	$96.89^{+0.05}_{-0.05}$	$0.76^{+0.02}_{-0.02}$	$0.45^{+0.02}_{-0.02}$	$96.46^{+0.05}_{-0.05}$	$0.44^{+0.02}_{-0.02}$
			backward endcap		
Baseline Finder	$60.5^{+0.1}_{-0.1}$	$1.08^{+0.04}_{-0.04}$	$0.03^{+0.01}_{-0.01}$	$58.0^{+0.1}_{-0.1}$	$4.08^{+0.07}_{-0.07}$
CAT Finder	$95.02^{+0.06}_{-0.06}$	$1.03^{+0.03}_{-0.03}$	$0.4^{+0.02}_{-0.02}$	$94.28^{+0.06}_{-0.06}$	$0.78^{+0.02}_{-0.03}$
Baseline Fitter	$58.8^{+0.1}_{-0.1}$	$0.92^{+0.03}_{-0.04}$	$0.02_{-0.01}$	$56.8^{+0.1}_{-0.1}$	$3.28^{+0.06}_{-0.06}$
CAT Fitter	$87.55^{+0.09}_{-0.09}$	$0.44^{+0.02}_{-0.02}$	$0.16^{+0.01}_{-0.01}$	$82.8^{+0.1}_{-0.1}$	$5.4^{+0.07}_{-0.07}$

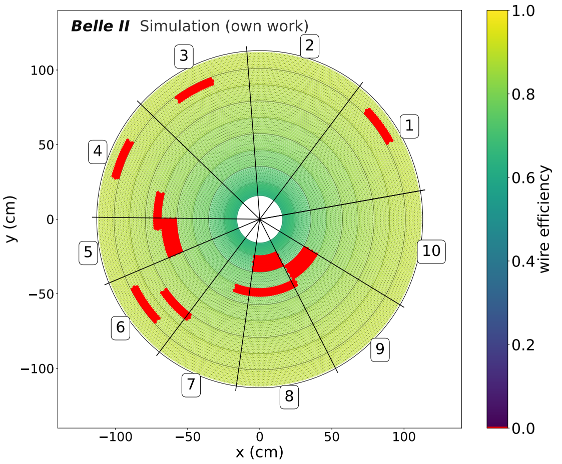
6.4.2 Detector Ageing Effects

With increasing luminosity at Belle II, hardware-related issues such as reduced wire efficiency and failing readout boards become more prominent, as discussed in Section 3.1.2. While I already analyzed the model's response to different wire efficiency maps in the previous section, I now investigate its robustness under future conditions.

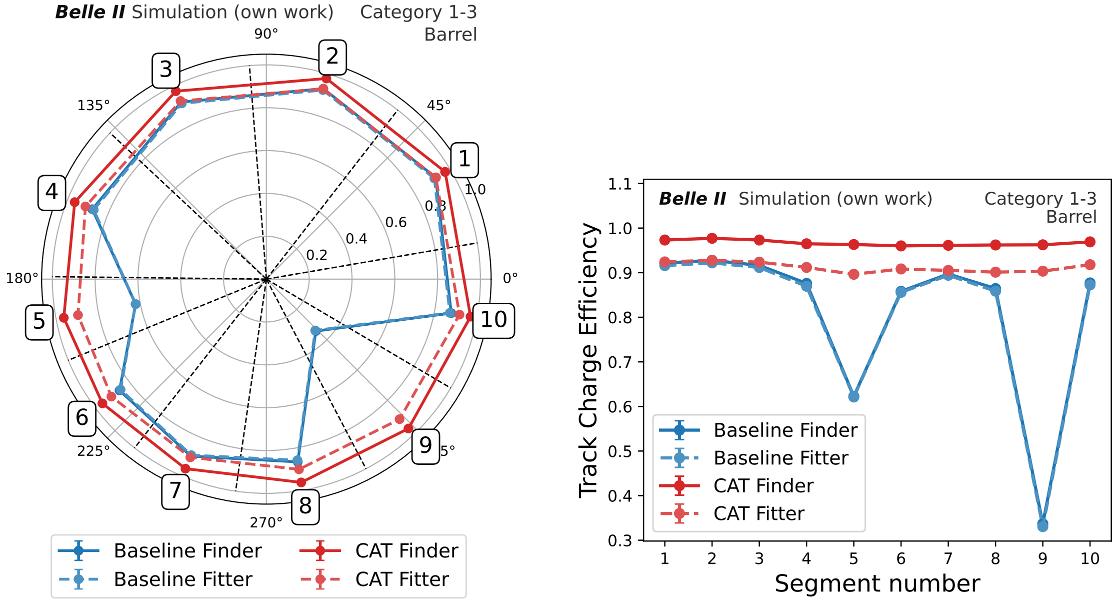
To this end, I approximate a realistic wire efficiency scenario expected in run II, as shown in Fig. 3.4d. I then evaluate both the *CAT Finder* and *Baseline Finder* on this new efficiency map using the technical samples from categories 1–3.

To study localized detector effects, the detector is divided into ϕ-regions aligned with the masked board positions, as shown in Fig. 6.19a. These include a stereo hole in superlayer U2 (segment 8), an axial hole in A3 (segment 9), and another axial hole in A5 (segment 5). The track charge finding and fitting charge efficiency drops massively for the *Baseline Finder* for the two segments 5 and 9 with the axial holes. For the *Baseline Finder*, the track fitting charge efficiency in segments without axial holes is 92%. However, it drops to 33.1% for the inner axial hole in segment 9 and to 62.2% for the outer axial hole in segment 5. The stereo hole in segment 8, on the other hand, does not seem to impact the *Baseline Finder* performance. This is expected as only the hits on axial superlayers are included in the *Baseline Finder* track finding algorithm (see Section 5.2). The *CAT Finder* remains robust against all superlayer holes, achieving a stable track fitting charge efficiency of approximately 92% over the full ϕ regions. These results show that the current *Baseline Finder* algorithm is not robust against hardware-related challenges. It introduces ϕ-dependent asymmetries, which are difficult for analysts to handle or correct. In contrast, the *CAT Finder* maintains stable performance under these conditions, showing strong robustness against hardware failures. Beyond the superior track charge efficiency, the *CAT Finder* model also shows improved hit efficiency and momentum resolution (see Section C.11).

While the tracking algorithms must work reliably even when boards fail during data taking, offline reconstruction can take advantage of known detector conditions. To further improve performance, in the thesis I supervised [35], the fine-tuning of the *CAT Finder* was specifically analyzed for runs with degraded hardware. In this work, the *CAT Finder* was fine-tuned on the exact wire efficiency and board configuration of each run. The results demonstrate that while the general *CAT Finder* already outperforms the *Baseline Finder* in all tested scenarios, fine-tuning leads to further improvements. Furthermore, training on a broad set of efficiencies and wire board maps improves generalization and adaptability to unknown detector conditions. This thesis also investigates worst-case scenarios. Wire efficiencies below 0.5 (see Fig. 3.3b) are found to significantly degrade the performance of the *Baseline Finder*, whereas the *CAT Finder* is able to recover. These findings are relevant for stable CDC operation before the next Belle II upgrade and demonstrate that the *CAT Finder* offers a reliable, future-proof solution for track reconstruction.



(a) Wire efficiency map sectioned in ϕ segments.



(b) Polar View.

(c) Unrolled, zoomed view.

Figure 6.19: The approximated wire efficiency map, described in detail in Section 3.1.2, sectioned in ϕ-segments in Fig. 6.19a. The track finding and fitting charge efficiency is given for the *Baseline Finder* (blue) and the *CAT Finder* (red) for the ϕ segments in the polar view (Fig. 6.19b) and the unrolled view (Fig. 6.19b).

Chapter 7

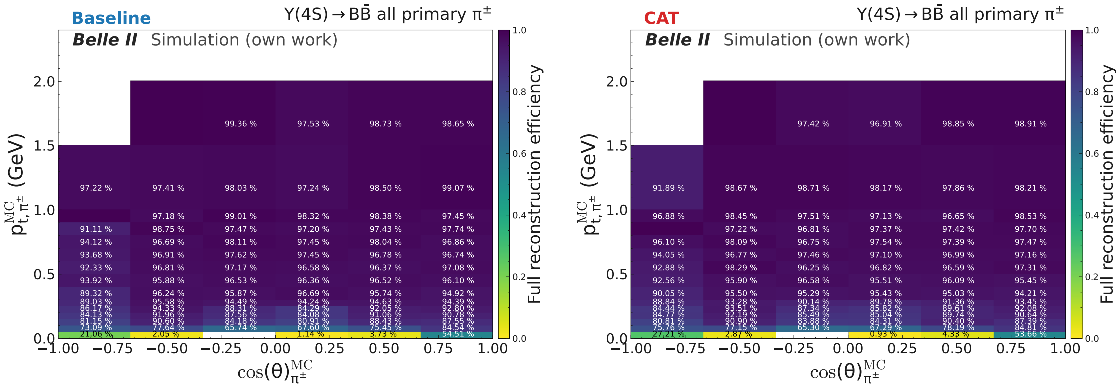
Validation of Full Detector Reconstruction

After analyzing the performance of my new GNN based track finder, the *CAT Finder*, for the CDC detector only (see Chapter 6), the next step is to validate the performance within the full Belle II reconstruction. While I validated and compared the *CAT Finder* with the *Baseline Finder* within the official Belle II validation framework in detail, I summarize the most important validation metrics in this chapter.
First, I validate the global track reconstruction performance, which is based on the combination of all three tracking detectors (Section 5.4), for both prompt and displaced tracks in Section 7.1. In the next step, I examine the track charge asymmetry in Section 7.2, resulting from reconstruction differences between positively and negatively charged particles. Lastly, I validate the implementation of my tracking algorithm into the Belle II framework by examining the particle identification probing both the correct implementation within the tracking framework and the correct extrapolation of tracks to the other subdetectors in Section 7.3. The samples in this chapter are statistically independent between the algorithms, as they have been independently produced for these studies and statistically independent from the CDC-only validation in the previous chapter.

7.1 Track Efficiency

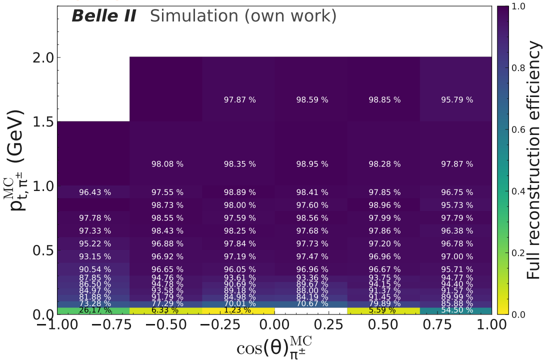
7.1.1 Prompt Tracks in $B\bar{B}$

I can now directly compare the CDC-only results with those obtained after full reconstruction. Particles are only counted as fully reconstructed if their charge is correctly identified. The different full reconstruction efficiencies (Eq. (4.9)) for pions for the three algorithms, together with their comparison, are presented in Fig. 7.1. For the *Baseline Finder*, the previously reported poor performance in the endcaps is recovered by the SVD track finding, as described in Section 5.4. In the case of the *CAT Finder*, it was shown in the previous chapter that while the tracks in the barrel region are successfully found (Fig. 6.8b) not

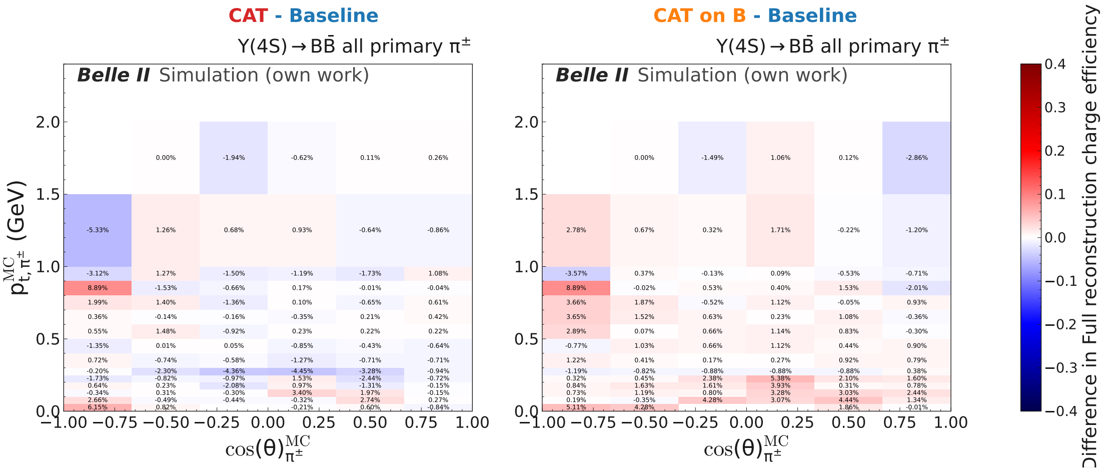


(a) *Baseline Finder*.

(b) *CAT Finder*.



(c) *CAT B Finder*.



(d) Difference in the full reconstruction charge efficiency for the *CAT Finder* and *CAT B Finder* vs the *Baseline Finder*.

Figure 7.1: The full reconstruction efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. 7.1a), *CAT Finder* (Fig. 7.1b), and *CAT B Finder* (Fig. 7.1c). The difference between the two new tracking algorithms to the *Baseline Finder* is given as difference in full reconstruction efficiency in Fig. 7.1d. The same plot but for the CDC-only track charge finding and fitting efficiency is given in Fig. 6.8 and Fig. 6.9.

all of them can be fitted (Fig. 6.9b), leading to a reduced track charge fitting efficiency compared to the *Baseline Finder*. This is due to curlers not being properly handled at this stage, as discussed in the previous chapter in Section 6.1.1. However, this performance degradation is also mitigated by the SVD track finding. The *CAT B Finder* model, on the other hand, already showed superior performance compared to both *Baseline Finder* and *CAT Finder* in the CDC-only track finding and fitting. When looking at the comparison, I now observe that the *CAT B Finder* model remains better by 1 percentage point in the barrel region than the *Baseline Finder*, in contrast to the 5 percentage points improvement observed in the previous chapter. This reduction in performance gap is due to the *flip and refit* module described in Section 5.4, which is applied only to the *Baseline Finder*. Tracks that previously had incorrect charge predictions are now mostly refitted and stored with the correct charge after running this module. Since both *CAT Finder* and *CAT B Finder* already predict the charge correctly, there is no need to apply the module to these algorithms, simplifying the overall optimization process.

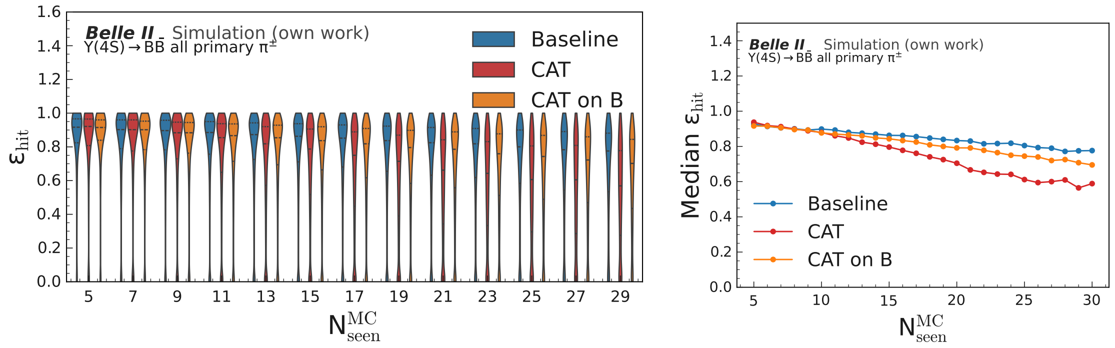
The total integrated values for prompt pion efficiency are summarized in Table 7.1. The *CAT Finder* performs 0.25 percentage point worse than the *Baseline Finder* in terms of efficiency and also exhibits a slightly higher fake rate of 4.08% compared to 3.44% for *Baseline Finder*. Additionally, the clone rate for *CAT Finder* is 4.76%, which is higher than the 3.29% for *Baseline Finder*. The *CAT B Finder* model achieves the highest efficiency overall, with improvements of one percentage point while also maintaining a lower fake rate compared to the *Baseline Finder*.

I observe the same trend in hit efficiency and hit purity over the number of $N_{\text{seen}}^{\text{MC}}$ in the full reconstruction in Fig. 7.2, as in the CDC-only case shown in Fig. 6.10. Both *CAT Finder* and *CAT B Finder* maintain the highest purity, but their hit efficiency drops more noticeably with increasing $N_{\text{seen}}^{\text{MC}}$ for more than 20 seen particles in the event, more so for *CAT B Finder* than for *Baseline Finder*, and most significantly for *CAT Finder*.

The results for electrons, muons, kaons, and protons are summarized in Section D.1. For both kaons and protons, the full reconstruction efficiency is reduced when using the *CAT Finder* compared to the *Baseline Finder*. The *CAT B Finder* achieves similar results as the *Baseline Finder*. The hit purity for all particle types is consistently higher for both *CAT Finder* algorithms than for the *Baseline Finder*, though the hit efficiency decreases with increasing $N_{\text{seen}}^{\text{MC}}$. As observed for pions, this decrease is more pronounced

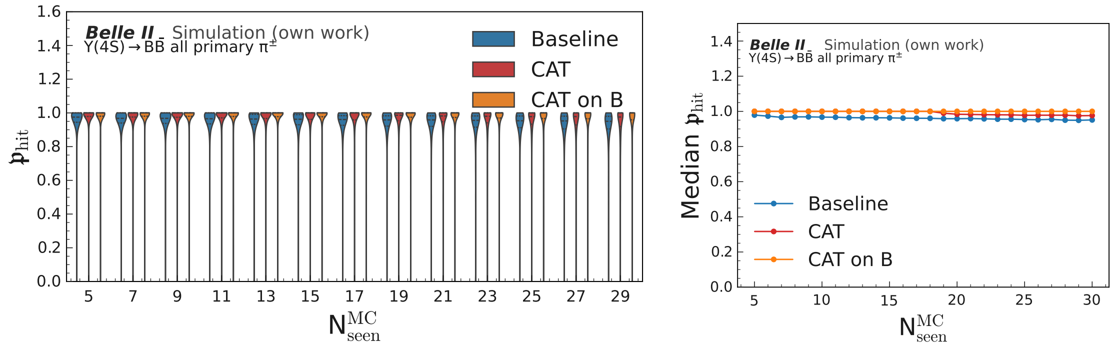
Table 7.1: Full reconstruction efficiency, fake and clone rate for prompt pions for the *Baseline Finder*, *CAT Finder* and *CAT B Finder* in $B\bar{B}$ events.

Algorithm	**π Efficiency**	**Fake Rate**	**π Clone Rate**
Baseline	$90.57 \pm 0.07\%$	$3.44 \pm 0.03\%$	$3.29 \pm 0.04\%$
CAT	$90.32 \pm 0.07\%$	$4.08 \pm 0.04\%$	$4.76 \pm 0.05\%$
CAT on B	$91.65 \pm 0.07\%$	$3.17 \pm 0.04\%$	$3.70 \pm 0.04\%$



(a) Hit efficiency ε_{hit} distributions.

(b) Median hit efficiency ε_{hit}.



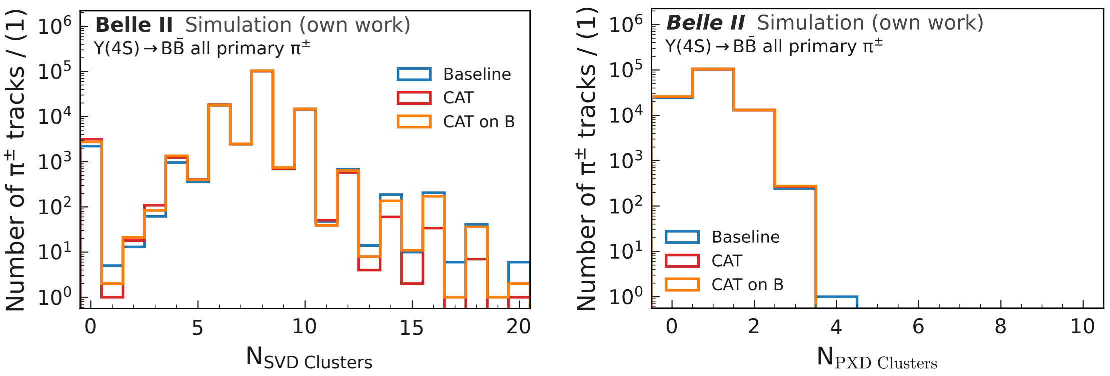
(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

Figure 7.2: Hit efficiency (top row) and hit purity (bottom row) over the number of particles seen in the CDC for all primary matched pions. The left column shows the distributions for the *Baseline Finder* in blue, *CAT Finder* in red, and *CAT B Finder* in orange, while the right column shows the median for all three algorithms. The same plot for the CDC-only reconstruction is given in Fig. 6.10.

for *CAT Finder* than for *CAT B Finder*. For muons, however, the hit efficiency remains unchanged between *CAT B Finder* and *Baseline Finder*, and the drop in efficiency for the *CAT Finder* is smaller than for the other particle types. This can be explained by the training setup, where the *CAT Finder* is trained exclusively on muon events and the *CAT B Finder* on a mixed dataset that includes 75% muon events (see Section 3.2). Since the input features vary across particle types, as shown in Fig. 5.5, particle-dependent performance is expected. Nevertheless, this dependence is less pronounced for the *CAT B Finder*, which achieves higher overall hit efficiency than the *CAT Finder* by incorporating multiple particle types in the training.

Finally, I examine the extrapolation to the two inner tracking detectors. The number of SVD and PXD clusters assigned to pion tracks is given in Fig. 7.3. For all three algorithms, most tracks are assigned eight SVD clusters, which corresponds to four hits from the four SVD layers (see Section 2.2). In the case of the PXD, the majority of tracks are assigned a single cluster, consistent with the PXD configuration during run I, where only one layer was installed (Section 2.2). I do not observe differences between the three tracking



(a) Number of SVD cluster assigned to tracks. (b) Number of PXD cluster assigned to tracks.

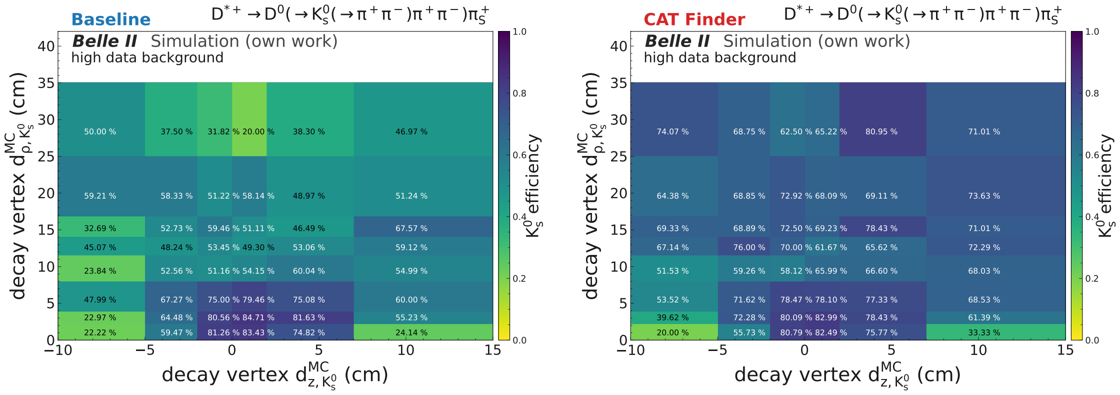
Figure 7.3: Distributions of the number of SVD clusters (left) and PXD clusters (right) assigned to the $\pi^\pm$ tracks. The *Baseline Finder* distribution is given in blue, the *CAT Finder* in red, and the *CAT B Finder* in orange.

algorithms, and the CKF is successfully able to extrapolate the CDC tracks to the inner tracking detectors as described in Section 5.4.

7.1.2 Displaced Tracks in $K_S^0 \to \pi^+\pi^-$ in High Multiplicity Events

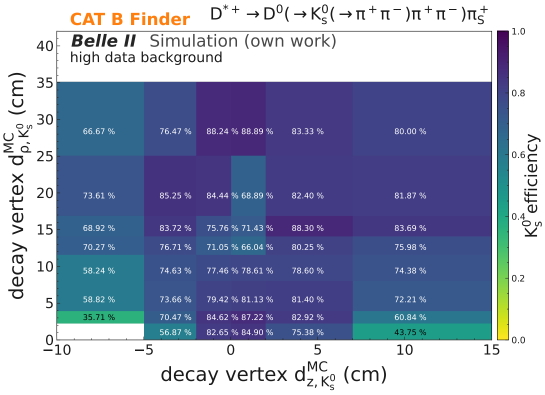
While the efficiency for very long-lived particles that decay further outwards within the CDC volume depends solely on the CDC track finding, the efficiency for particles originating from K_s^0 decays depends on the combined performance of the SVD and CDC. The overall K_S^0 efficiency, the K_S^0 fake rate, and the $\pi_{K_S^0}$ efficiency for the decay chain $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ are given in Table 7.2. Given the lifetime $c\tau(K_S^0) = 2.7$,cm [37] and a median (90% quantile) momentum of 0.55 GeV (1.38 GeV) in this sample, most K_S^0 decay within the two innermost tracking detectors. I observe that the reconstruction efficiency for K_S^0 increases by 5 percentage points for the *CAT Finder* and by 8 percentage points for the *CAT B Finder* model compared to the *Baseline Finder*. While the *CAT Finder* again exhibits a higher fake rate, the *CAT B Finder* model maintains the same fake rate as the *Baseline Finder* while achieving a higher efficiency.

Fig. 7.4d shows the difference in K_S^0 reconstruction efficiency between *CAT Finder* and *Baseline Finder* (left) and *CAT B Finder* and *Baseline Finder* (right). The binning is chosen according to the tracking detector acceptance, as described in Section 2.2. In the innermost region, the decays occur within the PXD, followed by the SVD. From 16.8 cm on, the decay vertex of the K_S^0 lies within the CDC, thus outwards going particles can only be reconstructed by the CDC track finding algorithms. Similar to prompt particles, I observe that K_S^0 decaying very close to the interaction point are reconstructed with decreased efficiency using the *CAT Finder* in comparison to the other two algorithms due to the handling of curlers. However, for K_S^0 with a decay vertex of more than 8 cm, which

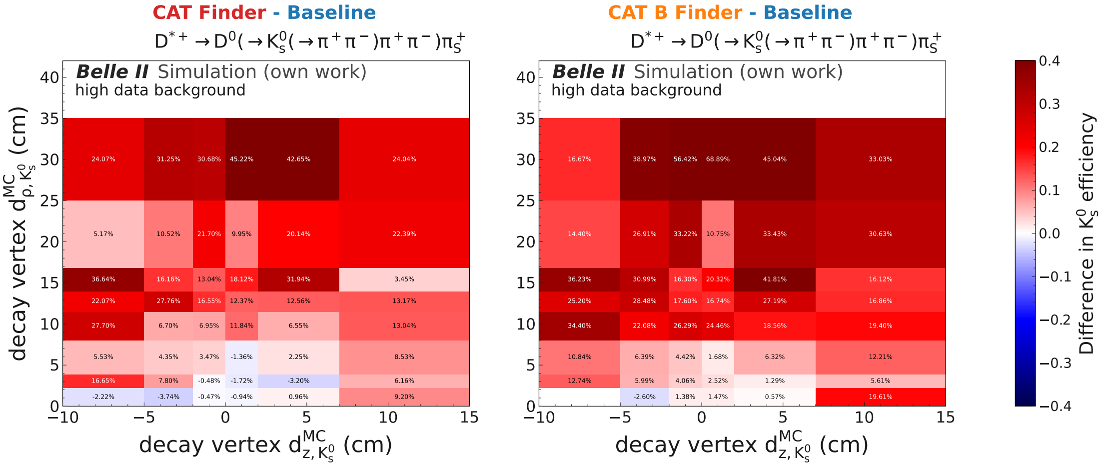


(a) *Baseline Finder*.

(b) *CAT Finder*.



(c) *CAT B Finder*.



(d) Difference in the full reconstruction charge efficiency for the *CAT Finder* and *CAT B Finder* vs the *Baseline Finder*.

Figure 7.4: The K_S^0 efficiency as a function of on the true decay vertex of K_S^0, with bins for the true radial distance d_ρ^{MC} over the true longitudinal distance d_z^{MC} for the *Baseline Finder* (Fig. 7.4a), *CAT Finder* (Fig. 7.4b), and *CAT B Finder* (Fig. 7.4c). The difference between the two new tracking algorithms to the *Baseline Finder* is given in Fig. 7.4d.

corresponds to the second SVD layer in ρ, the *CAT Finder* efficiency is improved compared the *Baseline Finder*. For the *CAT B Finder*, I observe higher efficiency across all regions.

Table 7.2: K_S^0 full reconstruction efficiency and fake rate, and the full reconstruction efficiency of $\pi_{K_S^0}$ from K_S^0 decays for the *Baseline Finder*, *CAT Finder* and *CAT B Finder*.

Algorithm	**K_S^0 Efficiency**	**K_S^0 Fake Rate**	**$\pi_{K_S^0}$ Efficiency**
Baseline	70.8±0.3%	5.8±0.1%	85.3±0.2%
CAT	75.2±0.3%	7.4±0.1%	88.4±0.1%
CAT on B	78.9±0.3%	5.8±0.1%	90.3±0.1%

The prompt performance is recovered with this model, and training on high-multiplicity events has further improved K_S^0 reconstruction even for more displaced K_s^0 compared to the *CAT Finder*.

7.2 Track Charge Asymmetry

One major research topic of Belle II is measuring the Charge Parity (CP) violation in multiple decay chains (i.e. [97–100]). A reconstruction asymmetry for positively and negatively charged particles can mimic the CP violation. While this effect is negligible for particles with $p_t > 0.5\,\text{GeV}$, it becomes relevant for low momentum particles. In the analysis measuring the CP asymmetry in $D^0 \to \pi^+\pi^-\pi^0$ [101], the flavor of the D^0 is determined from the soft pion emitted in the strong decay $D^{*+} \to D^0\pi^+$. For this analysis, the reconstruction asymmetry is the dominating systematic uncertainty. Thus, it is necessary to examine this effect in detail for a new tracking algorithm.

With $N^{\text{reco}}(\pi^+)$ and $N^{\text{reco}}(\pi^-)$ as the number of of correctly reconstructed prompt π^+ and π^-, and $N^{\text{MC}}(\pi^+)$ and $N^{\text{MC}}(\pi^-)$ as the number of generated prompt pions, the detection charge asymmetry A^{det} is defined as the difference between the reconstruction charge asymmetry

$$A^{\text{reco}} = \frac{N^{\text{reco}}(\pi^+) - N^{\text{reco}}(\pi^-)}{N^{\text{reco}}(\pi^+) + N^{\text{reco}}(\pi^-)} \tag{7.1}$$

and the generation asymmetry

$$A^{\text{MC}} = \frac{N^{\text{MC}}(\pi^+) - N^{\text{MC}}(\pi^-)}{N^{\text{MC}}(\pi^+) + N^{\text{MC}}(\pi^-)} \tag{7.2}$$

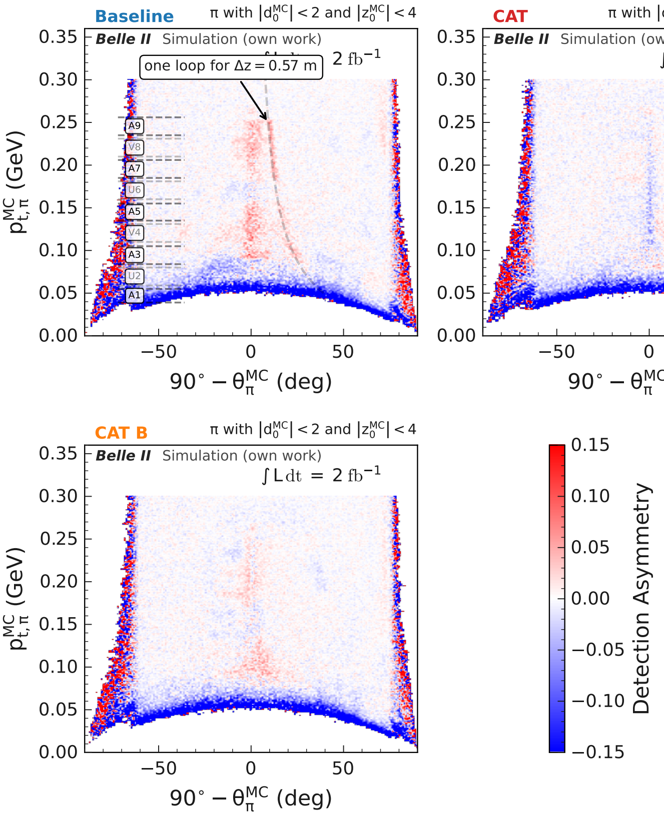
which results in

$$A^{\text{det}} = A^{\text{reco}} - A^{\text{MC}} \tag{7.3}$$

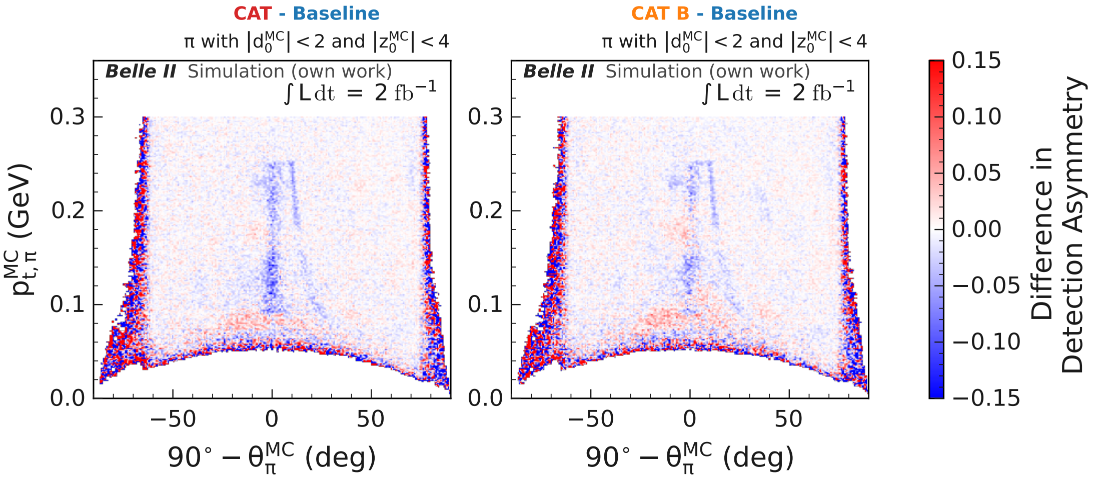
for all prompt pions.

The detection charge asymmetry describes only the effects due to the track reconstruction algorithms.

Fig. 7.5 shows the detection asymmetry for the *Baseline Finder*, the *CAT Finder*, and the *CAT B Finder*, as well as the difference between the *CAT Finder* (*CAT B Finder*) and the *Baseline Finder*.



(a) Detection asymmetry for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom left).



(b) Difference in detection asymmetry for *CAT Finder* and *Baseline Finder* (left) and *CAT B Finder* and *Baseline Finder* (right).

Figure 7.5: Detection asymmetry between positively charged and negatively charged prompt pions for all three track finding algorithms in full reconstruction. Fig. 7.5a shows the overall values. Red translates to a higher detection efficiency for positively charged pions, while blue translates to a higher detection efficiency for negatively charged pions. The dashed black curve shows the particle traversing exactly one loop (Eq. (7.4), explained in Fig. 7.7) for the respective transverse momentum and polar angle for the inner CDC boundary in forward direction of $z = 57\,\text{cm}$. The superlayer boundaries described in Section 6.1.1 are added as guidelines in the top left plot. Fig. 7.5b shows the difference of detection asymmetry.

I additionally give the reconstruction efficiency in Fig. 7.6, to simultaneously monitor that the changes in asymmetry are not due to an overall worse reconstruction efficiency. The reconstruction efficiencies split by charge are given in Section D.2.

While the *Baseline Finder* is often not able to provide correct information on the direction if the particle trajectory is exactly one loop in the CDC, the *flip and refit* module, explained in Section 5.4, improves the correct charge assignment significantly. While it basically removes any dependency on detector geometry in the backwards direction, the one loop in the forward direction is still visible, as more positively charged particles are being detected. The condition for a particle to complete exactly one loop in the CDC, as shown in Fig. 7.7, is obtained by requiring that, over a longitudinal distance of $\Delta z = 0.57$ m, the helix advances by 2π in the azimuthal plane. This distance corresponds to the inner boundary of the CDC in the forward direction. In this case, the transverse momentum p_t^{MC} and polar angle θ^{MC} are related by

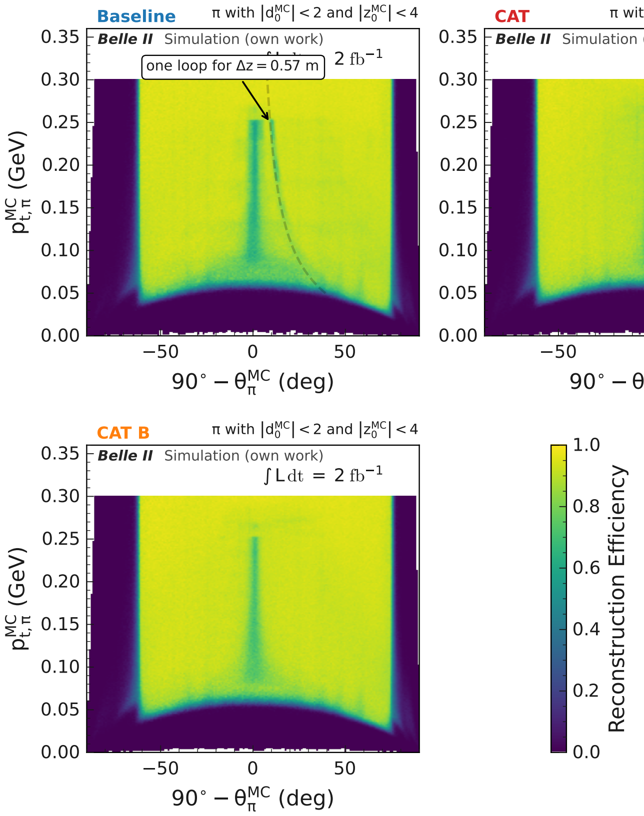
$$\theta^{\mathrm{MC}} = \arctan\left(\frac{\Delta z \cdot B}{2\pi \cdot p_t^{\mathrm{MC}}}\right), \tag{7.4}$$

for the magnetic field strength of $B = 1.5$ T. This relation demonstrates that the observed charge asymmetry in Fig. 7.5a coincides with the kinematics of a one-loop curler in the detector. I do not observe this geometrical dependence for either the *CAT Finder* nor the *CAT B Finder* model, which can correctly predict the direction of the loop.

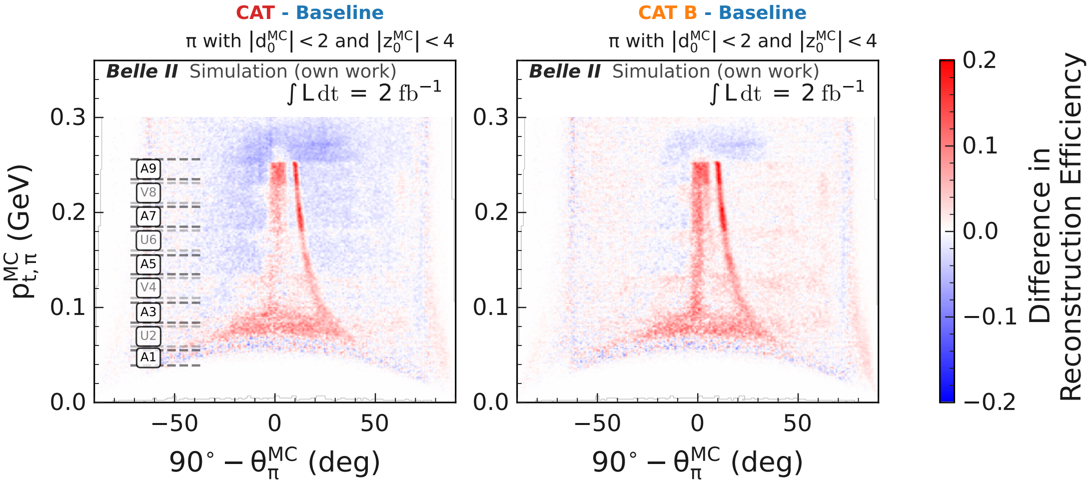
Another region with detection asymmetry is the region around $\theta^{\mathrm{MC}} = 90^\circ$. The direction of the particle can be inferred from the energy loss in the particle trajectory. However, especially within the inner CDC superlayers, the higher beam background occupancy will interfere with the signal hit measurements from the particle. While for angles differing from $\theta^{\mathrm{MC}} = 90^\circ$, the algorithms can assume that the particles come from the collision point and fix the direction according to the change in the z-coordinate, this is not possible for particles with $\theta^{\mathrm{MC}} = 90^\circ$.
All three track finding algorithms have a detection asymmetry within this region. The detection asymmetry is the most prominent for the *Baseline Finder*. Interestingly, the detection asymmetry flips sign for tracks with a transverse momentum high enough to reach the superlayer U6, but then flips back for higher momentum again.
For outer curlers with a transverse momentum of $p_t^{\mathrm{MC}} > 0.255\,\mathrm{GeV}$, no asymmetry is observed. This is because the energy loss in the outer CDC walls and surrounding detectors is larger than within the tracking volume, and even if the tracks re-enter the CDC, their trajectories before and after re-entry differ too much to be misidentified (see Fig. 7.7). The *CAT Finder* has less detection asymmetry than the *Baseline Finder*, and flips sign at the region $\theta^{\mathrm{MC}} = 90^\circ$. For the *CAT B Finder*, which was trained on $B\bar{B}$ events in addition, the structure is more similar to the *Baseline Finder*, but still overall less prominent.

One important next step is checking the detection asymmetry on measured data. For the *Baseline Finder*, the observed detection asymmetry on data has the opposite sign compared to the simulation [101]. It would be relevant to see whether this discrepancy is reduced with the *CAT Finder*, or has an even larger effect, but this is out of scope for this work.

Fig. 7.6 shows that the asymmetry is correlated with an overall reconstruction inefficiency. For the inner region at $\theta^{\mathrm{MC}} = 90^\circ$, all three algorithms have decreased reconstruction efficiency, due to the previously named reasons. This reconstruction inefficiency shows up for positively and negatively charged pions as seen in Section D.3. Furthermore, for the *Baseline Finder*, at the one loop line at $\Delta z = 0.57\,\mathrm{m}$ in the forward direction, the efficiency is also decreased. This is not the case for both *CAT Finder* and *CAT B Finder*. Decreased efficiencies for the *Baseline Finder* are observed in regions where the maximum ρ distance the particle can travel in the detector happens to be within an stereo superlayer U2, V4, or U6 in Fig. 7.6b. The GNN approach, in contrast to the *Baseline Finder*, accounts for detector geometry and does not exhibit such reconstruction inefficiencies. Unlike the *Baseline Finder*, where the optimization of detection asymmetry requires a new algorithm and involves a trade-off between the central region at $\theta^{\mathrm{MC}} = 90^\circ$ and the one-loop curlers, such that improving asymmetry in one region degrades it in the other, both *CAT Finder* algorithms avoid this dependency.

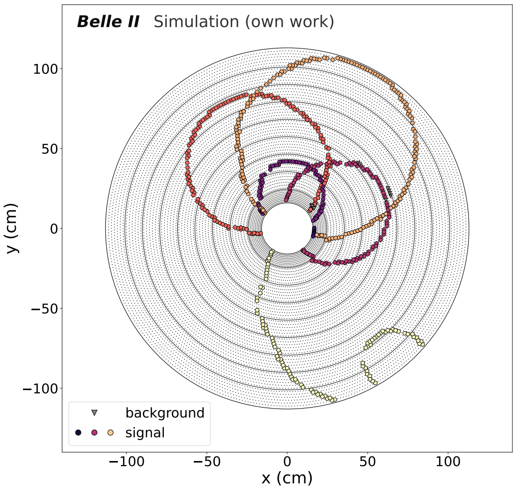


(a) Reconstruction efficiency for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom left).

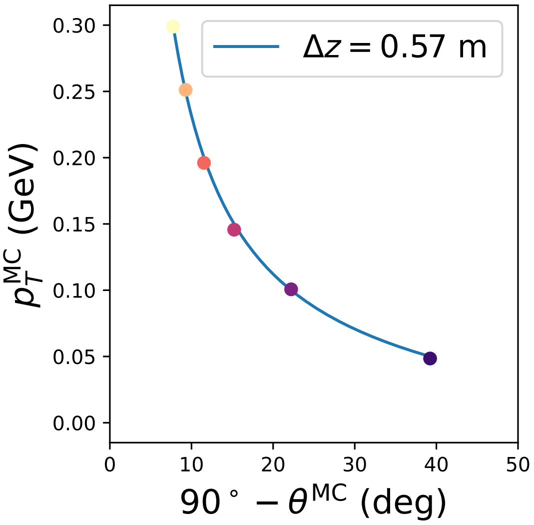


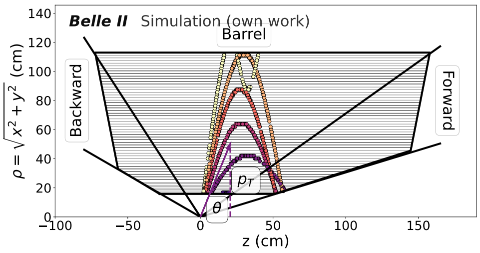
(b) Difference in reconstruction efficiency for *CAT Finder* and *Baseline Finder* (left) and *CAT B Finder* and *Baseline Finder* (right).

Figure 7.6: Reconstruction efficiency of prompt pions for all three track finding algorithms in full reconstruction. Fig. 7.6a shows the overall values. The dashed black curve shows the particle traversing exactly one loop (Eq. (7.4), explained in Fig. 7.7) for the respective transverse momentum and polar angle for the inner CDC boundary in forward direction of $z = 57\,\mathrm{cm}$. Fig. 7.6b shows the difference of detection asymmetry. The superlayer boundaries described in Section 6.1.1 are added as guidelines in the bottom left plot.



(b) x–y view.





(a) p_t^{MC} vs $90^\circ - \theta^{\mathrm{MC}}$ (Eq. (7.4)).

(c) ρ–z view.

Figure 7.7: The left shows the curve for Eq. (7.4) in blue for $\Delta z = 0.57\,\mathrm{m}$ and 6 points close to the curve, corresponding to the generated particle transverse momentum p_t^{MC} and θ^{MC} for the event displays on the right. The right shows an event display of different curlers in the $x - y$ view (top, Fig. 6.4a) and the $\rho - z$ view (bottom, Fig. 6.4b). Filled, colored circular markers represent *signal hits* belonging to the different particles on the left. The combination of p_t^{MC} and θ^{MC} always produces one single loop in the CDC, except for the particle with $p_t^{\mathrm{MC}} = 0.3\,\mathrm{GeV}$, that leaves the CDC, loses energy outside the CDC, and comes back with a small outer curler.

7.3 Particle Identification

To reliably perform analyses such as measuring rare decays where backgrounds need to be suppressed [102], lepton flavour violation [103], and lepton flavour universality [104], it is essential to accurately distinguish the particle types of FSPs (see Section 3.2.5). For each subdetector (see Section 2.2) except the PXD, a likelihood $\mathcal{L}_{\text{particle}}^{\text{detector}}$ is calculated for a given particle hypothesis, further described in [11, 105]. These individual likelihoods are then combined to form a global likelihood. For example, the combined likelihood for the pion hypothesis in Belle II is given by:

$$\log \mathcal{L}_\pi = \log \mathcal{L}_\pi^{\text{SVD}} + \log \mathcal{L}_\pi^{\text{CDC}} + \log \mathcal{L}_\pi^{\text{TOP}} + \log \mathcal{L}_\pi^{\text{ARICH}} + \log \mathcal{L}_\pi^{\text{ECL}} + \log \mathcal{L}_\pi^{\text{KLM}}. \quad (7.5)$$

Many of these likelihoods depend on the accuracy of the reconstructed track parameters, as the particle trajectories must be extrapolated to the corresponding subdetectors, as shown in Fig. 7.8, as well as the particle momentum.

SVD The PID calulcation of the SVD depends on the ionization loss in the SVD. This ionization loss depends on the particle type [30] and is sensitive to the correct number of SVD clusters assigned to the tracks to calculate. As the CDC tracks are extrapolated with a CKF to the SVD to assign clusters, this is dependent on the CDC track finding algorithms.

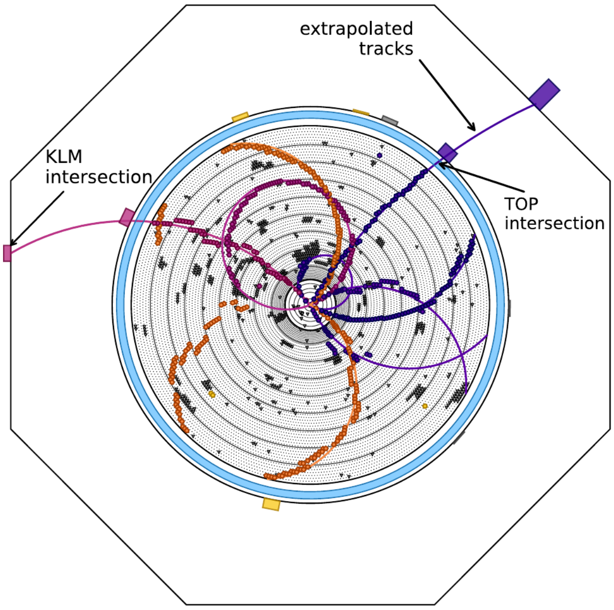


Figure 7.8: Event display for a $B\bar{B}$ event in the $x - y$ view at $z = 0$ for all subdetectors (Section 2.2). The hits in the tracking detectors are shown with colored circular markers for the signal and grey triangular markers for the background hits. The clusters of the ECL are shown as bars, following the color coding from the tracking detector. The clusters of the KLM are summarized in the first intersection. The figure is adapted from T. Brandes.

CDC As for the SVD, the ionization loss in the CDC is measured to identify the particle type. This is dependent on the correct number of hits assigned to the CDC tracks.

TOP and ARICH The PID of the TOP and ARICH detectors depends on the intersection of the tracks with the respective detectors as well as the particle momentum to determine the distribution of the photons from the Cherenkov radiation (see Section 2.2).

ECL The ECL needs to match clusters to the tracks and relies on the correct track extrapolation to distinguish between photons and electrons. Furthermore, the ratio between the measured energy in the ECL and the respective particle momentum that is dependent on the CDC distinguishes electrons from the other particle types.

KLM The outermost detector relies on the correct track extrapolation to distinguish muons from other particles, depending on their expected trajectory assuming minimal interaction.

Validating the particle identification performance tests not only the tracking quality, but also the coordination between subdetectors, to check the correctness of the overall implementation. The likelihoods of the subdetectors can differ due to the differences in the tracking algorithms. For example, the hit efficiency for *CAT Finder* is decreased for high multiplicity events, which results in a decreased momentum resolution for low momentum particles as shown in Fig. 6.14. While *CAT B Finder* and *Baseline Finder* have a very similar hit efficiency, *CAT B Finder* has an overall higher reconstruction efficiency, especially for displaced particles, therefore, it is necessary to test this in detail.

7.3.1 Ionization Energy Loss

One method to identify particles is by their specific ionization energy loss (dE/dx) in a medium. This energy loss is approximately proportional to the particle's charge and velocity [37], resulting in the same function for all FSPs. When the energy loss is expressed as a function of momentum, as particles with the same charge and velocity but different masses will have different momenta, this leads to distinct curves for each particle type. As a result, measuring dE/dx as a function of momentum allows for an estimation of the particle's mass, which in turn enables its identification.

For each CDC hit belonging to a track, the dE/dx in the drift cell is measured. The CDC hits with the lowest 5% and the highest 25% of dE/dx measurements of each track are excluded and for the remaining CDC hits the average dE/dx is calculated. The average is truncated to remove the effects of the non-Gaussian tails, especially towards higher dE/dx values. These two cut values of 5% and 25% are the result of the centralized calibration and will require re-calibration for future use with the new track-finding algorithm. This is because the dE/dx measurement depends on hit purity, which has been improved in both

the *CAT Finder* and *CAT B Finder* algorithms, and hit efficiency, which has decreased in the case of *CAT Finder*. I report the results without re-calibration in this work.
Fig. 7.9 shows the ionisation energy loss of pions, kaons, and protons in the CDC over the reconstructed particle momentum. The ionization energy losses for muons and electrons are shown in addition in Section D.3, but the separation power in the momentum range of Belle II is limited. Using the CDC dE/dx, pions, protons, and kaons for track momentum $p < 1\,\mathrm{GeV}$ can be distinguished. For larger momentum, the separation power decreases significantly. The *CAT Finder* distributions for the three particles are slightly broader than those from the *Baseline Finder* and *CAT B Finder* algorithms, which is expected due to the lower hit efficiency of the *CAT Finder*. This can be seen in detail in Fig. 7.10, where the same events are reconstructed with all three algorithms, and I select the intersecting sample between all three. While the distributions of the *CAT B Finder* and *Baseline Finder*

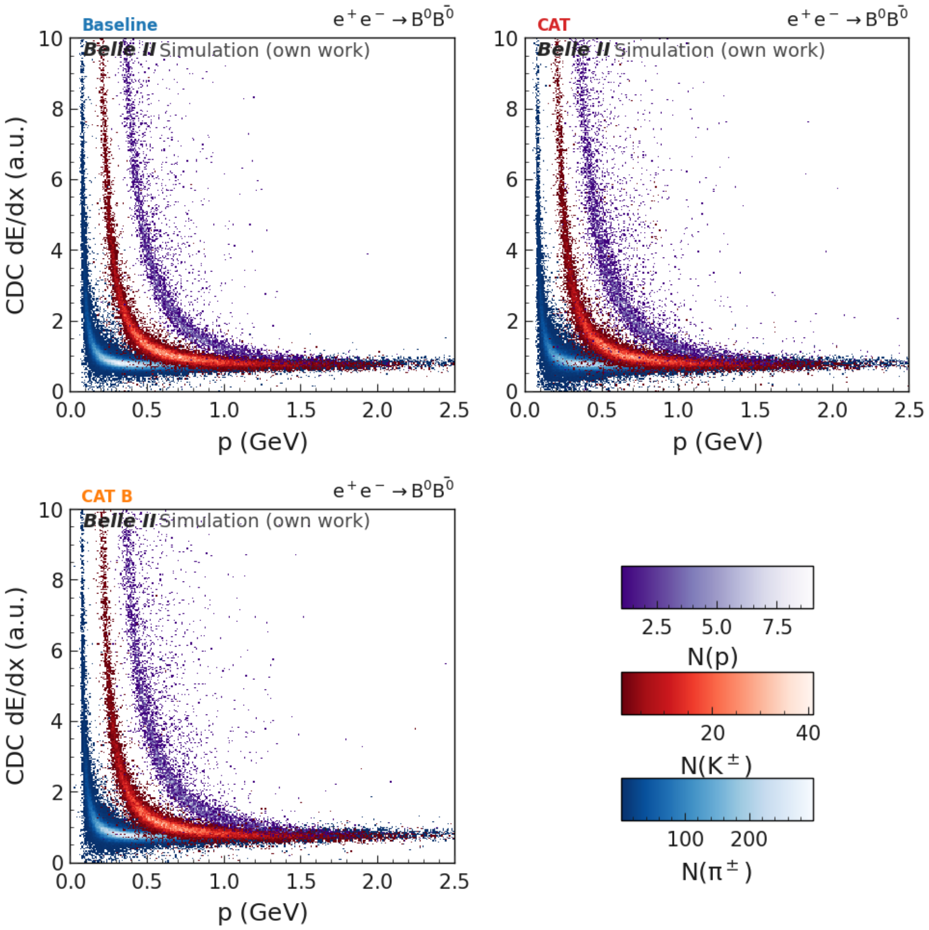
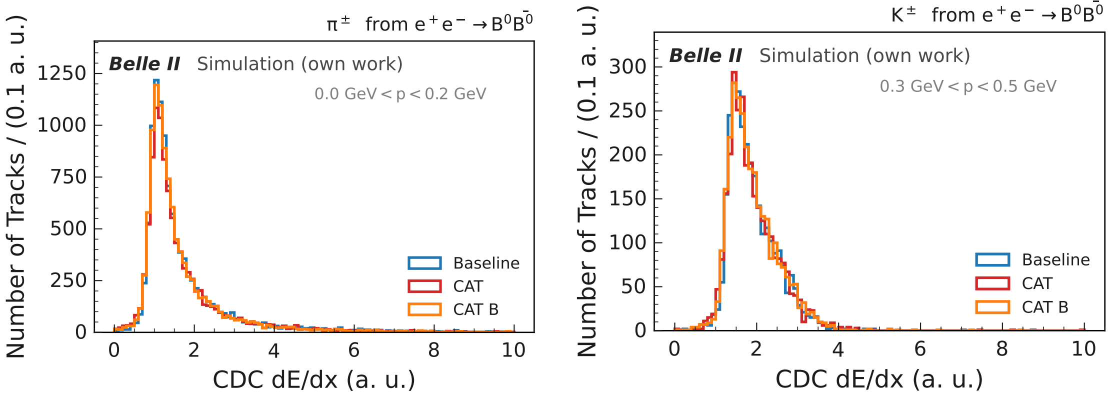
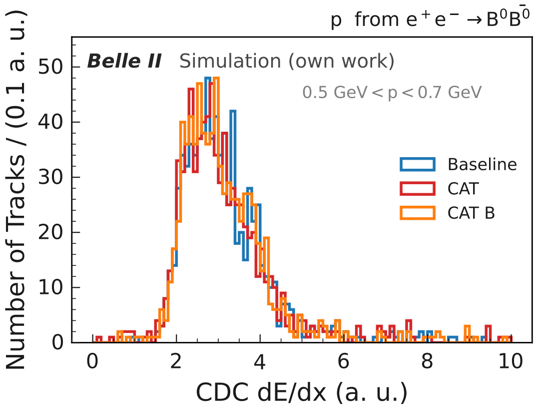


Figure 7.9: CDC dE/dx in arbitrary units (a.u.) over momentum for pions (blue), Kaons (red), and protons (purple) for the *Baseline Finder* (top left), *CAT Finder* (top right) and *CAT B Finder* (bottom left). The dE/dx values are normalized to the average energy loss for an electron at the Fermi plateau [105], as the absolute value is not necessary to distinguish between the different particle types.



(a) $\pi^{\pm}$ dE/dx for $0\,\text{GeV} < p < 0.2\,\text{GeV}$.

(b) $K^{\pm}$ dE/dx for $0.3\,\text{GeV} < p < 0.5\,\text{GeV}$.



(c) $p/\bar{p}$ dE/dx for $0.5\,\text{GeV} < p < 0.7\,\text{GeV}$.

Figure 7.10: CDC dE/dx distribution for a given true momentum range for pions (left), kaons (middle), and protons (right) for the *Baseline Finder* (blue), *CAT Finder* (red), and *CAT B Finder* (orange) for the intersecting sample between all three algorithms. No statistical uncertainties are shown, since the same samples are used for all three algorithms in these figures. See Fig. 7.9 for details.

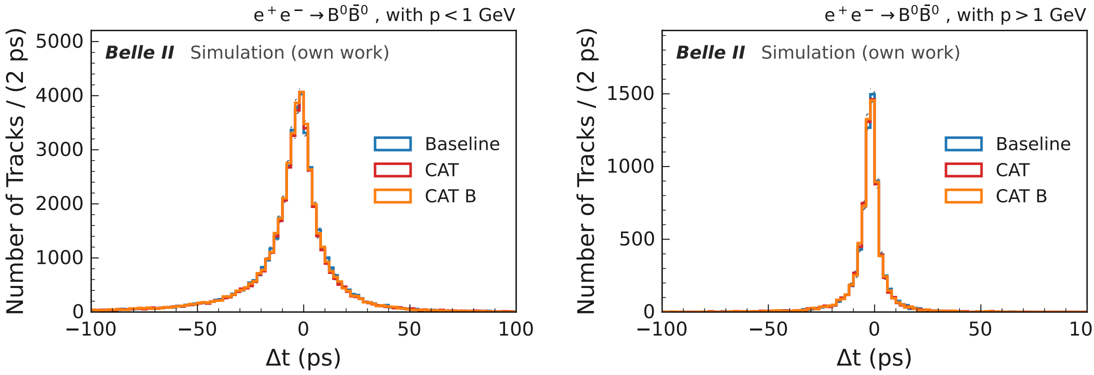
models are in good agreement, the distribution for pions in the *CAT Finder* model (see Fig. 7.10a) is shifted toward lower values at the left tail. This leads to a reduced dE/dx resolution for *CAT Finder*, which is also observed in *Baseline Finder* for tracks pointing into the endcaps, as these typically have fewer assigned hits. A similar comparison is shown for the SVD dE/dx in Section D.3. The ionization loss in the SVD is larger than in the CDC, but as I observe the same number of SVD clusters assigned to the tracks in Fig. 7.3 for all three algorithms, the differences also stem from the momentum resolution and not the dE/dx measurements, as seen in Fig. D.14.

7.3.2 Track Extrapolation

TOP and ARICH subdetector

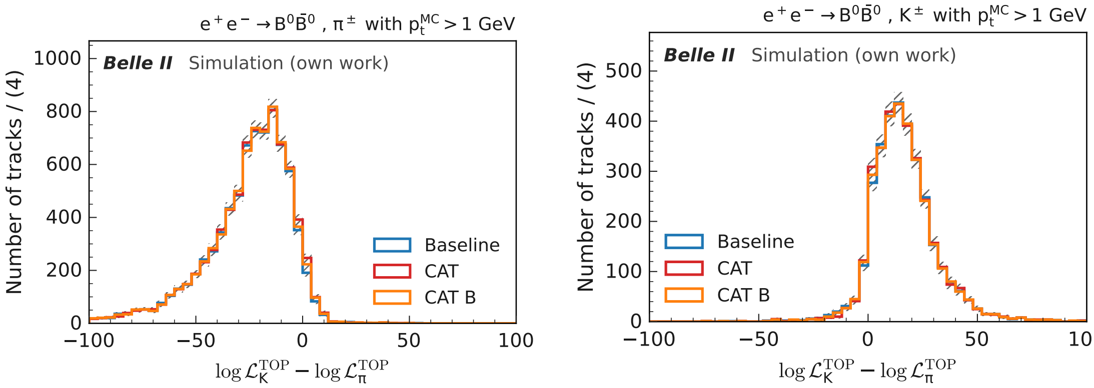
The TOP likelihood relies on precise momentum resolution and the correct extrapolation of the tracks. From the track extrapolation, the crossing point of the particle within the TOP detector is determined. This is necessary to compute the time-of-flight correctly. The time-of-flight resolution for both low- and high-momentum tracks is shown in Fig. 7.11. As expected, low-momentum tracks exhibit a broader resolution due to the reduced overall tracking resolution at lower momenta (see Fig. 6.14). Consequently, time differences for slower particles, affected by this lower resolution, tend to increase more significantly in proportion. All three tracking algorithms yield consistent results within uncertainties, confirming the correct implementation in basf2. The resulting photon distribution, produced via Cherenkov radiation, is combined with the time-of-flight measurement and particle momentum to compute a likelihood for each particle hypothesis. I compare the difference in log-likelihood $\log \mathcal{L}$ between the kaon and pion hypotheses for true pions (Fig. 7.12a) and true kaons (Fig. 7.12b), restricted to high-momentum tracks with $p \geq 1\,\mathrm{GeV}$. True pions predominantly yield negative likelihood differences, while kaons result in positive values, demonstrating a clear separation between the two particle types. Again, all three tracking algorithms agree within uncertainties.

The ARICH detector follows a similar approach, but in the forward direction. The ARICH-only PID for pions and kaons is given in Fig. 7.13, where all three algorithms agree within uncertainties.



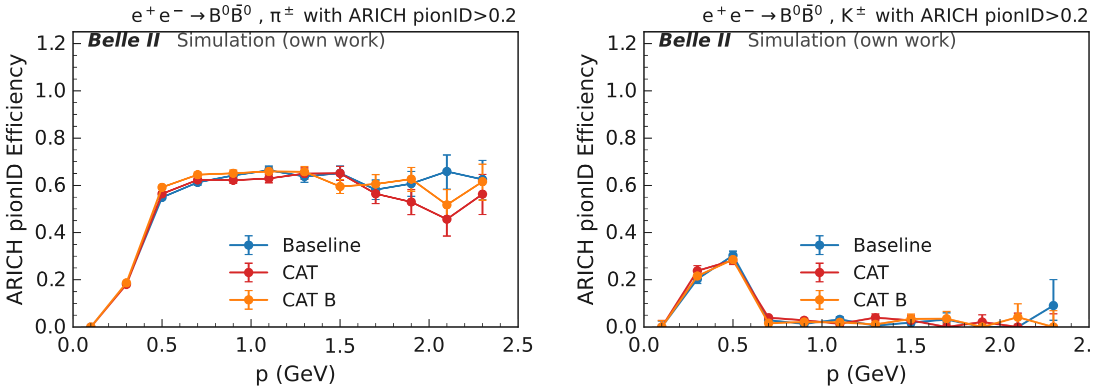
(a) $p < 1\,\mathrm{GeV}$. (b) $p \geq 1\,\mathrm{GeV}$.

Figure 7.11: Time-of-flight determined from the track extrapolation to the TOP detector for low momentum tracks with $p < 1\,\mathrm{GeV}$ (left) and high momentum tracks with $p \geq 1\,\mathrm{GeV}$ (right) for the *Baseline Finder* (blue), *CAT Finder* (red), and *CAT B Finder* (orange).



(a) True pions. (b) True kaons.

Figure 7.12: Difference in the logarithmic particle likelihood of the TOP between kaons $\mathcal{L}_K^{TOP}$ and pions $\mathcal{L}_\pi^{TOP}$ for pions (left) and kaons (right) with $p \geq 1\,\text{GeV}$ for the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange).

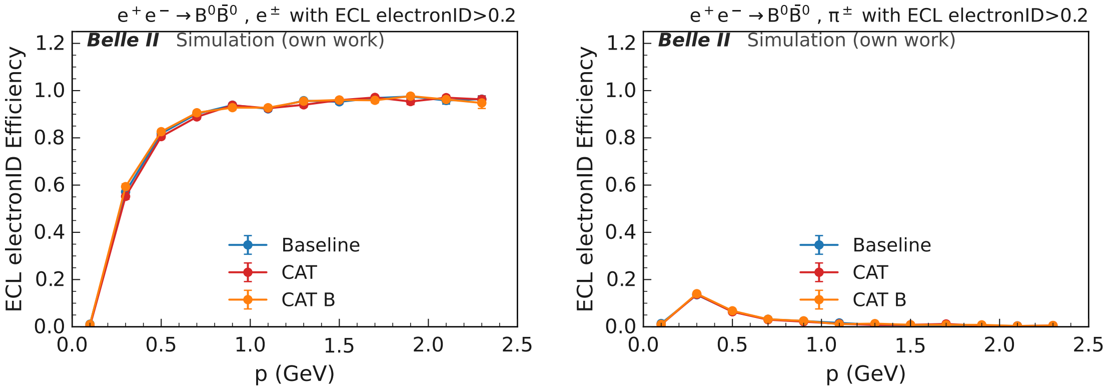


(a) ARICH *pionID* efficiency for true pions. (b) ARICH *pionID* efficiency for true kaons.

Figure 7.13: ARICH *pionID* efficiency with a cut value of 0.2 for true pions (left) and true electrons (right) for the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange) over bins of reconstructed particle momentum.

ECL Subdetector

Electron identification is primarily driven by the ECL. To probe this, I examine the ECL *electronID* efficiency as a function of particle momentum in Figure 7.14. This *electronID* relies, among others, on the correct matching of the ECL clusters to the tracks and the measurement of the cluster energy over the particle momentum. The cluster energy over momentum is expected to be close to one for electrons, as electrons deposit almost all energy within the cluster, and is significantly lower for other charged particles. Both of these depend on the momentum resolution. Electrons with transverse momentum above approximately 0.3 GeV have sufficient energy to reach the ECL barrel region. As shown in Fig. 7.14a, the ECL *electronID* efficiency for true electrons is low in the first momentum



(a) ECL *electronID* efficiency for true electrons. (b) ECL *electronID* efficiency for true pions.

Figure 7.14: ECL *electronID* efficiency with a cut value of 0.2 for true electrons (left) and true pions (right) for the *Baseline Finder* (blue), *CAT Finder* (red), and *CAT B Finder* (orange) over bins of reconstructed particle momentum.

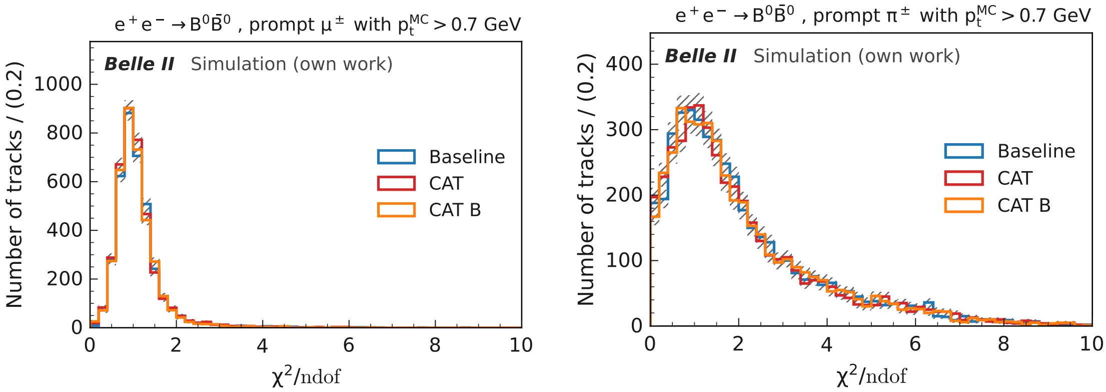
bin, where electrons do not reach the ECL. However, the efficiency improves significantly once electrons enter the ECL acceptance region in the second and third bin, and then plateaus at ≈ 95% efficiency. In comparison, Fig. 7.14b demonstrates that true pions are effectively suppressed by applying the same selection on the ECL *electronID* variable. All three algorithms agree within uncertainties.

KLM Subdetector

Muons are primarily identified by the KLM. Their distinguishing feature is the ability to penetrate thick absorber material without major energy loss compared to other particles. This property is used to construct the muon likelihood in the KLM. Particle trajectories are extrapolated from the inner subdetectors through the detector material and magnet to the KLM under the muon hypothesis, assuming minimal interaction. In the same way as the TOP, ARICH, and ECL particle identification, this approach is highly dependent on the precision of the track reconstruction. For this, the track is extrapolated to the KLM assuming the muon hypothesis. The extrapolated trajectory is compared with the positions of the measured clusters per layer, and a χ^2 test [106] is performed to assess the discrepancy between the two. The resulting χ^2 value is divided by the number of degrees of freedom (ndof) to obtain the reduced χ^2/ndof. Fig. 7.15 shows the distribution of this reduced value for true, prompt muons (Fig. 7.15a) and true, prompt pions (Fig. 7.15b).

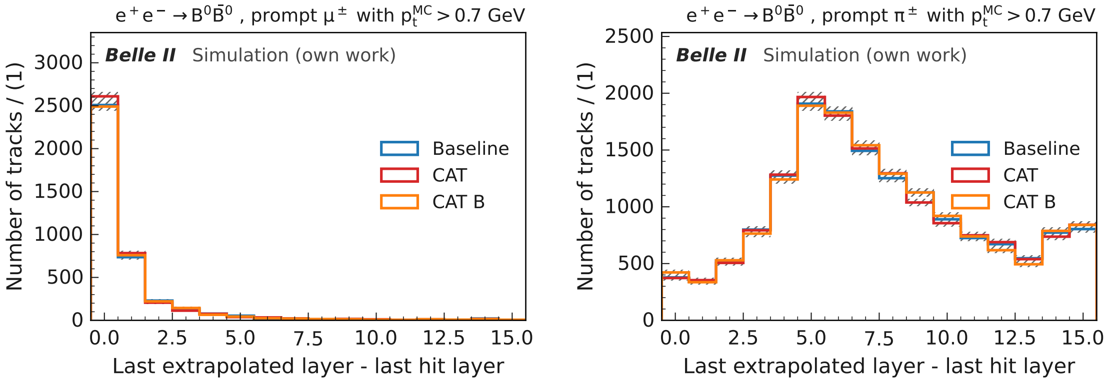
I require all prompt particles to have a transverse momentum of at least 0.7 GeV to be able to reach the KLM in the barrel region.

For muons, the extrapolated track under the muon hypothesis agrees well with the measured clusters, resulting in a distribution peaking around 1. In the case of pions, although the distribution also peaks near 1, significantly larger values are observed, showing the expected mismatches between the muon hypothesis and the actual cluster measurements from the



(a) Reduced χ^2/ndof for muons.

(b) Reduced χ^2/ndof for pions.



(c) Difference between extrapolated and true last layer reached by muons.

(d) Difference between extrapolated and true last layer reached by pions.

Figure 7.15: The relevant metrics to distinguish muons (left) and pions (right) reaching the KLM with a true transverse momentum of at least 0.7 GeV. The top row shows the reduced χ^2 over the number of degrees of freedom between the actual measured clusters in the KLM and the track extrapolation under muon hypothesis, while the bottom row shows the last extrapolated KLM layer the track should reach if it were a muon, in contrast to the last layer it reached.

pions. In addition to the reduced χ^2/ndof, the last extrapolated layer within the KLM is also determined. This layer corresponds to the maximum depth a particle is expected to reach under the muon hypothesis, based on its momentum. Figure 7.15c and Fig. 7.15d illustrate the difference between this extrapolated last layer and the actual last layer hit by the track. For muons, the extrapolation aligns closely with the measured hits. For pions, however, the difference is more pronounced, as their greater ionization energy loss causes them to stop earlier than predicted.

These values enter the likelihood calculation for the KLM, enabling the separation between muons and other particles. All algorithms agree within uncertainties.

7.3.3 Global PID

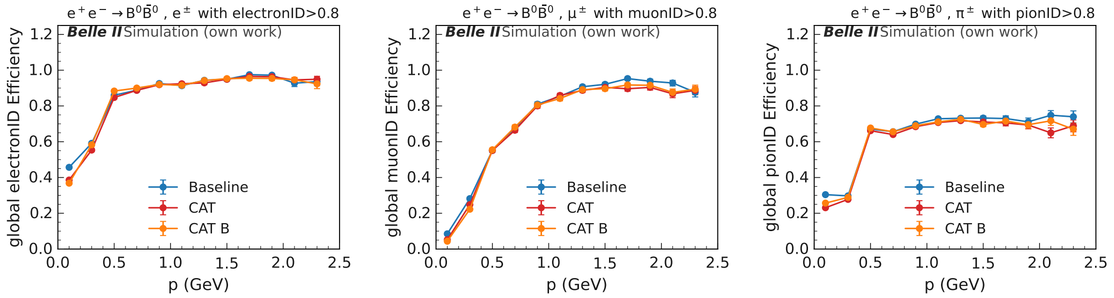
After testing the specific detector identification power for the CDC, directly dependent on the track finding algorithms, as well as the correct extrapolation to all subdetectors outside of the tracking detectors, I can confirm the correct implementation of my new GNN-based algorithm. One last validation is to check the resulting global PID from all subdetectors (Eq. (7.5)), defined as the normalized likelihood for each particle type. For pions, it is given as:

$$\mathrm{PID}(\pi) = \frac{\mathcal{L}_\pi}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K + \mathcal{L}_p + \mathcal{L}_d}. \tag{7.6}$$

The global PID efficiency for the charged particles e, μ, π, K, and p in $B^0\bar{B}^0$ events*, requiring a global PID value greater than 0.8, is shown in Fig. 7.16 as a function of the reconstructed momentum. This cut value does not directly translate to efficiency and needs to be optimized for each analysis. As described in this section, due to the differences in the hit efficiency of the tracking algorithms,re-optimization for different algorithms might be necessary. For low-momentum electrons and pions, the efficiency is decreased compared to *Baseline Finder*. In the momentum region between 1.5 and 2 GeV, the *Baseline Finder* has a slightly higher *muonID* efficiency, whereas the *CAT B Finder* model has a slightly higher *protonID* efficiency overall. For all other regions, the three track-finding algorithms show consistent results within uncertainties. These findings confirm that both *CAT Finder* and *CAT B Finder* are correctly implemented within basf2 and can be reliably used in the full reconstruction. Furthermore, I will use the same PID cuts for all three algorithm configurations in the following analyses, as they agree within uncertainties.

---

*Deuterons are not shown in this section, due to statistical limitations in $B^0\bar{B}^0$ events.



(a) Global *electronID* efficiency. (b) Global *muonID* efficiency. (c) Global *pionID* efficiency.

(d) Global *KaonID* efficiency. (e) Global *protonID* efficiency.

Figure 7.16: Global PID efficiency for the respective particles in $B^0\bar{B}^0$ events, for *Baseline Finder* (blue), *CAT Finder* (red), and *CAT B Finder* (orange), where the lines between points are only helps to guide the eye.

Chapter 8

Validation on Data

After confirming that I implemented the *CAT Finder* and *CAT B Finder* algorithms correctly within basf2 based on the simulation (see Chapter 7), the next important step is to validate them on measured data. For all three algorithms, I simulate and reconstruct all samples and measured raw data with the same software setup, besides the respective CDC tracking algorithms.

Running over the full Belle II measured data is computationally out of scope for this thesis. Instead, I use pre-selected data by the HLT, which is described in detail in Section 3.4. The HLT selection depends on the *Baseline Finder* as the reconstruction algorithm, which was run during data taking. For high track multiplicity events (see Section 3.4.3), the efficiency difference between the *Baseline Finder* and the new *CAT Finder* algorithm is negligible. Since these events typically contain an average of 11 charged tracks, and the selection only requires at least three, it is highly unlikely that *Baseline Finder* would fail to reconstruct a sufficient number of tracks. Low multiplicity events, on the other hand, depend on the successful reconstruction of the two tracks in the event to pass the HLT selection (see Section 3.4.1 and Section 3.4.2). As a result, they are more sensitive to differences between the track finding algorithms. In simulated $e^+e^- \to \mu^+\mu^-$ events, I find that while track charge efficiency is high, *CAT Finder* and *Baseline Finder* can miss tracks in independent events (see Section 6.1.1). This biases the low multiplicity evaluation against *Baseline Finder*, since *CAT Finder* cannot exceed *Baseline Finder* 's performance in cases where *CAT Finder* finds two tracks but *Baseline Finder* misses one. As of the writing of this thesis, no pre-selected data is available for low multiplicity events without relying on track reconstruction. Furthermore, no measured data is available based on a HLT selections for displaced tracks in low multiplicity events. One such example is $e^+e^- \to \Phi(\to K_S^0 K_L^0)\gamma$, where the K_S^0 decays into two pions can be probed, and which can be selected based on the photon energy. As no HLT selection is available, I am not able to evaluate the performance for displaced tracks in low multiplicity events.

In this chapter, I evaluate the performance of prompt muons and electrons in high mul-

tiplicity events in Section 8.1, followed by displaced pions in high multiplicity events in Section 8.2, probing three particle types in total.
It is necessary to check the model's robustness to different particle types, as they all have different material interactions, which could not be modelled correctly in simulation compared to measured data. I then study the reconstruction efficiency and resolution of prompt muons and electrons in low multiplicity events in Section 8.3.1 and Section 8.3.2. I summarize the results of this chapter in Section 8.4.

For high multiplicity, I use approximately $1.65\,\mathrm{fb}^{-1}$ of available pre-selected data, with a medium background scenario, called *runs from experiment 22* (similar to experiment 22, run 468 with 541 median beam background hits, see Fig. 3.2) and a high background scenario, from now on called *runs from experiment 26* (similar to experiment 26, run 1430 with 1216 median hits).

For low multiplicity, I use the available smaller subset of data from Belle II for the same runs as the high multiplicity studies. I use simulated samples with low simulated beam backgrounds to compare with experiment 22 and high data beam backgrounds for comparison with experiment 26. While the simulation is statistically independent between the three tracking algorithms, I reconstruct the same measured data. While the reconstruction and selections can differ due to the differences between the algorithms, the statistical uncertainties on measured data are correlated in this chapter.

Differences between simulation and measured data are expected, as simulation does not perfectly reproduce detector measurements, which is why corrections are applied to the simulation for analyses. Corrections between simulation and measured data are determined using *Baseline Finder*, and obtaining them for my new tracking algorithm is out of scope. The same applies to systematic uncertainties. Therefore, no corrections or systematic uncertainties are applied to either the *Baseline Finder* or the *CAT Finder* algorithms. The focus in this chapter is on comparing the CDC tracking algorithms and their respective differences between simulation and data. Applying *Baseline Finder*-based corrections would cancel out in a direct comparison and is thus omitted.

8.1 Prompt Tracks in High Multiplicity Events

In this section, I compare the tracking performance for prompt tracks from J/ψ decays in hadronic high multiplicity events, following the HLT selection described in Section 3.4.3.

Event Selection

By reconstructing the excited state $\psi(2S)$ from the decay chain $\psi(2S) \to J/\psi(\to l^+l^-)\pi^+\pi^-$, I can improve the purity of the J/ψ selection. The $\psi(2S)$ comes from B-meson decays (see

Section 3.4 for the cross section), where the branching fraction $B \to \psi(2S)X$, with X being anything, is $(3.06 \pm 0.30) \cdot 10^{-3}$ [37]. The branching fraction for the decay $\psi(2S) \to J/\psi\,\pi^+\pi^-$ is $34.69 \pm 0.34\%$, and the J/ψ has a branching fraction of $5.971 \pm 0.032\%$ for the electron and $5.961 \pm 0.033\%$ for the muon final state [37]. While this leads to a low expected number of decays, I can select a pure sample of prompt electron and muon tracks to compare my new track finding algorithm to the existing *Baseline Finder*. The analysis selection is presented below.

Lepton selection Electron and muon candidates must be prompt, with transverse and longitudinal impact parameters satisfying $d_{\mathrm{IP}} < 2\,\mathrm{cm}$ and $|z_{\mathrm{IP}}| < 4\,\mathrm{cm}$, respectively, and within the CDC acceptance. For electrons, a bremsstrahlung correction is applied where the electron track is merged with photons with an energy of at least 0.5 GeV and an opening angle between the photon and electron track of up to 11°. Particles are identified with a likelihood-based PID. The PID(e) and the cluster energy over particle momentum E_{cluster}/p are required to be larger than 0.8. For muons, the PID(μ) is required to be larger than 0.8.

J/ψ candidate The J/ψ candidate is reconstructed from two oppositely charged electrons or muons. The reconstructed mass $M_{l^+l^-}$ is required to be between 2.8 GeV and 3.4 GeV.

Pion selection Pion candidates also must be prompt, with transverse and longitudinal impact parameters satisfying $d_{\mathrm{IP}} < 0.5\,\mathrm{cm}$ and $|z_{\mathrm{IP}}| < 2\,\mathrm{cm}$, respectively, and within the CDC acceptance.

$\psi(2S)$ candidate The J/ψ is combined with two opposite charged pions to form the $\psi(2S)$ candidate. A vertex fit on the full decay chain is performed. The mass difference between the $\psi(2S)$ and the J/ψ candidate is required to be between 0.4 GeV to 0.7 GeV. By reconstructing the excited state $\psi(2S)$ from the decay chain $\psi(2S) \to J/\psi(\to l^+l^-)\pi^+\pi^-$, I can then chose best candidate for the reconstructed J/ψ by requiring difference between the reconstructed mass for the $\psi(2S)$ and J/ψ, $M^{\mathrm{reco}}_{\psi(2S)} - M^{\mathrm{reco}}_{J/\psi}$ and the difference between the true mass according to [37] with $M^{\mathrm{PDG}}_{\psi(2S)} = 3686.097 \pm 0.025\,\mathrm{MeV}$ and $M^{\mathrm{PDG}}_{J/\psi} = 3096.900 \pm 0.006\,\mathrm{MeV}$ for each candidate in the event to be minimal

$$\text{candidate} = \min_{i..N} |(M^{\mathrm{PDG}}_{i,\psi(2S)} - M^{\mathrm{PDG}}_{i,J/\psi}) - (M^{\mathrm{reco}}_{\psi(2S)} - M^{\mathrm{reco}}_{J/\psi})|. \tag{8.1}$$

The selection is summarized in Table 8.1.

Table 8.1: Selection summary for $\psi(2S) \to J/\psi(\to l^+l^-)\pi^+\pi^-$.

Analysis selection	**Selection criteria**	
	$J/\psi \to \mu^+\mu^-$	$J/\psi \to e^+e^-$
Lepton	$\lvert z_{\mathrm{IP}}\rvert < 4\,\mathrm{cm}$	
	$\lvert d_{\mathrm{IP}}\rvert < 2\,\mathrm{cm}$	
	$17^\circ < \theta < 150^\circ$	
	$\mathrm{PID}(\mu) > 0.8$	$\mathrm{PID}(e) > 0.8$
		$E_{\mathrm{cluster}}/p > 0.8$
J/ψ	$2.8\,\mathrm{GeV} < M < 3.4\,\mathrm{GeV}$	
π^+	$\lvert z_{\mathrm{IP}}\rvert < 2\,\mathrm{cm}$	
	$\lvert d_{\mathrm{IP}}\rvert < 0.5\,\mathrm{cm}$	
	$17^\circ < \theta < 150^\circ$	
$\psi(2S)$	$0.4\,\mathrm{GeV} < M_{\psi(2S)} - M_{J/\psi} < 0.7\,\mathrm{GeV}$	

After selecting the events following these selections, I then fit the reconstructed mass distribution of the J/ψ using zfit [107]. The signal and background parametrization is given in the following.

Signal Parametrization

The signal S mass distribution is modelled by a Double Sided Crystal Ball (DSCB) probability density function (PDF) [108, 109], defined as:

$$f(x;\mu,\sigma,\alpha_l,n_l,\alpha_r,n_r) = N_{\mathrm{sig}} \cdot \begin{cases} A_l\left(B_l - \frac{x-\mu}{\sigma}\right)^{-n_L}, & \text{if } \frac{x-\mu}{\sigma} < -\alpha_L \\ \exp\left(-\left(\frac{x-\mu}{2\sigma}\right)^2\right), & \text{if } -\alpha_l \le \frac{x-\mu}{\sigma} \le \alpha_r \\ A_r\left(B_r - \frac{x-\mu}{\sigma}\right)^{-n_r}, & \text{if } \frac{x-\mu}{\sigma} > \alpha_r \end{cases} \tag{8.2}$$

where

$$A_{l/r} = \left(\frac{n_{l/r}}{|\alpha_{l/r}|}\right)^{n_L} \exp\left(-\frac{|\alpha_{l/r}|^2}{2}\right) \tag{8.3}$$

$$B_{l/r} = \frac{n_{l/r}}{|\alpha_{l/r}|} - |\alpha_{l/r}| \tag{8.4}$$

and the normalisation N_{sig} for the signal yield. The function has at its core a Gaussian with the mean μ and width σ, and in addition, exponential tails on the right and left side. The tails are described by the parameters $\alpha_{l/r}$, giving the point for the transition between the Gaussian to the exponential function, which is described by the parameter $n_{l/r}$. As the transition points $\alpha_{l/r}$ and the exponential order parameter $n_{l/r}$ are strongly correlated, this leads to large uncertainties on the fit. To stabilize the fit, the exponential order parameters are fixed to $n_{l/r} = 2$, which is within the range of $1.5 - 2.5$ for the $n_{l/r}$ parameters if left floating, as well as consistent with the simulation results. The tails are then described by the transition points $\alpha_{l/r}$.

Background Parametrization

The background of the reconstructed mass is expected to be smooth within the fit range. I model the background using the Chebyshev polynomial PDF to the first order with the linear Chebyshev coefficient c_1 and the normalization factor for the background, given by the background yield estimator N_{bkg}.

The signal and background fit is therefore modelled with seven floating parameters.

8.1.1 $J/\psi \to \mu^+\mu^-$

Reconstructed J/ψ mass

In Fig. 8.1, I show the reconstructed J/ψ mass $M_{\mu^+\mu^-}$ for $J/\psi \to \mu^+\mu^-$ in runs from experiment 22. The fit is performed in the range 3.0–3.2 GeV and the fit results for measured data and simulation are given in Table 8.2. This process also probes the pion reconstruction efficiency in addition, as all four FSP of the decay chain need to be combined to reconstruct the $\psi(2S) \to J/\psi(\to \mu^+\mu^-)\,\pi^+\pi^-$.

For simulation, the background is smoothly distributed. The measured data agree with the shape of the simulation, but the signal yield is larger in the measured data compared to the simulation. For all algorithms, the fitted mean of the measured data agrees within uncertainties and is statistically consistent within two standard deviations of $M_{J/\psi}^{\mathrm{PDG}} = 3096.900 \pm 0.006\,\mathrm{MeV}$ [37], which probes the bias of the reconstruction.
The simulation is statistically independent, but the same measured data are reconstructed with the three algorithms. Although the algorithms do not reconstruct exactly the same events, they share a large fraction, leading to correlated results on measured data. The statistical uncertainties are common to all algorithms.
The signal yield N_{sig} is highest for *CAT B Finder*, but all results agree within the fit uncertainties, which shows that all three algorithms achieve a similar reconstruction efficiency. Background yield N_{bkg} is highest for *CAT Finder* and lowest for *CAT B Finder*, also within uncertainties. I calculate the significance of the measurement with $N_{\mathrm{sig}}/\sqrt{N_{\mathrm{bkg}}}$, given in Table 8.2, and the results all agree within statistical uncertainties for all three tracking algorithms. The width σ is similar for *CAT B Finder* and *Baseline Finder*, while *CAT Finder* has a slightly larger value ($6.7 \pm 0.5\,\mathrm{MeV}$ compared to $6.5 \pm 0.5\,\mathrm{MeV}$), still within uncertainties.
In contrast to the higher prompt particle efficiency (see Section 7.1.1), I do not see a

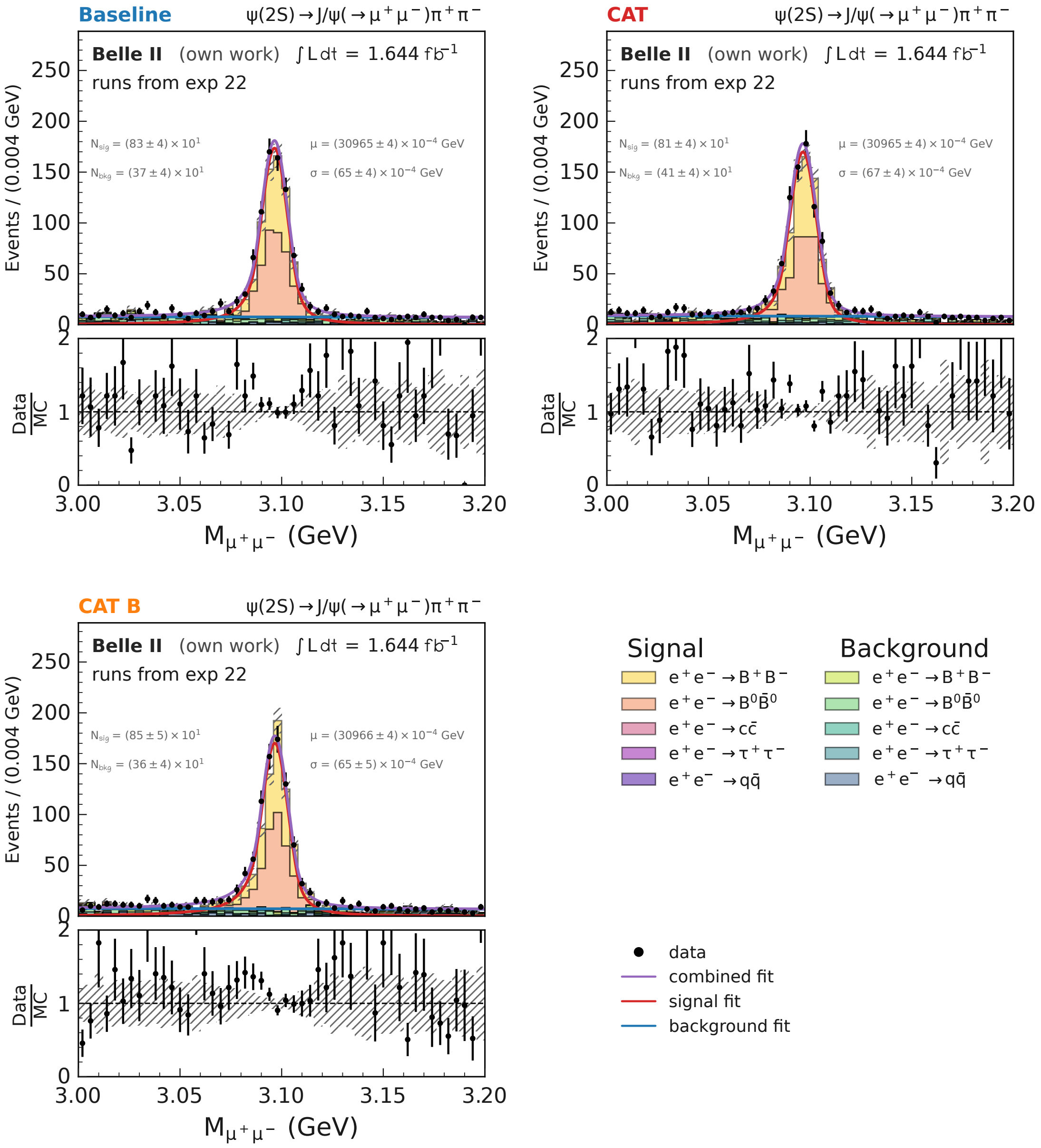


Figure 8.1: Reconstructed J/ψ mass from two opposite charged muons $M_{\mu^+\mu^-}$ in the signal region of 3.0 GeV to 3.2 GeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $M_{\mu^+\mu^-}$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ, and are shown as colored, stacked histograms. Data is shown in black. The combined fit to data (purple), as well as the individual signal (red) and background (blue) components, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.644\,\mathrm{fb}^{-1}$.

significant signal yield increase for *CAT B Finder* in either simulation or measured data, as the fit precision is limited by statistical uncertainties.

I repeat the analysis for runs from experiment 26 in Fig. 8.2 to study the performance on higher backgrounds. For all three algorithms, the simulation using high data beam backgrounds from experiment 26 with matching beam background conditions agrees with the measured data. The signal yield on measured data is lowest for *CAT Finder* (710 ± 50), higher for *Baseline Finder* (740 ± 50), and highest for *CAT B Finder* (760 ± 50), but all three agree within uncertainties. The peak position shifts from 3096.5 ± 0.4 MeV in experiment 22 to 3097.6 ± 0.5 MeV for *Baseline Finder*, with slightly different but statistically compatible values for *CAT Finder* and *CAT B Finder*. The peak width σ increases from 6.5 ± 0.4 MeV to 7.9 ± 0.5 MeV for *Baseline Finder*, with 7.4 ± 0.3 MeV for *CAT Finder* and 7.8 ± 0.4 MeV for *CAT B Finder*, which indicates that the momentum resolution decreases for the tracks in higher beam backgrounds. Despite the higher integrated luminosity compared to experiment 22, the overall signal and background yield is smaller. The reason here is that experiment 22 has fewer beam background hits and better CDC conditions compared to experiment 26. Overall, my new GNN-based algorithms reconstruct the J/ψ peak for prompt muons with performance matching *Baseline Finder* in experiment 22 and experiment 26.

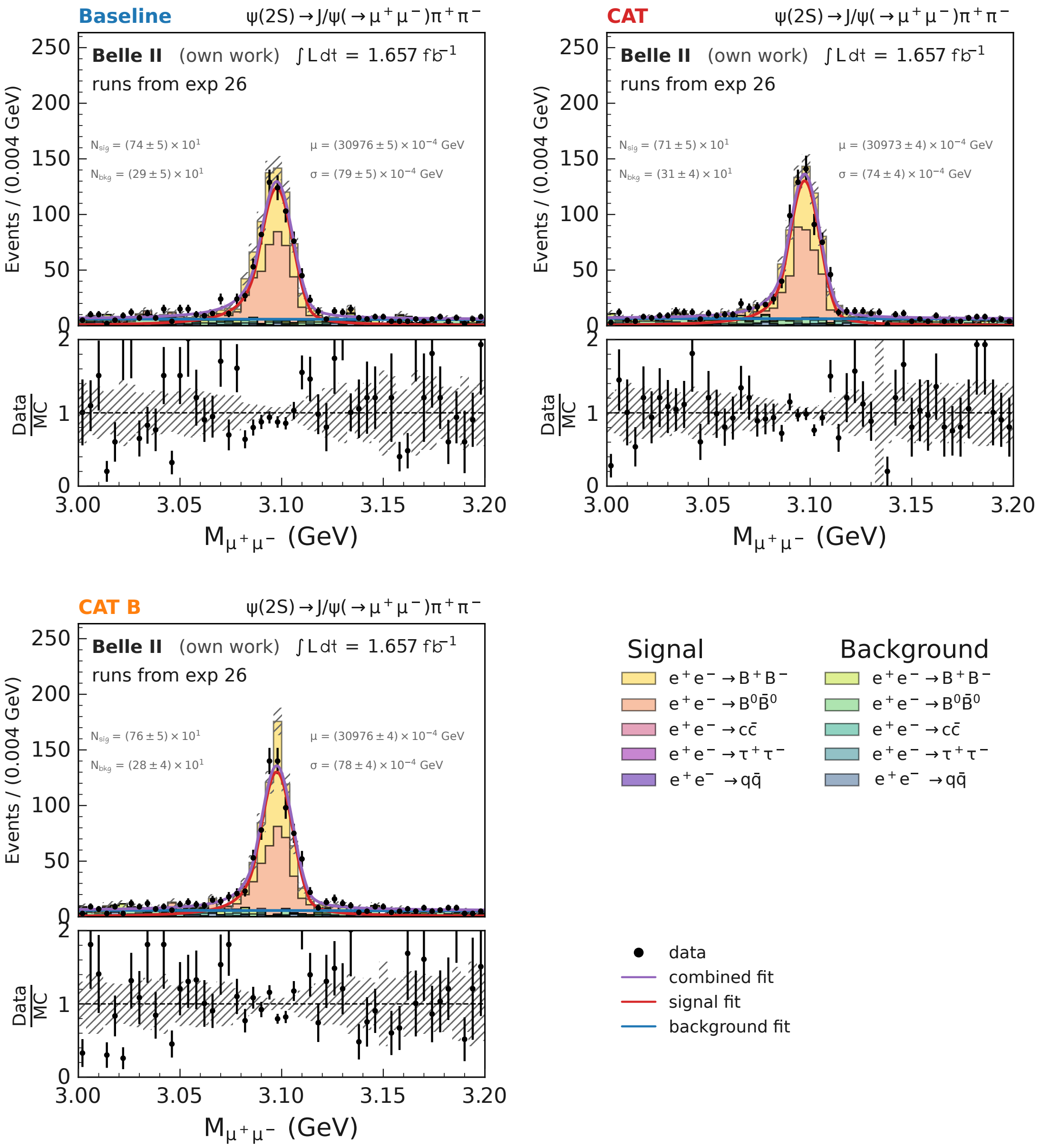


Figure 8.2: Reconstructed J/ψ mass from two opposite charged muons $M_{\mu^+\mu^-}$ in the signal region of 3.0 GeV to 3.2 GeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $M_{\mu^+\mu^-}$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ, and are shown as colored, stacked histograms. Data is shown in black. The combined fit to data (purple), as well as the individual signal (red) and background (blue) components, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.657\,\mathrm{fb}^{-1}$.

Algorithm		N_{sig}	N_{bkg}	μ (GeV)	σ (GeV)	$N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$
Runs from experiment 22						
Baseline Finder	data	$(83 \pm 4) \times 10^1$	$(37 \pm 4) \times 10^1$	$(30965 \pm 4) \times 10^{-4}$	$(65 \pm 4) \times 10^{-4}$	$(431 \pm 31) \times 10^{-1}$
	simulation	720 ± 25	342 ± 25	$(309696 \pm 30) \times 10^{-5}$	$(621 \pm 26) \times 10^{-5}$	$(389 \pm 20) \times 10^{-1}$
CAT Finder	data	$(81 \pm 4) \times 10^1$	$(41 \pm 4) \times 10^1$	$(30965 \pm 4) \times 10^{-4}$	$(67 \pm 4) \times 10^{-4}$	$(400 \pm 28) \times 10^{-1}$
	simulation	758 ± 22	328 ± 22	$(30973 \pm 4) \times 10^{-4}$	$(67 \pm 4) \times 10^{-4}$	$(419 \pm 19) \times 10^{-1}$
CAT B Finder	data	$(85 \pm 5) \times 10^1$	$(36 \pm 4) \times 10^1$	$(30966 \pm 4) \times 10^{-4}$	$(64 \pm 5) \times 10^{-4}$	$(450 \pm 40) \times 10^{-1}$
	simulation	747 ± 23	335 ± 23	$(309775 \pm 25) \times 10^{-5}$	$(608 \pm 26) \times 10^{-5}$	$(408 \pm 19) \times 10^{-1}$
Runs from experiment 26						
Baseline Finder	data	$(74 \pm 5) \times 10^1$	$(29 \pm 5) \times 10^1$	$(30976 \pm 5) \times 10^{-4}$	$(79 \pm 5) \times 10^{-4}$	43 ± 5
	simulation	$(75 \pm 6) \times 10^1$	$(30 \pm 6) \times 10^1$	$(30964 \pm 4) \times 10^{-4}$	$(75 \pm 5) \times 10^{-4}$	43 ± 6
CAT Finder	data	$(71 \pm 5) \times 10^1$	$(31 \pm 4) \times 10^1$	$(30973 \pm 4) \times 10^{-4}$	$(74 \pm 3) \times 10^{-4}$	40 ± 4
	simulation	$(74 \pm 4) \times 10^1$	$(29 \pm 4) \times 10^1$	$(30974 \pm 4) \times 10^{-4}$	$(72 \pm 4) \times 10^{-4}$	43 ± 4
CAT B Finder	data	$(76 \pm 5) \times 10^1$	$(28 \pm 4) \times 10^1$	$(30976 \pm 4) \times 10^{-4}$	$(78 \pm 4) \times 10^{-4}$	45 ± 4
	simulation	$(75 \pm 4) \times 10^1$	$(28 \pm 4) \times 10^1$	$(30974 \pm 6) \times 10^{-4}$	$(63 \pm 7) \times 10^{-4}$	45 ± 4

Table 8.2: Fit results for $J/\psi \to \mu^+\mu^-$ including statistical uncertainties on measured data and simulation for the three algorithms *Baseline Finder*, *CAT Finder* and *CAT B Finder* for runs from experiment 22 and runs from experiment 26. N_{sig} describes the signal yield, N_{bkg} the background yield and μ and σ mean and width of the DSCB (see Eq. (8.2)). The last column shows the signal significance estimate $N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$.

Number of hits in prompt muon tracks

I achieve a clean selection of signal muons in the mass window 3.07 GeV–3.12 GeV. One additional relevant metric to probe the track finding quality is the number of CDC hits per track $N_{\text{CDC Hits}}$. Fig. 8.3 compares $N_{\text{CDC Hits}}$ for μ^+ and μ^- in runs from experiments 22 and 26 for all three algorithms. I find more hits assigned to tracks reconstructed with *Baseline Finder* for both *CAT Finder* and *CAT B Finder*. Comparing simulation and measured data for experiment 22 for μ^+ (Fig. 8.4), *Baseline Finder* and *CAT Finder* agree within uncertainties, though *CAT Finder* assigns fewer hits than *Baseline Finder*. For *CAT B Finder*, I see significant discrepancies, with fewer hits in measured data than in simulation, indicating potential issues from the training on $B\bar{B}$ events.
For experiment 26 (Fig. 8.6), all algorithms show significant simulation-data differences.

For μ^-, discrepancies appear in experiment 22 (Fig. 8.5) and increase in experiment 26 (Fig. 8.7) for all algorithms between simulation and measured data. This opposite-sign asymmetry is a known issue[110], affecting track quality selections and indicating that simulation does not model hit behaviour accurately. I suspect this also drives the *CAT B Finder* simulation–data differences.
Furthermore, beam background conditions affect all algorithms. Although *CAT Finder* has fewer hits on average, the reconstructed mass width is the same across algorithms for both experiments. The width difference between experiments 22 and 26 cannot be explained solely by hit counts but also by hit quality, with stronger interference between overlapping hits from particles and hits from beam-background in experiment 26.

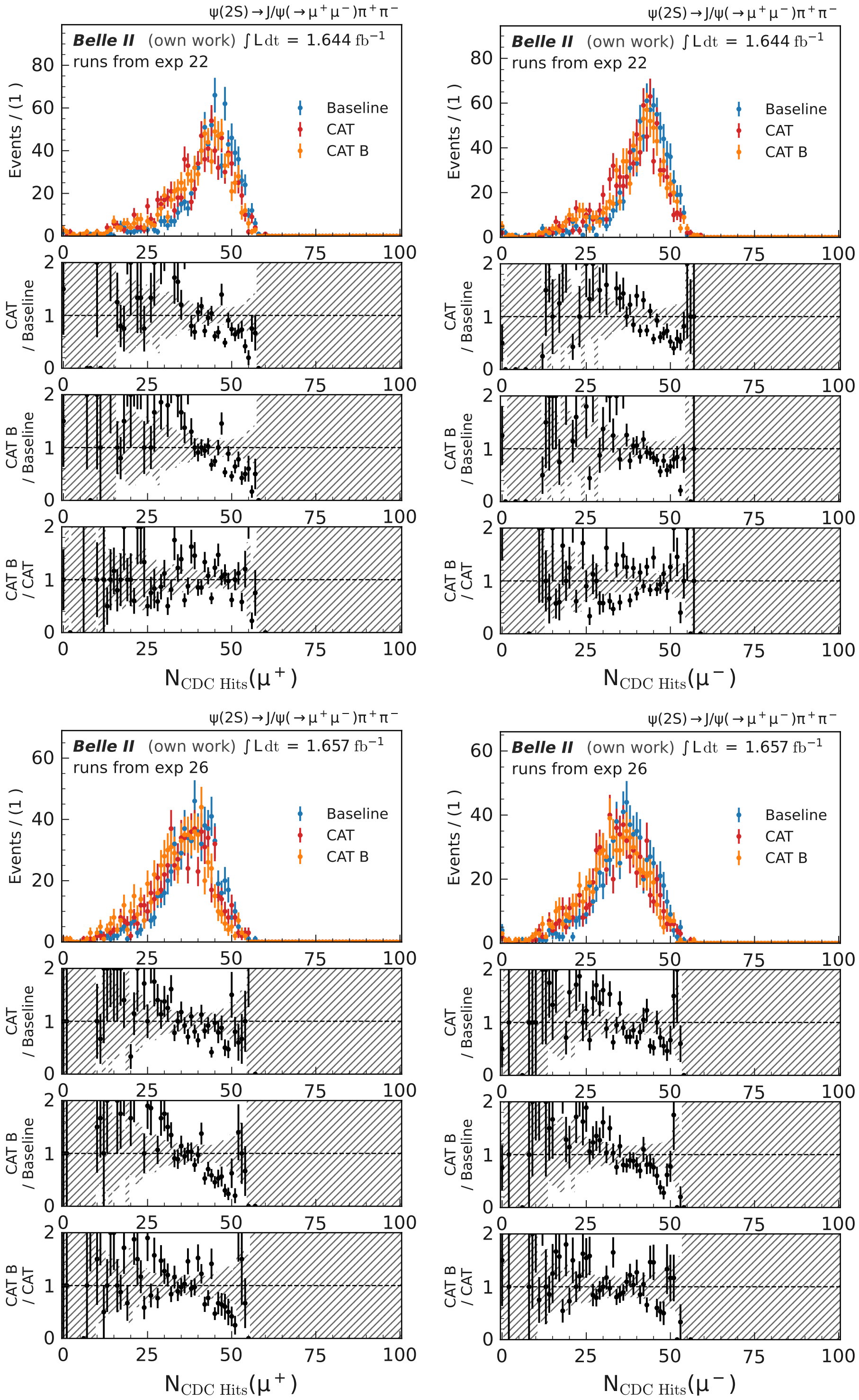


Figure 8.3: Number of CDC hits of the muons for the reconstructed J/ψ candidate from two opposite charged muons in the signal region of 3.07 GeV< $M_{\mu^+\mu^-}$ <3.12 GeV for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange). The number of CDC hits for μ^+ (left) and μ^- (right) is given for runs from experiment 22 (top) and runs from experiment 26 (bottom), including the ratio between each algorithm combination. Statistical uncertainties are correlated between all three algorithms, as the same measured data is used for reconstruction.

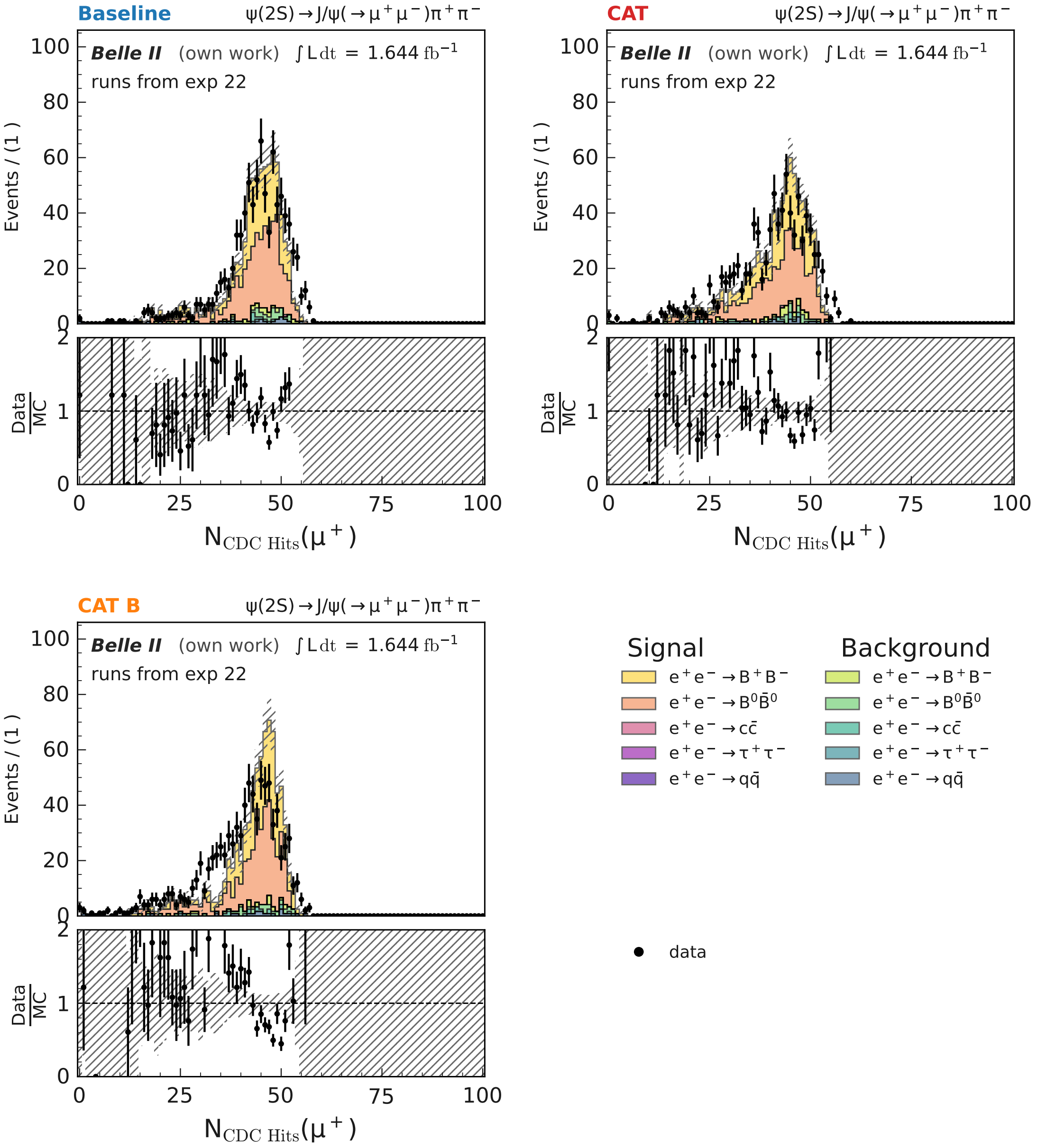


Figure 8.4: Number of CDC hits for the μ^+ of the reconstructed J/ψ candidate from two opposite charged muons in the signal region of $3.07\,\mathrm{GeV} < M_{\mu^+\mu^-} < 3.12\,\mathrm{GeV}$ for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(\mu^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

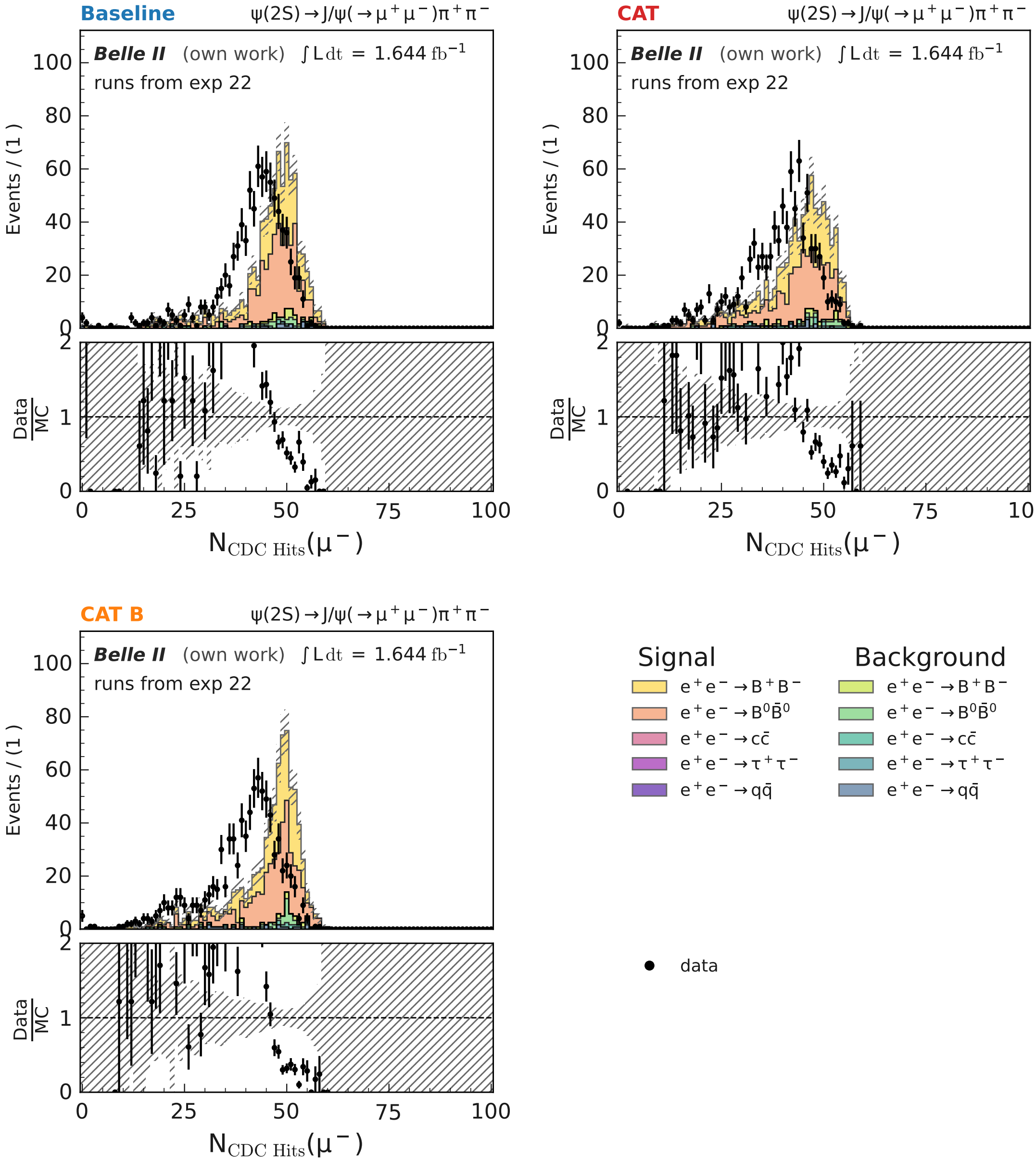


Figure 8.5: Number of CDC hits for the μ^- of the reconstructed J/ψ candidate from two opposite charged muons in the signal region of $3.07\,\mathrm{GeV} < M_{\mu^+\mu^-} < 3.12\,\mathrm{GeV}$ for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\mathrm{CDC\ Hits}}(\mu^-)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

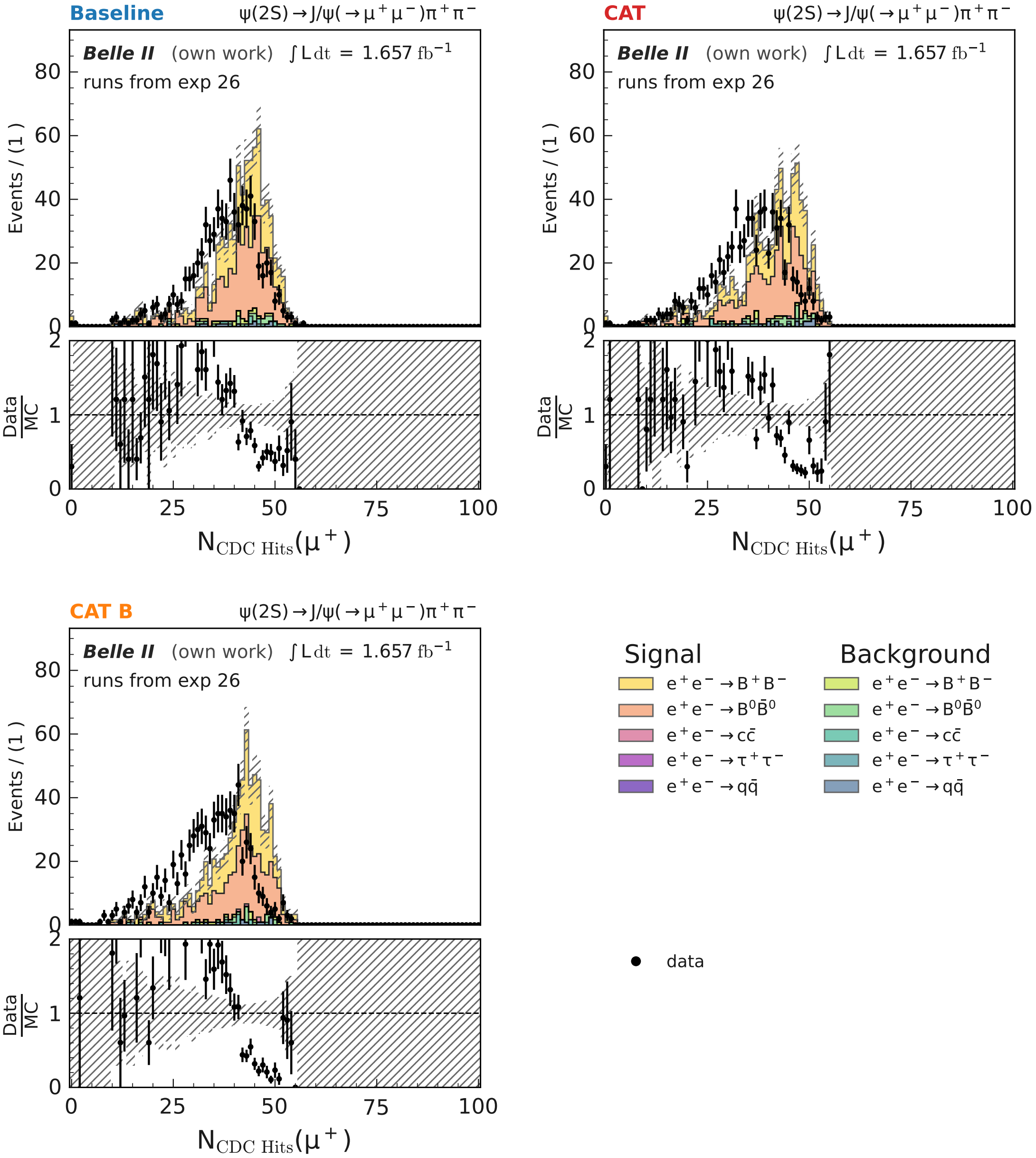


Figure 8.6: Number of CDC hits for the μ^+ of the reconstructed J/ψ candidate from two opposite charged muons in the signal region of 3.07 GeV< $M_{\mu^+\mu^-}$ <3.12 GeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(\mu^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.657\,\mathrm{fb}^{-1}$.

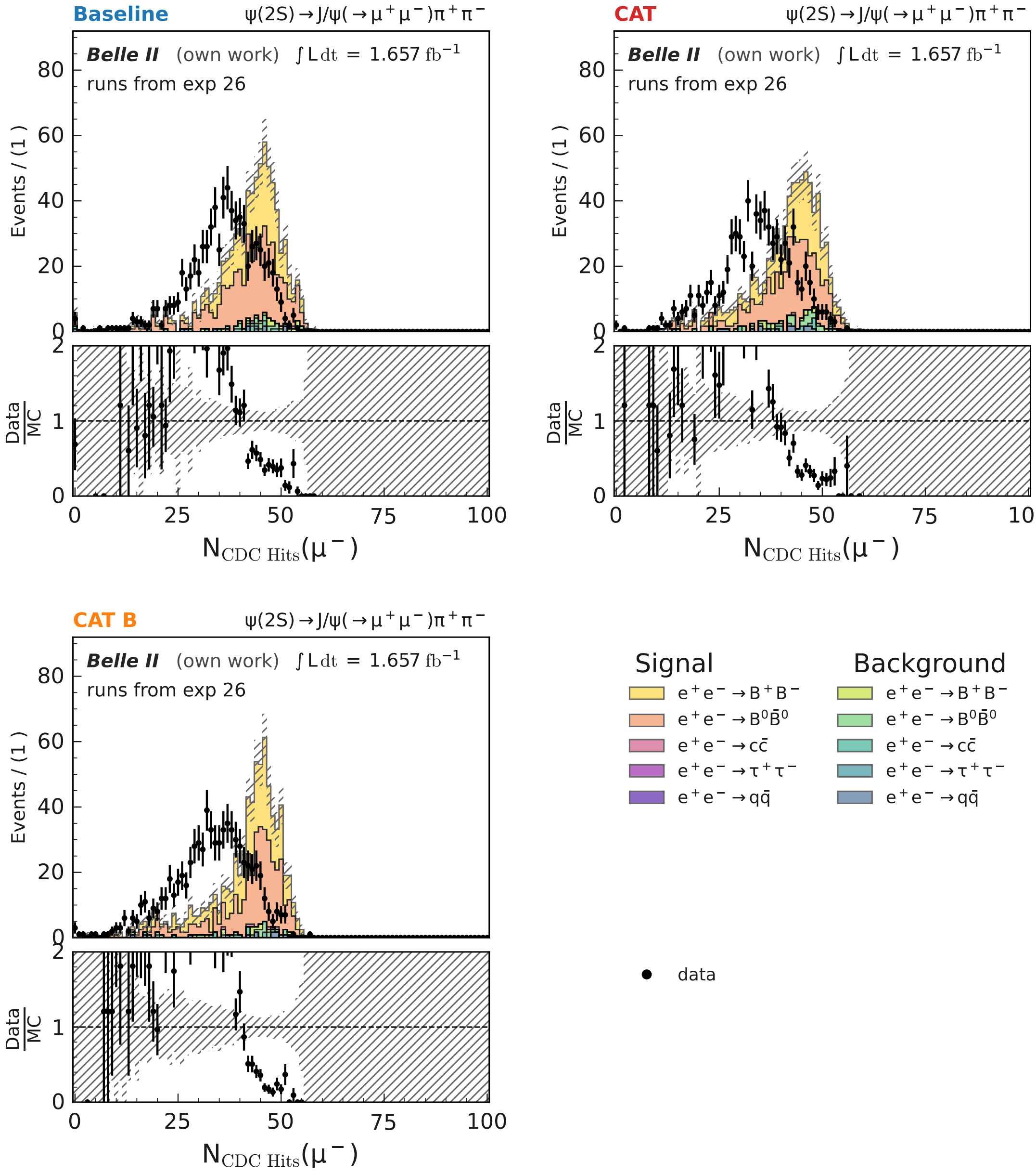


Figure 8.7: Number of CDC hits for the μ^- of the reconstructed J/ψ candidate from two opposite charged muons in the signal region of $3.07\,\text{GeV} < M_{\mu^+\mu^-} < 3.12\,\text{GeV}$ for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(\mu^-)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\text{d}t = 1.657\,\text{fb}^{-1}$.

8.1.2 $J/\psi \rightarrow e^+e^-$

Reconstructed J/ψ mass

I repeat the analysis with opposite-charged electrons to reconstruct the J/ψ. Electrons and positrons are particularly difficult to reconstruct in tracking because of bremsstrahlung. In this process, the energy loss is proportional to the particle's energy and alters its trajectory. Fig. 8.8 shows the fit results for experiment 22, and Fig. 8.9 for experiment 26, with the summary for both measured data and simulation given in Table 8.3. Compared to the $\mu^+\mu^-$ final state, the e^+e^- reconstructed J/ψ mass shows a pronounced low-mass tail due to electron (positron) bremsstrahlung. For experiment 22, the signal and background yields, as well as the peak position and width, are consistent across all three algorithms within uncertainties. For experiment 26, the signal yield is lowest for *Baseline Finder* at 471 ± 29, higher for *CAT Finder* at 540 ± 40, and highest for *CAT B Finder* at 550 ± 40. Backgrounds follow the opposite trend, largest for *Baseline Finder* and smaller for *CAT Finder* and *CAT B Finder*. Between *CAT B Finder* and *Baseline Finder* (*CAT Finder*), the peak position shifts from $3.0959 \pm 0.0011\,\text{GeV}$ to $3.0991 \pm 0.0014\,\text{GeV}$ ($3.0992 \pm 0.0009\,\text{GeV}$). In comparison to the $\mu^+\mu^-$ final state, I do not observe differences in the width between experiment 22 and experiment 26, as the tail towards lower values due to bremsstrahlung leads to larger fit uncertainties in comparison.

This demonstrates that my new algorithms successfully reconstruct both electron and muon tracks and reproduce the J/ψ mass in high-multiplicity events.

Number of hits in prompt electron tracks

I compare the number of hits in prompt electron tracks in the signal region 3.05–3.12 GeV for measured data in Fig. 8.10, with simulation comparisons in Section E.1. I observe the same trends as for muons: *Baseline Finder* assigns the most hits in experiment 22, with the difference decreasing in experiment 26. The *CAT B Finder* shows the fewest hits and the largest discrepancies between simulation and data, as observed for the $\mu^+\mu^-$ final state. The same positive–negative track asymmetry appears as for muons. Due to the large statistical uncertainties, I will discuss the discrepancy in the number of hits in detail in the next section using a decay that yields more events, resulting in higher statistics.

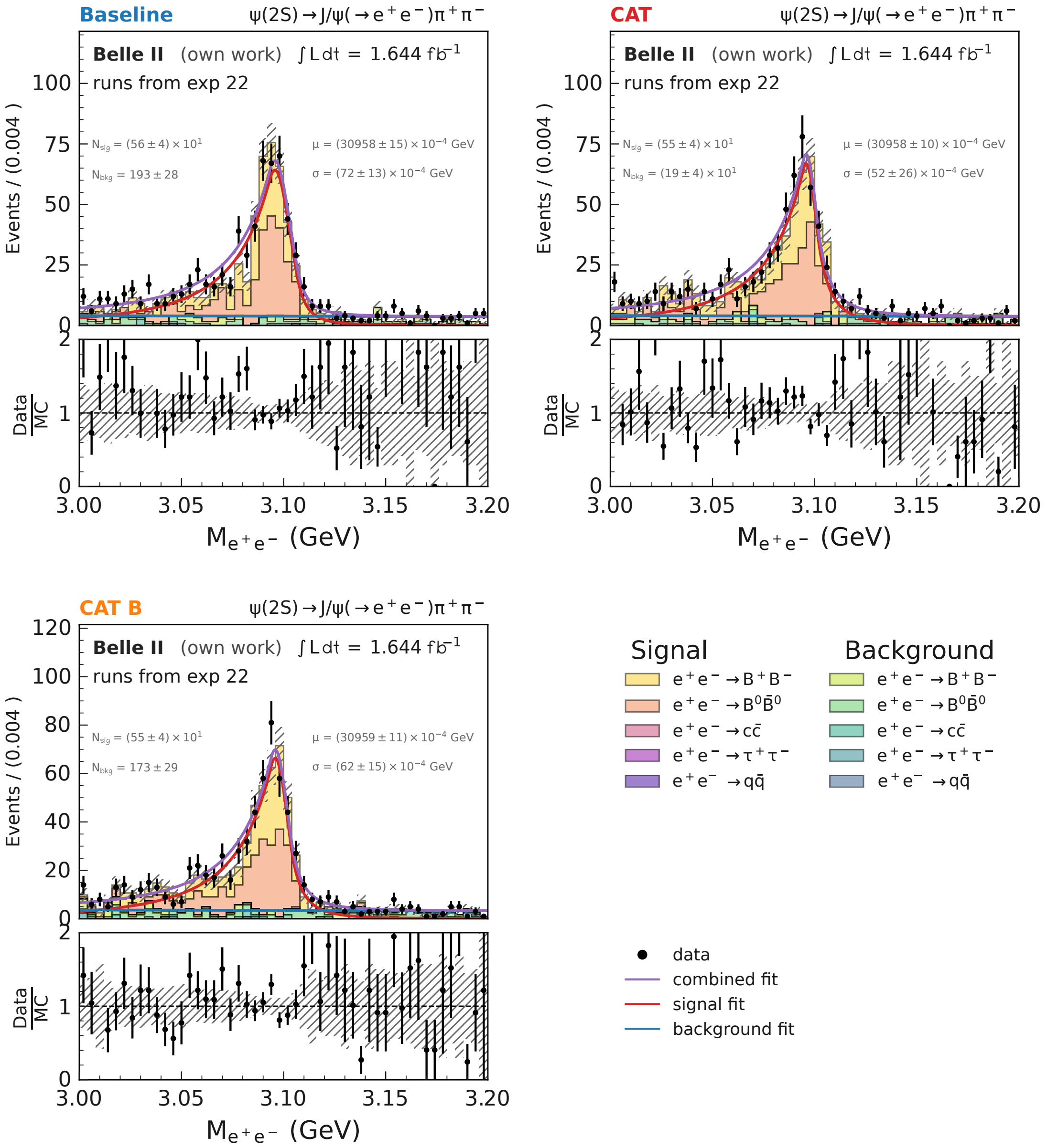


Figure 8.8: Reconstructed J/ψ mass from two opposite charged electrons $M_{e^+e^-}$ in the signal region of 3.0 GeV to 3.2 GeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $M_{e^+e^-}$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ, and are shown as colored, stacked histograms. Data is shown in black. The combined fit to data (purple), as well as the individual signal (red) and background (blue) components, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

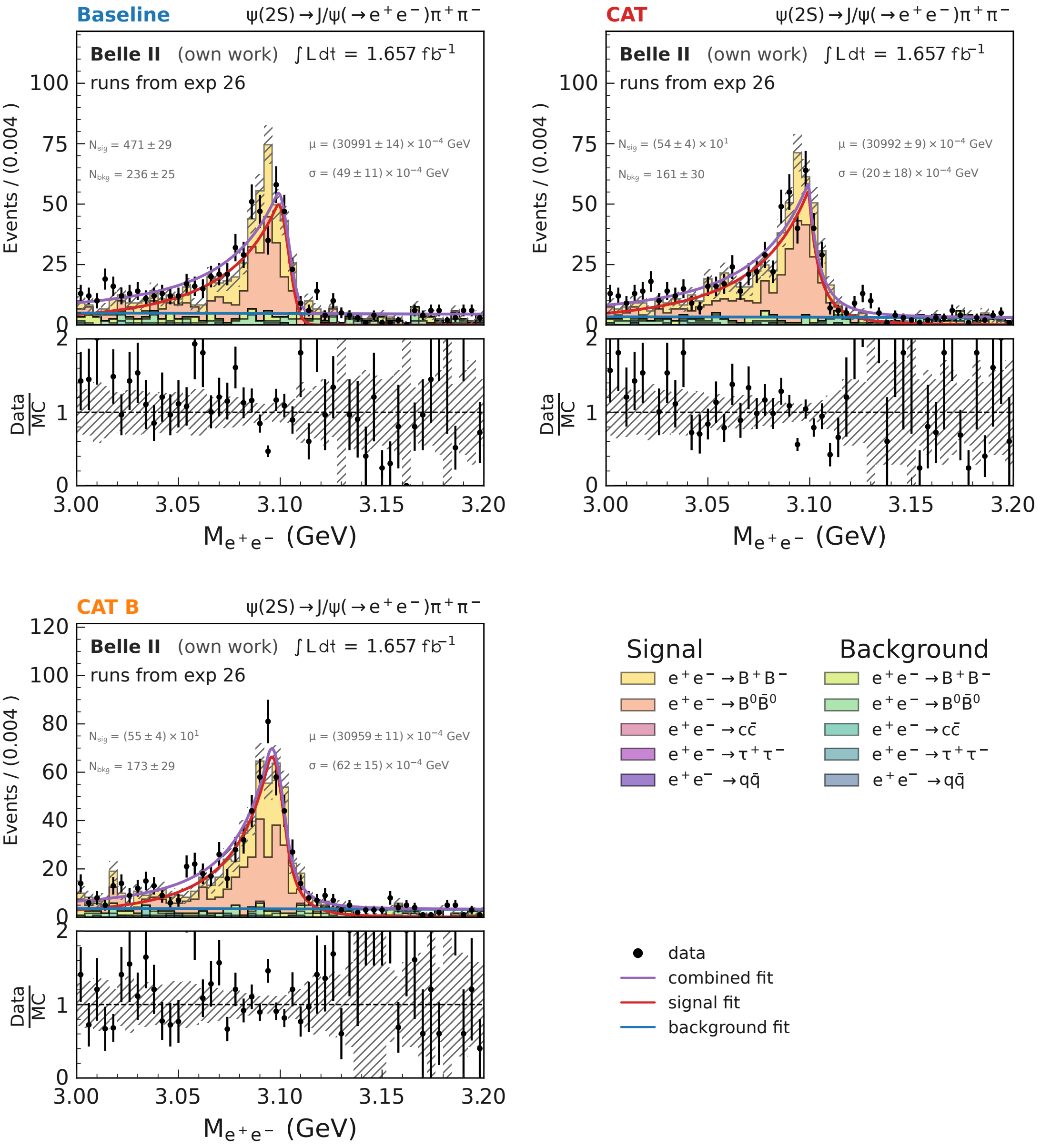


Figure 8.9: Reconstructed J/ψ mass from two opposite charged electrons $M_{e^+e^-}$ in the signal region of 3.0 GeV to 3.2 GeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $M_{e^+e^-}$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ, and are shown as colored, stacked histograms. Data is shown in black. The combined fit to data (purple), as well as the individual signal (red) and background (blue) components, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated. The simulation is scaled to the integrated luminosity of $\int \mathcal{L} dt = 1.657\,\text{fb}^{-1}$.

Algorithm		N_{sig}	N_{bkg}	μ (GeV)	σ (GeV)	$N_{\mathrm{sig}}/\sqrt{N_{\mathrm{bkg}}}$
Runs from experiment 22						
Baseline Finder	data	$(56 \pm 4) \times 10^{1}$	193 ± 28	$(30958 \pm 15) \times 10^{-4}$	$(72 \pm 13) \times 10^{-4}$	40 ± 4
	simulation	511 ± 23	124 ± 23	$(30960 \pm 8) \times 10^{-4}$	$(65 \pm 9) \times 10^{-4}$	46 ± 5
CAT Finder	data	$(55 \pm 4) \times 10^{1}$	$(19 \pm 4) \times 10^{1}$	$(30958 \pm 10) \times 10^{-4}$	$(52 \pm 26) \times 10^{-4}$	45 ± 5
	simulation	523 ± 22	153 ± 22	$(30976 \pm 10) \times 10^{-4}$	$(61 \pm 7) \times 10^{-4}$	$(423 \pm 35) \times 10^{-1}$
CAT B Finder	data	$(55 \pm 4) \times 10^{1}$	173 ± 29	$(30959 \pm 11) \times 10^{-4}$	$(62 \pm 15) \times 10^{-4}$	42 ± 5
	simulation	504 ± 23	179 ± 23	$(30982 \pm 7) \times 10^{-4}$	$(50 \pm 6) \times 10^{-4}$	$(377 \pm 30) \times 10^{-1}$
Runs from experiment 26						
Baseline Finder	data	471 ± 29	236 ± 25	$(30991 \pm 14) \times 10^{-4}$	$(49 \pm 11) \times 10^{-4}$	$(307 \pm 25) \times 10^{-1}$
	simulation	451 ± 24	194 ± 24	$(30958 \pm 8) \times 10^{-4}$	$(65 \pm 7) \times 10^{-4}$	$(325 \pm 26) \times 10^{-1}$
CAT Finder	data	$(54 \pm 4) \times 10^{1}$	161 ± 30	$(30992 \pm 9) \times 10^{-4}$	$(20 \pm 18) \times 10^{-4}$	41 ± 5
	simulation	517 ± 19	151 ± 19	$(30968 \pm 9) \times 10^{-4}$	$(77 \pm 8) \times 10^{-4}$	$(421 \pm 31) \times 10^{-1}$
CAT B Finder	data	$(55 \pm 4) \times 10^{1}$	173 ± 29	$(30959 \pm 11) \times 10^{-4}$	$(62 \pm 15) \times 10^{-4}$	42 ± 5
	simulation	540 ± 23	115 ± 23	$(30972 \pm 9) \times 10^{-4}$	$(61 \pm 8) \times 10^{-4}$	44 ± 4

Table 8.3: Fit results for $J/\psi \to e^+e^-$ including statistical uncertainties on measured data and simulation for the three algorithms *Baseline Finder*, *CAT Finder* and *CAT B Finder* for runs from experiment 22 and runs from experiment 26. N_{sig} describes the signal yield, N_{bkg} the background yield and μ and σ mean and width of the DSCB (see Eq. (8.2)). The last column shows the signal significance estimate $N_{\mathrm{sig}}/\sqrt{N_{\mathrm{bkg}}}$.

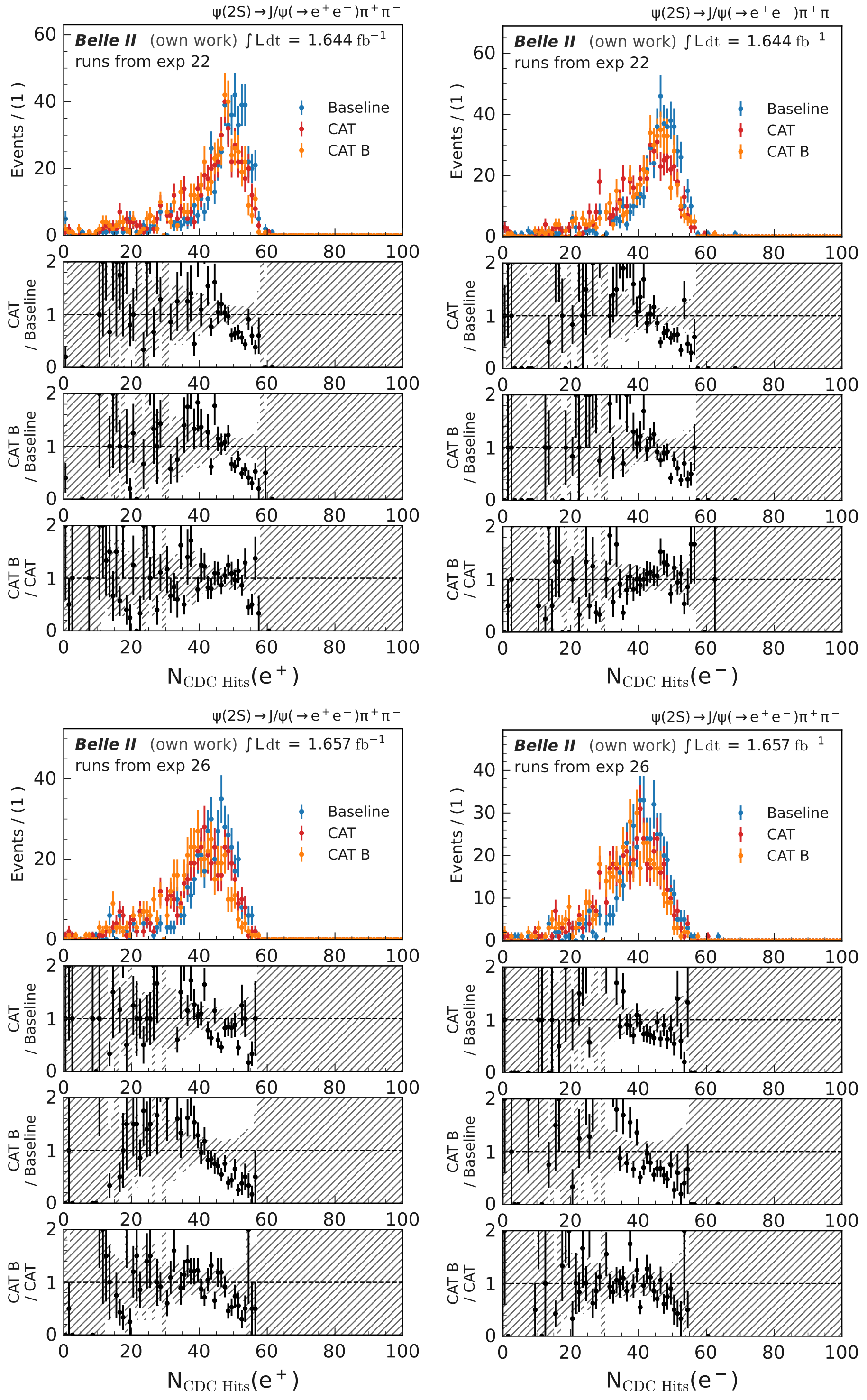


Figure 8.10: Number of CDC hits of the electrons for the reconstructed J/ψ candidate from two opposite charged electrons in the signal region of $3.05\,\text{GeV} < M_{e^+e^-} < 3.12\,\text{GeV}$ for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange). The number of CDC hits for e^+ (left) and e^- (right) is given for runs from experiment 22 (top) and runs from experiment 26 (bottom), including the ratio between each algorithm combination. Statistical uncertainties are correlated between all three algorithms, as the same measured data is used for reconstruction.

8.2 Displaced Tracks of $K_S^0 \to \pi^+\pi^-$ in High Multiplicity Events

Following the approach from the J/ψ decays, I reconstruct the $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ decay chain to increase the purity of the $K_S^0 \to \pi^+\pi^-$ selection. D^{*+} come from both B-mesons, with the branching fraction for $B \to D^{*+}X$ of $17.3 \pm 2.0\%$[37] and from $e^+e^- \to c\bar{c}$ events. The $D^{*+} \to D^0\pi^+$ branching fraction is $67.7 \pm 0.5\%$, with the $D^0 \to K_S^0\pi^+\pi^-$ branching fraction being $2.80 \pm 0.18\%$ [37]. The K_S^0 decays in $69.20 \pm 0.05\%$ into the $\pi^+\pi^-$ final state[37]. Especially due to the contribution from the $e^+e^- \to c\bar{c}$ events and the higher branching fractions, I expect a larger number of decays in comparison to the J/ψ decay.

Event Selection

I evaluate K_S^0 candidates in high multiplicity events on measured data using the same HLT filter as in the previous section (see Section 3.4.3). Following the J/ψ decay to select the best K_S^0 candidate and suppress backgrounds, the decay chain $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ is reconstructed. The analysis selection is given in the following

K_S^0 candidate Two oppositely charged particles without any further selections applied are combined in a vertex fit. The reconstructed mass of the two charged tracks M has to be between 480 MeV and 516 MeV. The significance of the displacement, defined as the ratio of d_r to its uncertainty given by the vertex fit, must exceed 3.

Pion selection D_0 Pion candidates π^+ must be prompt, with transverse and longitudinal impact parameters satisfying $d_{\text{IP}} < 2\,\text{cm}$ and $|z_{\text{IP}}| < 4\,\text{cm}$, respectively, with a transverse momentum above 0.1 GeV.

D^0 candidate The D^0 candidate is reconstructed from the K_S^0 candidate and two opposite charged pions. The reconstructed mass is required to be between 1.82 GeV and 1.91 GeV.

Pion selection D^{*+} This pion candidate π_S is also required to be prompt $d_{\text{IP}} > 2\,\text{cm}$ and $|z_{\text{IP}}| > 4\,\text{cm}$, but there is no requirement on the pion momentum as it is expected to be very small.

D^{*+} candidate The D^0 is combined with one charged pion to form the D^{*+} candidate. A vertex fit on the full decay chain is performed. The mass difference between the D^{*+} and the D^0 candidate is required to be between 0.144 GeV to 0.147 GeV. The center-of-mass momentum of the D^{*+}, $p^*_{D^{*+}}$, is required to be larger than 1.6 GeV. This suppresses the backgrounds from $B\bar{B}$ events at the cost of D^{*+} signal from $B\bar{B}$ events. As a large contribution of D^{*+} comes from $e^+e^- \to c\bar{c}$ events, this leads to a higher selection purity in total. I chose best candidate for the reconstructed

Analysis selection	Selection criteria
K_S^0	$0.48\,\mathrm{GeV} < M < 0.516\,\mathrm{GeV}$ significance of distance > 3
π^+	$\lvert z_{\mathrm{IP}}\rvert < 4\,\mathrm{cm}$ $\lvert d_{\mathrm{IP}}\rvert < 2\,\mathrm{cm}$ $p_t > 0.1\,\mathrm{GeV}$
D^0	$1.82\,\mathrm{GeV} < M < 1.91\,\mathrm{GeV}$
π_S^+	$\lvert dz\rvert < 4\,\mathrm{cm}$ $\lvert dr\rvert < 2\,\mathrm{cm}$
D^{*+}	$0.144\,\mathrm{GeV} < M_{D^{*+}} - M_{D^0} < 0.147\,\mathrm{GeV}$ $p^*_{D^{*+}} > 1.6\,\mathrm{GeV}$

Table 8.4: Selection summary for $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$.

K_S^0 by requiring the difference between the reconstructed mass for the D^{*+} and D^0, $M_{D^{*+}}^{\mathrm{reco}} - M_{D^0}^{\mathrm{reco}}$ and the difference between the true mass according to [37] with $M_{D^{*+}}^{\mathrm{PDG}} = 2010.26 \pm 0.05\,\mathrm{MeV}$ and $M_{D^0}^{\mathrm{PDG}} = 1864.84 \pm 0.05\,\mathrm{MeV}$ for each candidate in the event to be minimal.

$$\text{candidate} = \min_{i..N} \lvert (M_{i,D^{*+}}^{\mathrm{PDG}} - M_{i,D^0}^{\mathrm{PDG}}) - (M_{D^{*+}}^{\mathrm{reco}} - M_{D^0}^{\mathrm{reco}})\rvert. \tag{8.5}$$

The selection is summarized in Table 8.4.

Reconstruction

Compared to the $\psi(2S)$ reconstruction, there are now five instead of four final-state particles. The small mass difference between the D^{*+} and D^0 results in the π_s^+ having small momentum of $p_{\pi_s} < 0.15\,\mathrm{GeV}$. With a lifetime of around 2.7 cm[37], most K_S^0 decays are rather prompt. This challenges the *CAT Finder*, as many curlers in this momentum range are expected, and the reconstruction efficiency strongly depends on the SVD for all three algorithms. I find that at least one pion from the K_S^0 decay is reconstructed by SVD track finding in 38% of cases for the *Baseline Finder*, 41% for the *CAT Finder*, and 31% for the *CAT B Finder*.

Reconstructed K_S^0 mass

The reconstructed K_S^0 mass for experiment 22 is shown in Fig. 8.11. I fit the distribution using a DSCB function for the signal and a Chebyshev polynomial for the background, as described in Section 8.1. Here, $n_{l/r}$ is fixed to 3 as the range is between 2.5 to 3.5, both on simulation and if left floating on measured data. The fit results for measured data and simulation are given in Table 8.5. For experiment 22, the *Baseline Finder* achieves the lowest signal yield of 8850 ± 190 and the lowest background of 10130 ± 190. I then calculate the approximate significance of the measurement with $N_{\mathrm{sig}}/\sqrt{N_{\mathrm{bkg}}} = 87.9 \pm 2.1$.

The *CAT Finder* has a higher signal yield of 9060 ± 210, but also a larger background of 12470 ± 220, resulting in a slightly lower significance of 81.1 ± 2.0, consistent with simulation expectations due to higher fake and clone rates from SVD track recovery at low momentum. The *CAT B Finder* achieves the highest signal yield of 9900 ± 210, with a background similar to *CAT Finder* at 12590 ± 220, giving a significance of 88.9 ± 2. Within statistical uncertainties, the significance of the *Baseline Finder* and *CAT B Finder* agree, though the *CAT B Finder* peak in measured data is smaller than expected from simulation. For all three algorithms, the fitted mean obtained from the measured data is statistically consistent with $m_{K_S^0}^{\mathrm{PDG}} = 497.611 \pm 0.013\,\mathrm{MeV}$ [37].

For experiment 26, shown in Fig. 8.12, the *Baseline Finder* has a signal yield of 8410 ± 220 and background of 10810 ± 230, giving a reduced significance of 80.9 ± 2.3. The *CAT Finder* yields 7940 ± 240 with background 13340 ± 260, leading to a significance of 68.7 ± 2.2. The *CAT B Finder* again gives the highest signal yield 9010 ± 240 with background 13710 ± 250, resulting in a significance of 76.9 ± 2.2, consistent with *Baseline Finder* within uncertainties. For all algorithms, the signal yield on measured data is smaller than in simulation, and *CAT Finder* and *CAT B Finder* show increased background in the non-signal regions $M_{\pi^+\pi^-} < 0.49\,\mathrm{MeV}$ and $M_{\pi^+\pi^-} > 0.505\,\mathrm{MeV}$, particularly in experiment 26, between measured data and simulation. This again indicates that there is a mismatch in the simulation.

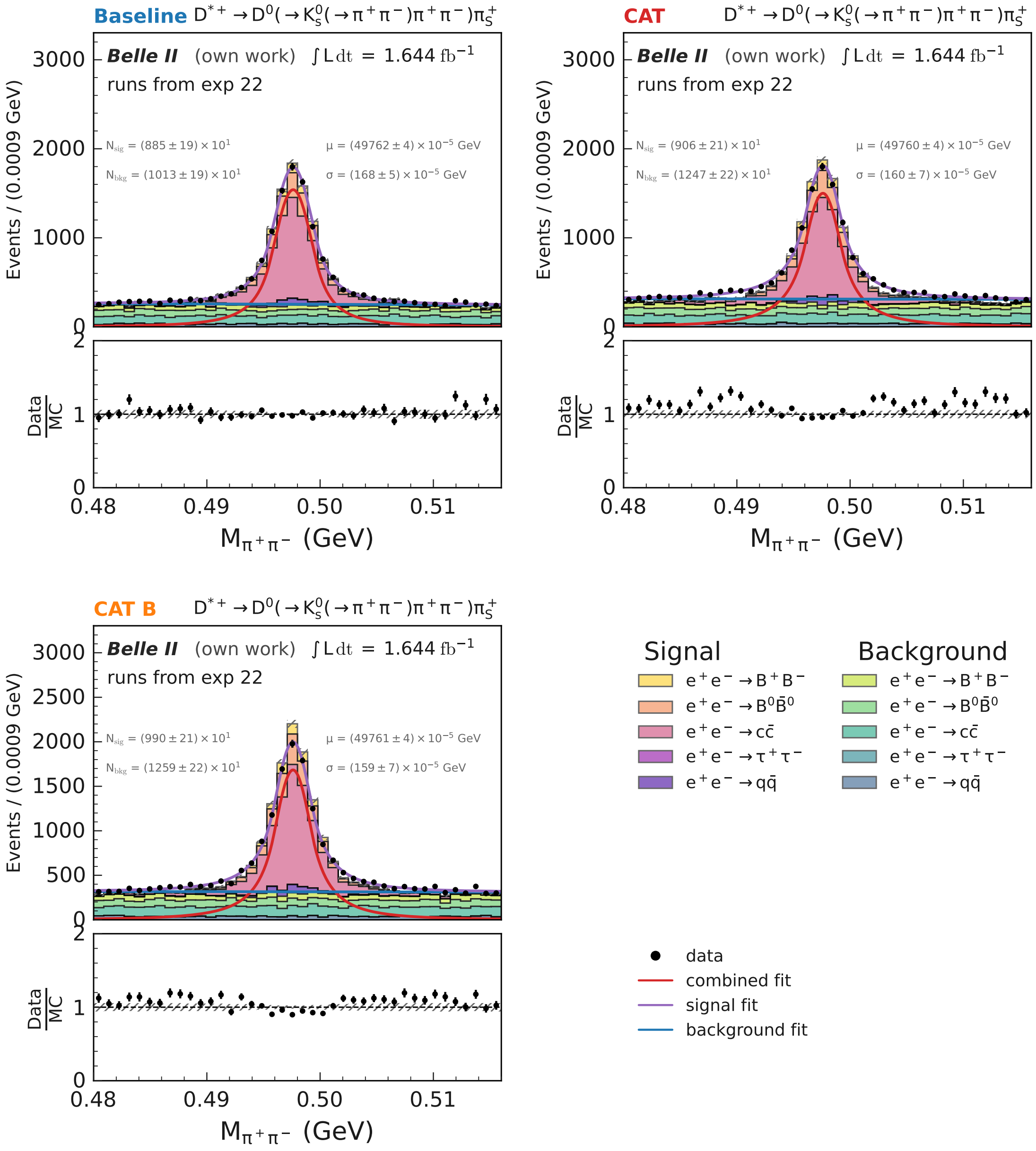


Figure 8.11: Reconstructed K_S^0 mass from two opposite charged pions $M_{\pi^+\pi^-}$ in the signal region of 0.48 GeV to 0.52 GeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $M_{\pi^+\pi^-}$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ, and are shown as colored, stacked histograms. Data is shown in black. The combined fit to data (purple), as well as the individual signal (red) and background (blue) components, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.644\,\text{fb}^{-1}$.

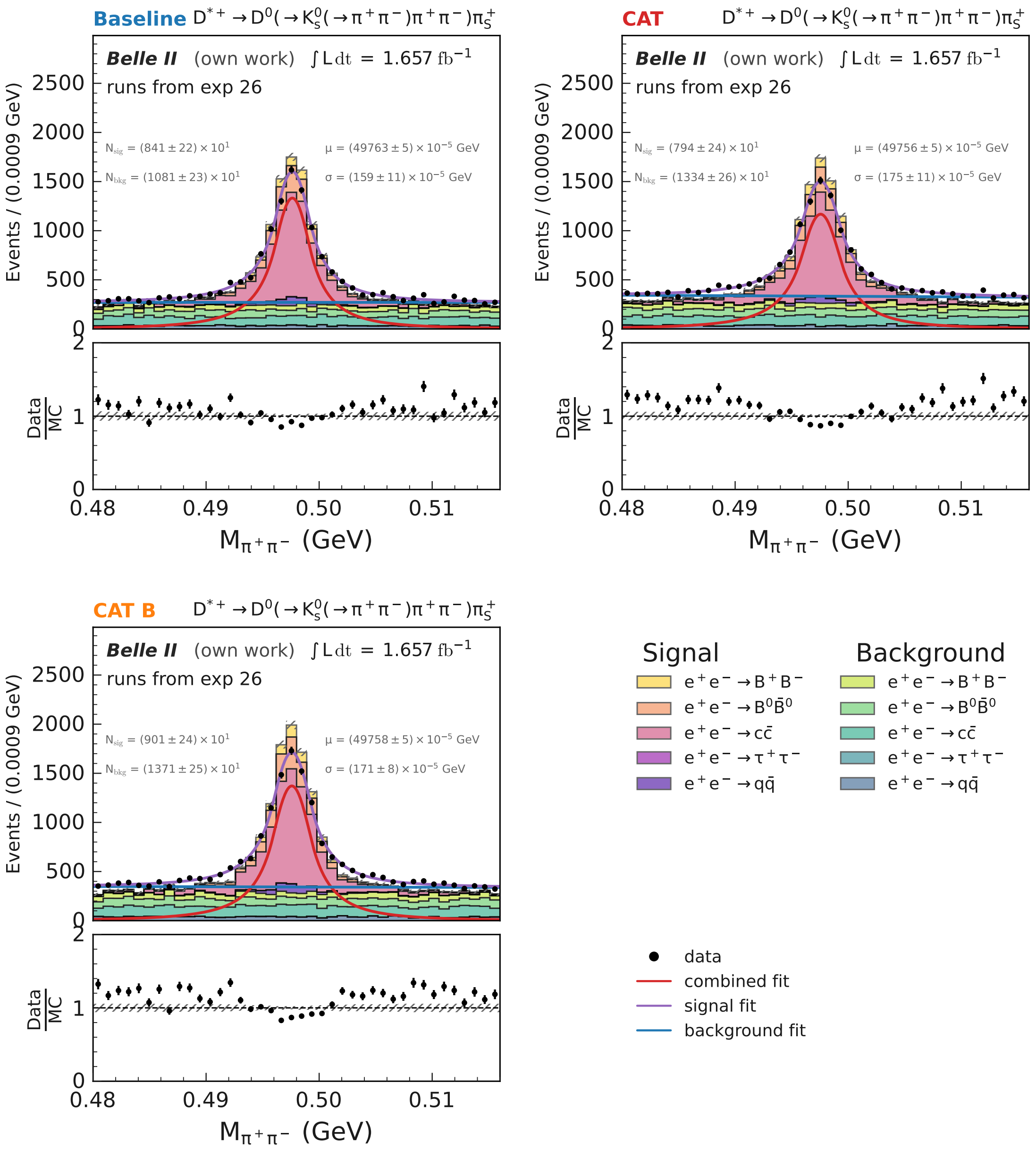


Figure 8.12: Reconstructed K_S^0 mass from two opposite charged pions $M_{\pi^+\pi^-}$ in the signal region of 0.48 GeV to 0.52 GeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $M_{\pi^+\pi^-}$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ, and are shown as colored, stacked histograms. Data is shown in black. The combined fit to data (purple), as well as the individual signal (red) and background (blue) components, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.644\,\text{fb}^{-1}$.

Algorithm		N_{sig}	N_{bkg}	μ (GeV)	σ (GeV)	$N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$
Runs from experiment 22						
Baseline	data	$(885 \pm 19) \times 10^1$	$(1013 \pm 19) \times 10^1$	$(49762 \pm 4) \times 10^{-5}$	$(168 \pm 5) \times 10^{-5}$	$(879 \pm 21) \times 10^{-1}$
	simulation	$(919 \pm 16) \times 10^1$	$(961 \pm 19) \times 10^1$	$(49761 \pm 4) \times 10^{-5}$	$(166 \pm 5) \times 10^{-5}$	$(937 \pm 19) \times 10^{-1}$
CAT Finder	data	$(906 \pm 21) \times 10^1$	$(1247 \pm 22) \times 10^1$	$(49760 \pm 4) \times 10^{-5}$	$(166 \pm 7) \times 10^{-5}$	$(811 \pm 21) \times 10^{-1}$
	simulation	$(932 \pm 21) \times 10^1$	$(1090 \pm 18) \times 10^1$	$(49754 \pm 4) \times 10^{-5}$	$(166 \pm 6) \times 10^{-5}$	$(893 \pm 22) \times 10^{-1}$
CAT B Finder	data	$(990 \pm 21) \times 10^1$	$(1259 \pm 22) \times 10^1$	$(49761 \pm 4) \times 10^{-5}$	$(159 \pm 7) \times 10^{-5}$	$(882 \pm 21) \times 10^{-1}$
	simulation	$(1042 \pm 18) \times 10^1$	$(1166 \pm 18) \times 10^1$	$(49762 \pm 3) \times 10^{-5}$	$(172 \pm 5) \times 10^{-5}$	$(965 \pm 19) \times 10^{-1}$
Runs from experiment 26						
Baseline	data	$(841 \pm 22) \times 10^1$	$(1081 \pm 23) \times 10^1$	$(49763 \pm 5) \times 10^{-5}$	$(159 \pm 11) \times 10^{-5}$	$(809 \pm 23) \times 10^{-1}$
	simulation	$(919 \pm 16) \times 10^1$	$(960 \pm 16) \times 10^1$	$(49761 \pm 4) \times 10^{-5}$	$(157 \pm 6) \times 10^{-5}$	727 ± 13
CAT Finder	data	$(794 \pm 24) \times 10^1$	$(1334 \pm 26) \times 10^1$	$(49756 \pm 5) \times 10^{-5}$	$(175 \pm 11) \times 10^{-5}$	$(687 \pm 22) \times 10^{-1}$
	simulation	$(949 \pm 19) \times 10^1$	$(1052 \pm 19) \times 10^1$	$(49761 \pm 4) \times 10^{-5}$	$(169 \pm 7) \times 10^{-5}$	$(925 \pm 21) \times 10^{-1}$
CAT B Finder	data	$(901 \pm 24) \times 10^1$	$(1371 \pm 25) \times 10^1$	$(49758 \pm 5) \times 10^{-5}$	$(171 \pm 8) \times 10^{-5}$	$(769 \pm 22) \times 10^{-1}$
	simulation	$(1012 \pm 17) \times 10^1$	$(1148 \pm 17) \times 10^1$	$(49760 \pm 3) \times 10^{-5}$	$(166 \pm 5) \times 10^{-5}$	$(945 \pm 18) \times 10^{-1}$

Table 8.5: Fit results for $K_S^0 \to \pi^+\pi^-$ including statistical uncertainties on measured data and simulation for the three algorithms Baseline, *CAT Finder* and *CAT B Finder* for runs from experiment 22. N_{sig} describes the signal yield, N_{bkg} the background yield and μ and σ mean and width of the DSCB. The last column shows the signal significance estimate $N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$.

Displacement

From the studies in Section 7.1.2, I expect the reconstruction efficiency for K_S^0 to increase for vertex displacements beyond 8 cm. Fig. 8.13 shows the measured data comparison for all three algorithms as a function of the reconstructed K_S^0 transverse vertex distance d_ρ, for both experiment 22 and 26 in the signal window $494.8\,\mathrm{MeV} < M_{\pi^+\pi^-} < 500.8\,\mathrm{MeV}$. I observe that *CAT B Finder* yields more entries for vertices beyond 8 cm compared to the *Baseline Finder* in both beam background conditions. For the *CAT Finder*, the distribution is similar to *Baseline Finder* for experiment 22, but for experiment 26, *CAT Finder* has fewer entries for vertices with $d_\rho \leq 8\,\mathrm{cm}$, while exceeding *Baseline Finder* for more displaced vertices. The comparison between simulation and measured data for d_ρ is shown in Fig. E.5 and Fig. E.6. For experiment 22 *Baseline Finder* and *CAT Finder* agree for simulation and measured data up to 16 cm, *CAT B Finder* has fewer events on measured data compared

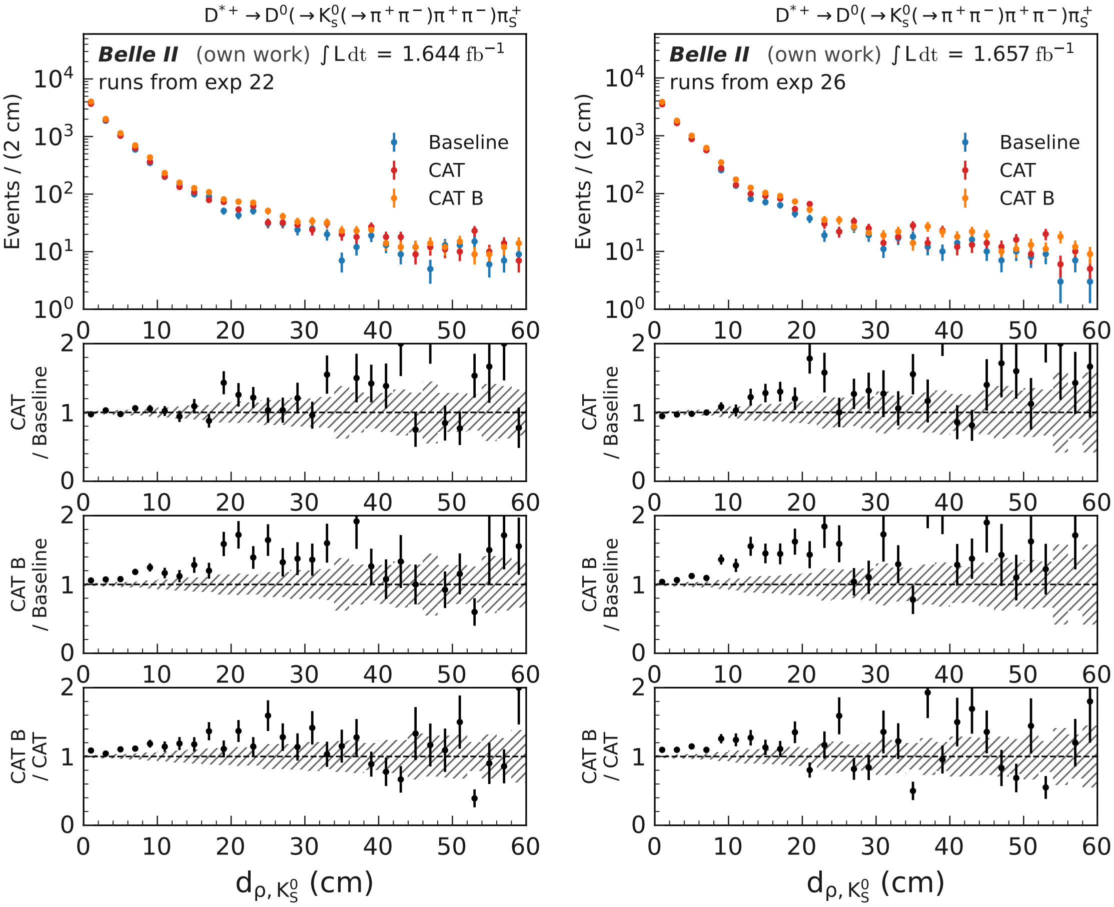


Figure 8.13: Transverse vertex distance $d_{\rho,K_S^0} = \sqrt{d^2_{x,K_S^0} + d^2_{y,K_S^0}}$ from the interaction point of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange). The distance is given for runs from experiment 22 (left) and runs from experiment 26 (right), including the comparison between each algorithm combination.

to simulation from 8 cm to 20 cm. For experiment 26 however, the simulation expects more events at higher displacements from 6 cm on for all three algorithms than is measured on data, with the largest impact for *CAT B Finder*.

Number of hits in displaced pion tracks

I also report the number of hits for the displaced pion tracks of the K_S^0 vertex in the reduced window $494.8\,\mathrm{MeV} < M_{\pi^+\pi^-} < 500.8\,\mathrm{MeV}$, to increase the fraction of signal to background. Multiple track parameters for π^+ and π^- are provided for completeness in Section E.2 for all three tracking algorithms. The comparison of measured data for all three tracking algorithms is shown in Fig. 8.14 for runs from experiment 22 and experiment 26, and the comparison with simulation is given in Fig. E.7-Fig. E.10.

For positively charged π^+ from experiment 22, I find that simulation mostly agrees with measured data for both *CAT Finder* and *Baseline Finder*, though the distributions in data are generally shifted to lower numbers of CDC hits. The first bin, representing SVD-only tracks with no CDC hits, is underestimated in the simulation for both algorithms. For *CAT B Finder*, I again see a large discrepancy: the algorithm performs much better in simulation than in measured data. In measured data, *CAT B Finder* only marginally outperforms *CAT Finder* and does not reach *Baseline Finder* performance.
For π^- from experiment 22, I observe the same disagreement between simulation and measured data, with all three algorithms showing a shift of roughly 10 hits. For experiment 26, both π^+ and π^- distributions are further shifted between simulation and measured data.
I investigated if hits that are attached during track finding are lost during the subsequent track fitting stages in the tracking chain (see Section 5.4). I calculate this as the difference between the number of CDC hits after track finding, $N_{\text{CDC Hits, reco}}$, and after full reconstruction, $N_{\text{CDC Hits}}$. Fig. 8.16 shows this for π^- in experiment 26. I find that *Baseline Finder* loses more hits in simulation between track finding and fitting, consistent with its lower hit purity compared to *CAT Finder* and *CAT B Finder* (Fig. E.14).
Interestingly, all three algorithms show a large difference in hits lost between simulation and measured data (Fig. 8.17). The *Baseline Finder* again loses the most hits during the track fitting steps in measured data, *CAT Finder* and *CAT B Finder* losses are comparable and less than *Baseline Finder*. The information on which hits are lost during the track fitting steps is currently not available within basf2. This makes further investigation on the differences between hits lost on simulation and measured data out of scope for my thesis. This study concludes that for *CAT B Finder*, the low number of assigned hits on measured data compared to *Baseline Finder* is not due to the track fitting step, but something different, which I will investigate next.

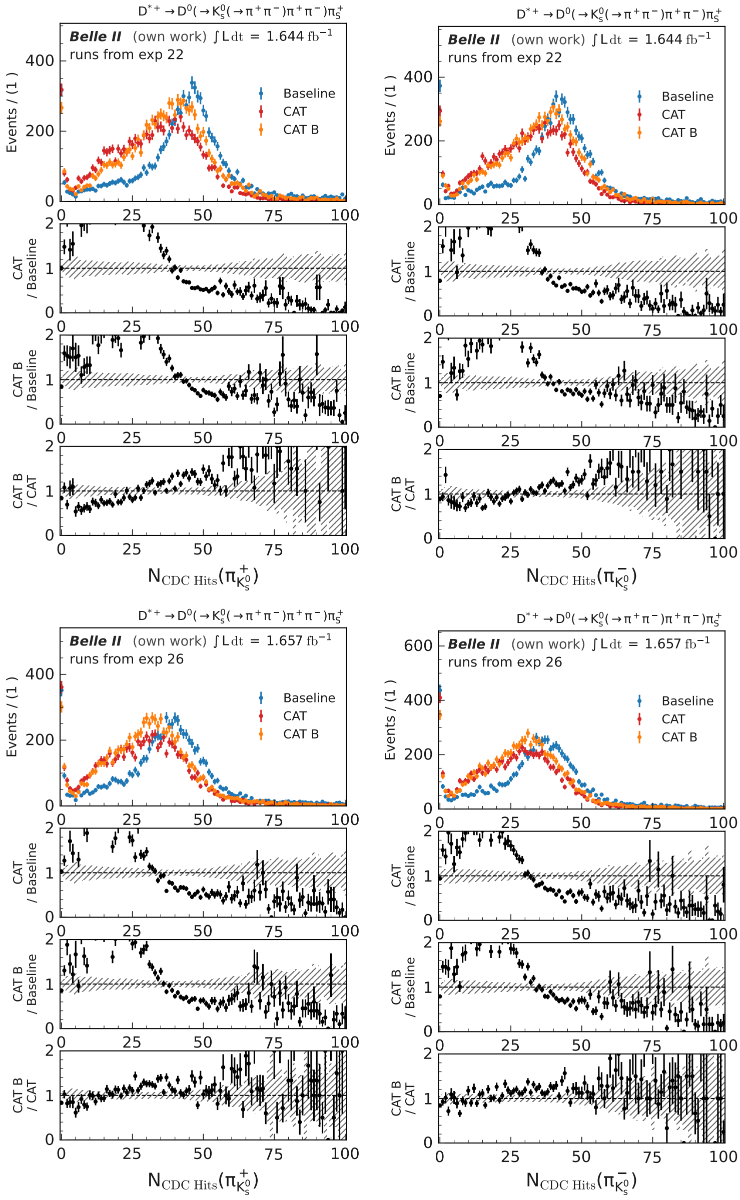


Figure 8.14: Number of CDC hits of the pions for the reconstructed K_S^0 mass from two opposite charged muons in the signal region of $494.8\,\text{MeV} < M_{\pi^+\pi^-} < 500.8\,\text{MeV}$ for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange). The number of CDC hits for π^+ (left) and π^- (right) is given for runs from experiment 22 (top) and runs from experiment 26 (bottom), including the ratio between each algorithm combination.

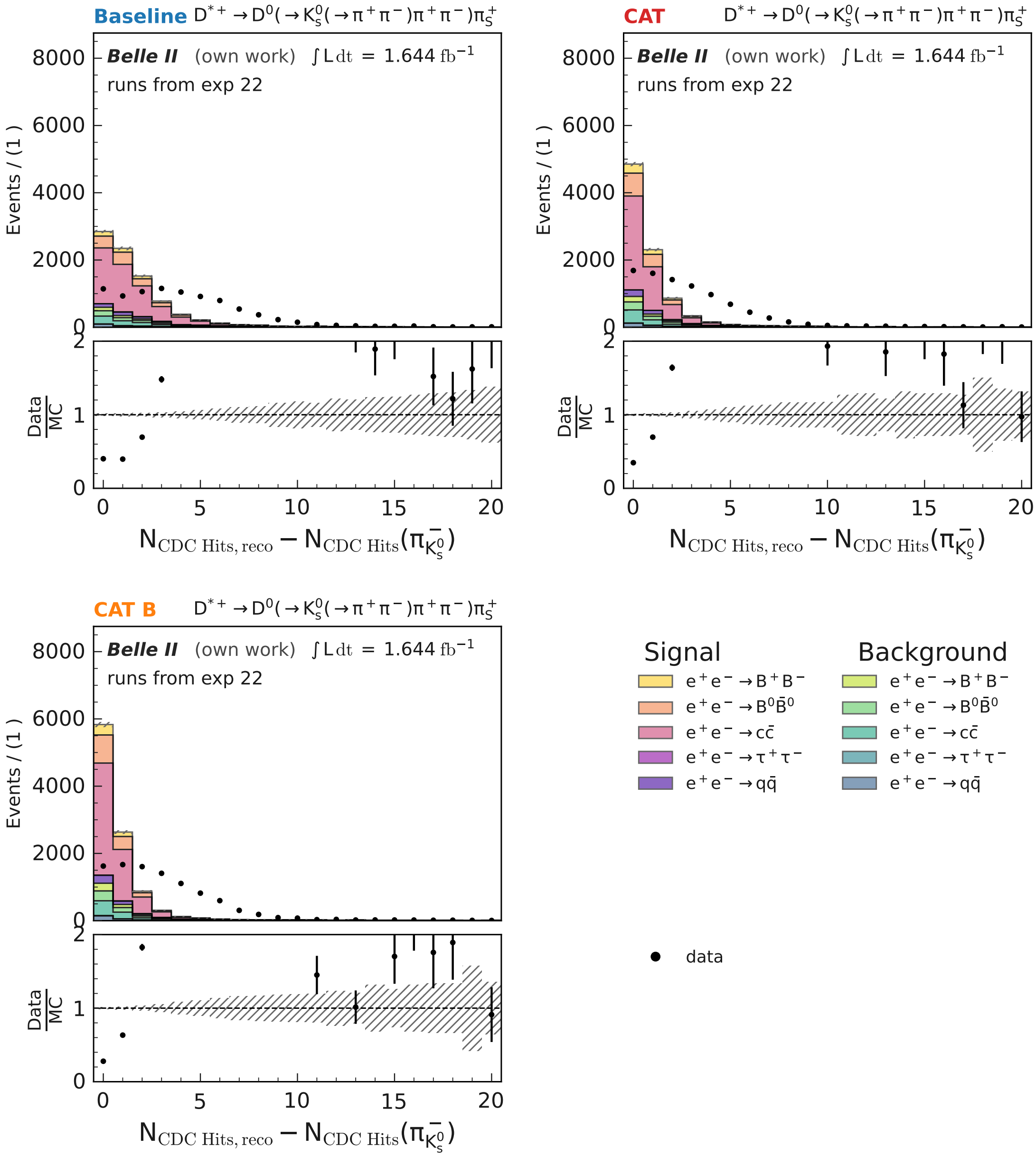


Figure 8.15: Difference of the number of CDC hits $N_{\text{CDC Hits, reco}}$ after the CDC track finding and after the full reconstruction $N_{\text{CDC Hits}}$ for the π^- of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show the difference in hits for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed K_S^0. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.644\,\text{fb}^{-1}$.

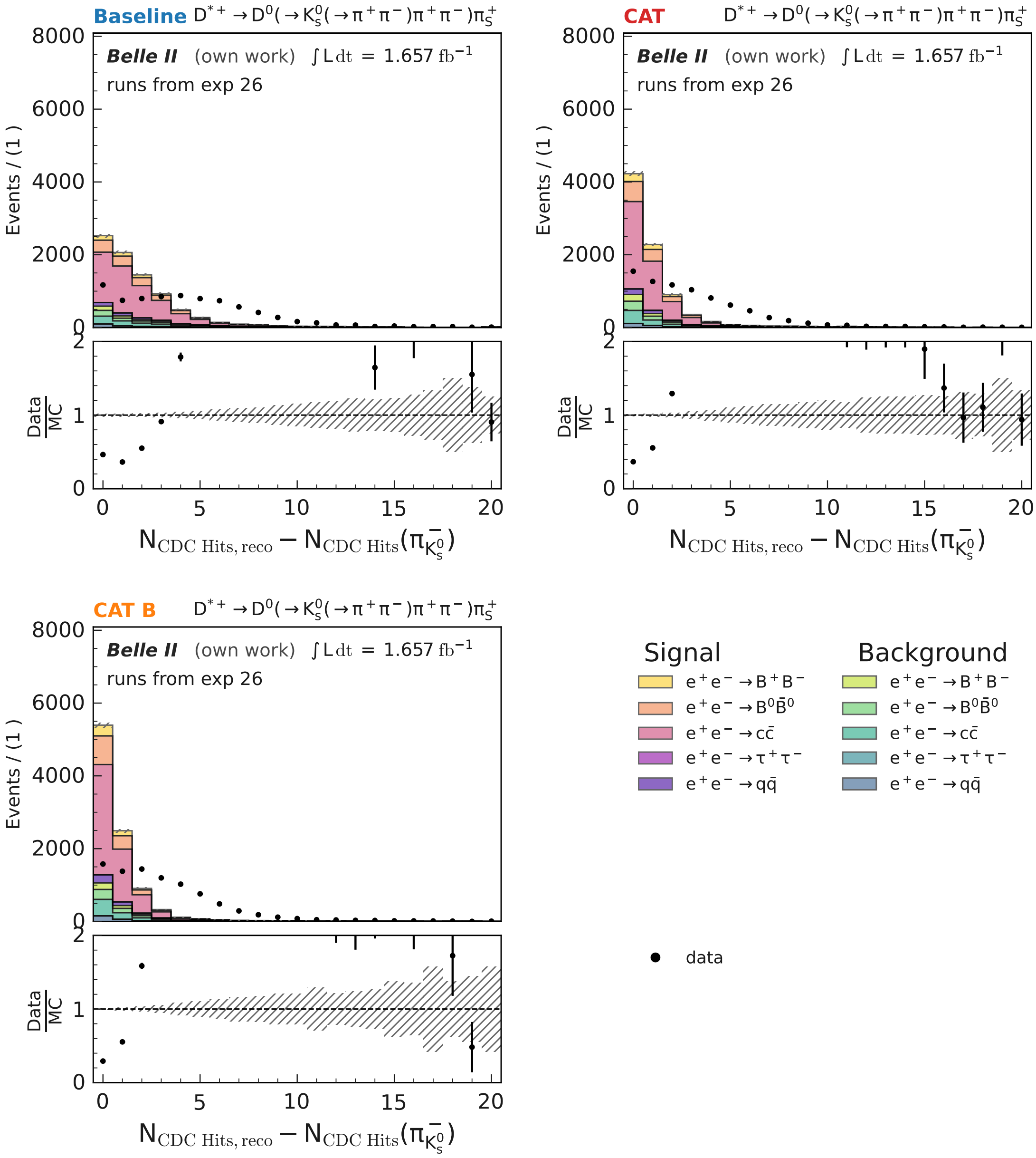


Figure 8.16: Difference of the number of CDC hits $N_{\text{CDC Hits, reco}}$ after the CDC track finding and after the full reconstruction $N_{\text{CDC Hits}}$ for the π^- of the reconstructed $K_S^{\bar{0}}$ in the signal region of 494.8 MeV to 500.8 MeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show the difference in hits for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed K_S^0. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.657\,\mathrm{fb}^{-1}$.

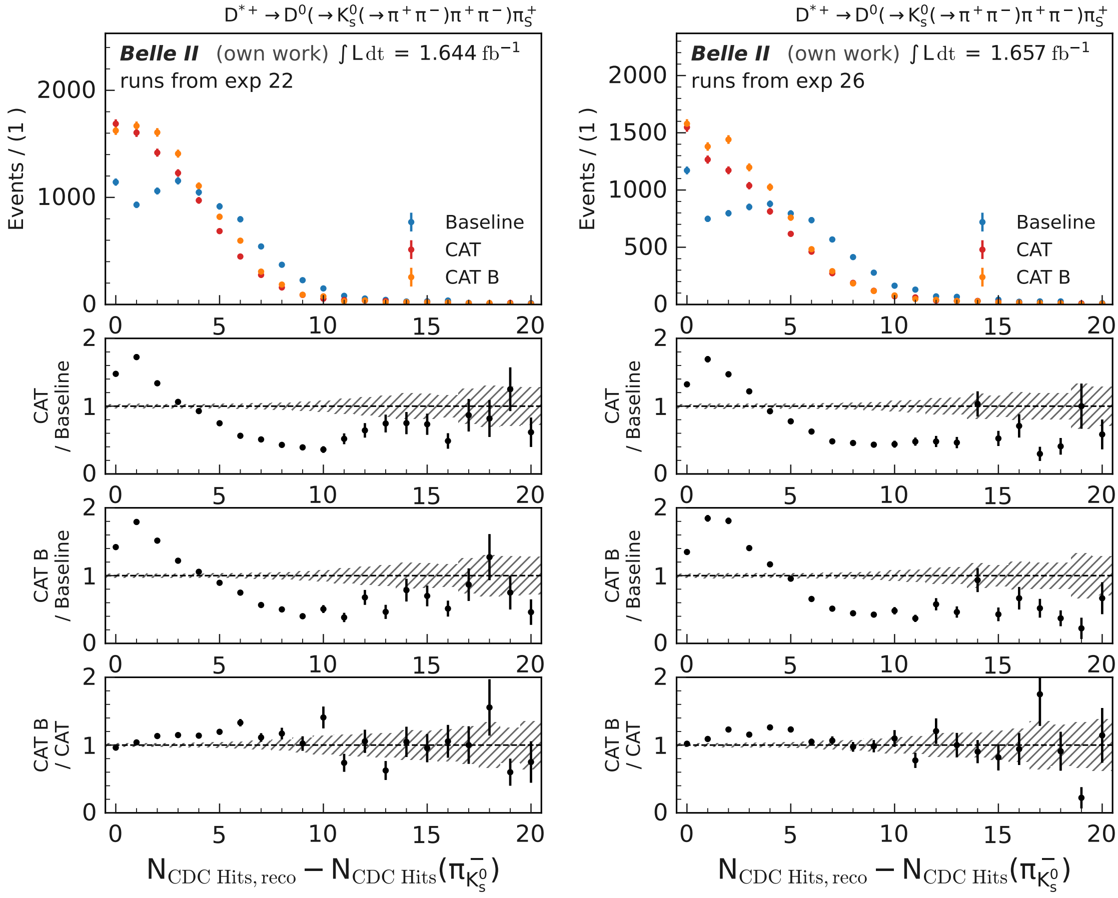


Figure 8.17: Difference of the number of CDC hits $N_{\mathrm{CDC\ Hits,\ reco}}$ after the CDC track finding and after the full reconstruction $N_{\mathrm{CDC\ Hits}}$ for the π^- of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for measured data for runs from experiment 22 (left) and experiment 26 (right). This plot shows the comparison between the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange). The statistical uncertainties between the three algorithms are correlated.

I aim to further investigate why *CAT B Finder* tends to attach fewer hits to tracks in data than in simulation, and why this discrepancy becomes more pronounced at higher beam background levels. A major challenge in this evaluation is that no ground truth is available for measured data, which prevents a direct validation on hit efficiency and hit purity. Since my goal is to improve upon the performance already achieved by *Baseline Finder*, the reconstructed tracks from *Baseline Finder* cannot be used as a baseline for comparison. This limitation is further reinforced by the fact that the discrepancy between simulation and data is already visible in the *Baseline Finder* results, especially for negatively charged particles and under high-background conditions. These observations make a direct comparison of measured data non-trivial and suggest that there may be unaccounted effects in the Belle II simulation. By investigating individual event displays in the signal window $494.8\,\text{MeV} < M_{\pi^+\pi^-} < 500.8\,\text{MeV}$ found by all three tracking algorithms, I observe two major issues:

- Low-momentum tracks are often found by the SVD track finding algorithm instead of the CDC track finding algorithms for *Baseline Finder*. If a track is skipped by the CDC, the SVD can recover it (see Section 7.1.1).
  In Fig. 8.18, I show a measured data event where all three algorithms reconstruct the $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ chain within the signal window $494.8\,\text{MeV} < M_{\pi^+\pi^-} < 500.8\,\text{MeV}$. I focus on the π^+ from the K_S^0. For both *CAT Finder* and *CAT B Finder*, part of the track is found by the respective CDC track finding algorithm. But only the outer part of the loop in superlayers A3, V4, and A5 is reconstructed here. The hits from superlayer A1 and U2, where the track starts, are missed. While this outer part π^+ track is fitted, the limited number of hits reduces the resolution, causing the CKF (see Section 5.4) to fail in extrapolating the CDC track into the SVD to attach SVD clusters.
  The SVD standalone tracking (see Section 5.4) finds the beginning of the π^+ track within the SVD. This SVD π^+ track then gets extrapolated to the CDC with another CKF. The CDC hits at the beginning of the track are assigned to this SVD track, but many hits are already assigned to the π^+ track found previously by the CDC track finding. In the end, the track is found twice by the CDC and SVD track finding, which results in a clone track (see Chapter 4) and leads to poor hit efficiency for either π^+ track.
  In this case, it would have been better not to find the track at all with the CDC track finding algorithms *CAT Finder* or *CAT B Finder*, as *Baseline Finder* did in Fig. 8.18. As mentioned earlier, at least on pion from the K_S^0 decay is reconstructed by the SVD in 38% of cases for the *Baseline Finder*, 41% for the *CAT Finder*, and 31% for the *CAT B Finder*. Optimizing this is challenging, and I propose an approach in Section 10.7.

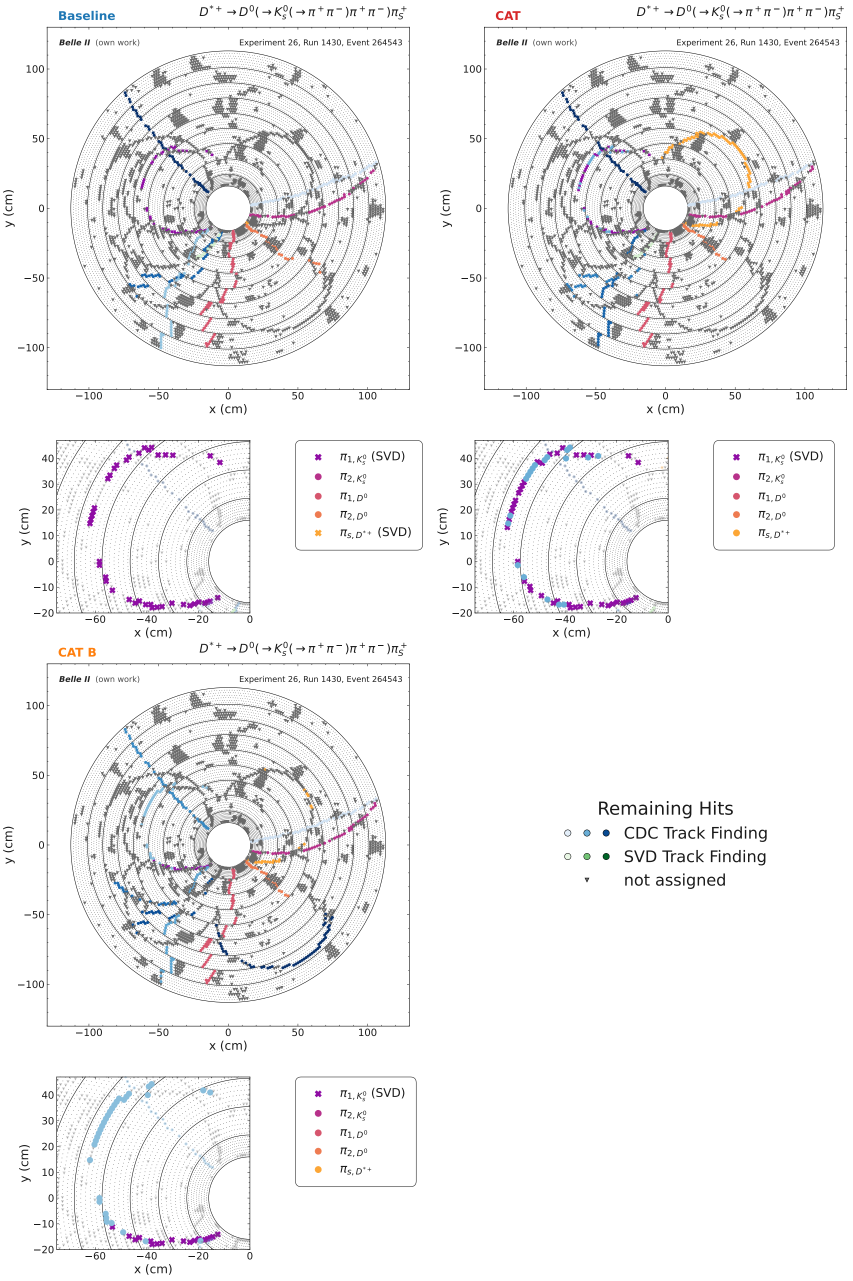


Figure 8.18: An event display showing $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ in the signal window 494.8 MeV $< M_{\pi^+\pi^-} <$ 500.8 MeV reconstructed by all three track finding algorithms. In the plot, tracks shown in the purple to yellow colormap correspond to the reconstructed D^{*+}. Circular markers show that the track is reconstructed by the CDC track finding algorithm, and cross-markers show that the track is reconstructed by the SVD track finding and extrapolated to the CDC, which is also annotated in the respective legend (see text for details). The blue tracks represent the remaining tracks found by the CDC track finding, and green tracks indicate those reconstructed by the SVD track finding and extrapolated back to the CDC. *Baseline Finder* is top left, *CAT Finder* in the top right, and *CAT B Finder* on the bottom. In the cropped, zoomed view of the plot, both *CAT Finder* and *CAT B Finder* show that part of the loop of the π^+_{1,K_S^0} (highlighted in purple) is reconstructed by the CDC track finding (highlighted in blue).

- For *CAT Finder* and *CAT B Finder*, hits from tracks that pass through cross-talk regions (see Fig. 3.1 and Section 3.1) are often completely unassigned. In Fig. 8.19, I highlight such a region, showing that both algorithms miss hits compared to *Baseline Finder*, for example, in the case of the π^- from K_S^0. Even *Baseline Finder* misses hits in these regions, in case of the π_2 from the D^0.

  Cross-talk increases with higher beam-background levels. However, cross-talk is currently not estimated correctly in the beam-background simulation as described in Section 3.1. Instead, beam backgrounds are measured on data (Section 3.1), but in this case, hits are already digitized. This means that the interference between the simulated signal and measured beam background might not be modelled correctly, as for example by wave-form addition of the ADC and TDC values described in Section 2.3.

  As a result, the number of hits associated with the reconstructed tracks is overestimated in simulation, especially at high backgrounds, leading to the discrepancies with measured data shown in Fig. E.4.

  These observations indicate that the simulation does not fully capture beam-background effects, resulting in systematic misalignments between simulation and real data. All input features used in my *CAT Finder* models are already employed in basf2 for multiple MVA modules and were considered safe in previous studies. Nonetheless, the GNN could capture correlations present in the simulation that were previously overlooked, which could result in the observed difference to the *Baseline Finder*. I discuss potential mitigation strategies for this behaviour in Section 10.1.

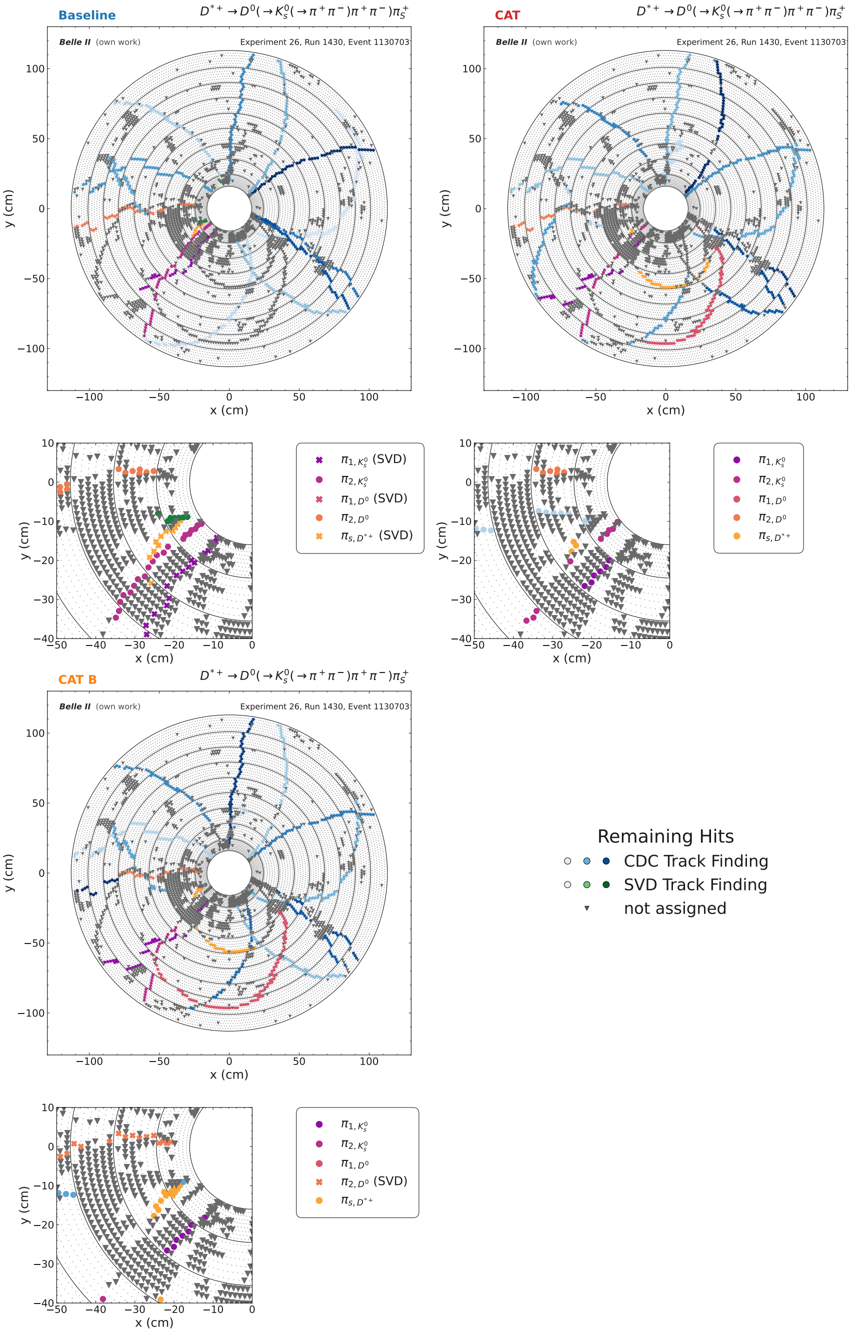


Figure 8.19: An event display showing $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ in the signal window 494.8 MeV $< M_{\pi^+\pi^-} <$ 500.8 MeV reconstructed by all three track finding algorithms. In the plot, tracks shown in the purple to yellow colormap correspond to the reconstructed D^{*+}. Circular markers show that the track is reconstructed by the CDC track finding algorithm, and cross-markers show that the track is reconstructed by the SVD track finding and extrapolated back to the CDC, which is also annotated in the respective legend (see text for details). *Baseline Finder* is top left, *CAT Finder* in the top right, and *CAT B Finder* on the bottom. In the cropped, zoomed view of the plot, a region with cross-talk (Fig. 3.1) is shown.

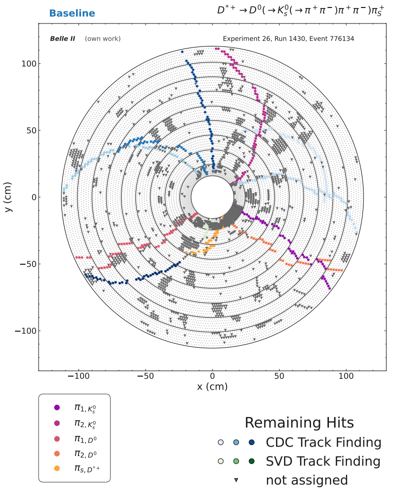
I performed one last investigation by directly examining the latent cluster space of the model output (see Section 5.3.2) for measured data events. I select an event in the signal window $494.8, \mathrm{MeV} < M_{\pi^+\pi^-} < 500.8, \mathrm{MeV}$ where the *Baseline Finder* seems to find all tracks and assigns approximately all hits to them. Then, I evaluate this event with the *CAT Finder* and *CAT B Finder* models, shown in Fig. 8.20.

For the *CAT Finder* cluster space, I see clear seperation between the hits assigned to tracks by *Baseline Finder* and the unassigned hits. However, I observe hits assigned to the tracks, that are far away from the condensation points. These hits form tails extending toward the unassigned hits cluster. These tails suggest that a different clustering algorithm in the latent cluster space, rather than the current fixed-size approach (see Section 5.3.3), could improve assignment. More in-depth studies, especially on the input features of the hits far away from the condensation points, should be performed in the future, but they were out of the scope of my thesis.

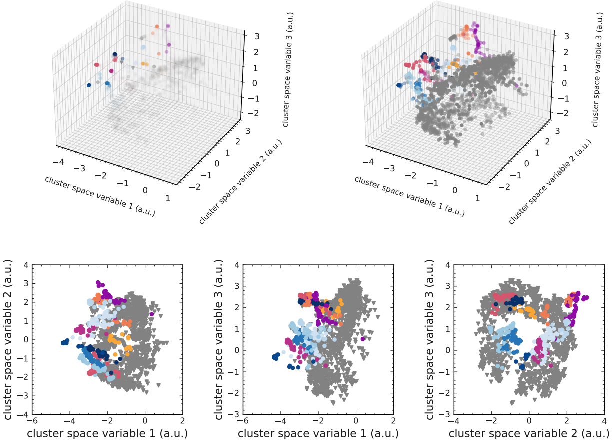
Nevertheless, these tails can overlap between tracks, requiring more work for separating the different tracks.

In contrast, the *CAT B Finder* cluster space, trained on higher multiplicity and different particle types versus the muon-only *CAT Finder*, appears to have less defined structures. More unassigned hits have high β-weights compared to *CAT Finder*, but not enough hits around to be able to form tracks (see Section 5.3.3).

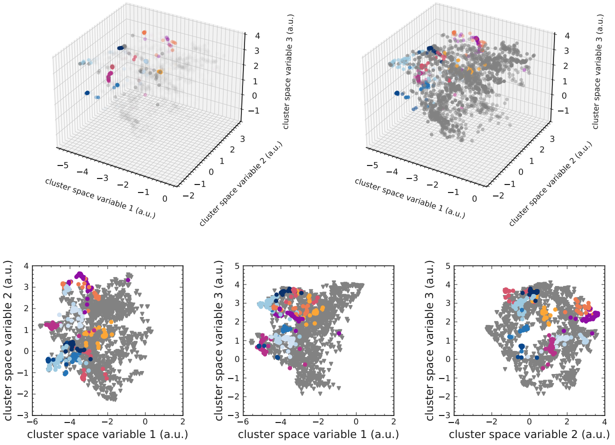
I tested different hit radii up to the cluster radius as well as adjusting the β-value on measured data, but this did not improve results. Exploring alternative clustering methods while monitoring measured data will be the next step, which I discuss in detail in Section 10.1 and Section 10.2.



(a) Event reconstructed by *Baseline Finder*.



(b) *CAT Finder* latent space.



(c) *CAT B Finder* latent space.

Figure 8.20: The event reconstructed by the *Baseline Finder* is shown in Fig. 8.20a (left). This event is also evaluated and given in the latent cluster space of the *CAT Finder* in Fig. 8.20b and *CAT B Finder* in Fig. 8.20c. All plots share the labels and colored hits according to tracks found by the *Baseline Finder*. The latent cluster space is shown with the hits α-value scaled according to the β-weight (see Section 5.3.3), highlighting the condensation points, as well as unscaled hits in the top row, together with the two-dimensional projections in the bottom row.

8.3 Prompt Tracks in Low Multiplicity Events

The last performance check on measured data I perform are prompt tracks in low multiplicity events for $e^+e^- \to \mu^+\mu^-(\gamma)$ and $e^+e^- \to e^+e^-(\gamma)$. This can be directly compared with the prompt tracks in high multiplicity events from the previous section Section 8.1. To calculate the momentum resolution on measured data, non-radiative events are selected by requiring two back-to-back tracks in the cms system. For these selections, the momentum of the muon or electron tracks is around 5 GeV, and the transverse momentum is high. For the hit efficiency discussed for prompt tracks in Section 7.1.1 and Section 8.1, it is also relevant to probe tracks with a smaller transverse momentum. For these, I select muon events with an additional radiated photon, with details given in the following.

8.3.1 Tracking Performance for $e^+e^- \to \mu^+\mu^-(\gamma)$

Event selection

The HLT-selection is given in Section 3.4.1, to differentiate the di-muon final state from Bhabha scattering. I apply further selections to the already pre-selected data that are more stringent. I differentiate, following the HLT selections, between having two muons in the event without a radiated photon $e^+e^- \to \mu^+\mu^-$, and the case where an additional photon is radiated $e^+e^- \to \mu^+\mu^-\gamma$.
The selection for only two muons, required to be back-to-back in the HLT selection in Section 3.4.1, in the final state is given in Table 8.6.

μ selection The μ candidates are selected with their tracks coming from the IP, $|d_0^{IP}| < 2.0$, $|z_0^{IP}| < 4.0$. A minimum cms transverse momentum of $p_t^* > 0.2$ GeV and momentum of $p^* > 0.5$ GeV are required. The polar angle θ of each track is required to be within the CDC barrel acceptance.

$\mu^+\mu^-$ I require exactly two reconstructed tracks in the event with the above without the requirement on the polar angle.

The selection summary is given in Table 8.6.

To select events with the additional radiative photon, the following selections are applied:

Table 8.6: Event selection criteria for $e^+e^- \to \mu^+\mu^-$ in the analysis.

Analysis selection	Selection criteria
μ^+	$p_t^* > 0.2$ GeV $\lvert d^{IP}\rvert < 2.0$ $\lvert z^{IP}\rvert < 4.0$ $p^* > 0.5$ GeV $45^\circ < \theta < 125^\circ$ (CDC barrel acceptance)
$\mu^+\mu^-$	Exactly two tracks in the event

Table 8.7: Selection criteria for $e^+e^- \to \mu^+\mu^-\gamma$ candidates in the analysis.

Analysis selection	Selection criteria
γ	$N_{\text{hits cluster}} \geq 1.5$ $(t_{\text{cluster}} - t_0) < 200\,\text{ns}$ $\lvert r^{\text{ECL}}_{\text{track}} - r^{\text{ECL}}_{\text{cluster}} \rvert > 50\,\text{cm}$ $E_{\text{cluster}} > 1\,\text{GeV}$
μ	$0.05\,\text{GeV} < E_{\text{cluster}} < 0.25\,\text{GeV}$ $17^\circ < \theta < 150^\circ$ $d^{IP} < 2$ and $\lvert z^{IP} \rvert < 4$
$\mu^+\mu^-\gamma$	$9.4\,\text{GeV} < M^{\text{reconstructed}}_{\mu^+\mu^-\gamma} < 10.8\,\text{GeV}$ $\lvert\theta_\gamma - \theta_{\text{recoil}}\rvert < 6^\circ$ $\lvert\phi_\gamma - \phi_{\text{recoil}}\rvert < 6^\circ$ $0.5 < p_\gamma/p_{\text{recoil}} < 1.5$

γ selection The photon candidate is reconstructed with the summed weights of the ECL crystal energy distribution of at least 1.5 and an energy greater than 1 GeV. The ECL cluster timing has to be within 200 ns of the event timing. The nearest track to the photon cluster has to be farther away than 50 cm.

μ selection The muon candidates are required to have low energy, with $0.05\,\text{GeV}E < 0.25\,\text{GeV}$, deposited in the matched ECL clusters, and the θ of the tracks has to be within the CDC acceptance, and they are required to come from the IP.

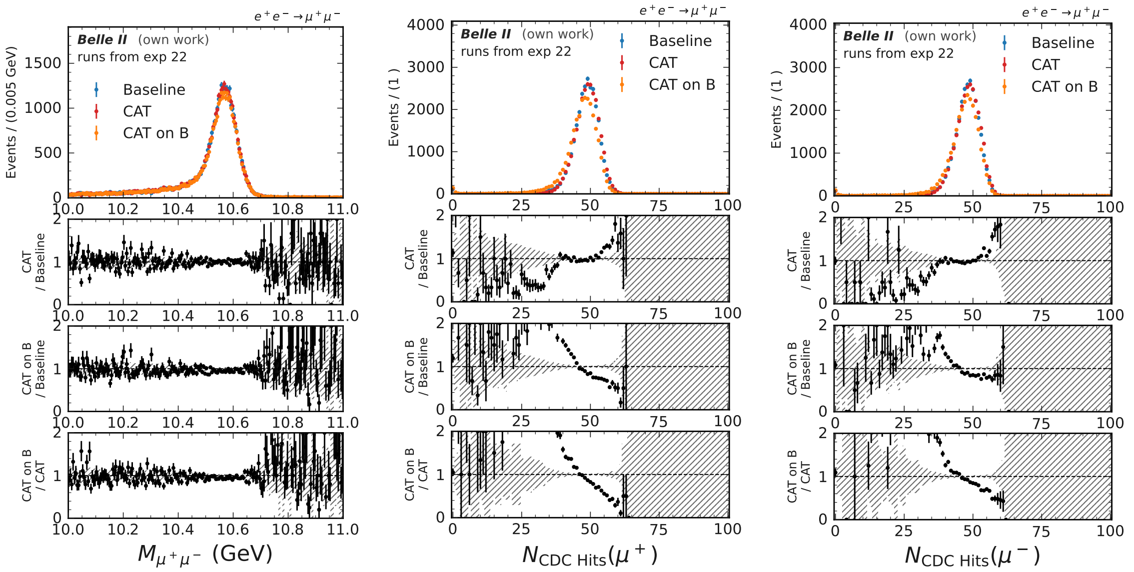
$\mu^+\mu^-\gamma$ candidate The $\mu^+\mu^-\gamma$ should have a reconstructed mass $M^{\text{reconstructed}}_{\mu^+\mu^-\gamma}$ close to the beam energy within 9.4 GeV to 10.8 GeV, and the difference of the recoil polar and azimuthal angle and the photon should be less than 6°. The photon momentum should be within 50% of the recoil momentum.

The selections are summarized in Table 8.7.

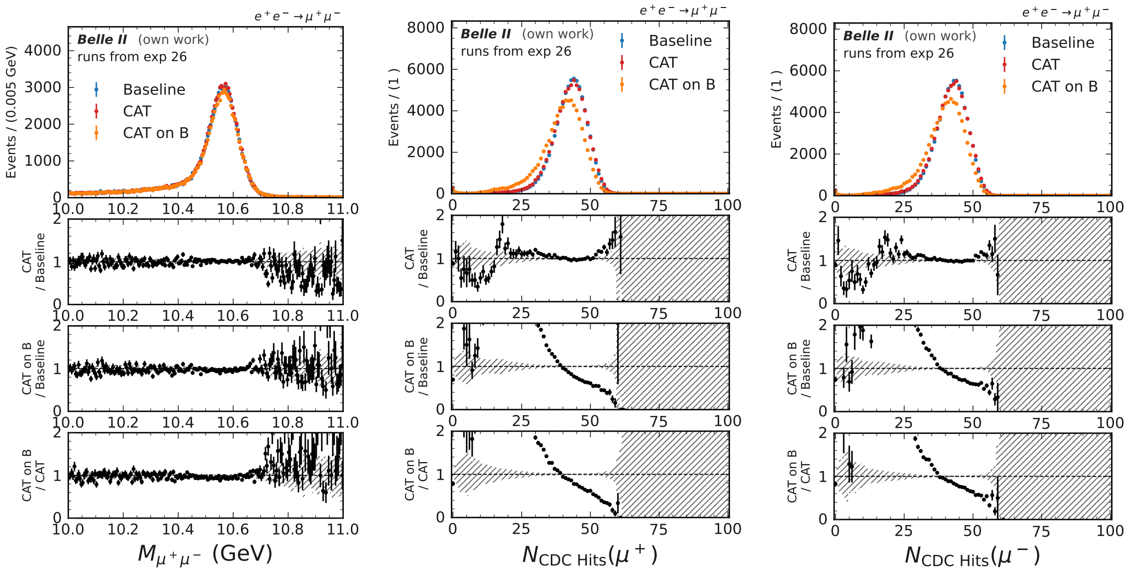
Reconstructed mass and number of hits

I compare the yield on measured data for all three algorithms in Fig. 8.21 for experiment 22 and runs from experiment 26, using the reconstructed mass of the two oppositely charged muons. While *CAT Finder* and *Baseline Finder* agree within uncertainties, the *CAT B Finder* model shows a significantly smaller signal yield at the beam energy. Although *CAT B Finder* performed best for high-multiplicity events, its performance decreases for low-multiplicity events.

When examining the number of CDC hits assigned per track in Figs. 8.21b, 8.21c, 8.21e and 8.21f, I see that *CAT Finder* now assigns even more hits than *Baseline Finder*, whereas *CAT B Finder* assigns significantly fewer hits compared to the other two algorithms. This trend is consistent across both experiments 22 and 26. While *CAT Finder* had fewer hits attached in high-multiplicity events, it now outperforms *CAT B Finder* in low-multiplicity



(a) Reconstructed mass of the two muons for experiment 22.
(b) Number of CDC hits in the μ^+ track for experiment 22.
(c) Number of CDC hits in the μ^- track for experiment 22.



(d) Reconstructed mass of the two muons for experiment 26.
(e) Number of CDC hits in the μ^+ track for experiment 26.
(f) Number of CDC hits in the μ^- track for experiment 26.

Figure 8.21: $e^+e^- \to \mu^+\mu^-$ comparison on data between *Baseline Finder*, *CAT Finder* and *CAT B Finder* for the reconstructed mass (Fig. 8.21a and Fig. 8.21d) and the number of CDC hits for each the μ^+ (Fig. 8.21b and Fig. 8.21e) and μ^- (Fig. 8.21b and Fig. 8.21e) for experiment 22 (top) and experiment 26 (botton).

scenarios.

I also checked whether this higher hit assignment is due to particle momentum by examining events with a radiative photon in $e^+e^- \to \mu^+\mu^-\gamma$ (Section E.3). The number of hits assigned by *CAT Finder* remains high even for lower transverse momentum.

In conclusion, *CAT Finder* performs very well on low-multiplicity events, which aligns with its original training target. Extending to higher multiplicities and diverse particle types with *CAT B Finder* is challenging and requires careful dataset construction while accounting for discrepancies between simulation and measured data.

Momentum Resolution

With this clean selection of two back-to-back muon tracks for $e^+e^- \to \mu^+\mu^-$, I compare the momentum resolution for muon tracks on measured data. I calculate the difference in transverse cms momentum p_t^* between μ^+ and μ^- as

$$\Delta p_t = \frac{1}{\sqrt{2}}(p_t^*(\mu^-) - p_t^*(\mu^+)), \tag{8.6}$$

as I expect them to be equal. I fit this distribution using a DSCB (Eq. (8.2)), without any parameters fixed, and extract the resolution using the Full Width Half Maximum of the fitted distribution to get the estimate for the resolution FWHM(Δp_t). The results, shown in Fig. 8.22 for experiments 22 and 26, show that the mean μ of the fit is consistent across all three algorithms. The uncertainties for the three algorithms are correlated, as the same measured data is reconstructed. For both experiments, the resolution FWHM(Δp_t) between *CAT Finder* and *Baseline Finder* is compatible within uncertainties. However, *CAT B Finder* shows worse resolution than the other two algorithms, especially for experiment 26. I also observe that the resolution degrades for all three algorithms with increasing beam background, rising from 64.3 MeV to 74.3 MeV for *Baseline Finder*, which is consistent with the wider signal mass peaks for increasing backgrounds I observed for both J/ψ and K_S^0 (see Section 8.1 and Section 8.2).

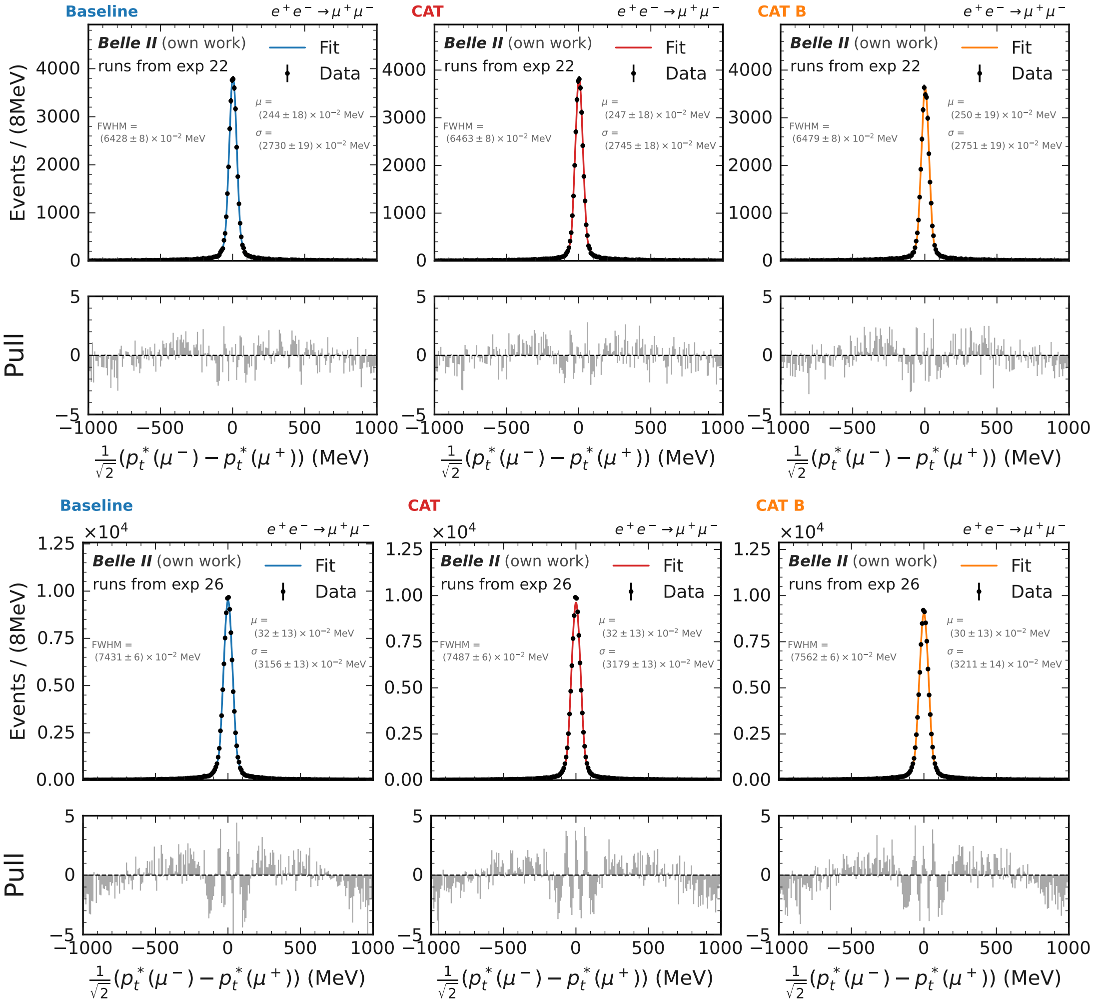


Figure 8.22: The momentum resolution as the difference between the transverse momentum in the center-of-mass frame between the μ^- and μ^+ for the three algorithms *Baseline Finder* (left), *CAT Finder* (middle) and *CAT B Finder* (right) for experiment 22 (top) and experiment 26 (bottom).

PXD and SVD extrapolation confirmation using d_0 and z_0 resolution

In addition, I probe the SVD and PXD extrapolation from the CDC tracks. In Fig. 8.23, I fit the distribution of (Δd_0) following Eq. (8.6) using a DSCB (Eq. (8.2)) and calculate FWHM(Δd_0), using the transverse distance from the interaction point $d_0^{\mathrm{IP}*}$. Similarly, I fit the distribution of Δz_0 and calculate FWHM(Δz_0) for the longitudinal coordinate z_0^{IP} in Fig. 8.24. Both of these metrics probe first the extrapolation to the SVD, followed by the correct PXD cluster assignment. I observe no significant differences, and all three algorithms are in good agreement, proving the implementation in the tracking chain.

---

[*] The interaction point is measured and calibrated using the official Belle II calibration, which relies on the *Baseline Finder*, and has not been recomputed for both *CAT Finder* algorithms.

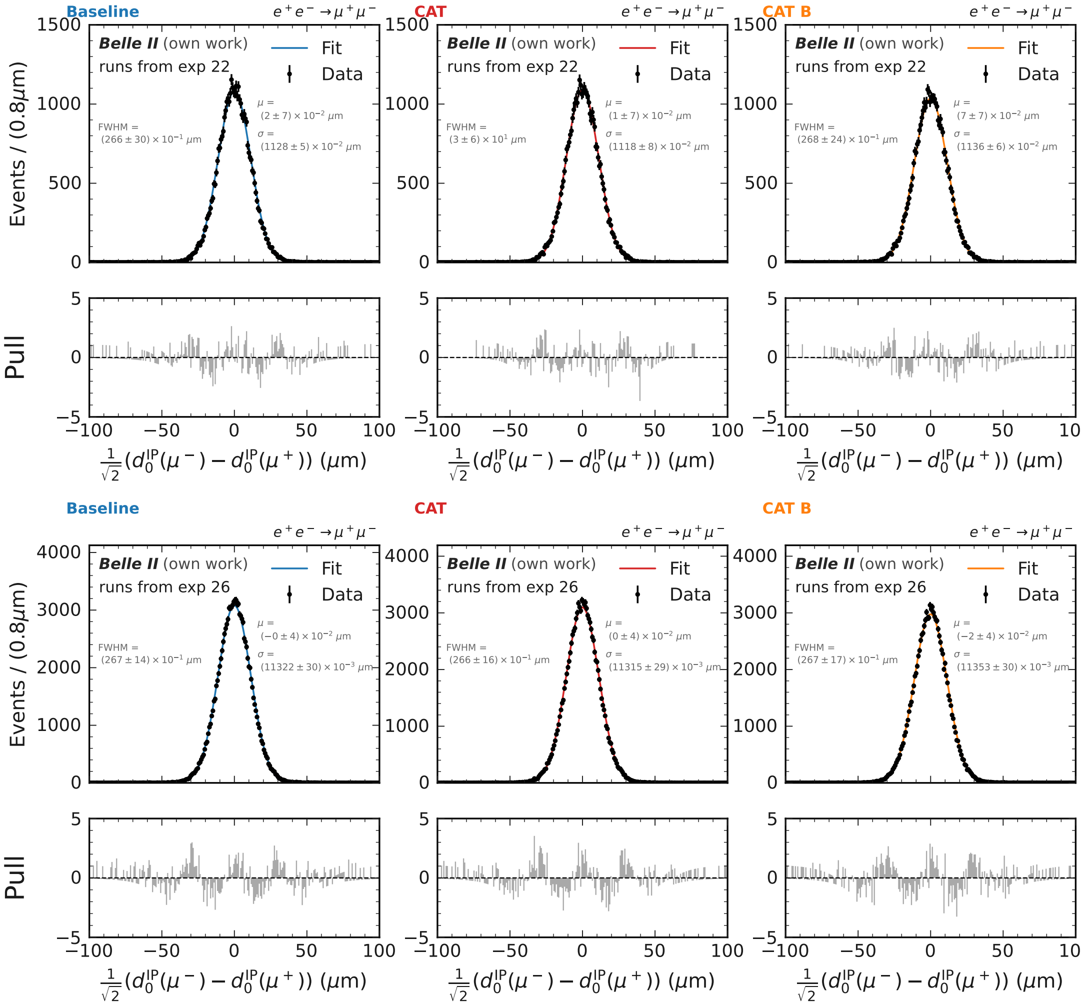


Figure 8.23: The d_0 difference of the μ^- and μ^+ for the three algorithms *Baseline Finder* (left), *CAT Finder* (middle) and *CAT B Finder* (left) for experiment 22 (top) and experiment 26 (bottom).

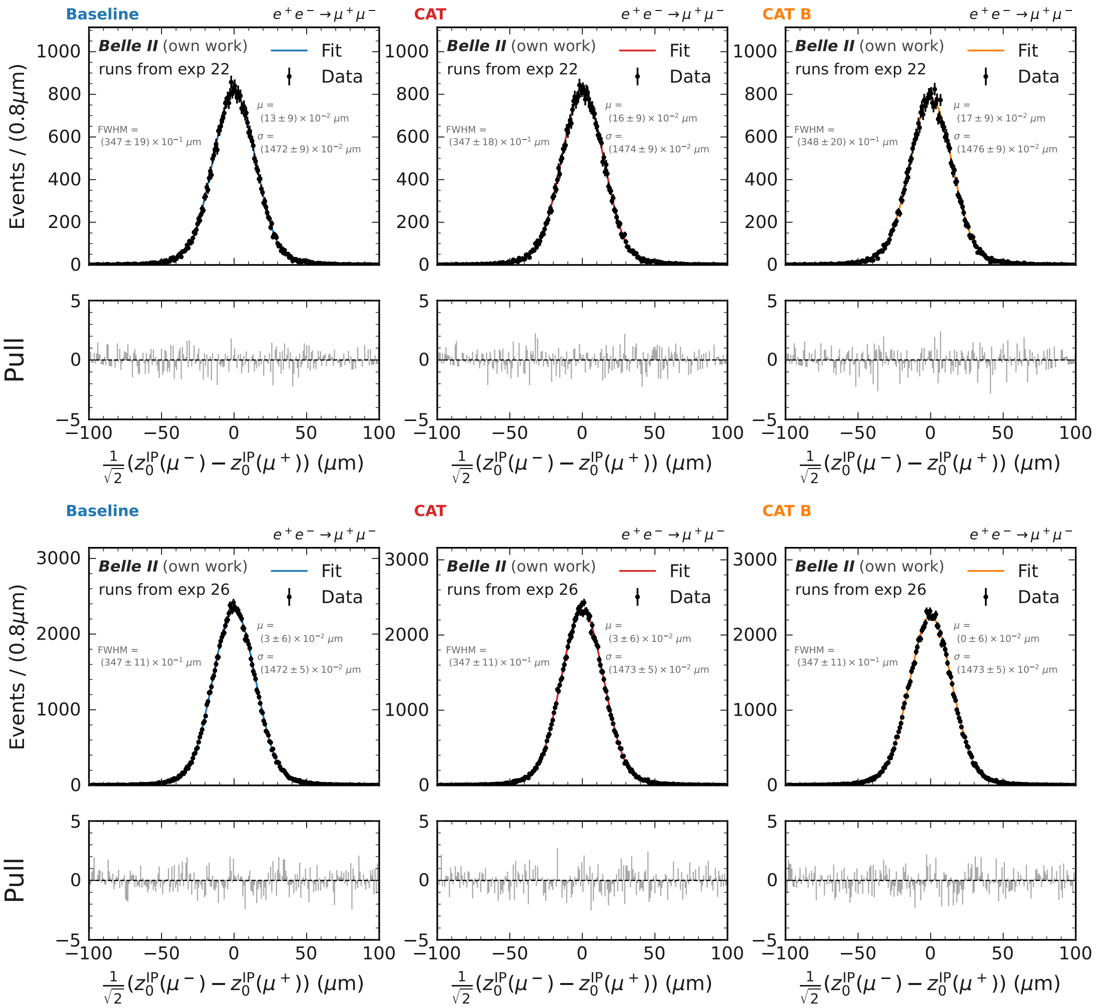


Figure 8.24: The z_0 difference of the μ^- and μ^+ for the three algorithms *Baseline Finder* (left), *CAT Finder* (middle) and *CAT B Finder* (left) for experiment 22 (top) and experiment 26 (bottom).

8.3.2 Tracking Performance for $e^+e^- \to e^+e^-(\gamma)$

In addition to the muons in low multiplicity events, I also want to compare the track finding reconstruction of the three algorithms following Section 8.1 for electrons. For this, I select Bhabha scattering events, to check the reconstructed mass and resolution on electrons too.

Event selection

The HLT-selection is given in Section 3.4.2. A similar selection as for the $\mu^+\mu^-$ events in the previous section is applied for the analysis of Bhabha scattering. The selection is given in the following:

e selection The e candidates are selected with their tracks coming from the IP, with $|d_0^{IP}| < 2.0$, $|z_0^{IP}| < 4.0$.A minimum cms transverse momentum of $p_t^* > 0.2\,\text{GeV}$ and momentum of $p^* > 0.5\,\text{GeV}$ are required. The polar angle θ of each track is required to be within the CDC barrel acceptance.

e^+e^- candidate I require exactly two reconstructed tracks in the event with the above without the requirement on the polar angle. The Bhabha selection for HLT includes radiative photons in comparison to the $e^+e^- \to \mu^+\mu^-$ selection. As I want to calculate the momentum resolution for the electrons as for the muons in the previous section, I require both electrons to be back-to-back in azimuthal and polar angle as for the μ in the HLT selection. This excludes a high-energy radiative photon in the final state.

The selection is summarized in Table 8.8.

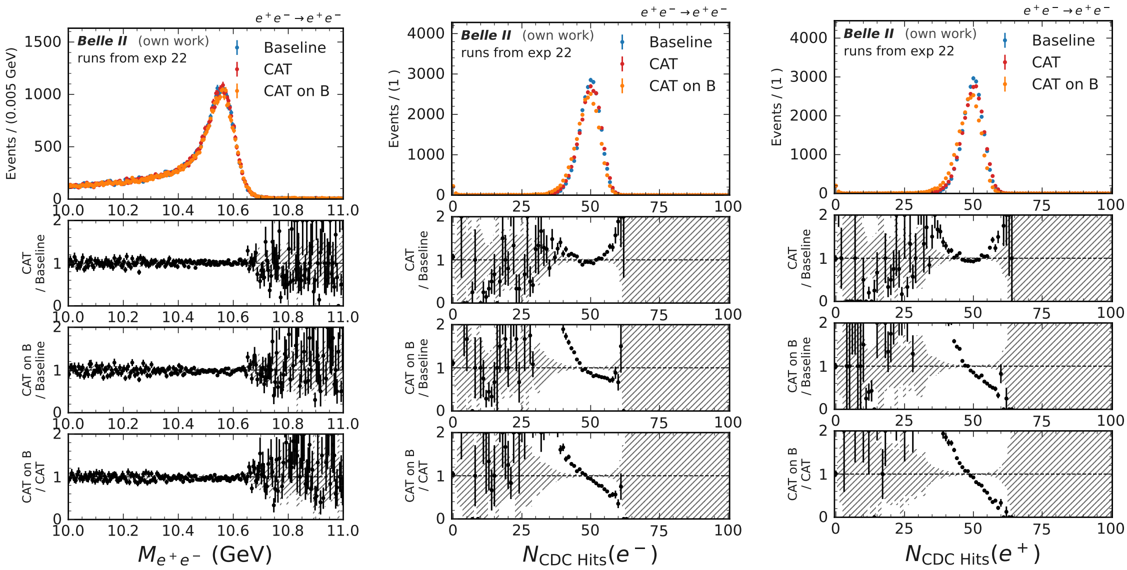
Reconstructed mass and number of hits

I repeat the previous evaluation for the $e^+e^- \to e^+e^-(\gamma)$ scattering and compare the peak yield of the collision energy in the cms frame for all three algorithms in Fig. 8.25. I observe the same behaviour as in the $\mu^+\mu^-$ final state: the *CAT B Finder* model underperforms on measured data compared to *CAT Finder* and *Baseline Finder*. The difference in the number of CDC hits assigned to tracks is also consistent with the previous results, where

Table 8.8: Event selection criteria for $e^+e^- \to e^+e^-$ in the analysis.

Analysis selection	Selection criteria
e	$p_t^* > 0.2\,\text{GeV}$ $\lvert d^{IP}\rvert < 2.0$ $\lvert z^{IP}\rvert < 4.0$ $p^* > 0.5\,\text{GeV}$ $45^\circ < \theta < 125^\circ$ (CDC barrel acceptance)
e^+e^-	Exactly two tracks in the event $\lvert\lvert\phi_1 - \phi_2\rvert - 180^\circ\rvert < 10^\circ$ $\lvert\lvert\theta_1 + \theta_2\rvert - 180^\circ\rvert < 10^\circ$

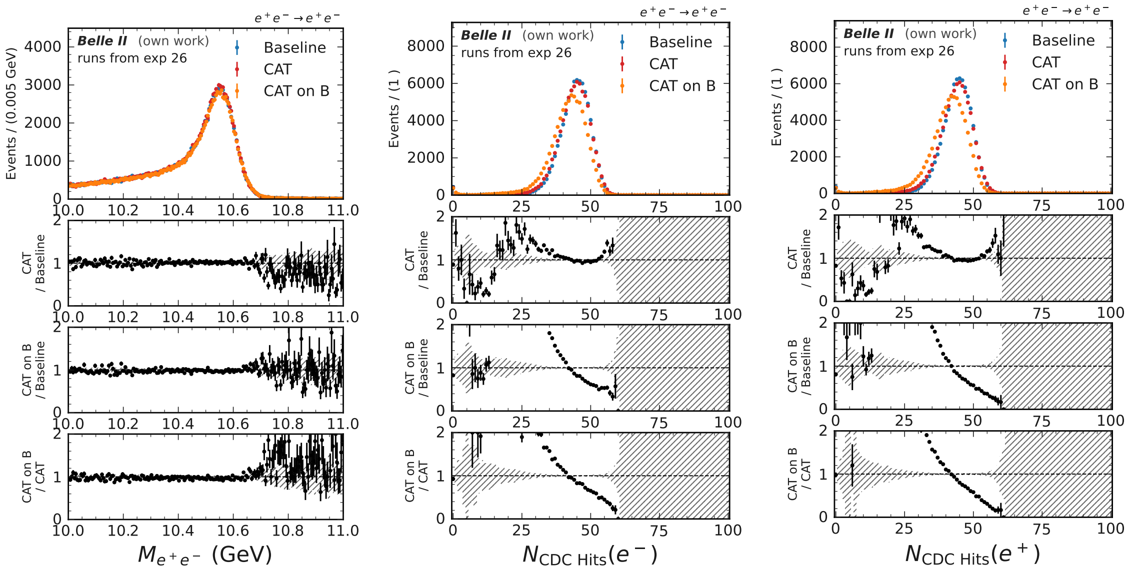
the *CAT Finder* assigns more hit and the *CAT B Finder* less than the *Baseline Finder*. The resolution for the transverse momentum and the SVD and PXD extrapolation show consistent results with the $\mu^+\mu^-$ final state and are given in Section E.4.



(a) Reconstructed mass of the two electrons for experiment 22.

(b) Number of CDC hits in the e^+ track for experiment 22.

(c) Number of CDC hits in the e^- track for experiment 22.



(d) Reconstructed mass of the two electrons for experiment 26.

(e) Number of CDC hits in the e^+ track for experiment 26.

(f) Number of CDC hits in the e^- track for experiment 26.

Figure 8.25: Bhabha comparison on data between *Baseline Finder*, *CAT Finder* and *CAT B Finder* for the reconstructed mass (Fig. 8.25a and Fig. 8.25d) and the number of CDC hits for each the e^+ (Fig. 8.25c and Fig. 8.25f) and e^- (Fig. 8.25b and Fig. 8.25e) for experiment 22 (top) and experiment 26 (botton).

8.4 Data Validation Summary

These findings conclude my analysis of the measured data. In this chapter, I have presented studies of the GNN-based tracking algorithms (*CAT Finder*, *CAT B Finder*) in comparison to the established *Baseline Finder* algorithm, using a variety of benchmark processes on measured data. The evaluation covers prompt tracks in both low- and high-multiplicity track environments, as well as displaced tracks in high-multiplicity track environments, with probing electrons, muons and pions, thereby providing a comprehensive assessment of tracking performance.

For high-multiplicity events, the reconstruction of prompt J/ψ decays shows that both *CAT Finder* and *CAT B Finder* can match or even surpass the performance of *Baseline Finder*. In the case of displaced K_S^0 candidates, however, only the *CAT B Finder* model shows improved performance, while *CAT Finder*, trained on events with on average 4.5 tracks, yields higher signal counts for the runs from experiment 22 but suffers from increased background, resulting in lower overall significance. In particular, *CAT B Finder* achieves the highest signal yields for both J/ψ and K_S^0, especially in challenging high beam-background scenarios, while maintaining comparable or improved significance. Nonetheless, all algorithms exhibit increased background and reduced signal in measured data compared to simulation in the high beam-background scenario, highlighting persistent mismodeling in the simulation.

The analysis of CDC hit assignment reveals that *CAT B Finder*, while highly effective in simulation, attaches about ten fewer CDC hits per track in measured data. More generally, all algorithms show a shift toward lower hit counts in data compared to simulation for high beam-background conditions.
Additional challenges arise in the combination of CDC and SVD track finding, where partially reconstructed low-momentum tracks in the CDC often fail to merge with SVD clusters. This makes it preferable not to reconstruct them with the CDC track finding algorithms at all. Furthermore, unassigned hits in cross-talk regions are more pronounced for the GNN-based algorithms, highlighting areas that require further improvement.

In low-multiplicity events, such as $e^+e^- \to \mu^+\mu^-(\gamma)$ and $e^+e^- \to e^+e^-(\gamma)$ scattering, *CAT Finder* and *Baseline Finder* deliver similar signal yields and resolution, with *CAT Finder* even assigning more CDC hits per track. In contrast, *CAT B Finder*, despite its strong performance in high-multiplicity environments, underperforms in low-multiplicity measured data, showing a significant drop in signal yield and fewer assigned hits. Resolution studies for momentum, d_0, and z_0 confirm that all algorithms achieve similar resolutions, except for *CAT B Finder* in high beam-background levels, where the performance decreases. These trends are observed consistently across both muon and electron final states.
Overall, the *CAT Finder* model proves robust across many event topologies and background

conditions, particularly in low-multiplicity data, but it does not reach the performance of *Baseline Finder* for displaced tracks below 8 cm in high beam-background scenarios. The *CAT B Finder* model, while outperforming both algorithms in simulation and yielding higher signals in high-multiplicity data, underperforms in low-multiplicity events and shows reduced hit assignment. These limitations, together with the discrepancies between data and simulation, indicate that both *CAT B Finder* and the detector simulation require further improvement. Despite its weaker performance in high-multiplicity events, I will focus on the *CAT Finder* model going forward, as it models measured data more reliably than *CAT B Finder*, which still needs further development, and will be discussed in Chapter 10.

Chapter 9

Sensitivity Study for $\boldsymbol{B^+ \to K^+ S}$

In this chapter, I compare the performance of the *CAT Finder* with the *Baseline Finder* in a sensitivity study of a long-lived spin-0 particle S in the decay $B^+ \to K^+ S$. This is described in Section 3.5.2, where the S can act as a mediator between the SM and DM with Yukawa-like couplings similar to the Higgs boson, leading to the radiation of S from heavy virtual particles. These couplings are smaller than those of a SM Higgs with the same mass and can lead to a displaced decays of S.

A search for this process was carried out in [13, 44] at the beginning of run I with an integrated luminosity of $\int L dt = 189\,\text{fb}^{-1}$. In this analysis, lifetimes from $10\,\mu\text{m}$ up to $4\,\text{m}$ for the final states that I consider were probed. One of the main conclusions of this search was that long-lived particles remain challenging at Belle II, as the standard tracking is not optimized for decay vertices displaced beyond the SVD. Improving reconstruction for these displaced particles is a main motivation for my new GNN-based algorithm. Therefore, I will probe lifetimes up to $100\,\text{m}$.

In contrast to the displaced vertices in low-multiplicity events discussed in Section 6.1.2, where I already showed superior performance at large displacements [12], this analysis tests the *CAT Finder* in a high-multiplicity $\Upsilon(4S) \to B\bar{B}$ environment, combined with displaced vertex signatures.

Following the approach of [13, 44] (summarized in Section 3.5.2), I analyse $B^+ \to K^+ S$ with $S \to \mu^+\mu^- / \pi^+\pi^- / K^+K^-$. The events are simulated with low simulated beam backgrounds and the event selection is given in Section 9.1. Following the selection, I present the signal efficiency in Section 9.2 and discuss the background estimation in Section 9.3. As in Chapter 8, systematic uncertainties and correction factors are not included, since obtaining them for the *CAT Finder* is beyond the scope of this thesis. While I cannot include them, the dominant systematics will be discussed in Section 9.4 when determining the expected upper limits. These results will be used for the model interpretations given in Section 9.5.

9.1 Event Selection

The event selection and reconstruction follows [44].

Displaced track selection: The S candidate is reconstructed from two oppositely charged tracks, with the decay vertex determined through a kinematic fit. Each track must have a transverse momentum $p_t > 0.25\,\text{GeV}$ and an extrapolated trajectory within the barrel region of the ECL, requiring $\theta^{\text{ECL}} > 30^\circ$, and be within the backward region of the CDC, requiring $\theta^{\text{CDC}} < 150^\circ$. Particle identification for the final-state muons and pions is performed using a combined likelihood variable, given the muon example

$$\text{PID}_{(\mu,\,\pi,\,e)}(\mu) = \frac{\mathcal{L}_\mu}{\mathcal{L}_e + \mathcal{L}_\mu + \mathcal{L}_\pi},$$

which must be greater than 0.4 for both tracks forming the S candidate. For kaon final states, the global PID^{K^+} (Eq. (7.5)), is required to be greater than 0.2.

***S* candidate:** The invariant S mass $M_{S\to x^+x^-}$ is calculated from the two charged tracks. To suppress SM background, the decay vertex displacement is required to satisfy $d_r > 0.05\,\text{cm}$. In resonant peaking regions of $M_{S\to x^+x^-}$ given in Table 9.2, this condition is tightened to $d_r > 0.2\,\text{cm}$. The significance of the displacement, defined as the ratio of d_r to its uncertainty given by the vertex fit, must exceed 3. Furthermore, the S decay direction must point back to the interaction point, enforced by requiring the cosine of the angle between the momentum and vertex of S in the $x-y$-plane $2d\ \cos\theta^{\text{S}}_{\text{point}} > 0.95$ for the muon and kaon final state and larger than 0.99 for the pion final state. Furthermore, events are vetoed if the invariant mass of the two tracks forming the S candidate, calculated under the pion mass hypothesis, falls within the K^0_S mass window of 0.489 GeV to 0.507 GeV.

Kaon candidate: The kaon track used to form the B candidate must satisfy the same PID requirement as for the displaced tracks with the global $\text{PID}^{K^+} > 0.2$. Additionally, it must be prompt, with transverse and longitudinal impact parameters satisfying $d_{\text{IP}} < 0.5\,\text{cm}$ and $|z_{\text{IP}}| < 2\,\text{cm}$, respectively, and must have $p_t > 0.15\,\text{GeV}$.

***B* candidate:** The B^+ candidate is reconstructed by combining the S and K^+ candidates. In the center-of-mass frame, the beam-constrained mass is defined as

$$M_{bc} = \sqrt{\frac{s}{4} - |p^*_B|^2},$$

where p^*_B is the reconstructed momentum of the B candidate, with s being the cms energy squared. The energy difference is defined as

$$\Delta E = E^*_B - \sqrt{s}/2,$$

Table 9.1: Event selection for $B^+ \to K^+ S$ with $S \to \mu^+\mu^-$, $S \to \pi^+\pi^-$ and $S \to K^+K^-$ taken from [44].

Particle	$S \to \mu^+\mu^-$	$S \to \pi^+\pi^-$	$S \to K^+K^-$
K^+	$\lvert z_{\mathrm{IP}}\rvert < 2\,\mathrm{cm}$ $\lvert d_{\mathrm{IP}}\rvert < 0.5\,\mathrm{cm}$ $p_t > 0.15\,\mathrm{GeV}$ $\mathrm{PID}(K^+) > 0.2$		
S	$d_r > 0.05\,\mathrm{cm}$ and $d_r > 0.2\,\mathrm{cm}$ for Table 9.2		
	2d $\cos\theta^{\mathrm{S}}_{\mathrm{point}} > 0.95$	2d $\cos\theta^{\mathrm{S}}_{\mathrm{point}} > 0.99$	2d $\cos\theta^{\mathrm{S}}_{\mathrm{point}} > 0.95$
	Sign. of distance > 3 $M_{\pi^+\pi^-} \notin (0.498, 0.507)$ GeV		
S tracks	$32^\circ < \theta^{\mathrm{ECL}}$ and $\theta^{\mathrm{CDC}} < 150^\circ$ $p_t > 0.25\,\mathrm{GeV}$		
	$\mathrm{PID}(\mu, \pi, e) > 0.4$		$\mathrm{PID}(K^+) > 0.2$
B^+	$\Delta M_{bc} > 5.27\,\mathrm{GeV}$		
	$\lvert\Delta E\rvert < 0.05\,\mathrm{GeV}$	$\lvert\Delta E\rvert < 0.035\,\mathrm{GeV}$	$\lvert\Delta E\rvert < 0.05\,\mathrm{GeV}$
Cont. sup.	$R2 < 0.45$	$R2 < 0.35$	$R2 < 0.45$

Table 9.2: Peaking backgrounds in $M_{S\to x^+x^-}$ in GeV where a tighter vertex d_r selection is applied, taken from [44].

Source	$S \to \mu^+\mu^-$	$S \to \pi^+\pi^-$	$S \to K^+K^-$
D^0	(1.70, 1.80)	(1.65, 1.85)	(1.75, 1.95)
J/ψ	(3.00, 3.15)	-	-
$\psi(2S)$	(3.65, 3.75)	-	-
η_c	-	(2.85, 3.15)	(2.80, 3.20)
$\chi_{c1}\eta_c(2S)$	-	(3.40, 3.80)	
ϕ	-	-	(1.00, 1.04)

where E^*_B is the energy of the B candidate in the center-of-mass frame. The selection requires $|\Delta E| < 0.05\,\mathrm{GeV}$ for muon and kaon final states and $|\Delta E| < 0.035\,\mathrm{GeV}$ for pion final states.

Continuum suppression: To suppress continuum background, the ratio of the second to the zeroth Fox–Wolfram moment[111, 112], R_2, is required to be less than 0.45 for muon and kaon final states, and less than 0.35 for pion final states.

The event selection is summarized in Table 9.1.

9.2 Signal Efficiency

Following the selections in the previous section, the signal efficiency is shown in Fig. 9.1. The signal efficiency is determined according to the fit on the signal in Section 9.4, and the

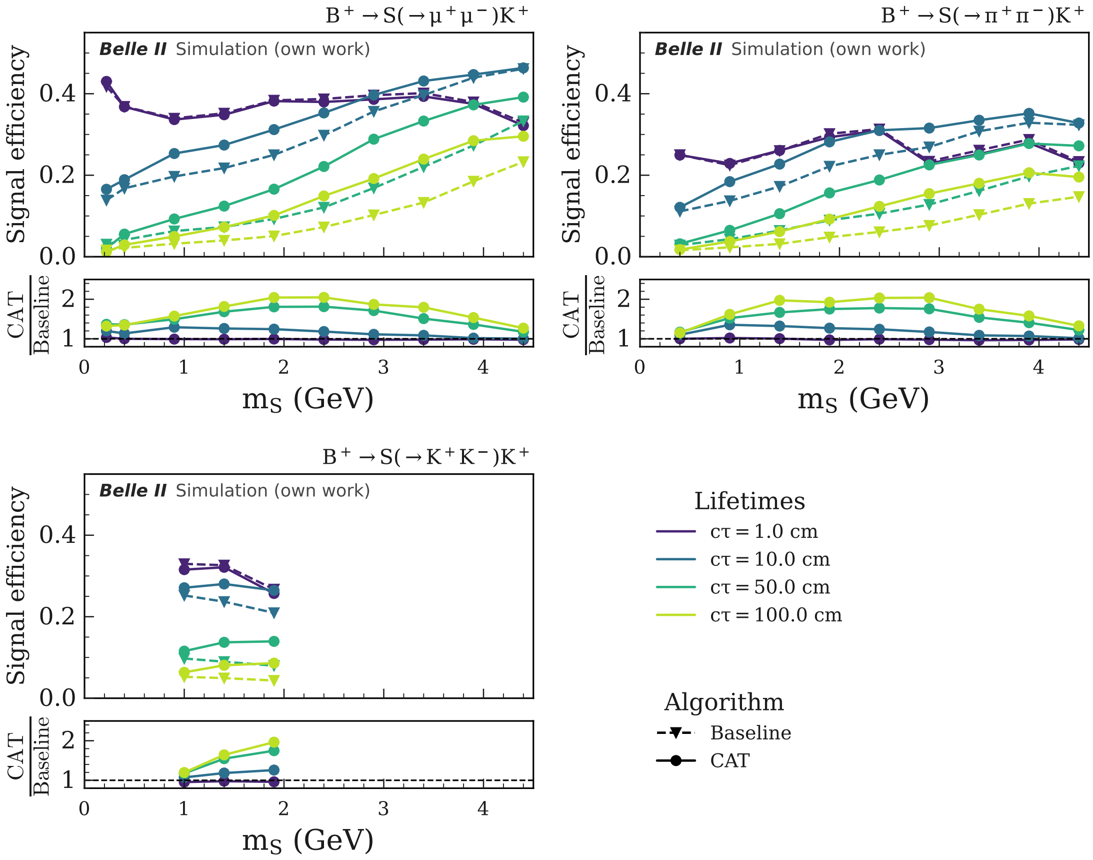


Figure 9.1: Signal efficiency for different lifetimes for the *Baseline Finder* (dashed line with triangular markers) and the *CAT Finder* (solid line with round markers). The lines are added to help guide the eye. The ratio between signal efficiency is given for each final state $S \to \mu^+\mu^-$ (top left), $S \to \pi^+\pi^-$ (top right) and $S \to K^+K^-$ (bottom).

resulting number of events in the mass range. For short lifetimes ($c\tau = 1\,\mathrm{cm}$), the performance of the *Baseline Finder* and *CAT Finder* is comparable, and for $m_S = 2$–$3.5\,\mathrm{GeV}$ the *Baseline Finder* is even slightly better, consistent with previous observations in Chapters 6 to 8 where the *CAT Finder* shows reduced performance in high-multiplicity events. As demonstrated for the *CAT B Finder*, this performance can be further improved by including $B\bar{B}$ events in the training samples, as shown in simulation in Chapter 7, but due to the differences observed in measured data in the previous chapter this requires additional work. At larger lifetimes ($c\tau = 10, 50, 100\,\mathrm{cm}$), however, the *CAT Finder* achieves up to a factor 2 higher efficiency than the *Baseline Finder*, consistently across all three S decay modes. Since higher efficiency alone does not guarantee better sensitivity, the next step is to evaluate the expected background yields.

9.3 Backgrounds

I use centrally produced simulated samples with $\int L dt = 500\,\mathrm{fb}^{-1}$, using a similar software release[*] to the *CAT Finder* implementation. Producing an equivalent sample for the *CAT Finder* is computationally out of scope for this thesis, instead, I generate $\int L dt = 10\,\mathrm{fb}^{-1}$ for *CAT Finder* and compare the background expectation to the centrally produced *Baseline Finder*. For cross-checks, I also compare background yields with the measured data and simulated results of [13], available on ***HEPData*** [113]. These were extracted with an earlier software release[†]. The biggest differences between the earlier and the current software release is the improved CDC simulation and the PID performance. Therefore, I expect different distributions for the simulation.

The observable used for signal extraction is the reduced reconstructed scalar mass,

$$M^{\mathrm{reduced}}_{S \to x^+ x^-} = \sqrt{M^2_{S \to x^+ x^-} - 4m_x^2}, \tag{9.1}$$

with $M^2_{S \to x^+ x^-}$ the reconstructed mass and m_x the mass of the daughters according to [37]. The reduced mass corresponds to twice the daughter momentum in the S rest frame, and is chosen to simplify the modelling close to the threshold given by the daughter masses. As the mass of the daughters is known with a higher precision than the reconstructed momentum, this introduces non-linear dependencies for the width of the mass distribution close to twice the daughters mass. Using the reduced mass instead results in an approximate linear behaviour.

The comparison of background yields as bins of $M^{\mathrm{reduced}}_{S \to x^+ x^-}$ is shown in Fig. 9.2 for $S \to \mu^+\mu^-$, $S \to \pi^+\pi^-$, and $S \to K^+K^-$.

For $S \to \mu^+\mu^-$, I observe only two events, consistent with the expectation in [13]. The centrally produced *Baseline Finder* background is lower than [13], reflecting differences in simulation versions. For $S \to \pi^+\pi^-$, the opposite trend appears: *Baseline Finder* yields exceed [13], while the *CAT Finder* shows higher background levels but with large statistical uncertainties. For $S \to K^+K^-$, *Baseline Finder* agrees well with [13], whereas the *CAT Finder* again exhibits higher background yields.

These higher background levels for the *CAT Finder* are not unexpected. The validation on $B\bar{B}$ events in Section 5.4 shows that the *CAT Finder* has a higher fake and clone rate than the *Baseline Finder*, which also showed in about 11% more background events in measured data for reconstructed $K^0_S \to \pi^+\pi^-$ decays in Section 8.2. I also demonstrated that this effect can be mitigated by retraining the *CAT Finder* on B-meson events, which is ongoing work but beyond the scope of this thesis.

To account for the different sample sizes, I calculate a scaling factor between the backgrounds

---

[*] Centrally produced samples use `release-08-00-07`, the *CAT Finder* is implemented in `release-08-01-08`.

[†] The software release used in [44] was `release-05`.

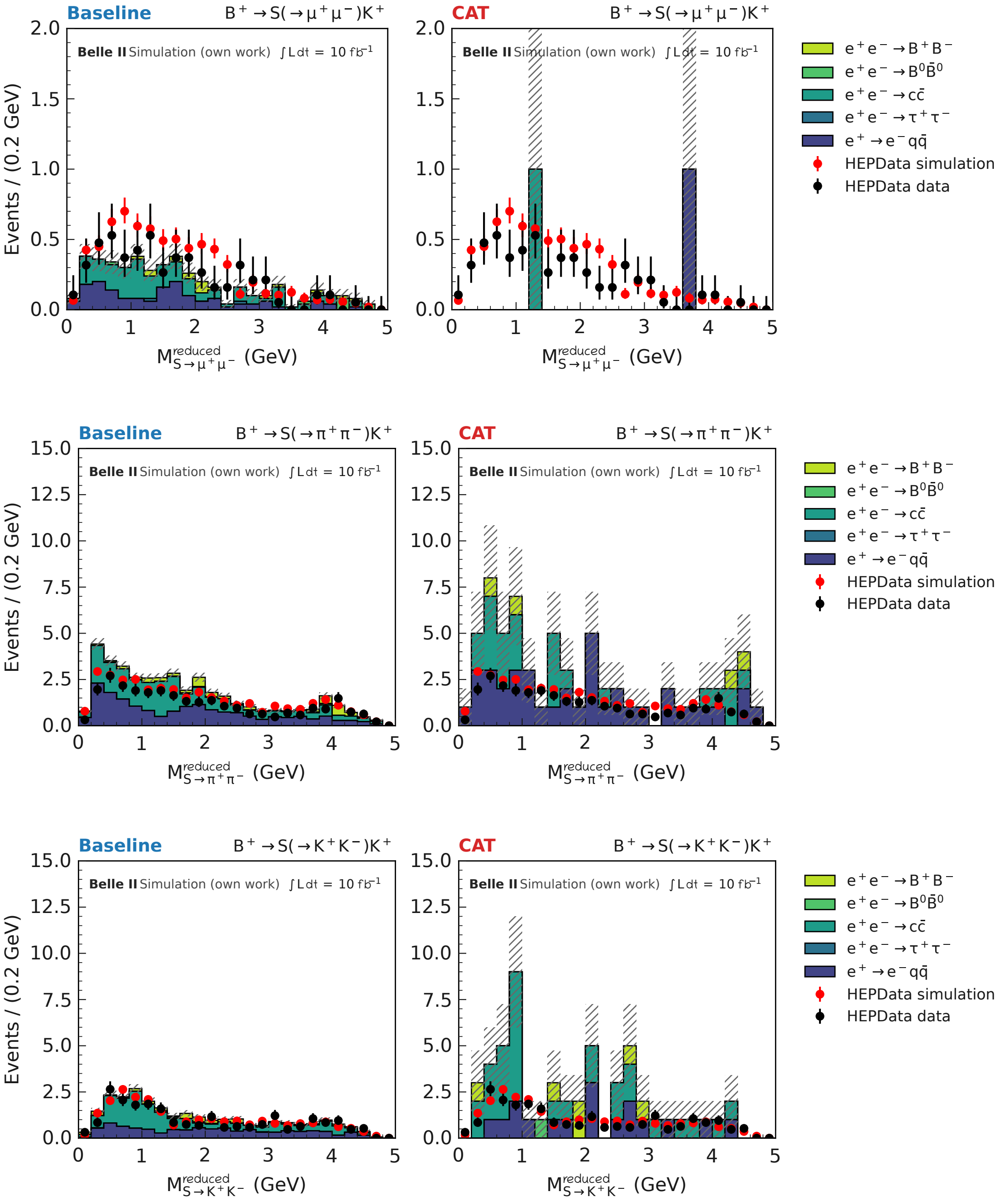


Figure 9.2: Comparison for the backgrounds for the observable $M^{\text{reduced}}_{S\to x^+x^-}$ including all event selections in Table 9.1 for the *Baseline Finder* (left) scaled down from $\int Ldt = 500\,\text{fb}^{-1}$ to $\int Ldt = 10\,\text{fb}^{-1}$ of the *CAT Finder* (right). The results from results of [13], available on ***HEPData*** [113], are overlayed for the simulation in red and for measured data in black, also scaled down to $\int Ldt = 10\,\text{fb}^{-1}$. Top shows $S \to \mu^+\mu^-$, middle $S \to \pi^+\pi^-$ and bottom $S \to K^+K^-$.

of *Baseline Finder* and *CAT Finder*. As I only observe two events for the $S \to \mu^+\mu^-$ final state for *CAT Finder*, I drop the PID requirement on the opposite-charge tracks forming

Table 9.3: Error-weighted average of the binned ratio of the $\int Ldt = 10\,\mathrm{fb}^{-1}$ *CAT Finder* background yields without the PID selections to the scaled down from $\int Ldt = 500\,\mathrm{fb}^{-1}$ *Baseline Finder* background yields.

Decay Channel	$\bar{x}$
$S \to \mu^+\mu^-$	1.40 ± 0.06
$S \to \pi^+\pi^-$	1.35 ± 0.14
$S \to K^+K^-$	1.32 ± 0.05

the S candidate (Table 9.1). This increases the available statistics to obtain an estimate on the scaling factor and results in the distributions in Fig. F.1. Building the ratio x between the scaled down *Baseline Finder* and *CAT Finder* for the binned distribution with i bins, I calculate the scaling factor as the error-weighted average $\bar{x}$ with

$$\bar{x} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \sigma_i^{-2}}. \tag{9.2}$$

Here, σ_i refers to the uncertainty on the ratio x_i. The statistical uncertainty on this error-weighted average is given with

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{\sum_{i=1}^{n} \sigma_i^{-2}}}. \tag{9.3}$$

These scaling factors, listed in Table 9.3, are applied to adjust the backgrounds used for the *CAT Finder* relative to the *Baseline Finder*, shown in Fig. F.2

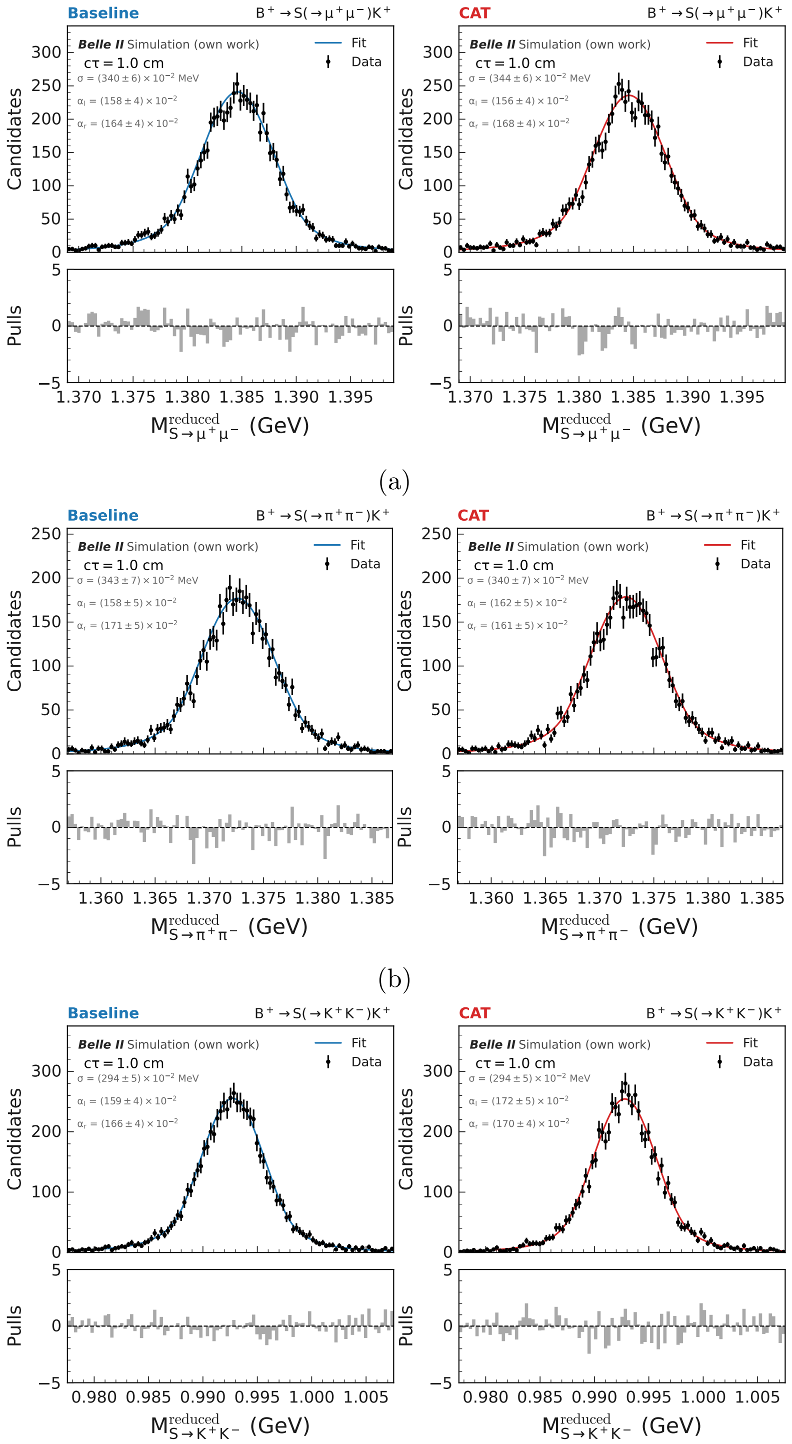
Overall, while the *CAT Finder* exhibits increased higher background levels, there is also a significant gain in signal efficiency at large displacements (up to a factor 2). I will determine the impact on the limits next.

9.4 Limits

To calculate the expected upper limits, I follow the approach in [44]. To search for an excess of the signal in the reconstructed mass, both the signal and background PDFs need to be modelled to describe the distribution. These are then fitted to the measured data to extract the number of signal and background candidates. As I only evaluate on simulated events in this work, I calculate the expected upper limits on the SM background-only distribution to estimate the sensitivity of the analysis between the *Baseline Finder* and *CAT Finder*.

9.4.1 Signal Fits

The signal distribution is modelled using a DSCB (see Eq. (8.2)), with the mean μ, the width σ, and $\alpha_{a/r}$ describing the tails of the distribution left floating for the fit. The parameters $n_{l/r}$ are fixed to 3 taken from fits where these values are left floating, following the approach described in Section 8.1. To determine the shape parameters of the DSCB PDF, I fit simulated signal samples, as shown in Fig. 9.3, for all combinations of lifetime, mass, and final state for both the *Baseline Finder* and *CAT Finder*. For short lifetimes ($c\tau = 1\,\mathrm{cm}$), the distributions and fit results are similar for both algorithms. At larger lifetimes, the distributions for both algorithms broaden, as illustrated in Fig. 9.4 for $c\tau = 1\,\mathrm{cm}$ (Fig. 9.4a) when comparing with $c\tau = 50\,\mathrm{cm}$ (Fig. 9.4b). The signal efficiency for the *CAT Finder* is increased for the large lifetime, leading to a higher yield. The fitted signal width also increases for larger lifetimes, with $\sigma = 2.73 \pm 0.28\,\mathrm{MeV}$ for the *Baseline Finder* and $\sigma = 3.1 \pm 0.4\,\mathrm{MeV}$ for the *CAT Finder*. Overall, the differences are small and mostly depend on the scalar mass m_S, with larger lifetimes leading to broader distributions and fewer candidates as seen in Fig. 9.5. This affects the combined signal+background fits, as a broader width slightly increases the background under the peak, which can affect the limits. This effect, however, is negligible, as the background and signal are modelled across the whole fit window and the added width is added to the background yield without affecting the signal contribution.

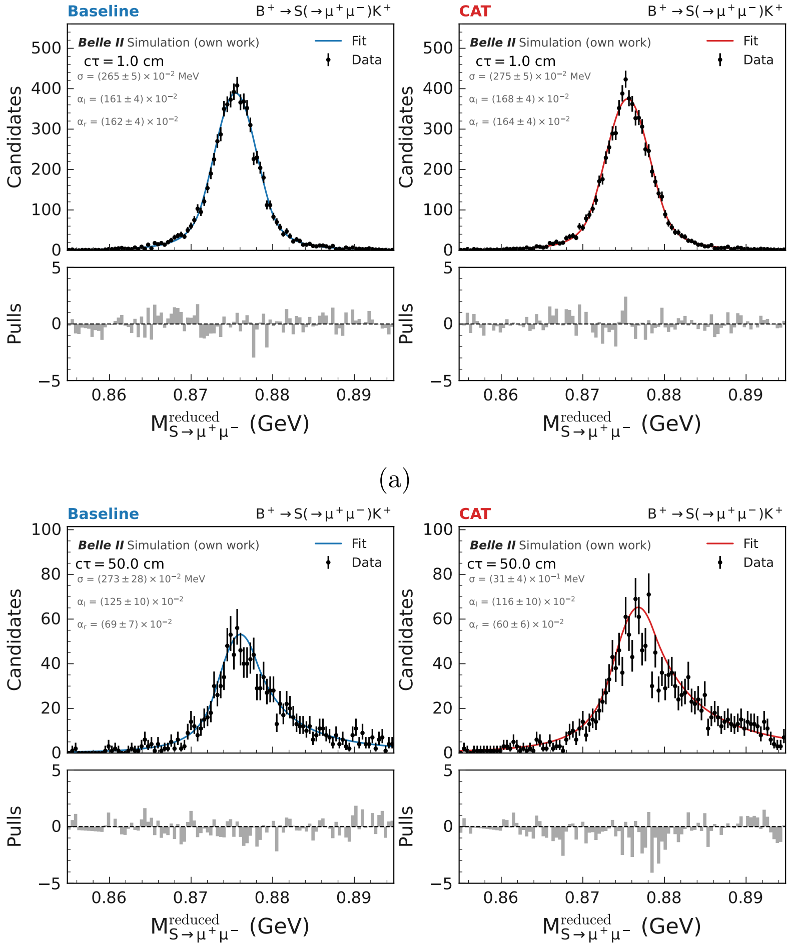


(a)

(b)

(c)

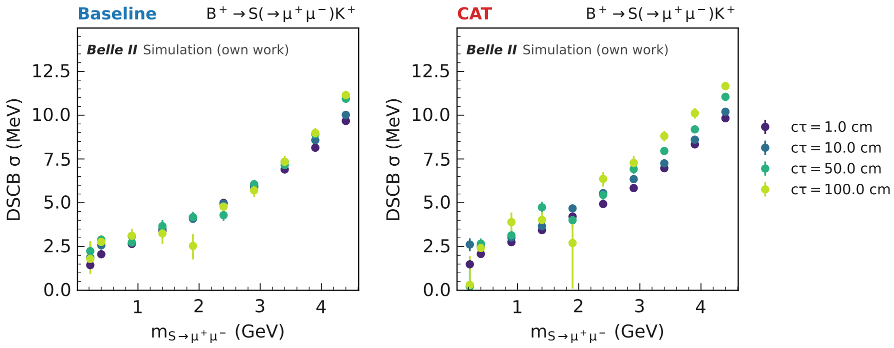
Figure 9.3: Signal fits for $M_{S\to x^+x^-} = 1.4\,\text{GeV}$ calculated in the observable $M^{\text{reduced}}_{S\to x^+x^-}$ for the final states $S \to \mu^+\mu^-$ (top), $S \to \pi^+\pi^-$ (middle) and $S \to K^+K^-$ (bottom) for a lifetime of $c\tau = 1\,\text{cm}$ for the *Baseline Finder* (left) and the *CAT Finder* (right).



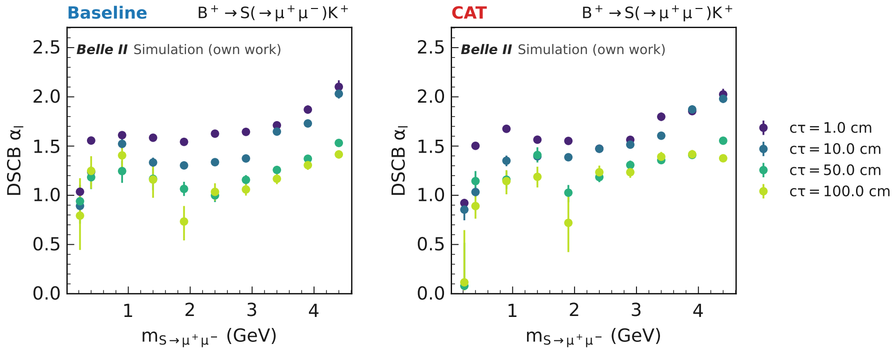
(a)

(b)

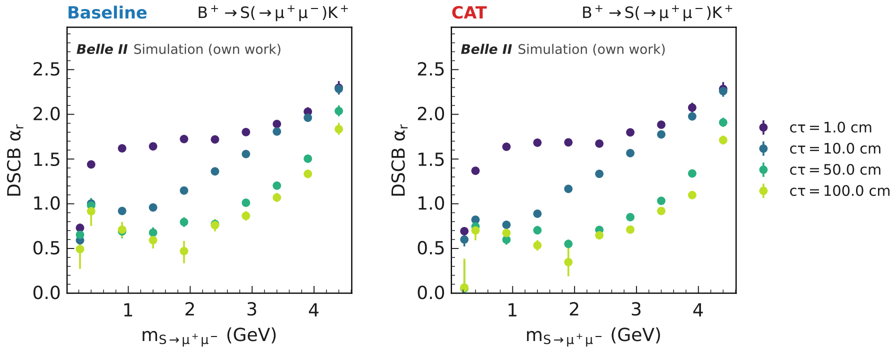
Figure 9.4: Signal fits for $M_{S\to x^+x^-} = 0.9\,\text{GeV}$ calculated in the observable $M^{\text{reduced}}_{S\to x^+x^-}$ for the final states $S \to \mu^+\mu^-$ for a lifetime of $c\tau = 1\,\text{cm}$ (top) and $c\tau = 50\,\text{cm}$ (bottom) for the *Baseline Finder* (left) and the *CAT Finder* (right).



(a) DSCB σ.



(b) DSCB α_l.



(c) DSCB α_r.

Figure 9.5: Signal fit result of the DSCB (Eq. (8.2)) on the signal sample $B^+ \to K^+S \to \mu^+\mu^-$ for the different tested masses $m_{S\to\mu^+\mu^-}$ for σ (top), α_l (middle) and α_l (bottom) for the *Baseline Finder* (left) and *CAT Finder* (right).

9.4.2 Combined Signal and Background Fits on Background Only

In addition to the signal, the background needs to be described. The reconstructed $M^{\text{reduced}}_{S\to x^+x^-}$ distributions of background events are smooth locally, and are modelled with a Chebyshev polynomial PDF with linear coefficient c_1, constrained to $[-1, 1]$ following [13]. The background distribution includes the normalization factor given by the background yield N_{bkg}. The signal yield estimator depends on the signal efficiency ϵ_{sig} determined in the previous section, the number of B-meson pairs N_{BB} determined for the run I dataset and scaled to the luminosity used in [13], and the signal branching fraction BF with

$$BF = BF(B^+ \to K^+S) \times BF(S \to x^+x^-), \tag{9.4}$$

given by the S production rate with the decay $B \to S$ and the decay rate $S \to x^+x^-$ of the final state x from theory calculations (discussed in Section 9.5). Only the number of charged B meson pairs is of interest for the signal yield, therefore $N_{BB} = (387 \pm 6) \times 10^6$[114] of run I is adjusted with $f^{\pm} = 0.5113^{+0.0073}_{-0.0108}$ [115] to take into account the number of charged B-meson pairs. This value is scaled down from the run I luminosity to the one used here. With this, the signal yield is given as

$$N_{\text{sig}} = 2N_{BB}f^{\pm}\epsilon_{\text{sig}}BF, \tag{9.5}$$

including the factor two as both charged B-mesons can decay in the signal process.

I fit the SM background only distribution (Fig. F.2) scaled to $\int L\mathrm{d}t = 189\,\mathrm{fb}^{-1}$ for the combined signal and background PDFs. For the combined signal and background fits, the minimum fit window is set to $\pm 20\sigma(M^{\text{reduced}}_{S\to x^+x^-})$ of the signal-only fit and iteratively increased by 10% until at least ten candidates are included to stabilize the fit following the procedure of [13]. The Gaussian mean of the DSCB is fixed to $\mu = M^{\text{reduced}}_{S\to x^+x^-}$. The remaining signal shape parameters of the signal only fits are fixed and only the signal yield estimator N_{sig}, the background yield estimator N_{bkg} and the background coefficient c_1 are floating for the fit. The fits on the background only distribution are repeated for all lifetime, mass and decay combinations given in Section 3.5.2.

One example fit result is shown in Fig. 9.6 for the *Baseline Finder* and the *CAT Finder*, where signal events are injected in addition to the background.

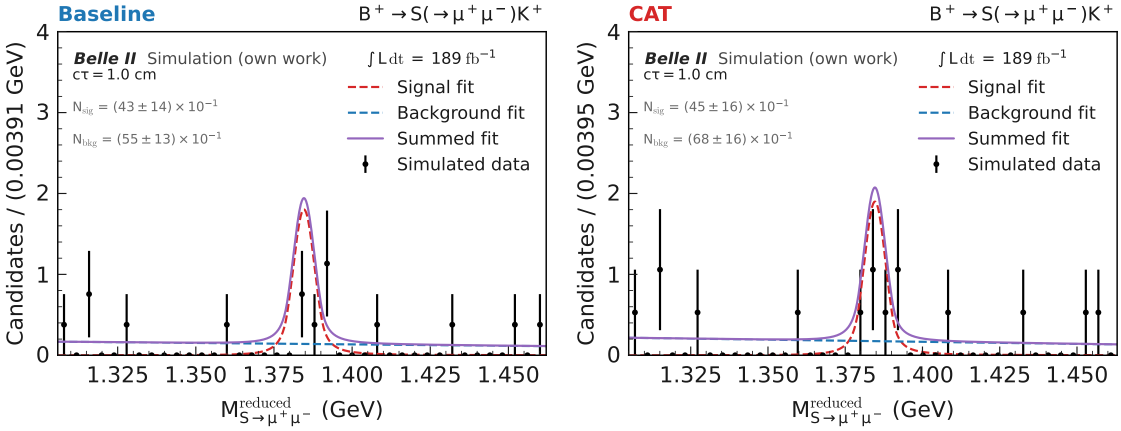


Figure 9.6: Combined signal+background fit to the simulated background samples scaled to the the target luminosity of $\int Ldt = 189\,\mathrm{fb}^{-1}$ for *Baseline Finder* (left) and *CAT Finder* (right) with four injected signal candidates.

9.4.3 Upper Limits

First, I calculate the sensitivity on the branching fraction BF without any theory assumptions. While this is still dependent on the model for a scalar or pseudoscalar, as the signal efficiency is derived for scalar kinematics (Section 3.5.2), these results have weaker model-dependency. The interpretation including branching fraction predictions from theory, which is fully model dependent, is discussed in the next section.
The upper limits on the branching fraction BF also follow [13] and use the `HEPSTATS`[116] framework which interfaces with the fit models of `zfit`[107]. The limit setting is done using the modified frequentist CL_S method[117] and asymptotic approximation[118] for the standard value of the community $\alpha = 0.5$ for the 95% confidence level (C.L.) upper limits.

While I can not apply any corrections or systematic uncertainties, I will discuss the most important one given in [44].
For small lifetimes, the systematic uncertainties on the signal efficiency in [44] are dominated by the number of $B\bar{B}$ pairs and the signal yield estimation. The systematic uncertainty is around 5 %. For low masses under 0.5 GeV, the veto on the $M_{\pi^+\pi^-}$ in the K^0_S region has the highest contribution, due to the reconstruction disagreement between simulation and measured data for K^0_S. This systematic is also the dominant one for long-lived particles with lifetimes larger than 10 cm. I discuss this effect in Section 8.2. While the disagreement did not improve between the publication of [13] and the time of this thesis, the estimate of the uncertainty improved, which results in a reduced contribution compared to the $\mathcal{O}(8\%)$ in [44]. Assuming similar uncertainties as in [13], the limits would weaken for around 2% for light S masses with large lifetimes and 1% for either heavier S or shorter lifetimes.

An example for $S \to \mu^+\mu^-$ is shown in Fig. 9.7. For short lifetimes ($c\tau = 1\,\text{cm}$), both algorithms yield similar limits. For long lifetimes ($c\tau = 100\,\text{cm}$), the *CAT Finder* achieves up to a factor 2 stronger limits, probing a larger branching fraction region. This improvement is equivalent to increasing the dataset by approximately a factor of four when considering the sensitivity improvements proportional to $N_{\text{sig}}/\sqrt{N_{\text{bkg}}}$, as discussed in Chapter 8.
Fig. 9.8* shows the comparison for all three final states, including the ratio between the *Baseline Finder* and *CAT Finder*, where values above 1 indicate weaker limits for the *Baseline Finder* compared to *CAT Finder*. While the *CAT Finder* performs slightly worse than *Baseline Finder* for short lifetimes of $c\tau = 1\,\text{cm}$, for longer lifetimes the *CAT Finder* consistently outperforms the *Baseline Finder*. Only for very small mass ($m_S = 0.22\,\text{GeV}$ in $S \to \mu^+\mu^-$), the *CAT Finder* underperforms, indicating that further improvements in track reconstruction for very small opening angles will be beneficial in future iterations.

---

[*] Only three values for the $S \to K^+K^-$ final state are probed that are relevant for the model-dependent interpretation in the next section.

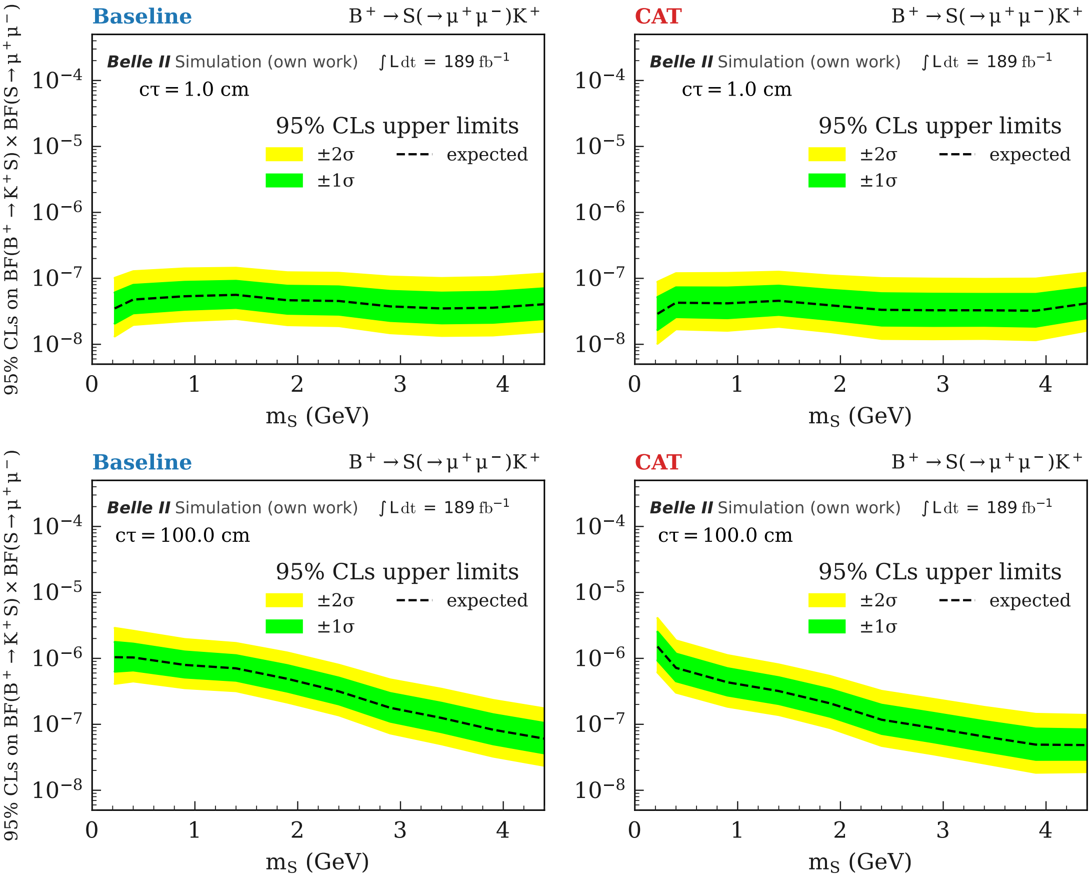


Figure 9.7: Expected 95% C.L. upper limits on simulation for the branching fraction $BF(B^+ \to K^+S) \times BF(S \to \mu^+\mu^-)$ for the lifetime $c\tau = 1\,cm$ (top) and $c\tau = 100\,cm$ (bottom) for the *Baseline Finder* (left) and *CAT Finder* (right).

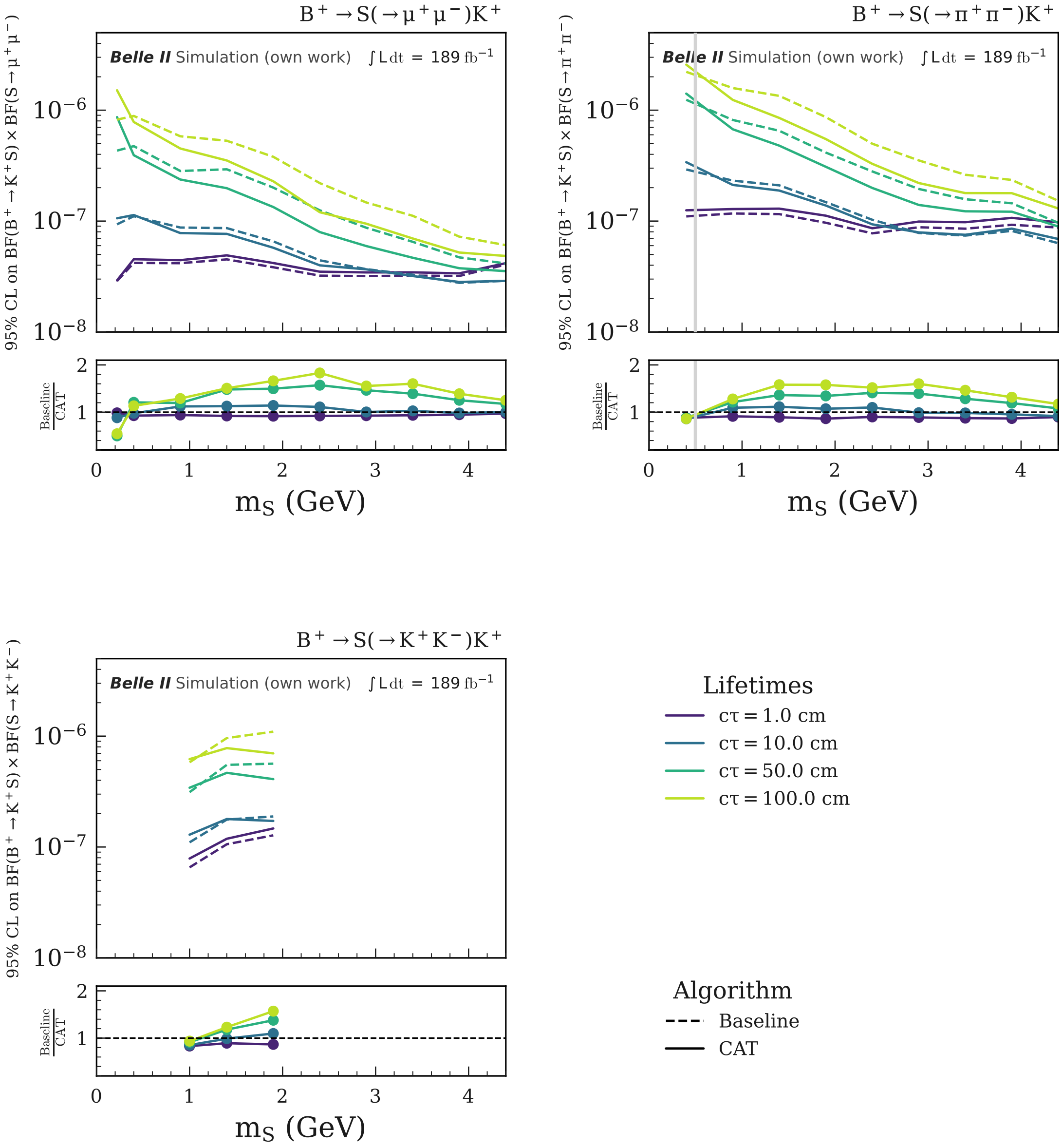
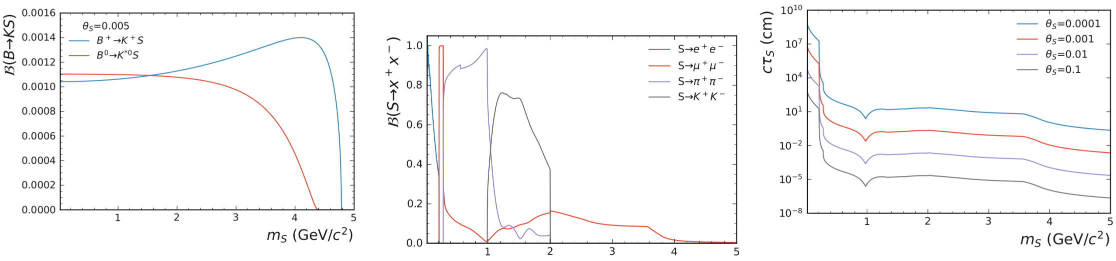


Figure 9.8: 95% C.L. expected upper limits on the product of the production branching fraction $BF(B^+ \to K^+S)$ and the final state branching fraction $BF(S \to \mu^+\mu^-)$ (top left), $BF(S \to \pi^+\pi^-)$ (top right), and $BF(S \to K^+K^-)$ (bottom). The region excluded due to the K_S^0 (Table 9.1) is shown in grey for $BF(S \to \pi^+\pi^-)$.



(a) Production rate of S in $B \to KS$.

(b) S decay branching fraction.

(c) Lifetimes over mass for different mixing angles.

Figure 9.9: Model predictions for the production branching fraction (left) of the dark scalar via the $B \to KS$ modes, the decay branching fraction in final state particles $S \to e^+e^-$, $S \to \mu^+\mu^-$, $S \to \pi^+\pi^-$, and $S \to K^+K^-$ (middle) and the lifetime for different mixing angles θ_S (right) over the dark scalar mass m_S, taken from [13].

9.5 Dark Higgs-like Scalar Model Interpretation

The spin-0 particle can be interpreted as either a scalar or pseudoscalar. In this work, I focus on the scalar case following [44], where the long-lived scalar S produced in $b \to s$ transitions mixes with the SM Higgs boson. Production occurs dominantly via radiation from the top quark or W boson in the loop. The model is characterized by the dark scalar mass m_S and the mixing angle θ_S, with couplings to SM particles proportional to those of the Higgs, scaled by $\sin\theta_S$. The mixing angle θ_S controls both the production and decay rates, and therefore the lifetime.

The branching fractions for individual decay modes are taken from [7, 119] (see Fig. 9.9) and are independent of θ_S. However production and total decay rates scale with $\sin^2\theta_S$, making the lifetime proportional to $\sin^{-2}\theta_S$. The upper limits from the previous section are now translated using the parameters for the dark scalar mass m_S and the mixing angle θ for the probed lifetimes for each scanned mass following [7, 119]. The translation into parameter space is shown in Fig. 9.10. The upper limits are interpolated between all tested lifetimes, and the intersection with the model prediction is computed. The region where the interpolated limits are below the model prediction is excluded. Performing this intersection for all tested masses results in the exclusion plots given in Fig. 9.11. Because the branching fraction decreases with increasing lifetimes, this model favours more prompt decays in the case of Belle II in which the *CAT Finder* has decreased performance due to the higher backgrounds, resulting in less parameter space being probed by the *CAT Finder* compared to *Baseline Finder*.

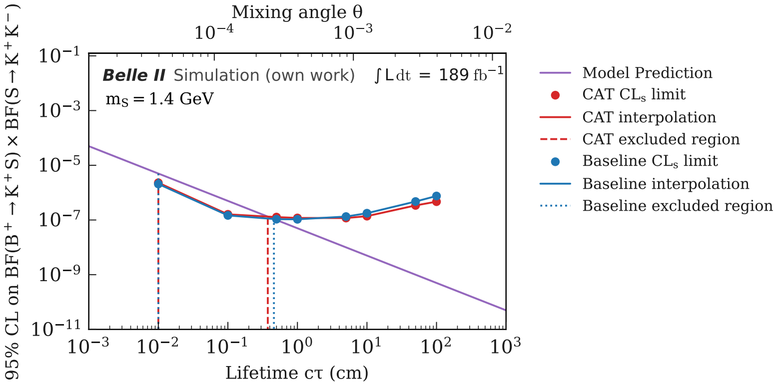


Figure 9.10: Translation of the model bound of the dark Higgs-like scalar model including the mixing angle θ for the $S \to K^+K^-$ final state with a dark scalar mass of $m_S = 1.4\,\text{GeV}$. The theoretical model prediction is shown in purple. The expected limits, including the interpolation and intersection with the model prediction, marking the excluded region, are shown for *Baseline Finder* (blue) and *CAT Finder* (red) for $\int Ldt$ =189 fb^{-1}. For *Baseline Finder*, the region between the two blue, dotted lines is excluded, while for the *CAT Finder* it is the region between the two red dashed lines.

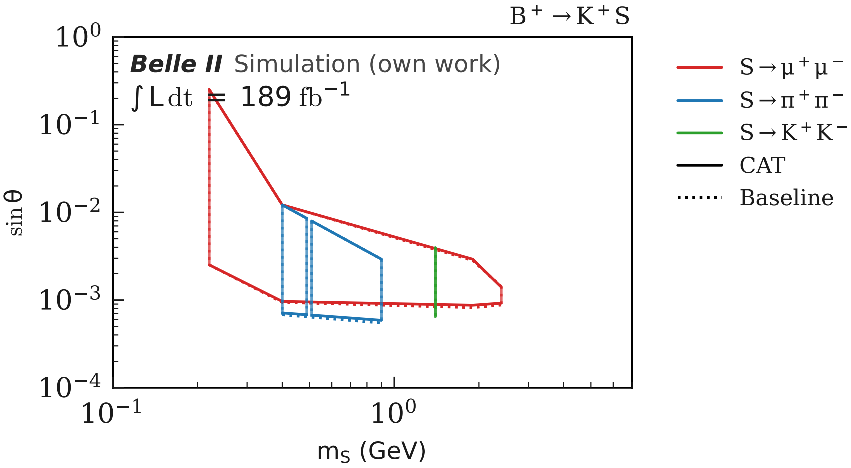


Figure 9.11: Bounds on the dark scalar model in $B^+ \to K^+S$ for the channels $S \to \mu^+\mu^-$ in blue, $S \to \pi^+\pi^-$ in blue and $S \to K^+K^-$ in green for $\int Ldt$ =189 fb^{-1}. This shows the excluded parameter space for the *Baseline Finder* (dotted line) and *CAT Finder* (solid line). There is only one value for $S \to K^+K^-$ due to model interpretation only yielding a result at this mass point.

For future Belle II searches, larger datasets will allow probing longer lifetimes. In this case, the search can be optimized more towards larger displacements as shown in $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ [12] (see Section 3.5.1). In Fig. 9.12, the backgrounds for *Baseline Finder* are shown depending on the displacement d_r of the vertex S. This is not available for the *CAT Finder* due to the low statistics. For all three final states, the backgrounds are reduced significantly for displacement higher than $d_r > 5\,\text{cm}$. While the signal efficiency also reduces, especially for small lifetimes, as shown in Fig. 9.13, this allows

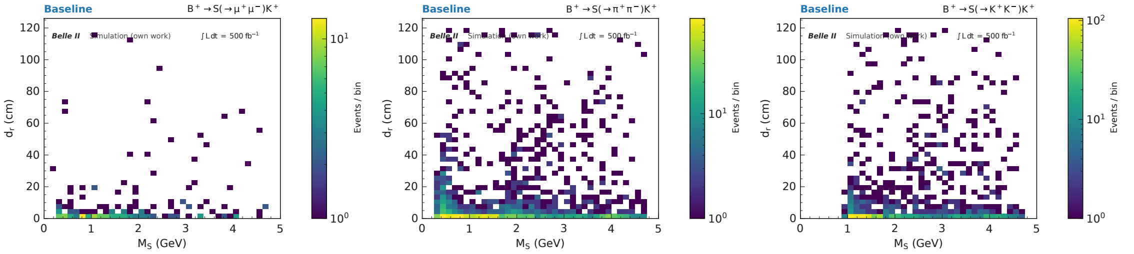


Figure 9.12: Background distribution for the *Baseline Finder* for the decay channel $S \to \mu^+\mu^-$ (left), $S \to \pi^+\pi^-$ (middle) and $S \to K^+K^-$ (right) for the displaced vertex d_r of the reconstructed scalar S against the reconstructed scalar mass M_S.

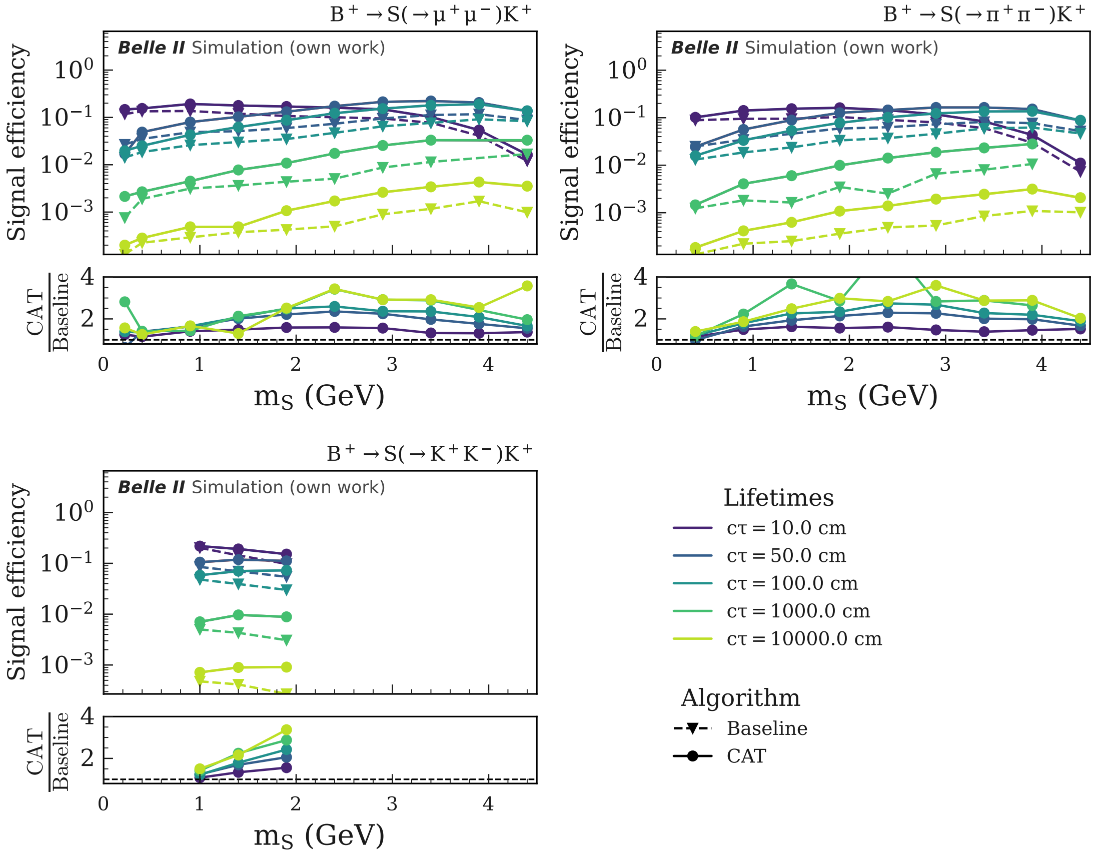


Figure 9.13: Signal efficiency for different lifetimes for the *Baseline Finder* (dashed line with triangular markers) and the *CAT Finder* (solid line with round markers) including the stricter cut on $d_r > 5.0\,\text{cm}$ compared to Table 9.1. The lines are added to help guide the eye. The ratio between signal efficiency is given for each final state $S \to \mu^+\mu^-$ (top left), $S \to \pi^+\pi^-$ (top right) and $S \to K^+K^-$ (bottom).

to suppress the backgrounds significantly. Furthermore, there is work ongoing at Belle II to improve background suppression through machine-learning techniques in reconstruction and analysis even further, for example improving PID with machine learning[105]. For an example future Belle II search, assuming an integrated luminosity of $\int Ldt = 50\,\text{ab}^{-1}$[2]

and negligible backgrounds, set to zero in this case, for S decays far from the interaction point with $d_r > 5\,\text{cm}$, the *CAT Finder* shows a clear advantage over the *Baseline Finder* (Fig. 9.14). The full result is given in Fig. 9.15, with the expected 95% C.L. upper limits calculated with the Bayesian Analysis Toolkit[120]. This improvement will be even more pronounced in analyses such as $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ [12], where the cross section is independent of the mixing angle and thus lifetime and lifetimes up to 100 m are already probed for the model-dependent parameter space.

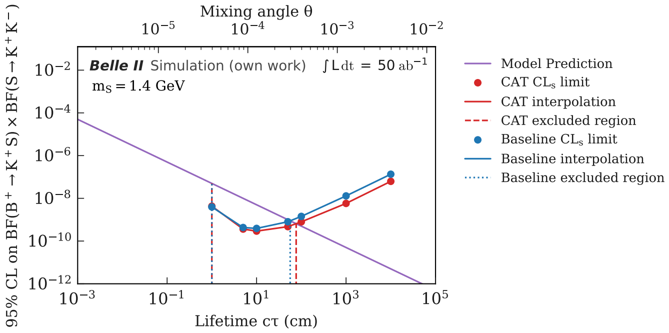


Figure 9.14: Translation of the model bound of the dark Higgs-like scaler model including the mixing angled θ for the $S \to K^+K^-$ final state with a dark scalar mass of $m_S = 1.4\,\text{GeV}$. The theoretical model prediction is shown in purple. The expected limits including the interpolation and intersection with the modelprediction, marking the excluded region are shown for *Baseline Finder* (blue) and *CAT Finder* (red) for $\int Ldt = 50\,\text{ab}^{-1}$. Lifetimes below 1 cm are not considered here due to the selection on $d_r > 5\,\text{cm}$.

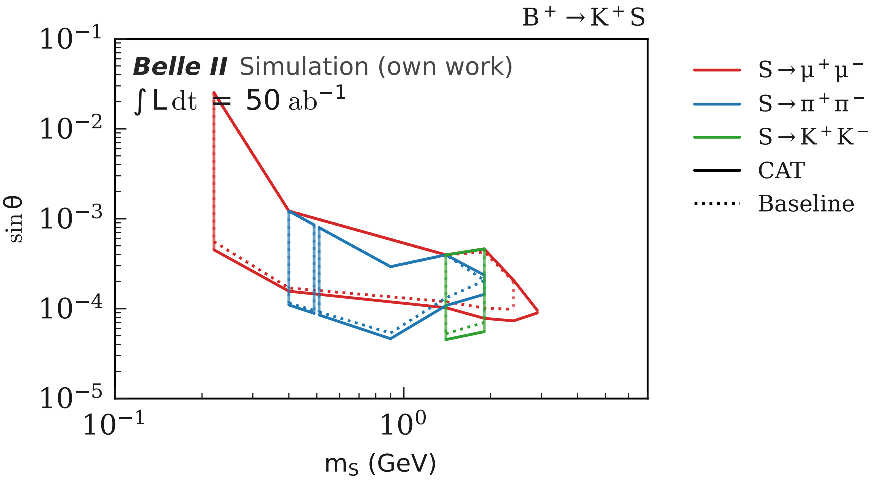


Figure 9.15: Bounds on the dark scalar model in $B^+ \to K^+S$ for the channels $S \to \mu^+\mu^-$ in blue, $S \to \pi^+\pi^-$ in blue and $S \to K^+K^-$ in green for $\int Ldt = 50\,\text{ab}^{-1}$. This shows the excluded parameter space for the *Baseline Finder* (dotted line) and *CAT Finder* (solid line).

Chapter 10

Ongoing Studies and Future Work

In this chapter, I report ongoing work on the development of the GNN tracking algorithm, and address the shortcomings of the algorithm that I found and reported in this thesis in the previous sections. As many findings in this chapter are part of the active development, I reference internal Belle II documentation which is available upon request.

While I briefly discussed some improvement strategies in the previous chapters during relevant passages, I summarize them in this chapter.

In Section 10.1, I summarize currently known shortcomings of the CDC simulation within basf2 and strategies to mitigate the differences between simulation and measured data. Following this discussion, I summarize my current findings on training on physics events (Section 3.2) in Section 10.2. Studies and next steps regarding the hyperparameter optimization are given in Section 10.3. I propose different next steps to improve the track finding, by running the *Baseline Finder* and *CAT Finder* together (Section 10.4), including hit filtering (Section 10.5) and tuning the track fitting algorithm (Section 10.6). Finally, I discuss a combined track finding of the CDC and SVD detectors in Section 10.7.

10.1 Discrepancy between CDC Simulation and Measured Data

Even though the *CAT B Finder* model performed the best on simulation, it underperforms on measured data (see Section 8.2 and Section 8.3.1). This suggests a possible mismatch in input features' modelling, likely stemming from limitations in the current CDC simulation in basf2. Supporting evidence includes an increasing discrepancy between the number of hits per track in simulation versus data with increasing background levels, which appears both for *Baseline Finder* and *CAT Finder* (see Section 8.1). Additionally, there is a charge-dependent bias, where positively charged tracks show more hits in simulation than in data, while the opposite holds for negatively charged tracks, as illustrated in Fig. 8.10. Different particles have different material interactions and ionization within the CDC volume, also between opposite charged particle types. This is more relevant for the *CAT B Finder* model, which includes more particle types than just muons in the simulation. It might be that particles with more interaction within the detector gas and material are not modelled correctly. For example, it is known that the simulation of low-energy photons stopped in the CDC gas is underestimated in the current simulation, and efforts to improve this are ongoing [121]. The *CAT B Finder* could pick up on these mismodeled effects, thus showing a better performance on simulation than on measured data.

One reason for the charge dependence of the number of CDC hits is the signed angle in which the particle crosses the drift cell [110]. There is work ongoing to determine correction factors depending on this crossing angle, that are applied after the detector simulation with GEANT4 when digitizing the CDC hits.

Another known issue is the $x - t$ drift time calculation from TDC (Section 2.3), where there was a disagreement found for large drift distances between simulation and measured data [121]. The drift time is then translated into the drift length and included in the CKF and final fits (Section 5.1). If this drift length is over- or underestimated, the hits can get excluded during the fitting stage [24]. Studies determined on data show that the translation factors differ for positively and negatively charged particles, which generates the same sign charge asymmetry in simulation as in measured data for the number of CDC hits. This effect is particularly sensitive to the composition of the CDC gas mixture, which was observed after the start of run II. The gas quality improved significantly compared to run I, which resulted in a larger fraction of hits being discarded due to the calibration and optimization still being based on the gas composition, including the contamination with vapor that happened during run I.

What might also be an issue is that the interference between the beam-background CDC hits and the signal hits is not handled correctly in simulation, due to more hits being lost during the fitting stage in measured data in comparison to simulation (see Fig. 8.16). The current approach to sum the digitized hits of signal and background if they come from the same wire. This requires major adjustments on how the beam-backgrounds are overlayed with the simulated signal, as a simple sum might no longer be sufficient, and might require a stronger focus on simulated beam-backgrounds. More in-depth studies go into updating the waveform simulation according to first-principle simulation for the ADC, TDC, and TOT simulation to gain a better understanding of the effects on the detector depending on the particle's momentum and gas conditions [122]. These studies might also improve the cross-talk simulations, which are currently not modelled correctly in the CDC (see Section 3.1.1 and [123]).

The choice of input features remains a challenge. The best input features to chose were investigated in [84] the bachelor thesis I supervised on simulated events. Including the x and y reference positions at $z = 0$ and adding TOT information significantly improves performance on simulated events, surpassing that of the *Baseline Finder* and *CAT B Finder*. However, the same setup underperforms on real data as shown in Fig. 10.1 All of the inputs used here are already used within the Belle II tracking chain. As these

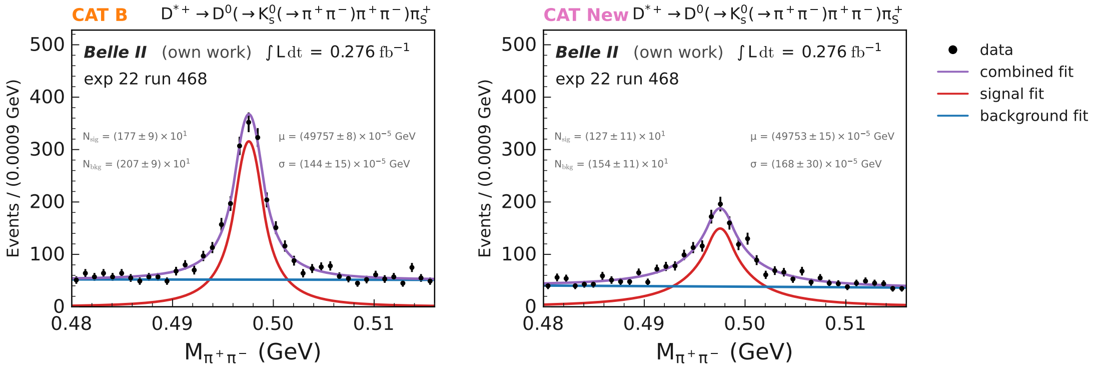


Figure 10.1: Reconstructed K_s^0 mass from two opposite charged pions $M_{\pi^+\pi^-}$ in the signal region of 0.48 GeV to 0.52 GeV for measured data for run 468 from experiment 22. The plots show $M_{\pi^+\pi^-}$ using the *CAT B Finder* (left), and the *CAT Finder* trained with updated inputs including the x and y position at $z = 0$ and TOT. The combined fit, as well as the individual signal and background parametrization, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated.

might not be correctly modelled in the simulation or the training picks up very specific correlations for the simulation, these effects need to be mitigated.

One option that should first be tested would be *noise injection* into the input features. By varying each value slightly, this can mitigate the overfitting to (spurious) correlations in the simulation [124]. Normally-distributed noise can also be directly interjected into the first or more hidden layers of the GNN model. This improves the model robustness against changes in the input features.

Another option is to add measured data to the training dataset, where the labels come from the reconstruction of the *Baseline Finder*. As the *Baseline Finder* mostly reconstructs trajectories that are close to the interaction point, this requires careful consideration of which events should be included. Furthermore, the hit purity of the *Baseline Finder* is only at 95% (see Fig. 6.10d), which means the tracks include fake hits, and the issue of the number of hits assigned to positive and negative charged tracks as well as the decreased hit efficiency with rising background levels has to be taken into account.

One more option that includes training on measured data is to train the model using a *domain adaptation classifier* [125] that distinguishes measured data from simulation. In addition to the track reconstruction, another classifier is added to predict if the feature is from simulation or measured data. The feature extractor is trained, often using adversarial techniques, to produce representations that are indistinguishable between simulation and measured data. This approach has been explored in jet tagging within LHC experiments [126, 127]. Performance is monitored via an additional domain adaptation loss, which is unsupervised and measures alignment between simulation and real data. Using either of these methods should be explored in the next step, to improve the agreement between simulation and measured data when training on multiple different particle types.

10.2 Training on Physics Samples

While the current models are successfully able to reconstruct multiple different particle types, the results also showed that the performance on simulation is the best for muons (see Section 7.1.1). This is due to the reason that all events for *CAT Finder* and 75% of events for *CAT B Finder* only consist of muons. The input feature study in Section 5.3.1 showed that the different particle types differ in their input features. Therefore, the next step is to increase the number of other particle types in the training dataset.

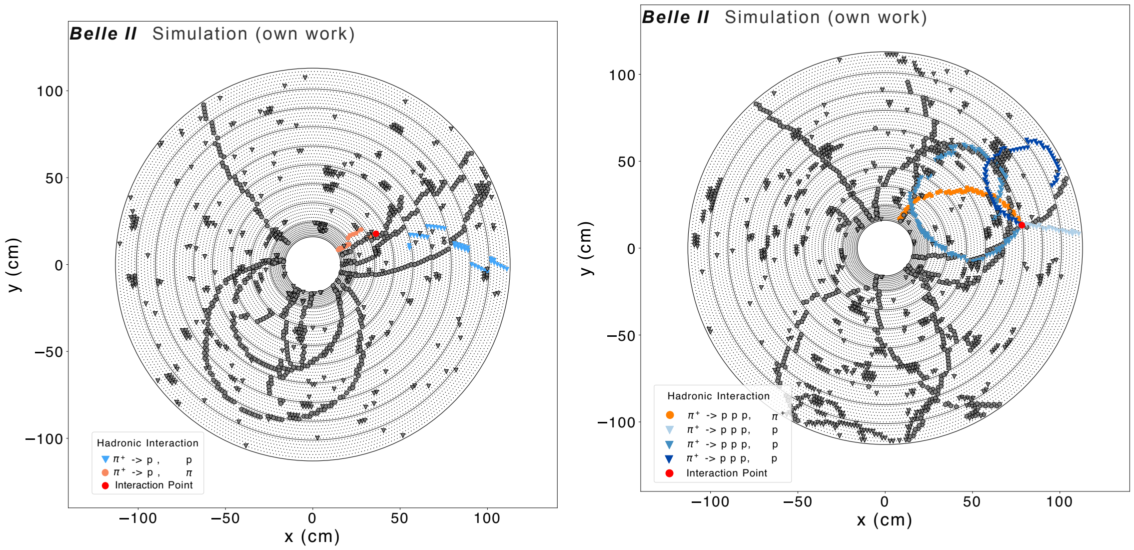
For particle species other than muons, incorporating them in the training dataset requires more careful consideration due to interactions in the detector. Unlike muons, which are minimum-ionizing particles with a lifetime long enough to make decays in flight negligible for Belle II, other particles can undergo processes that significantly impact track reconstruction. The most relevant effects for tracking include:

- ionization loss (see Section 7.3.1), which increases the momentum uncertainty;
- multiple Coulomb scattering, which is especially relevant for low momentum particles, as it can deflect the particles and change the trajectory;
- radiative processes such as bremsstrahlung, most relevant for electrons and positrons;
- decays in flight;
- hadronic nuclear interactions.

The last two processes are the primary sources of track reconstruction inefficiency and will be discussed in detail in the following. Almost every $B\bar{B}$ event involves at least one decay in flight or hadronic interaction. At the time of this thesis, I did not converge on a dataset that can describe the full Belle II physics space. I therefore focus on summarizing the approaches that I tested to address these challenges and propose potential directions for future work.

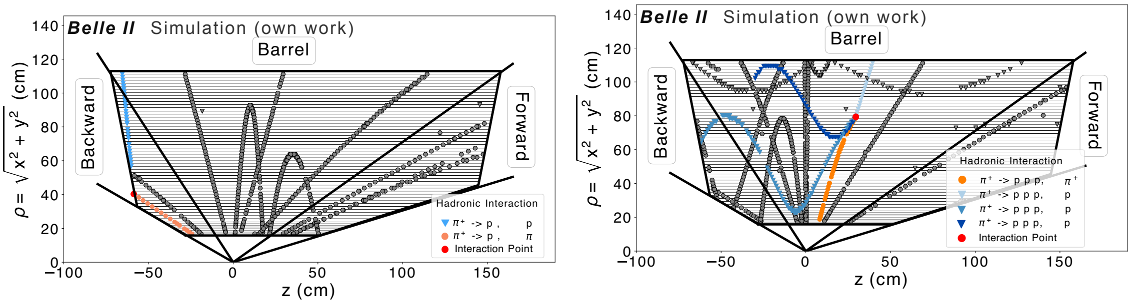
Fig. 10.2 shows particles originating from hadronic interaction with the CDC wall or from a hadronic interaction within the CDC, for example, with the drift or field wires.

If the extrapolated trajectories of the particle from the scattering pass close to the interaction point, they can still be reconstructed by the *Baseline Finder*. In the worst case, the *Baseline Finder* finds the original particle and the one resulting from the scattering as one track, if the changes in momentum and thus direction are small between both. These cases are particularly important to identify, since the particle from the scattering continues towards the TOP. The TOP then may register them as a different particle type than the originating one, leading to a PID misidentification. Only the proton going outwards the detector could be found in Fig. 10.2b. For the other 2 particles, they would not be found



(a) Particle from a hadronic interaction with the CDC wall, $x - y$ projection.

(b) Particle from a hadronic interaction within the CDC, $x - y$ projection.



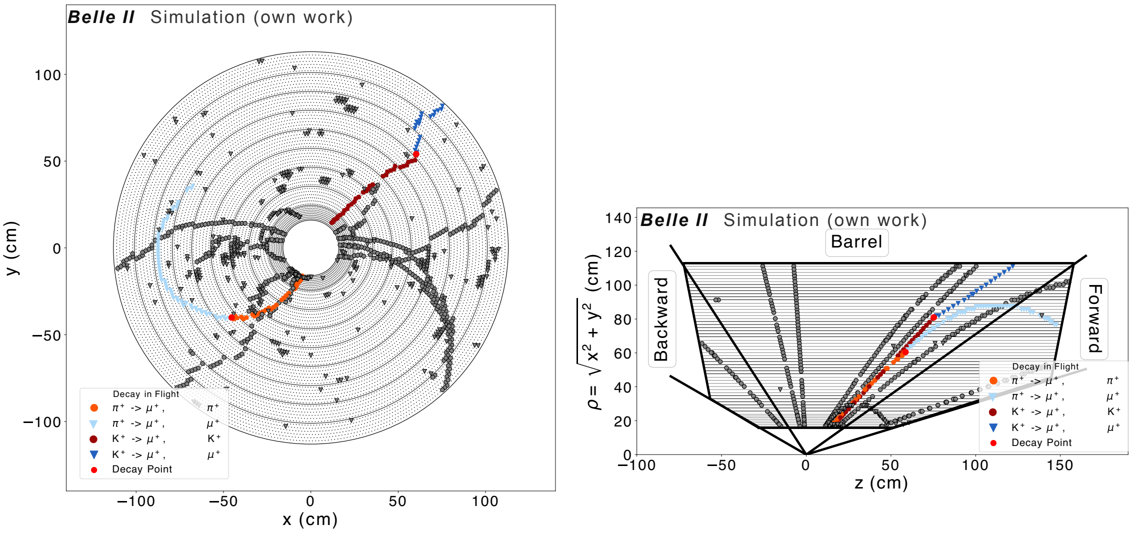
(c) Particle from a hadronic interaction with the CDC wall, $\rho - z$ projection.

(d) Particle from a hadronic interaction within the CDC, $\rho - z$ projection.

Figure 10.2: Particles from hadronic interaction with the detector wall (left) or within the CDC at wires (right) in the $x - y$ view (top) and $\rho - z$ view (bottom). The trajectories of the mother particles are in orange circular markers, and the products from the scattering are in blue triangular markers. The interaction point is marked with a red circle.

by the *Baseline Finder* algorithm. The reconstruction of these particles originating from detector scattering is undesired, as it would only contribute to an increased fake rate and not to the underlying physics process, unless specifically handled after the track finding step. Instead, the particles and their properties before the hadronic interaction need to be reconstructed correctly. There are three direct options for how these tracks can be handled:

- training on both the particles before and after hadronic interactions labelled as objects (see Section 5.3.2) and adding a new output layer to the GNN model architecture that classifies tracks as signal or background;
- labelling the unwanted hits from hadronic interactions as background hits during the training stage, making them similar in the model handling as the beam-background tracks (Section 3.1.1);



(a) Decay in flight, $x - y$ projection. (b) Decay in flight, $\rho - z$ projection.

Figure 10.3: Decay in flight for $\pi^+ \to \mu^+ \nu_\mu$ and $K^+ \to \mu^+ \nu_\mu$ in the $x - y$ view (left) and $\rho - z$ view (right) in the same event. Mother particles are shown in red circular markers and daughter particles in blue triangular markers, and the production vertex of the daughter particle is marked with a red circle.

- removing the unwanted hits from hadronic interaction from the training samples.

One additional method to handle these kinds of tracks was introduced with a recent basf2 release with the KinkFinder algorithm [128]. If a particle decays in flight, the particle's trajectory changes, thus creating a so-called kink. Fig. 10.3 shows two examples of a decay from a $\pi^+ \to \mu^+ \nu_\mu$ and $K^+ \to \mu^+ \nu_\mu$. For the *Baseline Finder*, often the mother-daughter combinations are found as a single track, especially if the trajectory only changes in the $\rho - z$ plane. The change in trajectory is thus only observed in the stereo layers, which are ignored during the first track finding step (see Section 5.2) and can be easily missed. One different option is that hits are mis-assigned to the other particle, which interferes with the matching requirement defined in Chapter 4. The new KinkFinder algorithm splits tracks that include a kink in their trajectory and can reassign hits between tracks. Additionally, this algorithm uses BDTs to determine if the track is coming from a hadronic interaction, providing the analyst with the option to exclude these types of tracks during the analysis. Comparing the performance of the *CAT Finder* on these kinds of tracks, as well as including hadronic interactions in the training dataset and evaluating it with the KinkFinder algorithm, is currently ongoing.

One additional challenge arises when training on high multiplicity events. In simulation, the model performs well for both low- and high-multiplicity events. However, when applied to measured data, the performance degrades in the low-multiplicity events (Section 8.3.1). The origin of this discrepancy still needs to be investigated in detail.

To address this, a new dataset must be constructed that provides a more representative coverage of the full Belle II phase space, including a balanced mix of different particle types and event multiplicities. Special focus should be put on the following signatures in addition to the samples described in Chapter 3 according to the findings of this thesis:

- high transverse momentum for small opening angles with the two trajectories often times having CDC hits in neighboring wires;
- balanced mix of different particle types, as the input features and material energy loss differ, which influences the particle trajectory. This needs to take into account the hadronic interactions within the detector, with solutions described earlier;
- treatment of curlers to optimize for track fitting. If the track fitting procedure is not adjusted, only the first loop of the curler should be identified by the track finding algorithm, as described in Section 6.1.1;

10.3 Hyperparameter Adjustment

Once such a dataset is available, re-running the hyperparameter optimization should be prioritized, since the increased complexity from overlapping and low-momentum tracks may require a higher model capacity than currently used. I explored training with a higher-dimensional condensation space that was increased from three to five. While this did not yield improvements within the current setup, it may become relevant in the future, especially in combination with more advanced clustering approaches in the latent space. This is especially relevant for high-multiplicity events, as clusters often overlap or extend into ellipsoidal shapes rather than the assumed spherical ones in Section 5.3.3. I also observed this on measured data as shown in Fig. 8.20.

Several options for latent space clustering are discussed in [91], but none resulted in improved performance. For example, [91] reported that for low transverse momentum particles with $p_t < 0.5\,\mathrm{GeV}$, the cluster size increased compared to higher transverse momentum. Trying to increase the size of the hit threshold t_h (see Section 5.3.3) according to the momentum prediction, however, did not yield improved results in [91]. Another observation from [91] is that the distance between background hits and condensation points decreases with the number of condensation points in an event. Similarly, the separation between condensation points themselves also decreases. Adjusting t_h according to the number of condensation points was tested but did not prove successful in [91]. These findings show that information about the structure of the cluster space can be gained from the other predictions. This information can be exploited when clustering in future studies.

There is also the option to use ML to assign the hits belonging to the cluster. One approach that will be explored in future work is to use algorithms in object detection, such as [58].

10.4 Running *Baseline Finder* and *CAT Finder* Together

As an intermediate step before finding a suitable dataset and improving the agreement between the detector simulation and the measured data, one option to include the *CAT Finder* in basf2 is to run both tracking algorithms. There are two possible strategies to combine the *Baseline Finder* and *CAT Finder* approaches:

1. run both algorithms independently and merge the resulting track candidates;
2. run *CAT Finder* only on hits that remain unused after the *Baseline Finder* reconstruction or vice versa.

The option to merge track candidates is available within basf2, but the investigation was out of scope for this work. For the second strategy, the results for the $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$ sample on data using the paper model are shown in Fig. 10.4. While the signal yield is increased and more displaced tracks are found as shown in Fig. 10.5, the background also increases to a similar level as the *CAT Finder* model.

A key challenge in this approach is to avoid an increase of the track clone rate. If tracks are already reconstructed by the *Baseline Finder* but only a fraction of hits is attached to the track, the remaining hits may be reused by the *CAT Finder* and produce duplicate track candidates. A dedicated training of the *CAT Finder* for this specific combined use case may help mitigate these issues. Although such studies are beyond the scope of this work, they are an important direction for future investigations. Both of these approaches could become especially valuable in scenarios where detector performance degrades further, since the *CAT Finder* has been shown to recover performance under challenging conditions such as axial holes or reduced wire efficiencies (see Section 6.4).

For running both algorithms, the computational load has to be considered. Currently, the track finding accounts for 30% of the reconstruction time for the HLT and for the 40% of the full reconstruction of high-multiplicity events. As of current no `C++` implementation is available of the *CAT Finder* model (discussed in Section 5.4), the track finding time of *CAT Finder* is not optimized due to change from `C++` to `Python` and back. For high multiplicity events, the time reported in [96] for the *CAT Finder* is 1.6 times higher than the *Baseline Finder*. This value is to be optimized once the *CAT Finder* is implemented in `ONNX` [129]. This makes the second strategy, where *CAT Finder* only runs on unused hits after the *Baseline Finder*, the more attractive option. However, the first strategy could also become feasible if the CDC hits are pre-filtered to reduce the input size. While the GNN model is already effective in suppressing background based on the input features, the track finding time improves with pre-filtered hits. I discuss this aspect further in the next section.

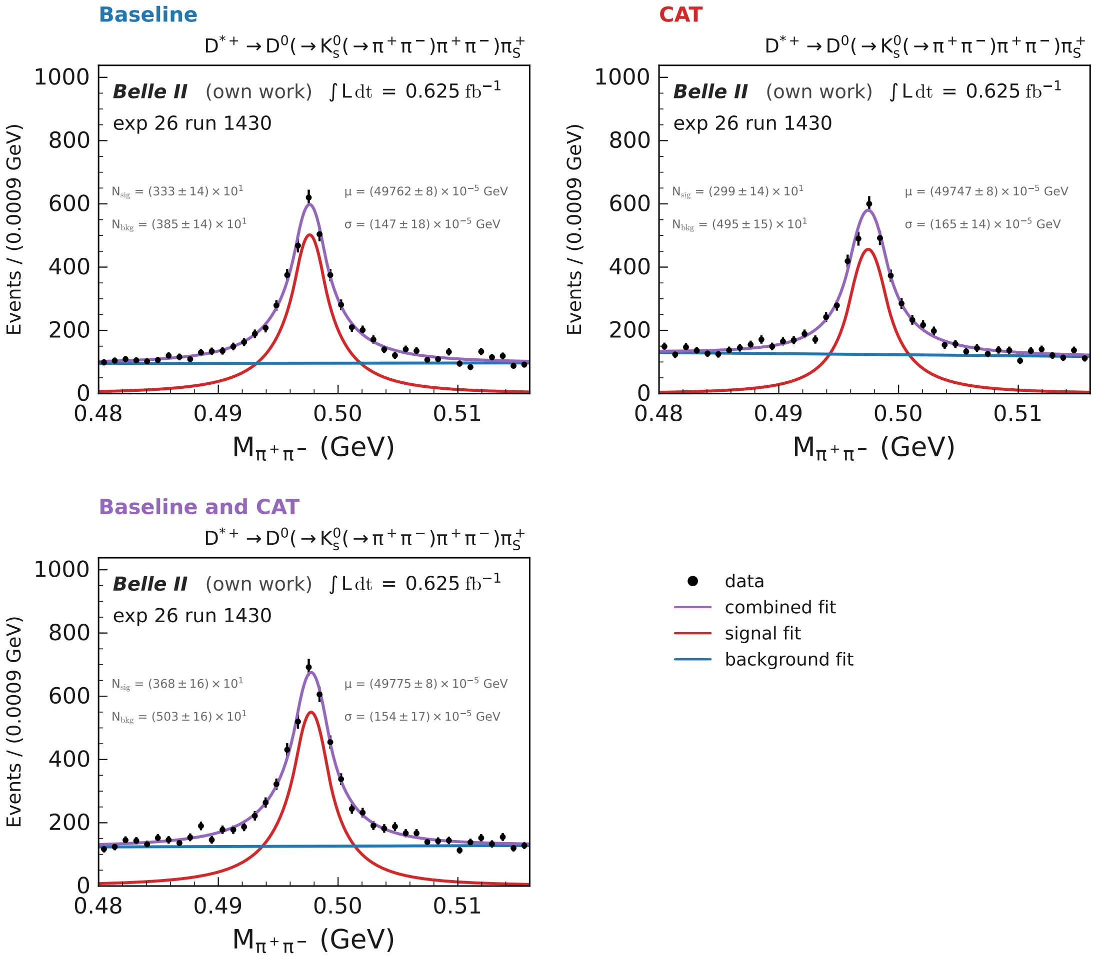


Figure 10.4: Reconstructed K_s^0 mass from two opposite charged pions $M_{\pi^+\pi^-}$ in the signal region of 0.48 GeV to 0.52 GeV for measured data for run 1430 from experiment 26. Top left shows the *Baseline Finder*, top right the *CAT Finder* and bottom shows the *Baseline Finder* followed by *CAT Finder*. The combined fit, as well as the individual signal and background parametrization, is added, and the most important fit parameters μ, σ, N_{sig} and N_{bkg} (see Section 8.1) are annotated.

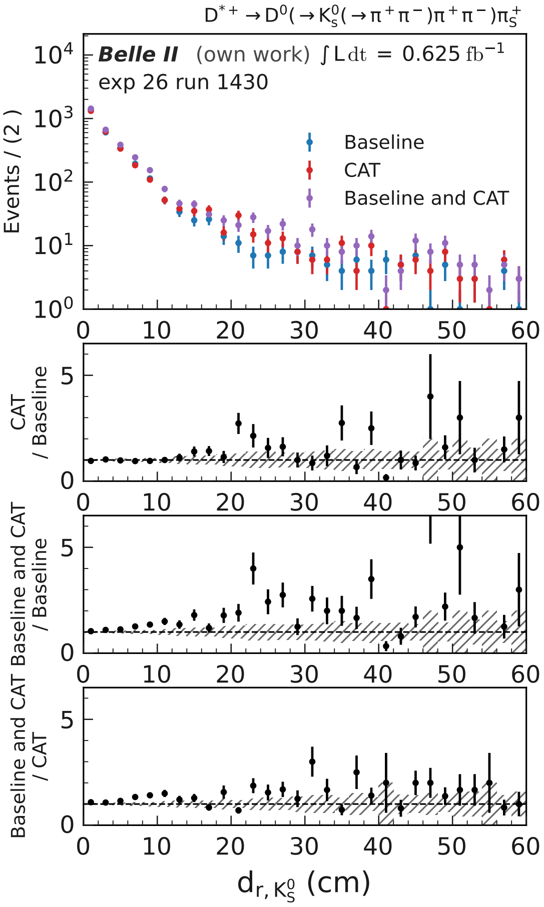


Figure 10.5: Transverse vertex distance $d_{\rho,K_s^0} = \sqrt{d_{x,K_s^0}^2 + d_{y,K_s^0}^2}$ from the interaction point of the reconstructed K_s^0 in the signal region of 494.8 MeV to 500.8 MeV for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *Baseline Finder* followed by *CAT Finder* (purple). The statistical uncertainties are correlated, as the same measured data is used for the different algorithms.

10.5 Hit Cleanup for *CAT Finder*

I describe the clipping of input features to better account for the signal distribution in Section 5.3.1. Instead of clipping the values, there is also the option of cutting at the specified value and removing the hits outside of the range from the model training and inference. This cut-based filtering approach is also used for the *Baseline Finder* (Section 5.2). Different cuts and the impact on the track fitting charge efficiency was investigated in detail in [84]. While the result shows that cutting does not improve the efficiency compared to clipping, it reduces training time of up to a factor of 2 depending on the cut values on high data beam backgrounds [84]. Beyond simple filtering using cut values on features of the CDC hits, more advanced filtering strategies can significantly reduce the computational load of track finding. Two approaches are currently available, FastBDT[34] based hit filtering introduced in `basf2` with `release-09`, and GNN-based cleanup [130].

For the expected high simulated beam background scenario (see Section 3.1.1), with on average 3180 beam background hits per event, applying the GNN-based cleanup prior to training reduces the training time by more than an order of magnitude. In addition to the speedup, both the *CAT Finder* and *Baseline Finder* algorithms show improved performance when the cleanup is applied. A detailed comparison is summarized in Table 10.1, with both algorithms having an improved full reconstruction efficiency for pions of 2 percentage points, with a decrease of 1 percentage point for the fake rate. While both the *CAT Finder* and GNN-based cleanup are using a `PyTorch Geometric` model, there is ongoing work to port both models to C++ using `ONNX`[129], which will reduce the execution time and memory consumption. These filtering methods are therefore crucial not only for improving reconstruction performance but also for to improve the training time for faster testing of different datasets and making combined track finding strategies computationally feasible.

in %	**π efficiency**	**$\mathfrak{r}_{\text{clone}}$**	**$\mathfrak{r}_{\text{fake}}$**	**training time/epoch**
Baseline Finder	86.2	3.5	7.5	-
with GNN cleanup	88.6	3.0	6.4	-
CAT Finder	86.7	5.6	7.8	≈ 120 min
with GNN cleanup	88.9	3.8	6.5	10.5 min

Table 10.1: Comparison of pion reconstruction efficiency, $\mathfrak{r}_{\text{clone}}$, and $\mathfrak{r}_{\text{fake}}$ metrics (in %) across different configurations for the high expected beam-background (see Section 3.1.1, for *Baseline Finder* and *CAT Finder* with and without the GNN-based hit cleanup comparison with Table 7.1). The training time per epoch is given for the *CAT Finder*.

10.6 Tuning the Subsequent CKF and Track Fitting Algorithm

One issue that I discussed in Section 8.2 is the extrapolation from the CDC tracks to the SVD. The CKFs responsible to attach the SVD clusters to the CDC track candidates (see Section 5.4) and should be tuned in future work.
Additionally, there are options within GENFIT2 that can reorder the hits assigned to a track candidate. While this is currently not implemented in basf2, this can be promising to to investigate to improve the fitting efficiency of curlers (see Section 6.1.1).
Other options are to replace GENFIT2 with A Common Tool for Tracking (ACTS)[131], which promises faster fitting in comparison. This would allow more iterations of the track fitting with loser requirements at the start, which might help in case the track finding seeds differ too much from the true track properties.

10.7 Combined SVD and CDC track finding

One possible approach to improve the incorporation of SVD clusters with CDC tracks is by tuning the CKF as discussed in the previous section. However, this does not address the more fundamental issue that it is often preferable to reject poor-quality track candidates altogether during the CDC track finding stage, as the SVD track finding can recover these tracks (see Fig. 8.18). This is particularly relevant for very low-momentum particles that only reach the first few superlayers and information about the trajectory from only the CDC hits is limited.

Optimizing this trade-off is difficult, as reflected in the complex design of the current tracking pipeline shown in Fig. 5.13. One simplification presented at [132] is instead combining the CDC and SVD track finding in a multi-modal GNN-based approach.
As the PXD depends on the CDC and SVD track finding to define the regions-of-interest for reading out clusters at the DAQ (see Section 2.2) and assign clusters (see Section 5.4), PXD clusters are not included in the track-finding stage. There are plans to include predictions on the region-of-interest for the PXD for readout and cluster attachment in future work.
The proposed tracking pipeline is shown in Fig. 10.6, combining the CDC track finding, the CKF extrapolation to attach the SVD clusters, the SVD track finding and the SVD extrapolation to attach CDC hits in one single step. The multi-modal GNN model architecture is shown in Fig. 10.7, keeping the structure of the current *CAT Finder* model and adding 3 additional layer after the combined input matrix of the CDC and SVD inputs.

This new algorithm is called *Belle II AI Track Finder* (*BAT Finder*). For the CDC, the updated inputs described in Section 10.1 are used. In the case of the SVD, the input consists of the three-dimensional hit information reconstructed by combining clusters from

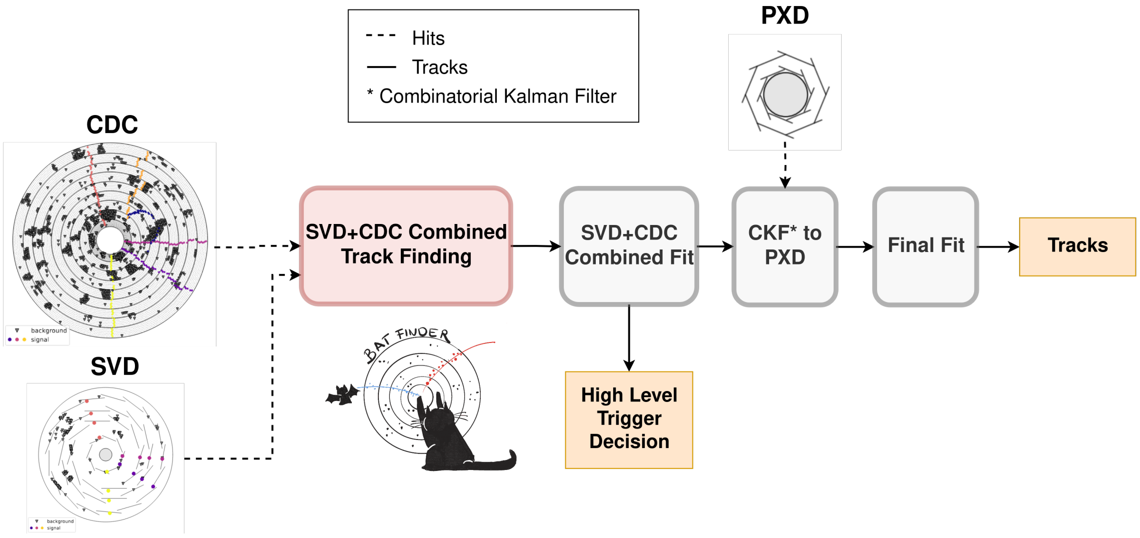


Figure 10.6: Proposed updated tracking pipeline using one step for combined SVD and |glscdc track finding compared to Fig. 5.13. Presented at [132].

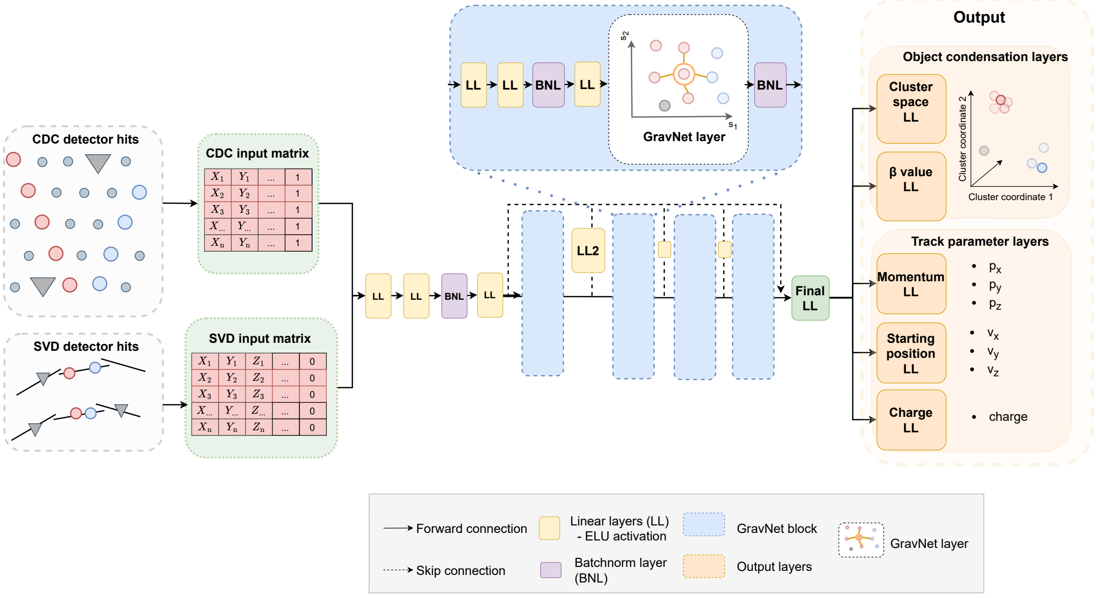
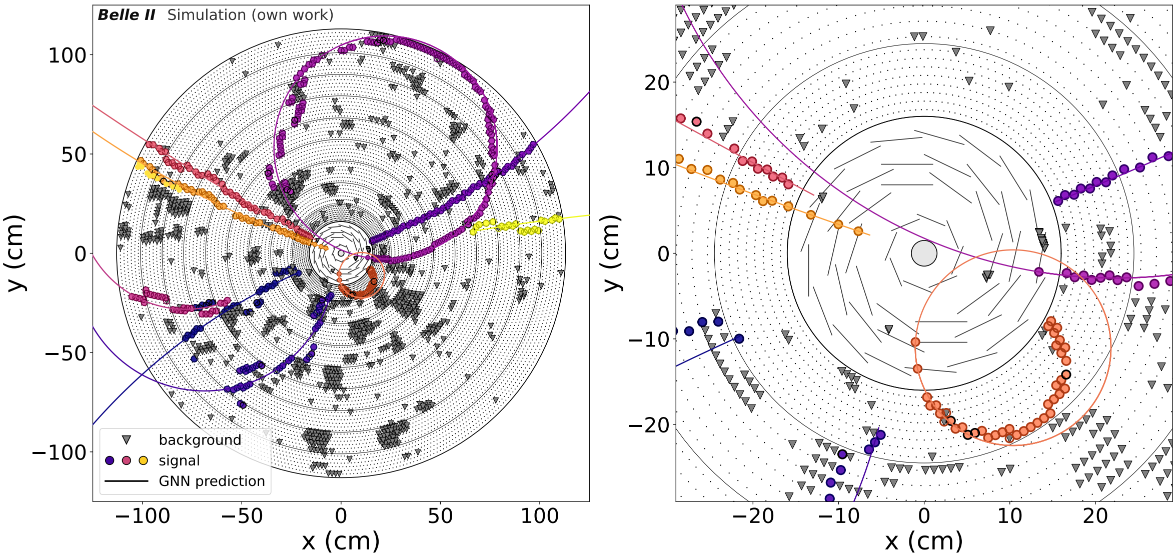


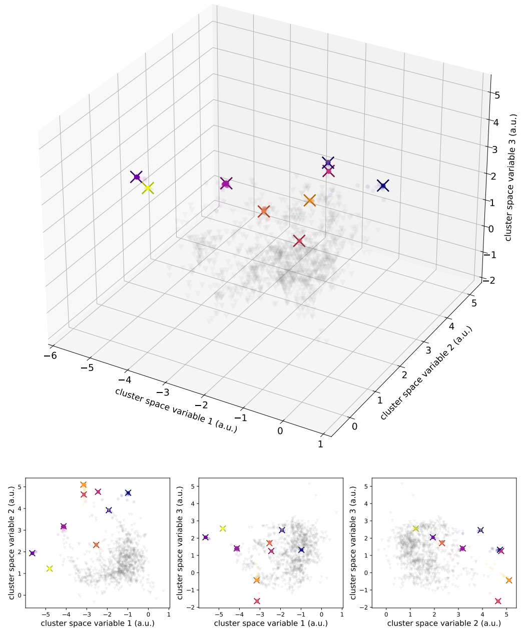
Figure 10.7: Illustration of the multi-modal GNN architecture (see Fig. 5.7 for comparison) including SVD and CDC inputs. Presented at [132].

opposite-side strips (Section 2.2). Each cluster contributes with its timing information to the corresponding strip hit. Additionally, the deposited charge and the signal-to-noise ratio of the clusters are included as input features [22]. One example prediction including the latent space is shown in Fig. 10.8.

Here, the low-momentum track only entering the first axial superlayer is found by the *BAT Finder*, showing that including the SVD is especially helpful for these very low momentum curlers that are also an issue for the extrapolation. Incorporating SVD information is also relevant for the reconstruction of curling tracks, as the z-coordinate



(a) Example event display in the x–y plane. Filled colored circular markers show signal hits, filled gray triangular markers show background hits. Markers with colored outlines are found by the GNN to belong to the same track object. The GNN predictions (colored lines) are drawn using the predicted starting point and three-momentum for the predicted particle charge, and the corresponding condensation point is marked by a colored cross.



(b) Top: 3D cluster space with condensation points marked by crosses. Bottom: 2D projections. Colors match Fig. 10.8a.

Figure 10.8: Event display (Fig. 10.8a) and latent cluster space (Fig. 10.8b) representation of one example event from category 11 (Table 3.2). Presented at [132].

provided by the SVD allows for a more reliable identification of the first trajectory arm and the track direction. Fig. 10.9 shows the track reconstruction results on category 8

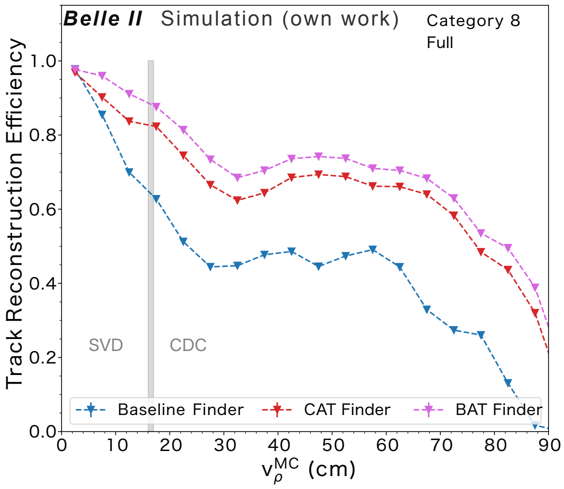


Figure 10.9: Track reconstruction efficiency over the displacement of *muons* from the category 8 samples (Table 3.2). *Baseline Finder* and *CAT Finder* results are for the full reconstruction, *BAT Finder* results are without including the PXD. Presented at [132].

Table 10.2: Pion reconstruction efficiency, clone rate, fake rate and wrong charge rate for category 5-7 fitted tracks over the full tracking acceptance. *Baseline Finder* and *CAT Finder* are included in the full reconstruction, *BAT Finder* results are without the PXD. Presented at [132].

(in %)	π **efficiency**	$\mathfrak{r}_{\textbf{clone}}$	$\mathfrak{r}_{\textbf{fake}}$	$\mathfrak{r}_{\textbf{wrong ch.}}$
Baseline Finder	$49.73^{+0.06}_{-0.06}$	$6.44^{+0.04}_{-0.04}$	$0.16^{+0.01}_{-0.01}$	$1.11^{+0.02}_{-0.02}$
CAT Finder	$69.35^{+0.05}_{-0.05}$	$4.62^{+0.03}_{-0.03}$	$0.20^{+0.01}_{-0.01}$	$2.87^{+0.02}_{-0.02}$
BAT Finder	$75.13^{+0.05}_{-0.05}$	$0.29^{+0.01}_{-0.01}$	$0.06^{+0.00}_{-0.00}$	$3.64^{+0.03}_{-0.03}$

samples (Table 3.2) and the summarized values are given in Table 10.2. The *BAT Finder*[*] achieves the π reconstruction efficiency of 75.13 % outperforming the full reconstruction including the *CAT Finder*[†] with 69.35 %. While the fake rate for the *BAT Finder* is the smallest, the wrong charge rate is increased, which might also be because the *BAT Finder* has the highest efficiency and finds hard to identify tracks. Physics evaluation and the implementation into basf2 are necessary next steps, however this first evaluation shows promising results for the combined SVD and CDC track finding.

Feasability Studies for Tracking with Degradating Detector

As discussed in Section 3.1.2, the performance of the CDC may degrade in the near future due to the Malter effect. Since this effect depends on the accumulated charge, the innermost axial layer with its smaller drift cells and shorter wires is affected the most. In the worst

---

[*] The PXD is excluded for the *BAT Finder*, which can influence the hit efficiency and hit purity requirements for track matching described in Chapter 4 as on average two additional clusters are included.

[†] The SVD track finding of basf2 is used here as described in Section 5.4.

case, the entire first axial superlayer may become unusable because the gain drops too low or the occupancy becomes too high to distinguish signal from noise.

Without the first axial superlayer, the *Baseline Finder* performance is massively reduced, and track finding would need to rely on the SVD. This, however, comes with the drawback of reduced purity unless the SVD track finding is further optimized. Replacing parts of the CDC is both costly and time-consuming, requiring a long shutdown of the experiment as was the case between run I and run II. This would also come with replacing the first axial superlayer with additional silicon layers (extending the SVD) to cope with the high beam background occupancy [34].

Demonstrating that tracking performance can be maintained or even improved with advanced algorithms is therefore crucial to avoid major unplanned downtimes. While I studied reduced CDC efficiency in Section 6.4.2, the complete loss of the first axial superlayer remains particularly challenging for the *CAT Finder*, especially for low-momentum particles or those leaving the detector early in the endcaps.

Promisingly, combining SVD and CDC information within the *BAT Finder* yields robust track finding even without the first axial superlayer, as shown in Fig. 10.10. Moreover, the multi-modal design of the *BAT Finder* allows for straightforward substitution of the first CDC inputs with SVD measurements, enabling studies of upgrade scenarios where the innermost CDC superlayer is replaced entirely by silicon.

This demonstrates that advanced tracking approaches such as the *BAT Finder* are not only essential for handling detector ageing but also provide a viable path for future Belle II tracking upgrades.

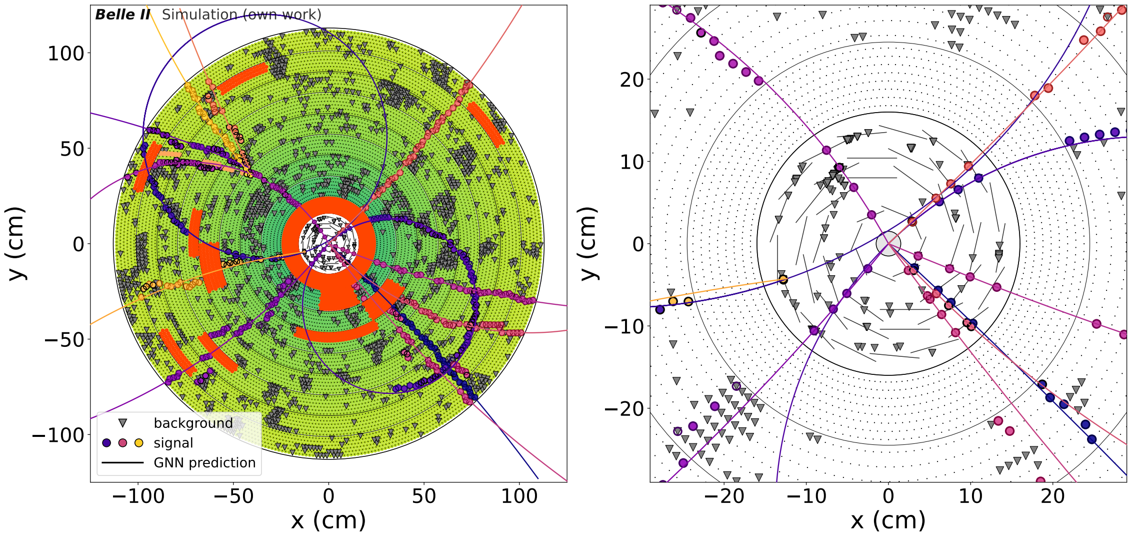


Figure 10.10: Example event display in the x–y plane showcasing the wire efficiency map where green stand for decreased efficiency and red for wires that cannot measure anymore (see Fig. 3.4 for details). Filled colored circular markers show signal hits, filled gray triangular markers show background hits. Markers with colored outlines are found by the GNN to belong to the same track object. The GNN predictions (colored lines) are drawn using the predicted starting point and three-momentum for the predicted particle charge. Presented at [132].

Chapter 11

Conclusion

In this thesis, I present the first end-to-end graph neural network-based tracking algorithm for a drift chamber detector in a realistic particle physics environment, evaluated on both simulated events and measured data from Belle II. I evaluate two models, the *CAT Finder* trained on a simulated dataset featuring rather low track multiplicity, and the *CAT B Finder*, which additionally includes $B\bar{B}$ events in the training to optimize for high track multiplicity.

The evaluation for the Belle II drift chamber (CDC) shows that the *CAT Finder* achieves superior performance for both displaced tracks and tracks leaving in the detector endcaps. For a long-lived, massive particle in the GeV range with a lifetime of 21.5 cm, the *CAT Finder* reconstructs both tracks in the event in 87.2% compared to 44.9% for the *Baseline Finder*. In addition, the *CAT Finder* achieves a smaller fake rate of 2.5% compared to the 3.3% of the *Baseline Finder*.

The *CAT Finder* also improves hit efficiency for curling tracks that leave and re-enter the CDC multiple times. This is important because beam-background occupancy per event is estimated from the number of hits remaining after track finding, which is a relevant quantity in many physics analyses.

The *CAT B Finder*, optimized for high track multiplicity, shows even better performance for prompt particles originating from the interaction point. Across the full detector acceptance, it achieves an integrated efficiency of 84.2% for prompt pions from $B\bar{B}$ events, compared to 74.7% achieved by both the *CAT Finder* and the *Baseline Finder*.

Both *CAT Finder* models are robust against detector aging effects, which are already present in the CDC and may worsen in the near future, and they recover performance relative to the *Baseline Finder*. This demonstrates that the *CAT Finder*-based algorithms provide a reliable and future-proof solution for track reconstruction at Belle II.

In addition to the CDC-only evaluation, I demonstrate the successful implementation and validation within the full reconstruction chain of Belle II. The *CAT B Finder* outperforms the *Baseline Finder* by 1 percentage point on the most relevant tracking metric, the prompt reconstruction efficiency from $B\bar{B}$ events. It exceeds the *Baseline Finder* by 8 percentage points for reconstructed, displaced $K_S^0 \to \pi^+\pi^-$ decays while achieving a lower fake rate. The `C++` implementation [96] of this algorithm will be integrated in the next full release of the Belle II reconstruction software, enabling its use in addition to the existing algorithm. After validating the track quality and implementation, I demonstrate that the algorithm can be used for physics analysis. Using my algorithm in combination with the other subdetectors on measured data, both the *CAT Finder* and *CAT B Finder* successfully reconstruct electron and muon tracks from J/ψ decays. For displaced $K_S^0 \to \pi^+\pi^-$ decays, the *CAT B Finder* achieves the largest signal yield on data out of all three algorithms, and an especially high yield for displaced tracks. However, the increased background yield for the *CAT B Finder* leads to a slightly lower approximated significance compared to the *Baseline Finder*.

For further improvements, I identify discrepancies between simulation and measured data for all three algorithms. For my algorithm, future work can address this by either incorporating measured data into training or by adding noise to the input features. I also show that applying filtering methods to the CDC hits reduces training time, enabling my proposed studies to mitigate simulation–data differences more efficiently. The *CAT Finder* model is currently being implemented in `ONNX` for a full `C++` implementation, which allows faster track finding compared to the current implementation. This enables running the *Baseline Finder* and the *CAT Finder* together, allowing the *CAT Finder* to focus entirely on recovering performance due to detector ageing and reconstructing displaced tracks. Lastly, the graph neural network-based algorithm presented here is suitable for irregular detector geometries. This setup enables the combination of the detectors used in Belle II track finding, which reduces background yields [132] and is a key direction for future work.

Finally, I study the sensitivity of the *CAT Finder* for a long-lived spin-0 particle S in $B \to K^+ S$. For lifetimes above 10 cm, the expected limits improve by up to a factor of 2, which is approximately equivalent to collecting four times more data when using the *Baseline Finder*.

To conclude this thesis, I report the training setup, model architecture, training loss, and the pre- and postprocessing for the track candidates, with the software code available in [133]. The algorithm and part of the CDC-only evaluation are published in [1]. The *CAT Finder* shown in this thesis is the first end-to-end machine learning tracking algorithm utilized in a realistic particle physics environment.

Danksagung

Bei so einem großen Projekt wie einer Doktorarbeit ist die Unterstützung durch die Referenten und Kollegen unglaublich wichtig. Deshalb bin ich sehr glücklich, dass Torben und Markus gegen Ende meiner Masterarbeit an das ETP gewechselt sind und mir die Promotion angeboten haben. Torben hat mich immer unterstützt und ich konnte ihn jederzeit um Rat fragen. Er hat mir die notwendigen Möglichkeiten gegeben, stetig mehr Verantwortung zu übernehmen. Dafür bin ich ihm sehr dankbar. Markus als Korreferent danke ich für alle Diskussionen, die mich vorangebracht haben und die Outreach Events, die sowohl mir als auch den Besuchern immer viel Spaß gemacht haben.

Meinen Kolleginnen und Kollegen gebührt ebenfalls ein großer Dank: für die Zusammenarbeit und fachliche Unterstützung und die wunderbare Zeit während der Promotion, die gemeinsamen Urlaube und vor allem den nächtlichen Karaoke-Sessions.
Ganz besonders möchte ich Giacomo danken, der mich kontinuierlich unterstützt hat, insbesondere bei der gemeinsamen Implementierung des Software-Codes. In stressigen Momenten war er jederzeit für mich da. Dadurch konnte ich ruhig und gelassen weiterarbeiten. Als nächstes möchte ich Isabel danken, die im selben Zeitraum promovierte und mit mir durch alle Phasen, auch die Tiefs, ging. Ihre Hilfe hat meinen Einstieg in das neue Thema unglaublich erleichtert. Gemeinsam konnten wir über alles reden, egal ob es um die Arbeit ging oder darüber hinaus. Unsere gemeinsamen Kaffee-Sessions waren ein elementarer Bestandteil für das Gelingen dieser Promotion. Mit Alex habe ich zusammen auf der viel zu warmen Südseite des ETP-Hochhauses gelitten, bevor wir dann endlich auf die kühlere Nordseite wechseln durften. Danke für die schöne gemeinsame Bürozeit, aber auch die gemeinsamen Erlebnisse während der CSC und den Aufenthalten in Japan. Meinen beiden anderen Bürokollegen - Sally und Patrick - möchte ich ebenso für die ermunternden Diskussionen und fachliche Unterstützung danken. Bei Computing-Fragen war Matthias immer eine große Hilfe. Jonas, Tristan, Lennard und Greta danke ich für die Diskussionen, die gemeinsame Zeit während der Kaffeepausen und Dienstreisen sowie das Korrekturlesen meiner Arbeit. Ich danke all meinen Kollegen am ETP für die wirklich schöne Zeit, die ich nicht missen möchte.

Der größte Dank gilt meiner Familie, meinen Eltern Carmen und Alfred, die mein In-

teresse für Physik geweckt haben und mir mein Studium erst ermöglicht haben. Zusammen mit meinem Bruder Michael haben sie mich jederzeit unterstützt. Zuletzt möchte ich Christoph danken, der immer für mich da war, insbesondere in den schwierigen Phasen. Er hat mich motiviert, wenn ich am Zweifeln war und mich davon überzeugt, dass ich alles schaffen kann.

Bibliography

[1] L. Reuter et al. “End-to-End Multi-track Reconstruction Using Graph Neural Networks at Belle II”. In: *Comput. Softw. Big Sci.* 9 (2025), p. 6. DOI: `10.1007/s41781-025-00135-6`.

[2] H. Aihara et al. “The Belle II Detector Upgrades Framework Conceptual Design Report”. In: (2024). arXiv: `2406.19421 [hep-ex]`.

[3] P. Ecker. “Search for a dark Higgs boson produced in association with inelastic dark matter at the Belle II experiment”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2024. 215 pp. DOI: `10.5445/IR/1000176696`.

[4] N. Arkani-Hamed et al. “A theory of dark matter”. In: *Phys. Rev. D* 79 (1 2009), p. 015014. DOI: `10.1103/PhysRevD.79.015014`.

[5] L. Roszkowski, E. M. Sessolo, and S. Trojanowski. “WIMP dark matter candidates and searches—current status and future prospects”. In: *Rep. Prog. Phys.* 81 (2018), p. 066201. DOI: `10.1088/1361-6633/aab913`.

[6] T. Ferber, C. Garcia-Cely, and K. Schmidt-Hoberg. “Belle II sensitivity to long–lived dark photons”. In: *Phys. Lett. B* 833 (2022), p. 137373. DOI: `10.1016/j.physletb.2022.137373`.

[7] A. Filimonova, R. Schäfer, and S. Westhoff. “Probing dark sectors with long-lived particles at Belle II”. In: *Phys. Rev. D* 101 (9 2020), p. 095006. DOI: `10.1103/PhysRevD.101.095006`.

[8] B. Batell, M. Pospelov, and A. Ritz. “Probing a secluded U(1) at B factories”. In: *Phys. Rev. D* 79 (11 2009), p. 115008. DOI: `10.1103/PhysRevD.79.115008`.

[9] M. Duerr et al. “Invisible and displaced dark matter signatures at Belle II”. In: *J. High Energy Phys.* 02 (2020), p. 039. DOI: `10.1007/JHEP02(2020)039`.

[10] M. Duerr et al. “Long-lived Dark Higgs and Inelastic Dark Matter at Belle II”. In: *J. High Energy Phys.* 04 (2021), p. 146. DOI: `10.1007/JHEP04(2021)146`.

[11] E. Kou et al. “The Belle II Physics Book”. In: *Prog. Theor. Exp. Phys.* 2019 (2019). [Erratum: Prog. Theor. Exp. Phys. 2020, 029201 (2020)], p. 123C01. DOI: `10.1093/ptep/ptz106`.

[12] **Belle II Collaboration**, I. Adachi et al. “Search for a dark Higgs boson produced in association with inelastic dark matter at the Belle II experiment”. In: *preprint* (2025). arXiv: `2505.09705 [hep-ex]`.

[13] **Belle II Collaboration**, I. Adachi et al. “Search for a long-lived spin-0 mediator in b→s transitions at the Belle II experiment”. In: *Phys. Rev. D* 108 (2023), p. L111104. DOI: `10.1103/PhysRevD.108.L111104`.

[14] J. Kieseler. “Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data”. In: *Eur. Phys. J. C* 80 (2020), p. 886. DOI: `10.1140/epjc/s10052-020-08461-2`.

[15] **SuperKEKB accelerator team**, K. Akai, K. Furukawa, and H. Koiso. “SuperKEKB collider”. In: *Nucl. Instrum. Methods Phys. Res. A* 907 (2018), pp. 188–199. DOI: `10.1016/j.nima.2018.08.017`.

[16] T. Abe et al. “Belle II Technical Design Report”. In: *preprint* (2010). arXiv: `1011.0352 [physics.ins-det]`.

[17] **SuperB**, M. Bona et al. “SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report”. In: *preprint* (2007). arXiv: `0709.0451 [hep-ex]`.

[18] *SuperKEKB/Belle II Complete 2024 Operations*. URL: `https://www2.kek.jp/ipns/en/news/7015/?simply_static_page=1642969` Access date: 2025-09-01.

[19] D. Matvienko. “The Belle II experiment: status and physics program”. In: *EPJ Web Conf.* 191 (2018), p. 02010. DOI: `10.1051/epjconf/201819102010`.

[20] H. Ye et al. “Commissioning and performance of the Belle II pixel detector”. In: *Nucl. Instrum. Methods Phys. Res. A* 987 (2021), p. 164875. DOI: `10.1016/j.nima.2020.164875`.

[21] P. Ahlburg et al. “The new and complete Belle II DEPFET pixel detector: Commissioning and previous operational experience”. In: *Nucl. Instrum. Methods Phys. Res. A* 1068 (2024), p. 169763. DOI: `10.1016/j.nima.2024.169763`.

[22] **Belle II SVD Collaboration**, K. Adamczyk et al. “The design, construction, operation and performance of the Belle II silicon vertex detector”. In: *J. Inst.* 17 (2022), P11042. DOI: `10.1088/1748-0221/17/11/P11042`.

[23] N. Taniguchi. “Central Drift Chamber for Belle-II”. In: *J. Inst.* 12 (2017), p. C06014. DOI: `10.1088/1748-0221/12/06/C06014`.

[24] T. Dong et al. “Calibration and alignment of the Belle II central drift chamber”. In: *Nucl. Instrum. Methods Phys. Res. A* 930 (2019), pp. 132–141. DOI: `10.1016/j.nima.2019.03.072`.

[25] H. Atmacan et al. “The imaging Time-of-Propagation detector at Belle II”. In: *Nucl. Instrum. Meth. A* 1080 (2025), p. 170627. DOI: `10.1016/j.nima.2025.170627`.

[26] K. Uno. “Operation and performance of the Belle II Aerogel RICH detector”. In: *Nucl. Instrum. Methods Phys. Res. A* 1056 (2023), p. 168635. DOI: `10.1016/j.nima.2023.168635`.

[27] A. Kuzmin. “Electromagnetic calorimeter of Belle II”. In: *Nucl. Instrum. Methods Phys. Res. A* 958 (2020), p. 162235. DOI: `10.1016/j.nima.2019.05.076`.

[28] T. Aushev et al. “A scintillator based endcap K_L and muon detector for the Belle II experiment”. In: *Nucl. Instrum. Methods Phys. Res. A* 789 (2015), pp. 134–142. DOI: `10.1016/j.nima.2015.03.060`.

[29] C. Ketter et al. “Design and commissioning of readout electronics for a K_L^0 and μ detector at the Belle II experiment”. In: *Nucl. Instrum. Methods Phys. Res. A* 1082 (2026), p. 170893. DOI: `10.1016/j.nima.2025.170893`.

[30] H. Kolanoski and N. Wermes. *Particle detectors: fundamentals and applications.* English. 1st ed. Oxford University Press, 2020, p. 950. DOI: `10.1093/oso/9780198858362.001.0001`.

[31] Z. J. Liptak et al. “Measurements of beam backgrounds in SuperKEKB Phase 2”. In: *Nucl. Instrum. Methods Phys. Res. A* 1040 (2022), p. 167168. DOI: `10.1016/j.nima.2022.167168`.

[32] A. Natochii et al. “Beam Background Expectations for Belle II at SuperKEKB”. In: *preprint* (2022). arXiv: `2203.05731 [hep-ex]`.

[33] L. Malter. “Thin Film Field Emission”. In: *Phys. Rev.* 50 (1 1936), pp. 48–58. DOI: `10.1103/PhysRev.50.48`.

[34] B-factory Programme Advisory Committee. *Full report of the annual review meeting — 3–5 march 2025, tsukuba, japan.* BELLE2-REPORT-2025-002. Belle II Collaboration, 2025. URL: `https://docs.belle2.org/files/4563/BELLE2-REPORT-2025-002/1/BELLE2-REPORT-2025-002.pdf`.

[35] F. Herdtweck. *Robustness Studies for GNN-tracking at Belle II.* Karlsruhe Institute of Technology (KIT). 2025.

[36] Belle II Collaboration. *CDC Operational Issues.* Version 1. 2024. URL: `https://agenda.infn.it/event/43143/contributions/242975/attachments/125922/185809/240926%20CDC%20Operational%20Issues%20v1%20.pdf`.

[37] **Particle Data Group Collaboration**, S. Navas et al. “Review of Particle Physics”. In: *Phys. Rev. D* 110 (3 2024), p. 030001. DOI: `10.1103/PhysRevD.110.030001`.

[38] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Methods Phys. Res. A* 462 (2001), pp. 152–155. DOI: `10.1016/S0168-9002(01)00089-4`.

[39] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: `10.1016/j.cpc.2015.01.024`.

[40] S. Jadach, B. F. L. Ward, and Z. Was. “The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions”. In: *Comput. Phys. Commun.* 130 (2000), p. 260. DOI: `10.1016/S0010-4655(00)00048-5`.

[41] S. Jadach, J. H. Kuhn, and Z. Was. “TAUOLA: A library of Monte Carlo programs to simulate decays of polarized tau leptons”. In: *Comput. Phys. Commun.* 64 (1990), p. 275. DOI: `10.1016/0010-4655(91)90038-M`.

[42] G. Balossini et al. “Photon pair production at flavour factories with per mille accuracy”. In: *Phys. Lett. B* 663 (2008), pp. 209–213. DOI: `https://doi.org/10.1016/j.physletb.2008.04.007`.

[43] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *J. High Energy Phys.* 07 (2014), p. 079. DOI: `10.1007/JHEP07(2014)079`.

[44] S. S. Dreyer. *Search for a long-lived spin-0 particle in $b \to s$ quark transitions at the Belle II experiment.* Universitaet Hamburg (Germany), 2023.

[45] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak interactions with lepton-hadron symmetry”. In: *Phys. Rev. D* 2 (7 1970), pp. 1285–1292. DOI: `10.1103/PhysRevD.2.1285`.

[46] **Belle II Tracking Group**, V. Bertacchi et al. “Track finding at Belle II”. In: *Comput. Phys. Commun.* 259 (2021), p. 107610. DOI: `10.1016/j.cpc.2020.107610`.

[47] P. Billoir. “Progressive track recognition with a Kalman-like fitting procedure”. In: *Comput. Phys. Commun.* 57 (1989), pp. 390–394. DOI: `10.1016/0010-4655(89)90249-X`.

[48] C. Hoppner et al. “A Novel Generic Framework for Track Fitting in Complex Detector Systems”. In: *Nucl. Instrum. Methods Phys. Res. A* 620 (2010), pp. 518–525. DOI: `10.1016/j.nima.2010.03.136`.

[49] J. Rauch and T. Schlüter. “GENFIT - a Generic Track-Fitting Toolkit”. In: *J. Phys. Conf. Ser.* 608 (2015), p. 012042. DOI: `10.1088/1742-6596/608/1/012042`.

[50] T. Bilka et al. “Implementation of GENFIT2 as an experiment independent track-fitting framework”. In: *preprint* (2019). arXiv: `1902.04405 [physics.data-an]`.

[51] Jojosito et al. *GenFit/GenFit.* `https://doi.org/10.5281/zenodo.10301439`.

[52] E. Lund et al. “Track parameter propagation through the application of a new adaptive Runge-Kutta-Nyström method in the ATLAS experiment”. In: *Journal of Instrumentation* 4 (2009), P04001. DOI: `10.1088/1748-0221/4/04/P04001`.

[53] T. Alexopoulos et al. “Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers”. In: *Nucl. Instrum. Methods Phys. Res. A* 592 (2008), pp. 456–462. DOI: `10.1016/j.nima.2008.04.038`.

[54] A. Glazov et al. "Filtering tracks in discrete detectors using a cellular automaton". In: *Nucl. Instrum. Methods Phys. Res. A* 329 (1993), pp. 262–268. DOI: `10.1016/0168-9002(93)90945-E`.

[55] N. Braun. "Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment". PhD thesis. Karlsruhe Institute of Technology, 2019. DOI: `10.5445/IR/1000089317`.

[56] HEP ML Community. *A Living Review of Machine Learning for Particle Physics.* `https://iml-wg.github.io/HEPML-LivingReview/`. 2025.

[57] Z. Zou et al. "Object Detection in 20 Years: A Survey". In: *Proceedings of the IEEE* 111 (2023), pp. 257–276. DOI: `10.1109/JPROC.2023.3238524`.

[58] J. Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *Conf. Comput. Vis. Pattern Recognit.* 2016, pp. 779–788. DOI: `10.1109/CVPR.2016.91`.

[59] N. Carion et al. "End-to-end object detection with transformers". In: *European conference on computer vision.* 2020, pp. 213–229. arXiv: `2005.12872`.

[60] P. Reiser et al. "Graph neural networks for materials science and chemistry". In: *Commun. Mater.* 3 (2022), p. 93. DOI: `10.1038/s43246-022-00315-6`.

[61] D. Ahmedt-Aristizabal et al. "Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future". In: *Sensors* 21 (2021), p. 4758. DOI: `10.3390/s21144758`.

[62] J. Zhou et al. "Graph neural networks: A review of methods and applications". In: *AI Open* 1 (2020), pp. 57–81. DOI: `10.1016/j.aiopen.2021.01.001`.

[63] Y. Wang et al. "Dynamic Graph CNN for Learning on Point Clouds". In: *preprint* (2018). arXiv: `1801.07829 [cs.CV]`.

[64] S. R. Qasim et al. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". In: *Eur. Phys. J. C* 79 (2019), p. 608. DOI: `10.1140/epjc/s10052-019-7113-9`.

[65] J. Pata et al. "MLPF: efficient machine-learned particle-flow reconstruction using graph neural networks". In: *Eur. Phys. J. C* 81 (2021), p. 381. DOI: `10.1140/epjc/s10052-021-09158-w`.

[66] I. Haide. "A Real-Time Graph Neural Network Trigger Algorithm for the Belle II Electromagnetic Calorimeter". PhD thesis. Karlsruher Institut für Technologie (KIT), 2025. 251 pp. DOI: `10.5445/IR/1000184927`.

[67] F. Wemmer et al. "Photon reconstruction in the belle ii calorimeter using graph neural networks". In: *Comput. Softw. Big Sci.* 7 (2023). DOI: `10.1007/s41781-023-00105-w`.

[68] J. Kahn et al. “Learning tree structures from leaves for particle decay reconstruction”. In: *Mach. Learn. Sci. Techn.* 3 (2022), p. 035012. DOI: `10.1088/2632-2153/ac8de0`.

[69] N. Choma et al. “Track Seeding and Labelling with Embedded-space Graph Neural Networks”. In: *preprint* (2020). arXiv: `2007.00149 [physics.ins-det]`.

[70] S. Caillou et al. “Novel fully-heterogeneous GNN designs for track reconstruction at the HL-LHC”. In: *EPJ Web Conf.* 295 (2024), p. 09028. DOI: `10.1051/epjconf/202429509028`.

[71] X. Ju et al. “Performance of a geometric deep learning pipeline for HL-LHC particle tracking”. In: *Eur. Phys. J. C* 81 (2021), p. 876. DOI: `10.1140/epjc/s10052-021-09675-8`.

[72] K. Lieret et al. “High Pileup Particle Tracking with Object Condensation”. In: *preprint* (2023). arXiv: `2312.03823 [physics.data-an]`.

[73] S. Amrouche et al. “The Tracking Machine Learning challenge : Accuracy phase”. In: *The NeurIPS ’18 Competition.* 2019, pp. 231–264. DOI: `10.1007/978-3-030-29135-8_9`.

[74] S. Amrouche et al. “The Tracking Machine Learning Challenge: Throughput Phase”. In: *Comput. Softw. Big Sci.* 7 (2023), p. 1. DOI: `10.1007/s41781-023-00094-w`.

[75] A. Correia et al. “Graph Neural Network-Based Track Finding in the LHCb Vertex Detector”. In: (2024). arXiv: `2407.12119 [physics.ins-det]`.

[76] A. Akram and X. Ju. “Track Reconstruction using Geometric Deep Learning in the Straw Tube Tracker (STT) at the PANDA Experiment”. In: *preprint* (2022). arXiv: `2208.12178 [hep-ex]`.

[77] X. Jia et al. “BESIII track reconstruction algorithm based on machine learning”. In: *EPJ Web Conf.* 295 (2024), p. 09006. DOI: `10.1051/epjconf/202429509006`.

[78] F. Kaneko et al. “Extracting Signal Electron Trajectories in the COMET Phase-I Cylindrical Drift Chamber Using Deep Learning”. In: (2024). arXiv: `2408.04795 [hep-ex]`.

[79] T.-Y. Lin et al. “Focal Loss for Dense Object Detection”. In: *preprint* (2018). arXiv: `1708.02002 [cs.CV]`.

[80] Z. Tian et al. “FCOS: Fully Convolutional One-Stage Object Detection”. In: *preprint* (2019). arXiv: `1904.01355 [cs.CV]`.

[81] C. Zhu, Y. He, and M. Savvides. “"feature selective anchor-free module for single-shot object detection"”. In: *preprint* (2019). arXiv: `1903.00621 [cs.CV]`.

[82] A. Huang et al. “A Language Model for Particle Tracking”. In: (2024). arXiv: `2402.10239 [hep-ph]`.

[83] S. Caron et al. “TrackFormers: In Search of Transformer-Based Particle Tracking for the High-Luminosity LHC Era”. In: (2024). arXiv: `2407.07179 [hep-ex]`.

[84] J. Bauer. *Input Feature Optimization for Track Reconstruction using Graph Neural Networks at Belle II.* Karlsruhe Institute of Technology (KIT). 2024.

[85] Belle II Collaboration. *Tracking Documentation — Belle II Software.* `https://software.belle2.org/release-08-01-10/sphinx/tracking/doc/index.html`. Accessed: 2025-07-09. 2025.

[86] M. Fey and J. E. Lenssen. "Fast Graph Representation Learning with PyTorch Geometric". In: (2019). arXiv: `1903.02428 [cs.LG]`.

[87] S. Ioffe and C. Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *preprint* (2015). arXiv: `1502.03167 [cs.LG]`.

[88] M. Lin, Q. Chen, and S. Yan. "Network In Network". In: *preprint* (2014). arXiv: `1312.4400 [cs.NE]`.

[89] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)". In: *preprint* (2016). arXiv: `1511.07289 [cs.LG]`.

[90] K. He et al. "Deep residual learning for image recognition". In: *Conf. Comput. Vis. Pattern Recognit.* 2016, pp. 770–778. arXiv: `1512.03385 [cs.CV]`.

[91] J. Ortner. *Latent Space Clustering for Tracking in Belle II.* Karlsruhe Institute of Technology (KIT). 2025.

[92] F. Ohlheiser. *Track reconstruction with Hough transformation for low momentum particles at Belle II.* Karlsruhe Institute of Technology (KIT). 2024.

[93] L. Li et al. "Hyperband: a novel bandit-based approach to hyperparameter optimization". In: *preprint* (2018). arXiv: `1603.06560[cs]`.

[94] L. Biewald. *Experiment Tracking with Weights and Biases.* Software available from `https://wandb.ai/site/`. 2020.

[95] R. Frühwirth et al. "Track finding in silicon trackers with a small number of layers". In: *Nucl. Instrum. Methods Phys. Res. A*. Vienna Conference on Instrumentation 2013 732 (2013), pp. 95–98. DOI: `10.1016/j.nima.2013.06.035`.

[96] Y. Klügl. *Development of a C++/Python Interface for a New Track Finding Algorithm in Belle II.* Karlsruhe Institute of Technology (KIT). 2024. DOI: `10.5281/zenodo.10895311`.

[97] **Belle II Collaboration**, I. Adachi et al. "Measurement of the branching fraction, polarization, and time-dependent CP asymmetry in $B^0 \to \rho+\rho$- decays and constraint on the CKM angle $\phi 2$". In: *Phys. Rev. D* 111 (2025), p. 092001. DOI: `10.1103/PhysRevD.111.092001`.

[98] **Belle and Belle II Collaborations**, I. Adachi et al. “Measurement of the time-integrated CP asymmetry in $D^0 \to K^0_S K^0_S$ decays using opposite-side flavor tagging at Belle and Belle II”. In: *Phys. Rev. D* 112 (2025), p. 012017. DOI: `10.1103/8x1h-39dp`.

[99] **Belle II Collaboration**, I. Adachi et al. “Measurement of the time-integrated CP asymmetry in $D^0 \to \pi^0\pi^0$ decays at Belle II”. In: *Phys. Rev. D* 112 (2025), p. 012006. DOI: `10.1103/vg9c-xvdc`.

[100] **Belle II Collaboration**, I. Adachi et al. “Measurement of the CP asymmetry in $D^+ \to \pi^+\pi^0$ decays at Belle II”. In: *Phys. Rev. D* 112 (3 2025), p. L031101. DOI: `10.1103/3kqk-1dm4`.

[101] **Belle-II**, M. Abumusabh et al. “Measurement of the CP asymmetry in $D^0 \to \pi^+\pi^-\pi^0$ decays at Belle II”. In: (2025). arXiv: `2510.21224 [hep-ex]`.

[102] **Belle II Collaboration**, I. Adachi et al. “Search for $B^0 \to K^{*0}\tau^+\tau^-$ decays at the Belle II experiment”. In: (2025). arXiv: `2504.10042 [hep-ex]`.

[103] **Belle and Belle II Collaborations**, I. Adachi et al. “Search for lepton flavor-violating decay modes $\mathrm{B}^0 \to \mathrm{K}^{*0}\tau^\pm\ell^\mp$ ($\ell = \mathrm{e}, \mu$) with hadronic B-tagging at Belle and Belle II”. In: *J. High Energy Phys.* 08 (2025), p. 184. DOI: `10.1007/JHEP08(2025)184`.

[104] **Belle II Collaboration**, L. Aggarwal and others. “Test of light-lepton universality in the rates of inclusive semileptonic B-meson decays at belle ii”. In: *Phys. Rev. Lett.* 131 (5 2023), p. 051804. DOI: `10.1103/PhysRevLett.131.051804`.

[105] **Belle II Collaboration**, I. Adachi et al. “Charged-hadron identification at Belle II”. In: (2025). arXiv: `2506.04355 [hep-ex]`.

[106] K. Pearson. “On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”. In: *Lond. Edinb. Dublin Philos. Mag. J. Sci* 50 (1900), pp. 157–175.

[107] J. Eschle et al. “zfit: Scalable pythonic fitting”. In: *SoftwareX* 11 (2020), p. 100508. DOI: `https://doi.org/10.1016/j.softx.2020.100508`.

[108] J. E. Gaiser. *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'*. Stanford University, 1983.

[109] T. Skwarnicki. “A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances”. Cracow, INP. PhD thesis. 1986.

[110] Y. Xiao, X. Shi, and Y. Guo. *Study of charge asymmetry in CDC*. Belle2 Internal CDC meeting. 2024. URL: `https://indico.belle2.org/event/14021/contributions/86478/attachments/32037/47357/2024_12_19_CDC_meeting.pdf`.

[111] G. C. Fox and S. Wolfram. “Observables for the analysis of event shapes in e+ e- annihilation and other processes”. In: *Phys. Rev. Lett.* 41 (1978), pp. 1581–1585. DOI: `10.1103/PhysRevLett.41.1581`.

[112] G. C. Fox and S. Wolfram. “Event Shapes in e+ e- Annihilation”. In: *Nucl. Phys. B* 149 (1979), pp. 413–496. DOI: `10.1016/0550-3213(79)90120-2`.

[113] *Search for a long-lived spin-0 mediator in $b \to s$ transitions at the Belle II experiment.* HEPData (collection). 2024. DOI: `10.17182/hepdata.147283`.

[114] **Belle II Collaboration**, I. Adachi et al. “Measurement of the branching fraction and CP-violating asymmetry of the decay $B^0 \to \pi^0\pi^0$ using 387 million $\Upsilon(4S)$ decays in Belle II data”. In: *Phys. Rev. D* 111 (7 2025), p. L071102. DOI: `10.1103/PhysRevD.111.L071102`.

[115] S. Banerjee et al. “Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023”. In: *preprint* (2024). arXiv: `2411.18639 [hep-ex]`.

[116] M. Marinangeli et al. *scikit-hep/hepstats: v0.9.2.* Version v0.9.2. 2025. DOI: `10.5281/zenodo.14973858`.

[117] A. L. Read. “Presentation of search results: the CLs technique”. In: *J. Phys. G Nucl. Part. Phys.* 28 (2002), p. 2693. DOI: `10.1088/0954-3899/28/10/313`.

[118] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: `10.1140/epjc/s10052-011-1554-0`.

[119] M. W. Winkler. “Decay and detection of a light scalar boson mixing with the higgs boson”. In: *Phys. Rev. D* 99 (2019), p. 015018. DOI: `10.1103/PhysRevD.99.015018`.

[120] A. Caldwell, D. Kollár, and K. Kröninger. “BAT – The Bayesian analysis toolkit”. In: *Comput. Phys. Commun.* 180 (2009), pp. 2197–2209. DOI: `10.1016/j.cpc.2009.06.026`.

[121] H. Ozaki. *Updates for driftime calculation.* Belle2 Internal CDC meeting. 2024. URL: `https://indico.belle2.org/event/12262/contributions/77798/attachments/28588/42350/20240523_v2.pdf`.

[122] M. Sumihama and H. Ozaki. *Status of waveform simulation.* Belle2 Internal CDC meeting. 2024. URL: `https://indico.belle2.org/event/11080/contributions/71842/attachments/26376/38933/20240118.pdf`.

[123] Slavomira Stefkova. *Implementation of the MC information for beam backgrounds.* Version 4. 2023. URL: `https://indico.global/event/6461/contributions/52695/attachments/26538/45990/CDCMCBKG_MLTriggerMeeting_8_12_2022_Stefkova_v4.pdf`.

[124] C. Shorten and T. M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning”. In: *J Big Data* 6 (2019), pp. 1–48. DOI: `10.1186/s40537-019-0197-0`.

[125] A. Farahani et al. “A brief review of domain adaptation”. In: *CoRR* abs/2010.03978 (2020). arXiv: `2010.03978`.

[126] C. Shimmin et al. “Decorrelated jet substructure tagging using adversarial neural networks”. In: *Phys. Rev. D* 96 (2017), p. 074034. DOI: `10.1103/PhysRevD.96.074034`.

[127] A. Stein et al. “Improving Robustness of Jet Tagging Algorithms with Adversarial Training”. In: *Comput. Softw. Big Sci.* 6 (2022), p. 15. DOI: `10.1007/s41781-022-00087-1`.

[128] D. Bodrov. *Kink Finder at Belle II*. Technical Note BELLE2-NOTE-TE-2024-020. 2024.

[129] *Open Neural Network Exchange (ONNX) open-source project*. `https://github.com/onnx/onnx`. Accessed: 2025-09-07.

[130] G. De Pietro et al. *GNN-based Hit Clean-up in the CDC*. BELLE2-NOTE-TE-2025-005. 2025.

[131] X. Ai et al. “A Common Tracking Software Project”. In: *Comput. Softw. Big Sci.* 6 (2022), p. 8. DOI: `10.1007/s41781-021-00078-8`. arXiv: `2106.13593 [physics.ins-det]`.

[132] L. Reuter et al. “Multi-Modal track reconstruction using Graph Neural Networks at Belle II”. In: *Workshop on Advanced Computing and Analysis Techniques in Physics Research*. Hamburg, Germany, 2025. URL: `https://indico.cern.ch/event/1488410/contributions/6561505/`.

[133] Reuter, L. and others. *The Belle II CDC AI Track (CAT) Finder*. `https://github.com/leareuter/catfinder`.

[134] C. Carloni Calame et al. “NNLO massive corrections to Bhabha scattering and theoretical precision of BabaYaga@NLO”. In: *Nucl. Phys. B* 225-227 (2012), pp. 293–297. DOI: `10.1016/j.nuclphysbps.2012.02.061`.

[135] **GEANT4 Collaboration**, S. Agostinelli et al. “GEANT4—a simulation toolkit”. In: *Nucl. Instrum. Methods Phys. Res. A* 506 (2003), p. 250. DOI: `10.1016/S0168-9002(03)01368-8`.

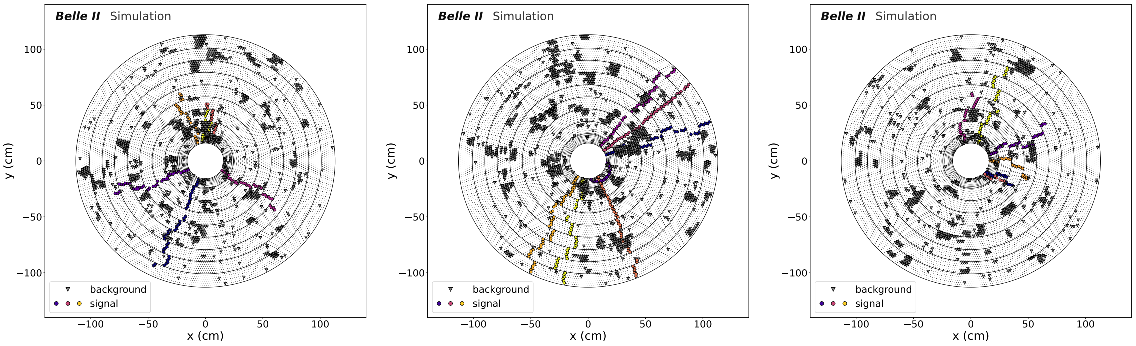
Appendices

Appendix A

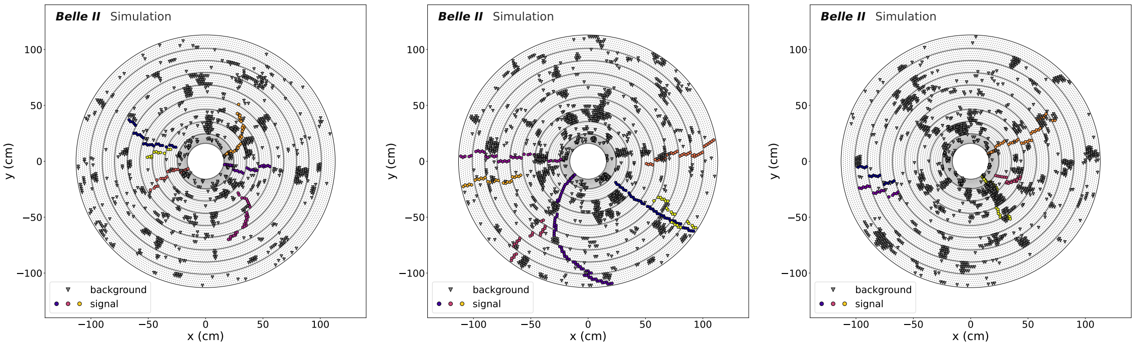
Dataset Appendix

A.1 Trainings dataset event displays

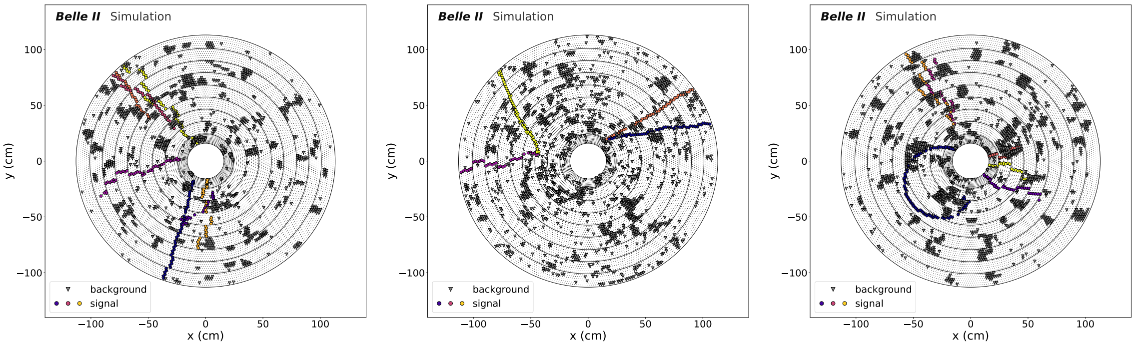
Figure A.1 shows typical example event displays of the different training samples described in Section 3.2 and Table 3.2.



(a) Prompt forward (category 1) (b) Prompt barrel (category 2) (c) Prompt backward (category 3)



(d) Displaced forward (category 4) (e) Displaced barrel (category 5) (f) Displaced backward (category 6)



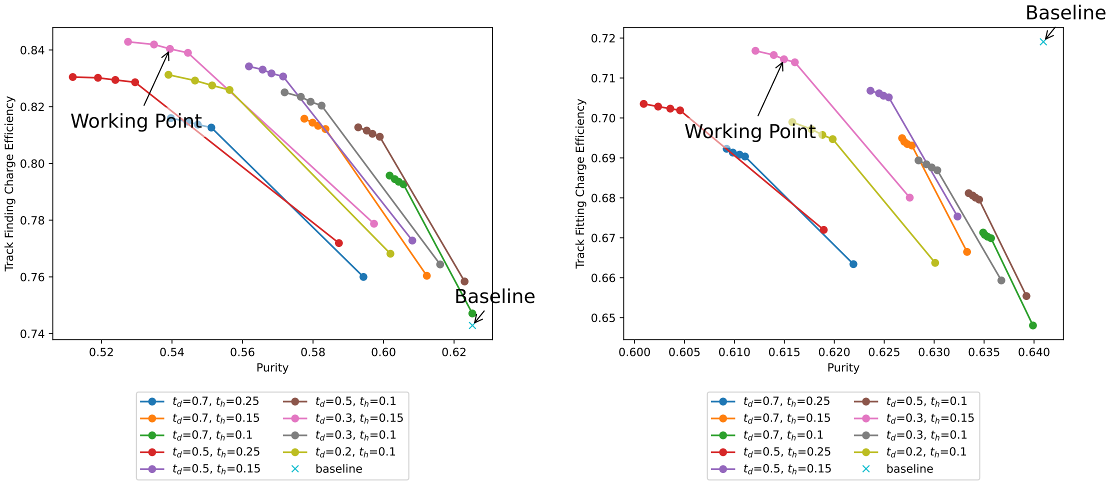
(g) Displaced angled (category 7) (h) Vertex (category 9+10) (i) Mix (category 11)

Figure A.1: Figure and text taken from [1]: event displays showing examples of the different training samples for *high data beam backgrounds*. Filled colored circular markers show signal hits, filled gray triangular markers show background hits. The markers correspond to the locations of the sense wires at the z position of the center of the wire for the wires with recorded ADC signals.

Appendix B

Hyperparameter for BB events

The hyperparameter optimization for the track finding hyperparameters described in Section 5.3.4 is given in Fig. B.1.



(a) Finding charge efficiency.

(b) Fitting charge efficiency.

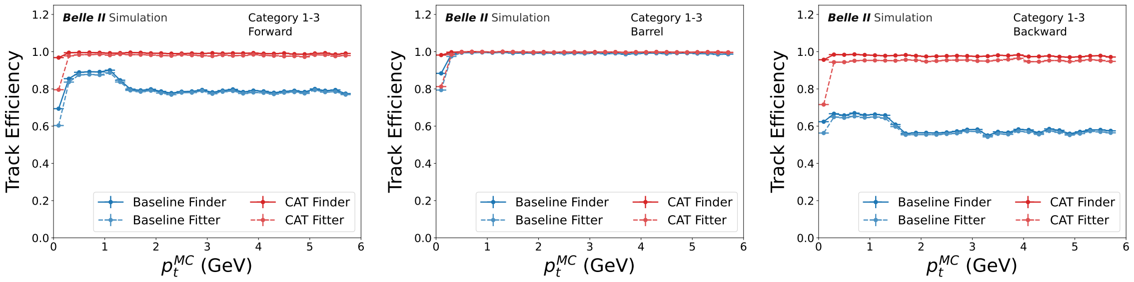
Figure B.1: Track finding (top) and fitting (bottom) charge efficiency as function of purity for the *CAT Finder*, and the respective value for the *Baseline Finder* for $B\bar{B}$ events for *high data beam backgrounds*.

Appendix C

CDC only results

C.1 Track efficiency for category 1-3

The track finding efficiencies, as well as the combined track finding and track fitting efficiencies for the *Baseline Finder* and the *CAT Finder* are shown in Fig. C.1 for non-curling tracks from category 1-3.

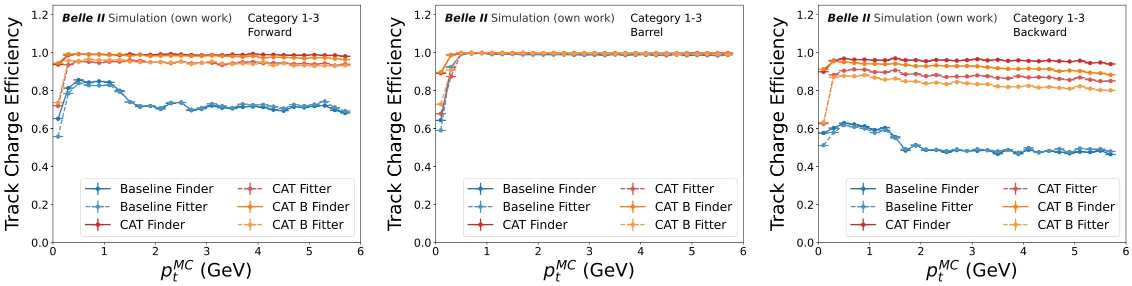


(a) Forward endcap. (b) Barrel. (c) Backward endcap.

Figure C.1: Figure and text taken from [1]: Track finding (markers connected by solid lines to guide the eye) and combined track finding and fitting efficiency (markers connected by dashed lines to guide the eye) for the prompt evaluation samples (category 1-3, *high data beam backgrounds*, see Table 3.2) with curler tracks removed, as function of simulated transverse momentum p_t^{MC} for the *Baseline Finder* (blue) and the *CAT Finder* (red) in the (a) forward endcap, (b) barrel, and (c) backward endcap. The vertical error bars that show the statistical uncertainty are smaller than the marker size. The horizontal error bars indicate the bin width. The uncertainties of the two track finding algorithms are correlated, since they use the same simulated events.

C.2 The *CAT B Finder* in Category 1-3

The track finding and combined track fining and fitting charge efficiency for all tracks from category 1-3 is shown in Fig. C.2 with the additional *CAT B Finder* algorithm. This includes curling tracks. While the performance in the barrel region is improved for *CAT B Finder*, this model has decreased efficiency compared to *CAT Finder* for the backwards region. The performance metrics for *CAT B Finder* are given in Table C.1



(a) Forward endcap.
(b) Barrel.
(c) Backward endcap.

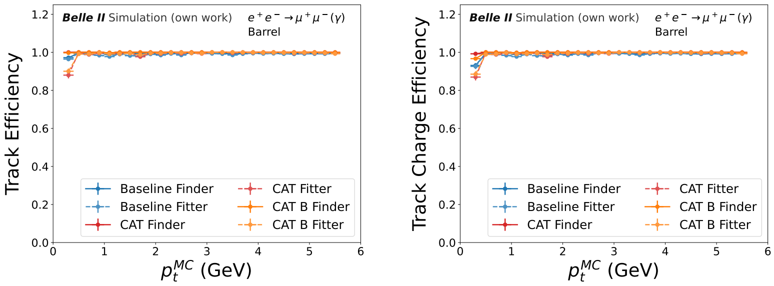
Figure C.2: Track finding (markers connected by solid lines to guide the eye) and combined track finding and fitting charge efficiency (markers connected by dashed lines to guide the eye) for the prompt evaluation samples (category 1-3, *high data beam backgrounds*, see Table 3.2), as function of simulated transverse momentum p_t^{MC} for the *Baseline Finder* (blue), the *CAT Finder* (red), and the *CAT B Finder* (orange) in the (a) forward endcap, (b) barrel, and (c) backward endcap. The vertical error bars that show the statistical uncertainty are smaller than the marker size. The horizontal error bars indicate the bin width. The uncertainties of the three track finding algorithms are correlated, since they use the same simulated events.

Table C.1: The performance metrics for the category 1-3 evaluation samples for the *CAT B Finder* in different detector regions for high data beam backgrounds, including curling tracks. Uncertainties below <0.01% are not shown in the table. Comparison with Table 6.1

(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
forward endcap					
Baseline Finder	$75.4^{+0.2}_{-0.2}$	$0.85^{+0.04}_{-0.04}$	$0.01_{-0.01}$	$73.7^{+0.2}_{-0.2}$	$2.33^{+0.06}_{-0.06}$
CAT Finder	$98.52^{+0.04}_{-0.04}$	$3.82^{+0.07}_{-0.07}$	$0.35^{+0.02}_{-0.02}$	$98.25^{+0.05}_{-0.05}$	$0.27^{+0.02}_{-0.02}$
CAT B Finder	$98.19^{+0.05}_{-0.05}$	$2.32^{+0.06}_{-0.06}$	$0.11^{+0.01}_{-0.01}$	$97.75^{+0.05}_{-0.05}$	$0.45^{+0.02}_{-0.03}$
Baseline Fitter	$73.5^{+0.2}_{-0.2}$	$0.66^{+0.03}_{-0.04}$	0.01	$72.5^{+0.2}_{-0.2}$	$1.33^{+0.05}_{-0.05}$
CAT Fitter	$93.71^{+0.09}_{-0.09}$	$0.64^{+0.03}_{-0.03}$	$0.07^{+0.01}_{-0.01}$	$91.9^{+0.1}_{-0.1}$	$1.93^{+0.05}_{-0.05}$
CAT B Fitter	$94.08^{+0.09}_{-0.09}$	$0.44^{+0.02}_{-0.03}$	$0.04^{+0.01}_{-0.01}$	$92.3^{+0.1}_{-0.1}$	$1.89^{+0.05}_{-0.05}$
barrel					
Baseline Finder	$96.59^{+0.03}_{-0.03}$	$1.26^{+0.02}_{-0.02}$	$0.77^{+0.01}_{-0.01}$	$94.1^{+0.04}_{-0.04}$	$2.58^{+0.03}_{-0.03}$
CAT Finder	$99.47^{+0.01}_{-0.01}$	$1.96^{+0.02}_{-0.02}$	$0.65^{+0.01}_{-0.01}$	$98.44^{+0.02}_{-0.02}$	$1.03^{+0.02}_{-0.02}$
CAT B Finder	$99.62^{+0.01}_{-0.01}$	$2.13^{+0.02}_{-0.02}$	$1.53^{+0.02}_{-0.02}$	$98.49^{+0.02}_{-0.02}$	$1.13^{+0.02}_{-0.02}$
Baseline Fitter	$95.12^{+0.04}_{-0.04}$	$0.89^{+0.02}_{-0.02}$	$0.49^{+0.01}_{-0.01}$	$93.43^{+0.04}_{-0.04}$	$1.77^{+0.02}_{-0.02}$
CAT Fitter	$94.37^{+0.04}_{-0.04}$	$0.69^{+0.01}_{-0.01}$	$0.13^{+0.01}_{-0.01}$	$94.07^{+0.04}_{-0.04}$	$0.31^{+0.01}_{-0.01}$
CAT B Fitter	$95.53^{+0.03}_{-0.03}$	$0.55^{+0.01}_{-0.01}$	$0.46^{+0.01}_{-0.01}$	$95.13^{+0.04}_{-0.04}$	$0.43^{+0.01}_{-0.01}$
backward endcap					
Baseline Finder	$55.6^{+0.2}_{-0.2}$	$1.73^{+0.06}_{-0.06}$	$0.02^{+0.01}_{-0.01}$	$53.0^{+0.2}_{-0.2}$	$4.7^{+0.09}_{-0.09}$
CAT Finder	$95.92^{+0.06}_{-0.06}$	$2.04^{+0.04}_{-0.05}$	$0.11^{+0.01}_{-0.01}$	$95.13^{+0.07}_{-0.06}$	$0.82^{+0.03}_{-0.03}$
CAT B Finder	$93.71^{+0.07}_{-0.07}$	$1.83^{+0.04}_{-0.04}$	$0.12^{+0.01}_{-0.01}$	$92.52^{+0.08}_{-0.08}$	$1.27^{+0.03}_{-0.04}$
Baseline Fitter	$54.0^{+0.2}_{-0.2}$	$1.24^{+0.05}_{-0.05}$	$0.02_{-0.01}$	$52.1^{+0.2}_{-0.2}$	$3.56^{+0.08}_{-0.08}$
CAT Fitter	$89.9^{+0.09}_{-0.09}$	$1.2^{+0.04}_{-0.04}$	$0.04^{+0.01}_{-0.01}$	$84.9^{+0.1}_{-0.1}$	$5.57^{+0.07}_{-0.07}$
CAT B Fitter	$86.7^{+0.1}_{-0.1}$	$0.9^{+0.03}_{-0.03}$	$0.05^{+0.01}_{-0.01}$	$81.7^{+0.1}_{-0.1}$	$5.73^{+0.08}_{-0.08}$

C.3 Efficiencies, fake rates, and clone rates for $\mu^-\mu^+(\gamma)$

The track finding efficiencies, and the combined track finding and track fitting efficiency and charge efficiency for the *Baseline Finder* in comparison with the *CAT Finder* are shown in Fig. C.3.



(a) Track finding and fitting efficiency.

(b) Track finding and fitting charge efficiency.

Figure C.3: Figure and text adapted from [1]: Track finding (empty markers, connected by lines to guide the eye) and combined track finding and fitting charge efficiency (filled markers) for $\mu^-\mu^+(\gamma)$ evaluation sample in the barrel with *high data beam backgrounds*. See Fig. C.1 caption for details.

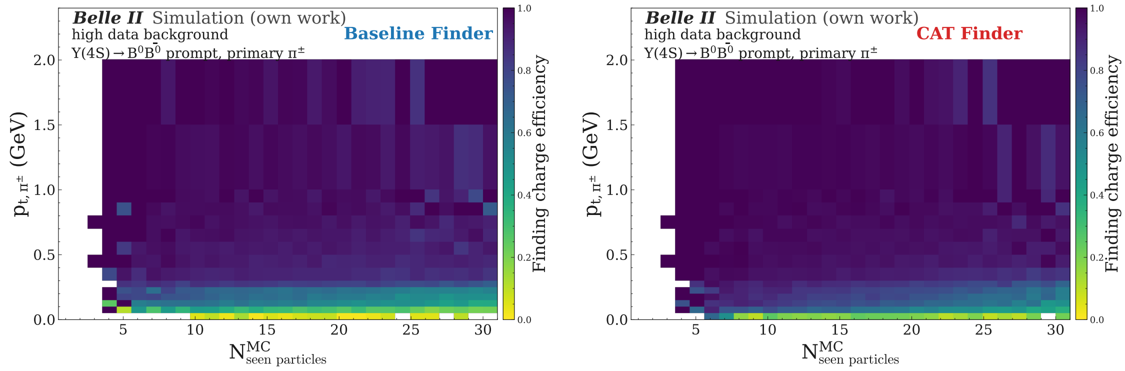
Track finding and fitting efficiency ε_{trk}, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, track charge efficiency $\varepsilon_{\text{trk,ch}}$ and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$ integrated over the full p_t for $\mu^-\mu^+(\gamma)$ events are shown in Table C.2.

Table C.2: Table and text adapted from [1]:The performance metrics for the $\mu^-\mu^+(\gamma)$ evaluation samples for different track finding algorithms in different detector regions for high data beam background. Uncertainties below <0.01% are not shown in the table.

(in %)	ε_{trk}	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\varepsilon_{\text{trk,ch}}$	$\mathfrak{r}_{\text{wrong ch.}}$
forward endcap					
Baseline Finder	$80.2^{+0.3}_{-0.3}$	$1.4^{+0.1}_{-0.1}$	$0.01^{+0.01}_{-0.01}$	$78.6^{+0.4}_{-0.4}$	$2.0^{+0.1}_{-0.1}$
CAT Finder	$99.81^{+0.04}_{-0.03}$	$8.5^{+0.2}_{-0.2}$	$0.07^{+0.02}_{-0.02}$	$99.78^{+0.04}_{-0.04}$	$0.04^{+0.01}_{-0.02}$
CAT B Finder	$99.4^{+0.07}_{-0.06}$	$6.0^{+0.2}_{-0.2}$	$0.01^{+0.01}_{-0.01}$	$99.3^{+0.07}_{-0.07}$	$0.1^{+0.02}_{-0.03}$
Baseline Fitter	$79.3^{+0.4}_{-0.3}$	$1.07^{+0.1}_{-0.1}$	$0.0_{-0.01}$	$78.2^{+0.4}_{-0.4}$	$1.4^{+0.1}_{-0.1}$
CAT Fitter	$99.08^{+0.08}_{-0.08}$	$1.19^{+0.09}_{-0.1}$	$0.01^{+0.01}_{-0.01}$	$97.0^{+0.2}_{-0.1}$	$2.1^{+0.1}_{-0.1}$
CAT B Fitter	$98.0^{+0.1}_{-0.1}$	$0.77^{+0.07}_{-0.08}$	$0.01^{+0.01}_{-0.01}$	$95.7^{+0.2}_{-0.2}$	$2.3^{+0.1}_{-0.1}$
barrel					
Baseline Finder	$99.42^{+0.03}_{-0.03}$	$3.0^{+0.07}_{-0.07}$	$0.04^{+0.01}_{-0.01}$	$99.3^{+0.04}_{-0.04}$	$0.12^{+0.01}_{-0.02}$
CAT Finder	100.0	$5.32^{+0.09}_{-0.09}$	$0.03^{+0.01}_{-0.01}$	$99.97^{+0.01}_{-0.01}$	$0.03^{+0.01}_{-0.01}$
CAT B Finder	$99.99^{+0.01}$	$5.28^{+0.09}_{-0.09}$	$0.11^{+0.01}_{-0.02}$	$99.94^{+0.01}_{-0.01}$	$0.05^{+0.01}_{-0.01}$
Baseline Fitter	$99.15^{+0.04}_{-0.04}$	$2.05^{+0.06}_{-0.06}$	$0.03^{+0.01}_{-0.01}$	$99.07^{+0.04}_{-0.04}$	$0.08^{+0.01}_{-0.01}$
CAT Fitter	$99.49^{+0.03}_{-0.03}$	$1.78^{+0.06}_{-0.06}$	0.01	$99.48^{+0.03}_{-0.03}$	$0.02_{-0.01}$
CAT B Fitter	$99.57^{+0.03}_{-0.03}$	$1.32^{+0.05}_{-0.05}$	$0.03^{+0.01}_{-0.01}$	$99.53^{+0.03}_{-0.03}$	$0.03^{+0.01}_{-0.01}$
backward endcap					
Baseline Finder	$60.6^{+0.4}_{-0.4}$	$4.7^{+0.3}_{-0.3}$	$0.04^{+0.02}_{-0.03}$	$57.0^{+0.4}_{-0.4}$	$5.9^{+0.3}_{-0.3}$
CAT Finder	$99.68^{+0.05}_{-0.05}$	$8.4^{+0.2}_{-0.2}$	$0.04^{+0.02}_{-0.02}$	$99.58^{+0.06}_{-0.06}$	$0.11^{+0.03}_{-0.03}$
CAT B Finder	$97.4^{+0.1}_{-0.1}$	$7.6^{+0.2}_{-0.2}$	$0.04^{+0.02}_{-0.02}$	$97.1^{+0.2}_{-0.1}$	$0.3^{+0.05}_{-0.05}$
Baseline Fitter	$60.0^{+0.4}_{-0.4}$	$3.8^{+0.2}_{-0.2}$	$0.01^{+0.01}_{-0.02}$	$57.2^{+0.4}_{-0.4}$	$4.6^{+0.2}_{-0.2}$
CAT Fitter	$97.6^{+0.1}_{-0.1}$	$4.4^{+0.2}_{-0.2}$	$0.02^{+0.01}_{-0.02}$	$92.4^{+0.2}_{-0.2}$	$5.4^{+0.2}_{-0.2}$
CAT B Fitter	$93.8^{+0.2}_{-0.2}$	$2.9^{+0.2}_{-0.2}$	$0.03^{+0.01}_{-0.02}$	$88.1^{+0.3}_{-0.3}$	$6.0^{+0.2}_{-0.2}$

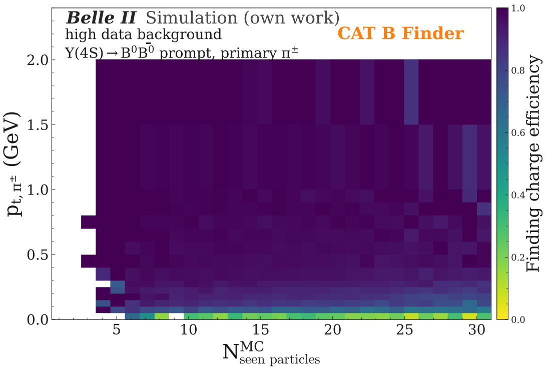
C.4 Efficiencies, fake rates and clone rates for $B^0\bar{B}^0$

C.4.1 Efficiencies for prompt pions according to the number of seen particles



(a) *Baseline Finder*.

(b) *CAT Finder*.



(c) *CAT B Finder*.

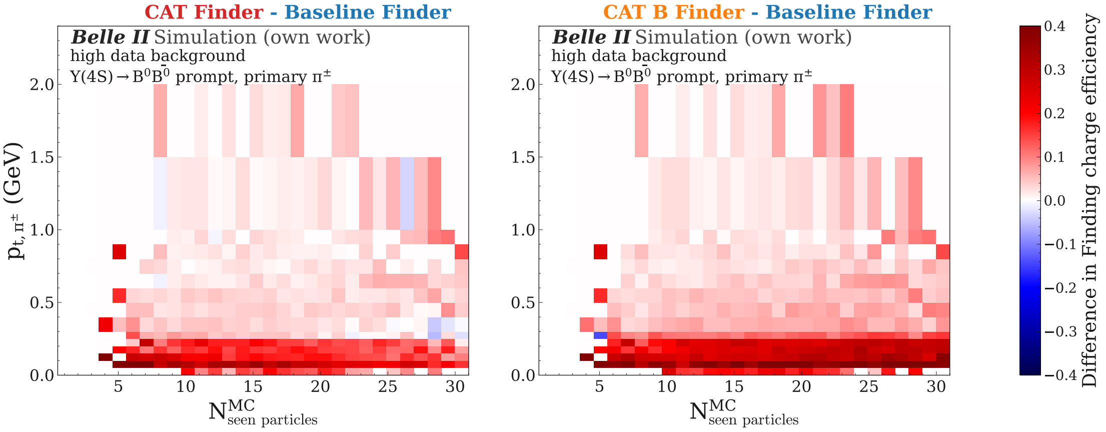
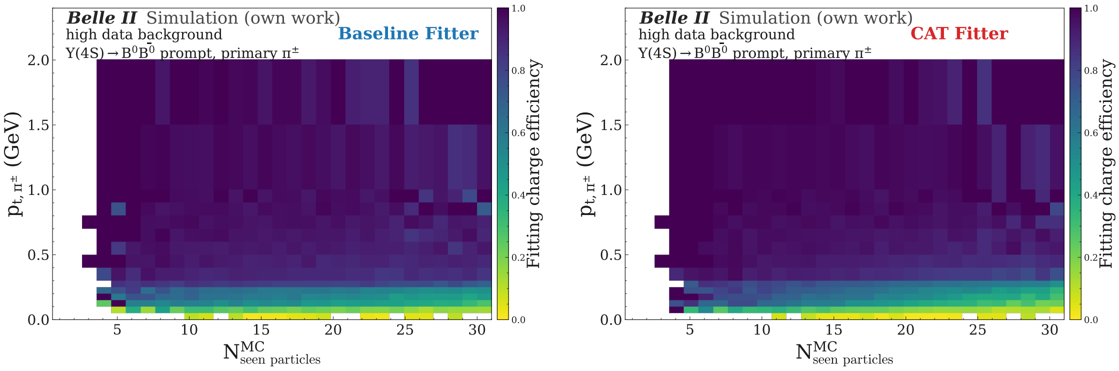
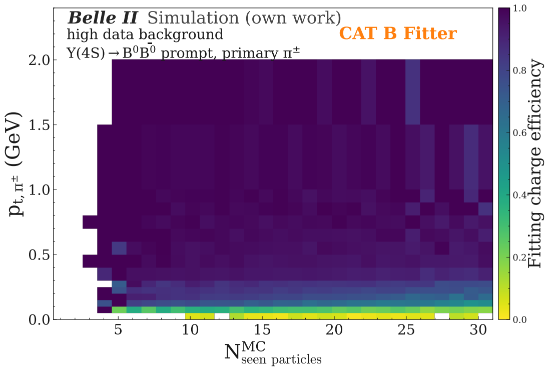


Figure C.4: The track finding charge efficiency for *Baseline Finder*, *CAT Finder* and *CAT B Finder*, as well as the differences between both *CAT Finder* algorithms to *Baseline Finder* given over the true transverse momentum and the number of seen particles in CDC.



(a) *Baseline Finder*. (b) *CAT Finder*.



(c) *CAT B Finder*.

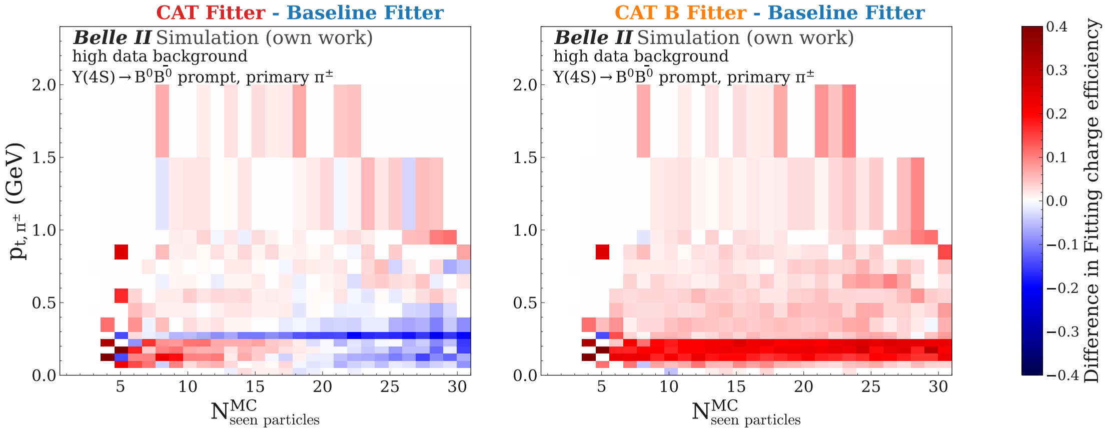
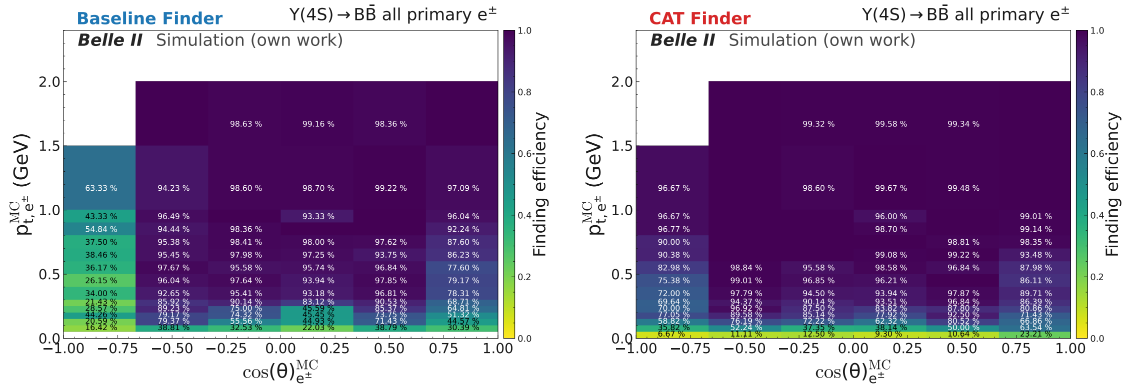
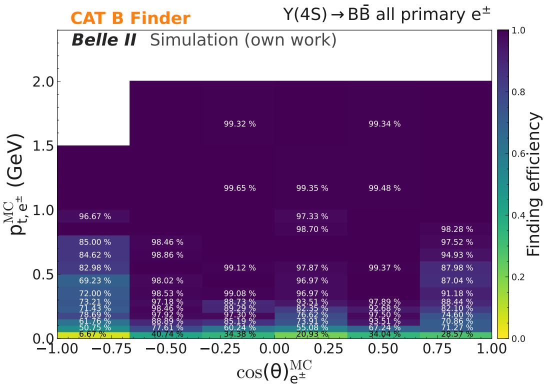


Figure C.5: The track finding and fitting charge efficiency for *Baseline Finder*, *CAT Finder* and *CAT B Finder*, as well as the differences between both *CAT Finder* algorithms to *Baseline Finder* given over the true transverse momentum and the number of seen particles in the CDC.

C.4.2 Prompt $e^\pm$



(a) *Baseline Finder.* (b) *CAT Finder.*



(c) *CAT B Finder.*

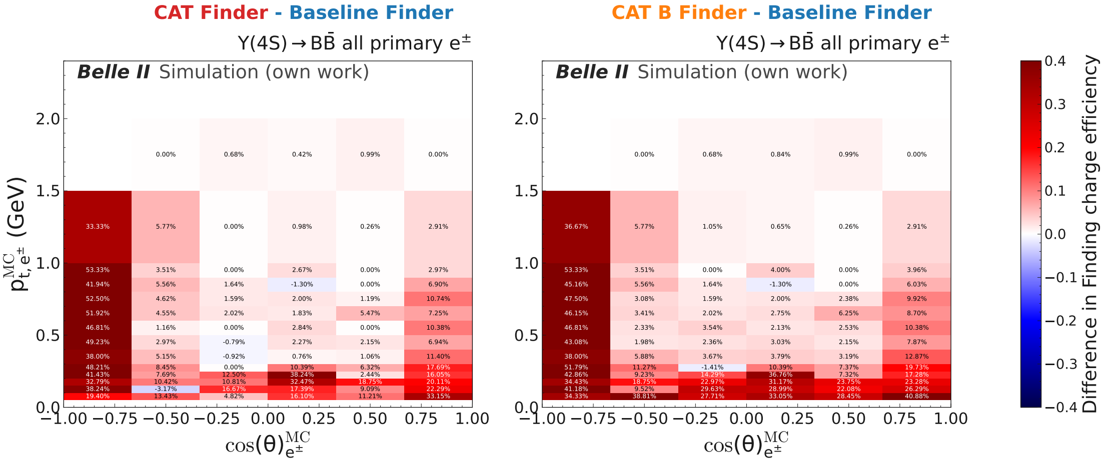
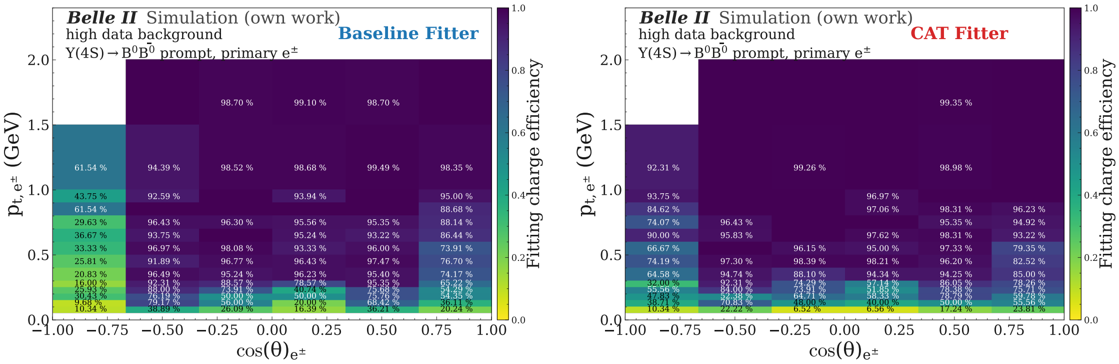
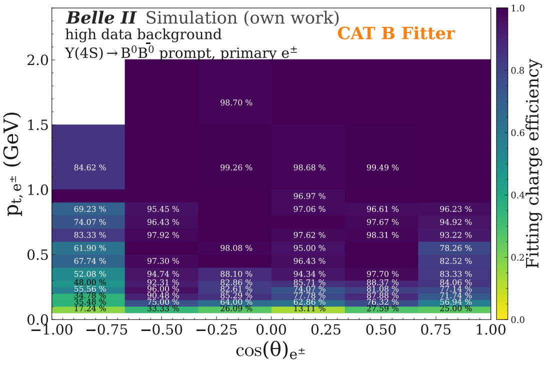


Figure C.6: Track finding charge efficiency for $e^{\pm}$.



(a) *Baseline Finder*.

(b) *CAT Finder*.



(c) *CAT B Finder*.

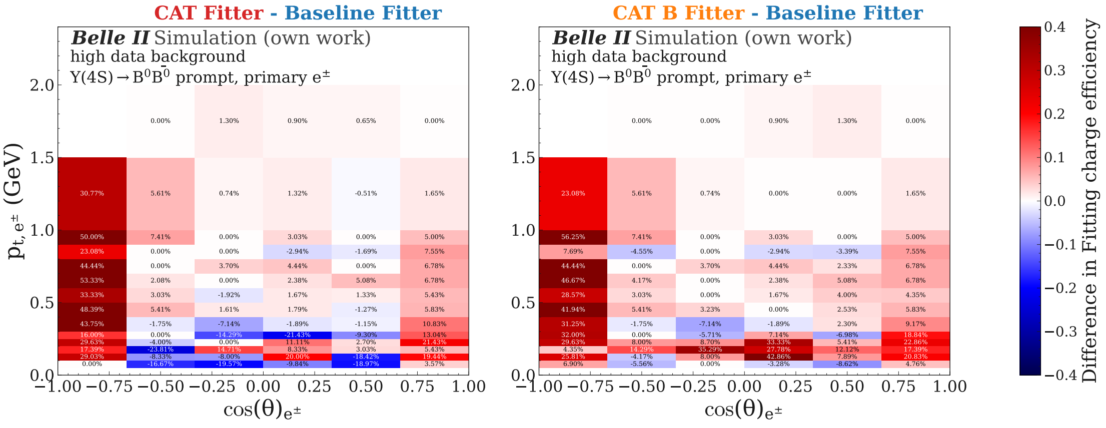
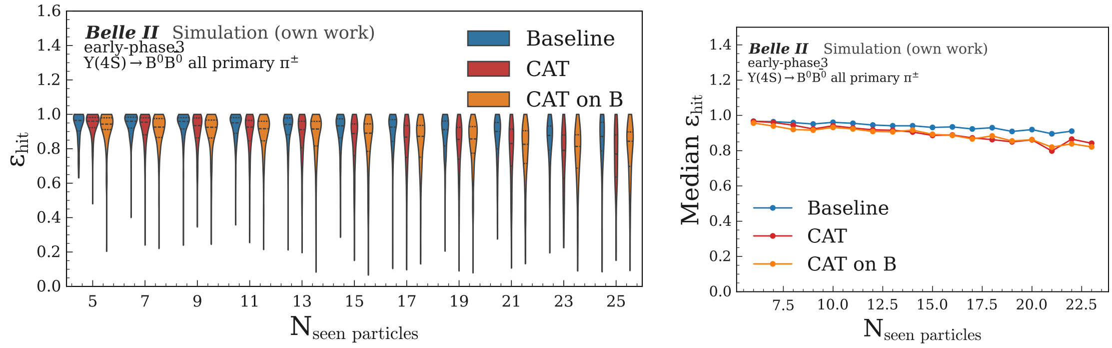
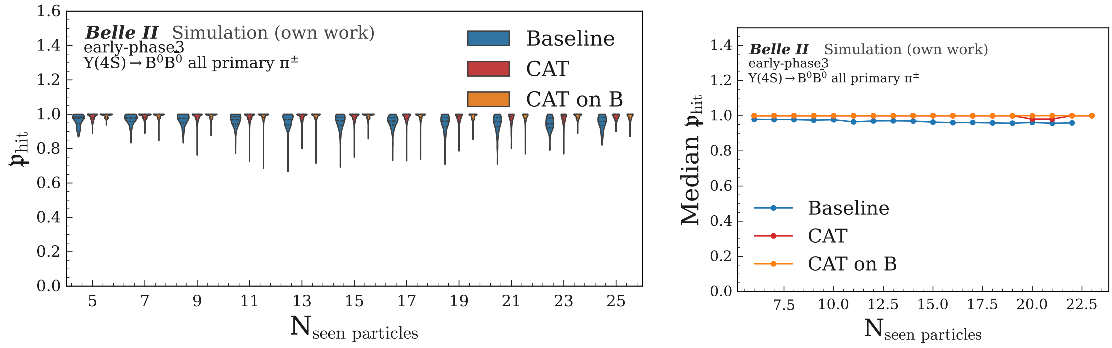


Figure C.7: Track fitting charge efficiency for $e^{\pm}$.



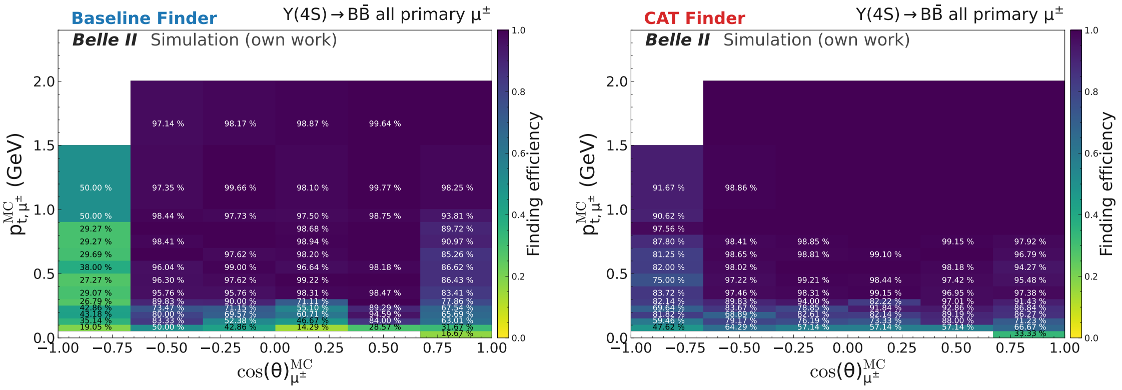
(a) Hit efficiency ε_{hit} distributions. (b) Median hit efficiency ε_{hit}.



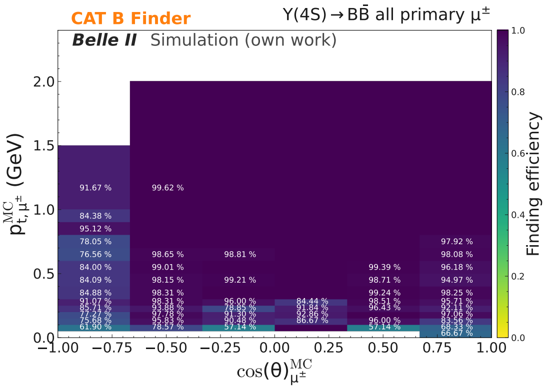
(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions. (d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

Figure C.8: Hit efficiency (top) and hit purity (bottom) vs. N_{seen} for all primary matched $e^{\pm}$ tracks. Left: distributions; right: medians.

C.4.3 Prompt $\mu^\pm$



(a) *Baseline Finder*. (b) *CAT Finder*.



(c) *CAT B Finder*.

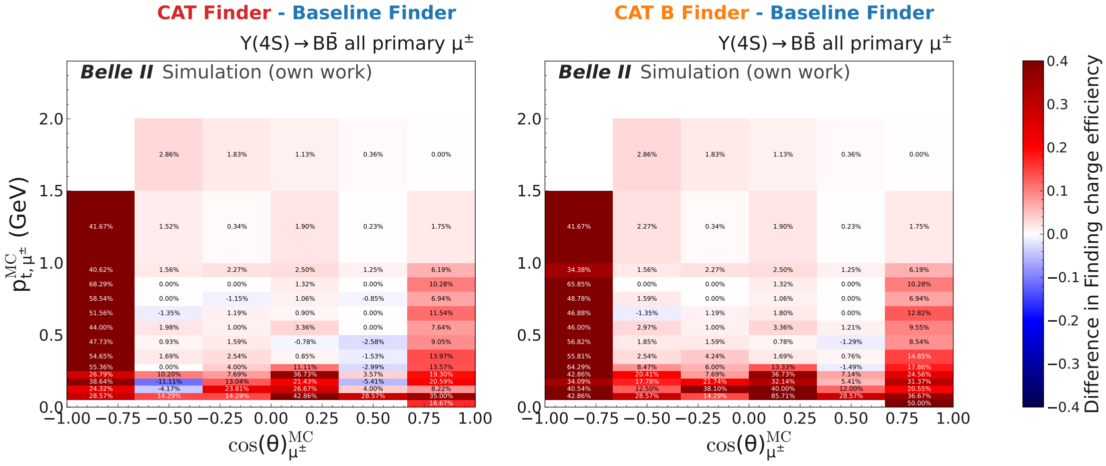
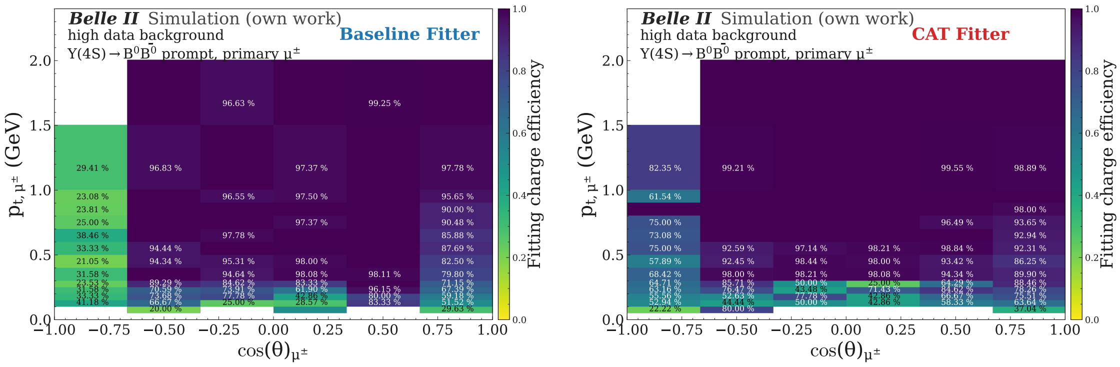
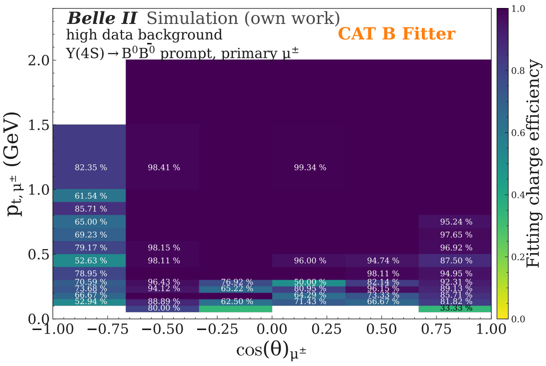


Figure C.9: Track finding charge efficiency for $\mu^\pm$.



(a) *Baseline Finder*. (b) *CAT Finder*.



(c) *CAT B Finder*.

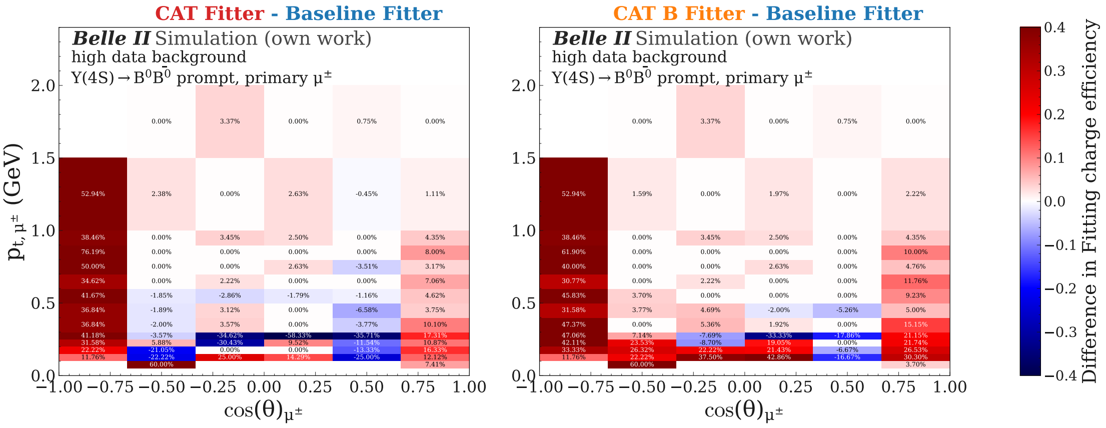
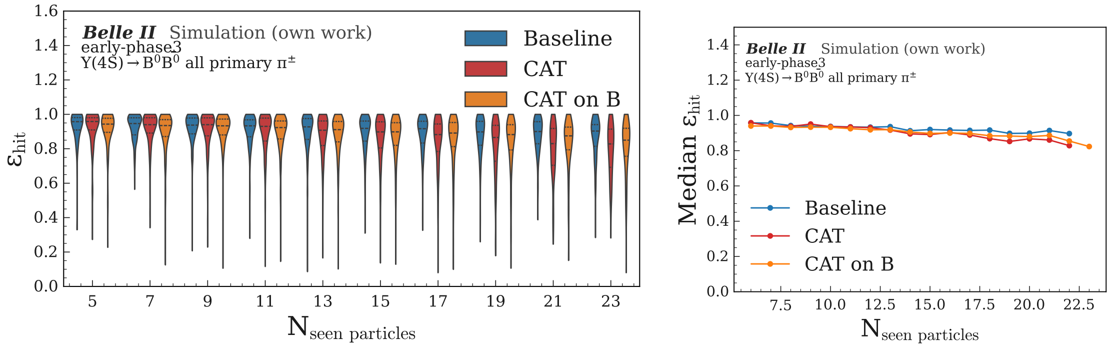
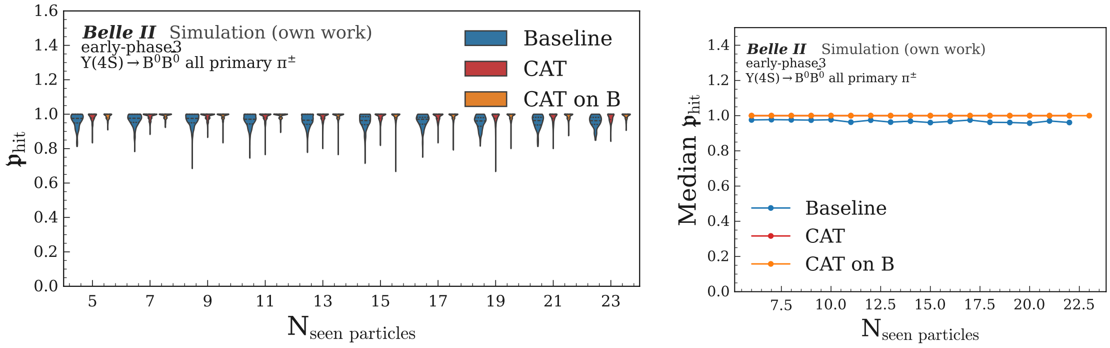


Figure C.10: Track fitting charge efficiency for $\mu^{\pm}$.



(a) Hit efficiency ε_{hit} distributions.

(b) Median hit efficiency ε_{hit}.

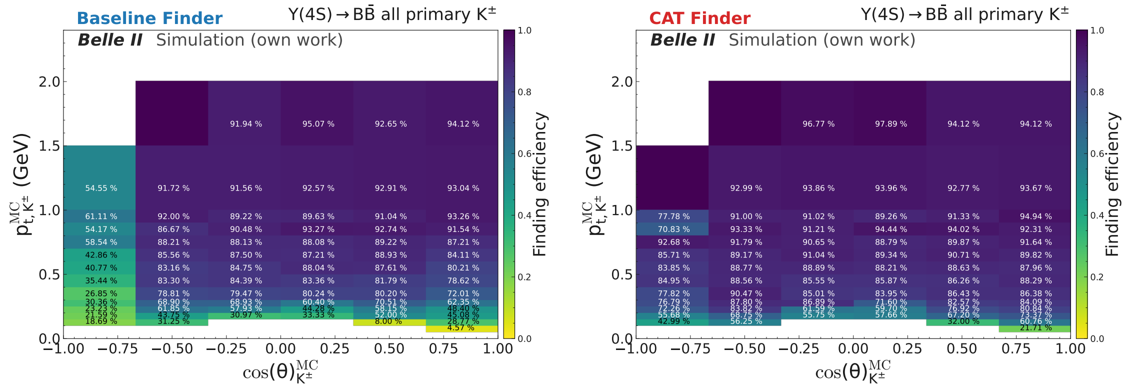


(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

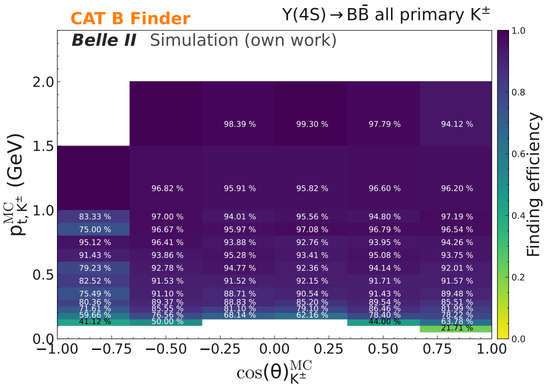
(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

Figure C.11: Hit efficiency (top) and hit purity (bottom) vs. N_{seen} for all primary matched $\mu^{\pm}$ tracks. Left: distributions; right: medians.

C.4.4 Prompt $K^\pm$



(a) *Baseline Finder*. (b) *CAT Finder*.



(c) *CAT B Finder*.

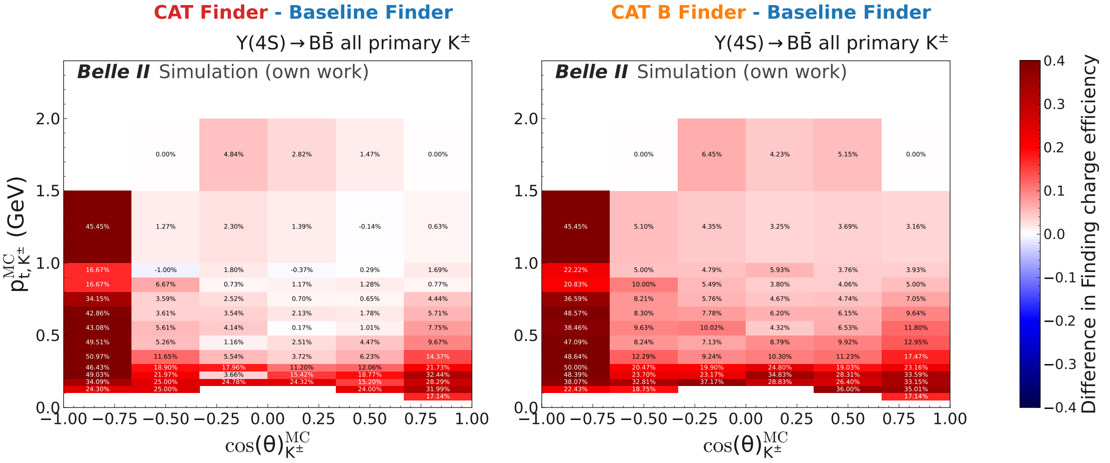
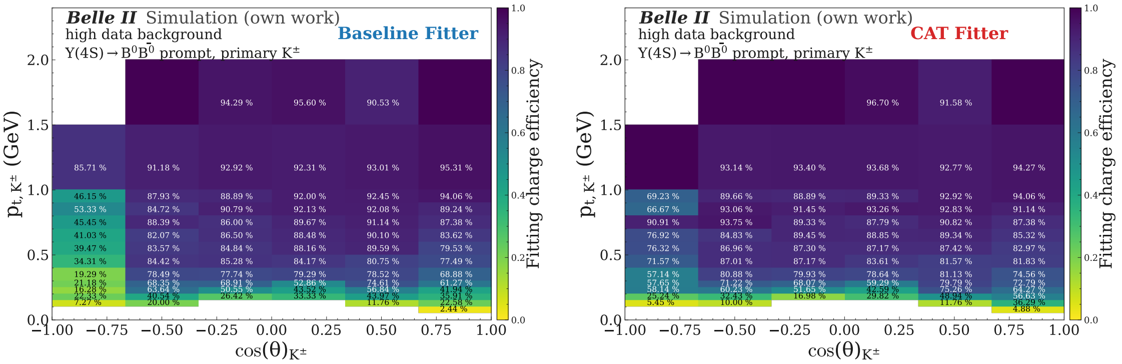
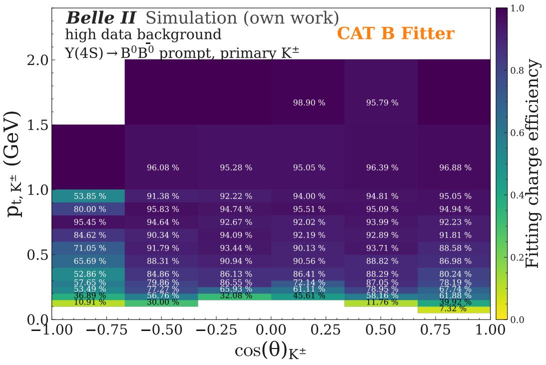


Figure C.12: Track finding charge efficiency for $K^\pm$.



(a) *Baseline Finder*. (b) *CAT Finder*.



(c) *CAT B Finder*.

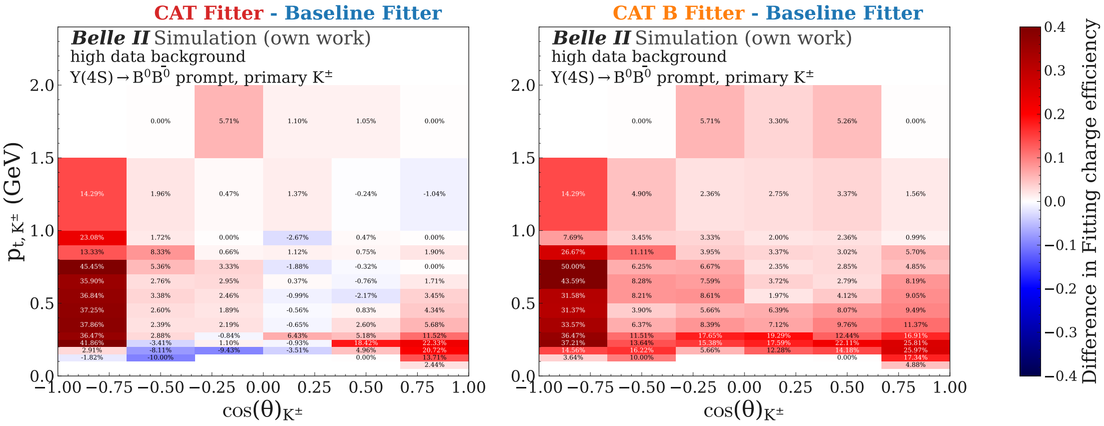
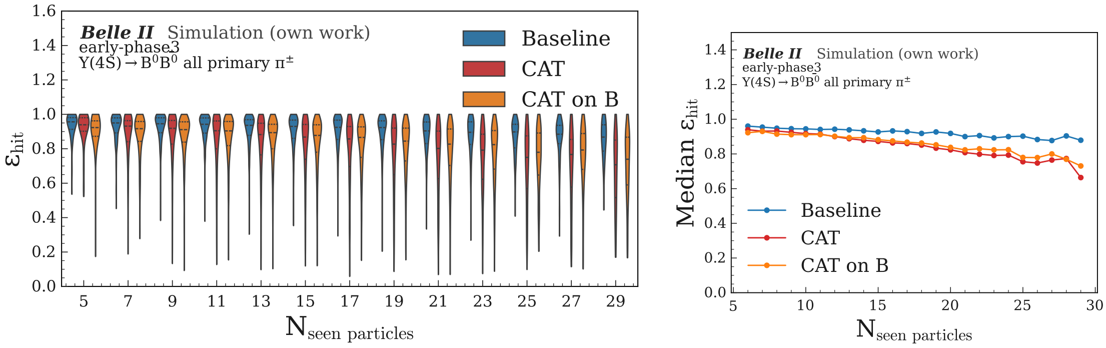
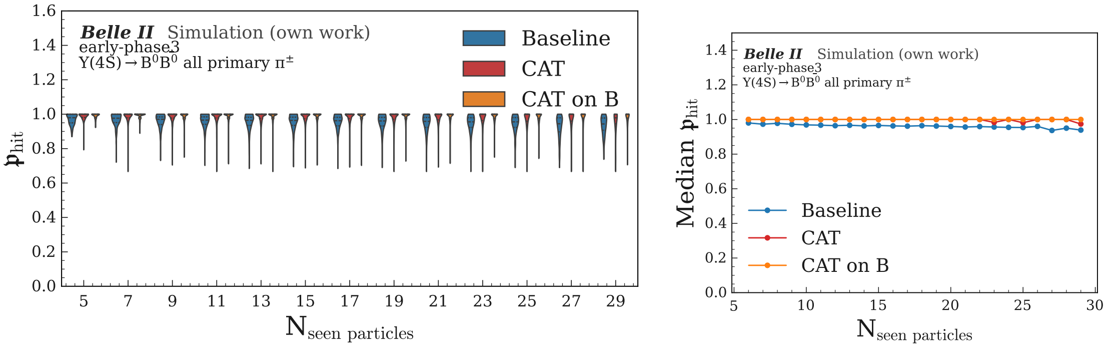


Figure C.13: Track fitting charge efficiency for $K^\pm$.



(a) Hit efficiency ε_{hit} distributions.

(b) Median hit efficiency ε_{hit}.

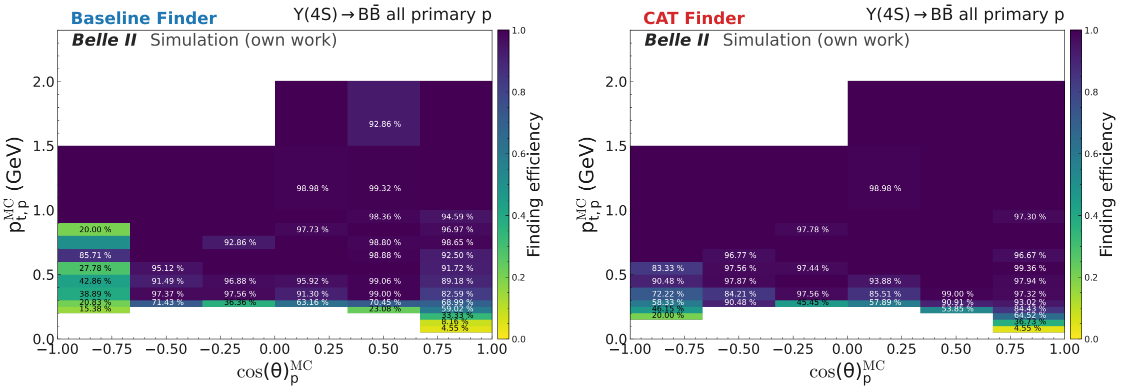


(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

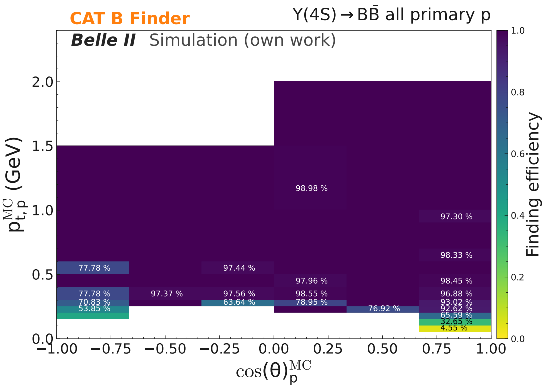
(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

Figure C.14: Hit efficiency (top) and hit purity (bottom) vs. N_{seen} for all primary matched $K^{\pm}$ tracks. Left: distributions; right: medians.

C.4.5 Prompt $p^{\pm}$



(a) *Baseline Finder*. (b) *CAT Finder*.



(c) *CAT B Finder*.

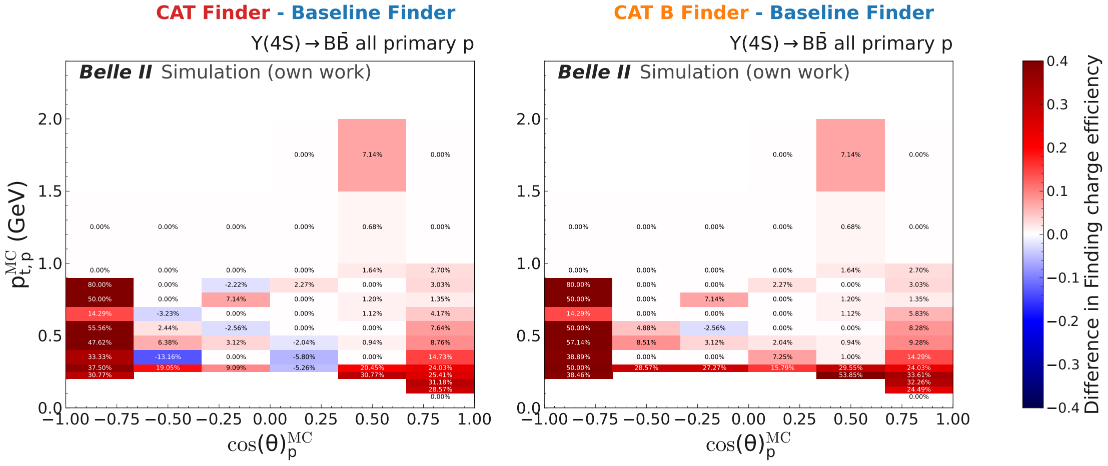
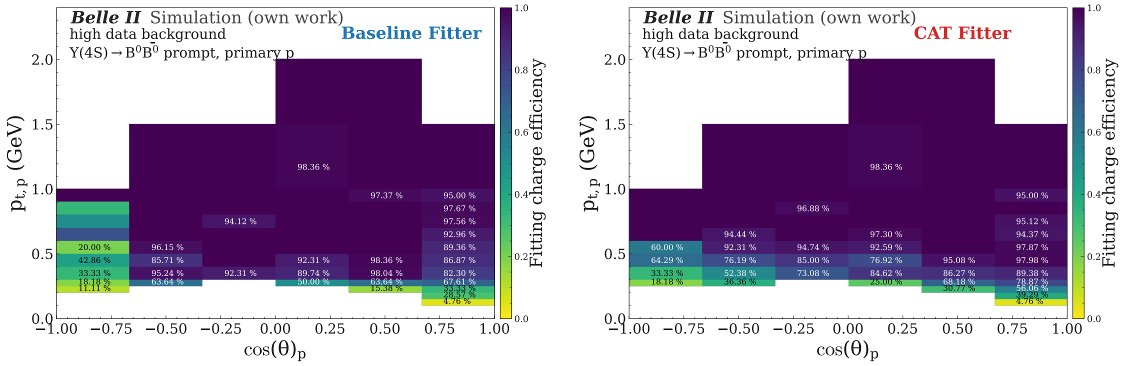
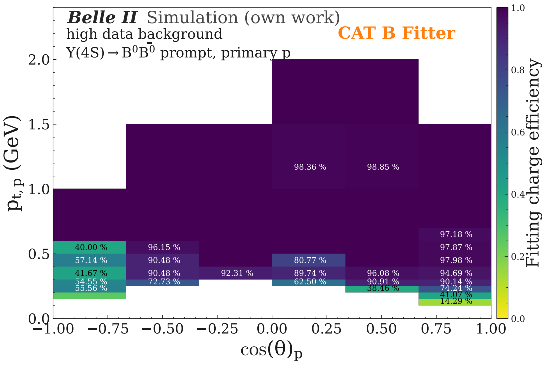


Figure C.15: Track finding charge efficiency for $p^{\pm}$.



(a) *Baseline Finder*.

(b) *CAT Finder*.



(c) *CAT B Finder*.

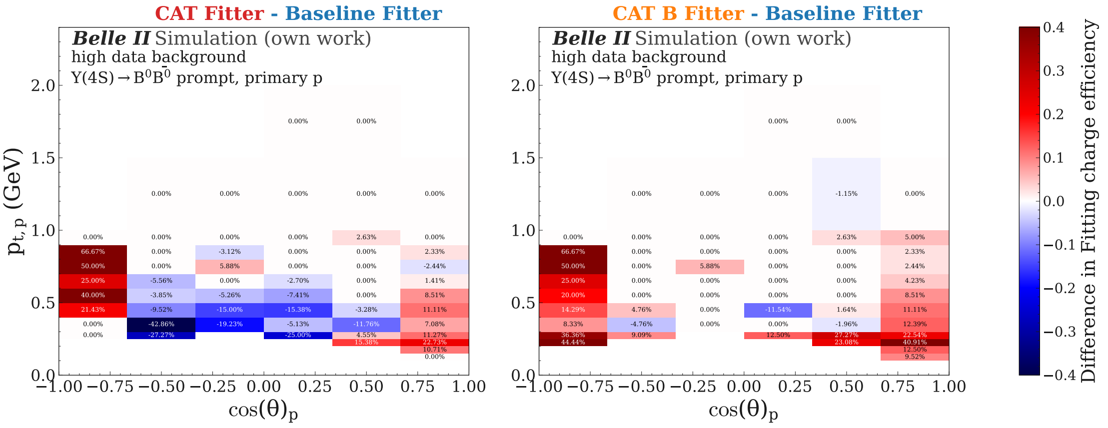
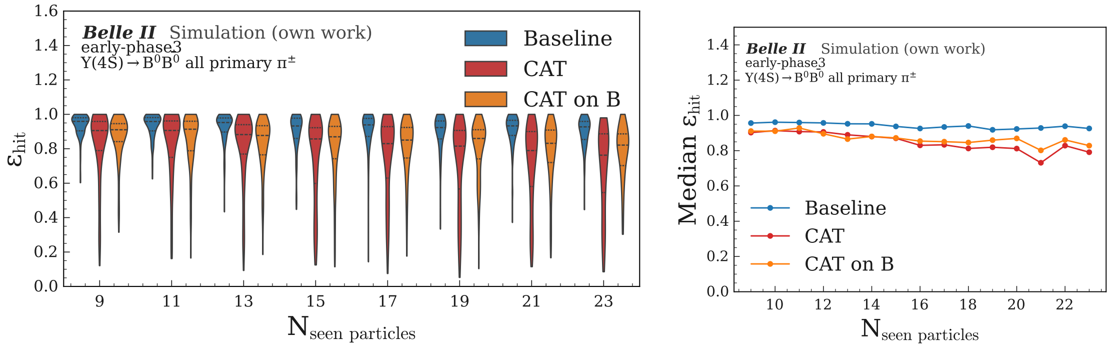
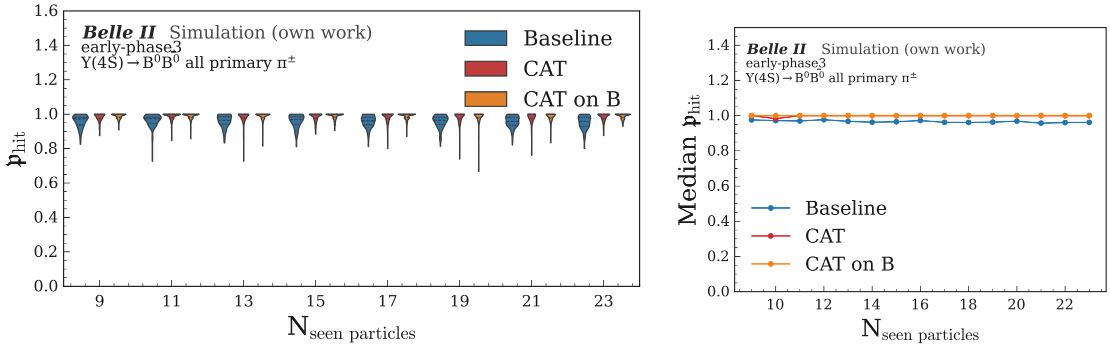


Figure C.16: Track fitting charge efficiency for $p^{\pm}$.



(a) Hit efficiency ε_{hit} distributions. (b) Median hit efficiency ε_{hit}.

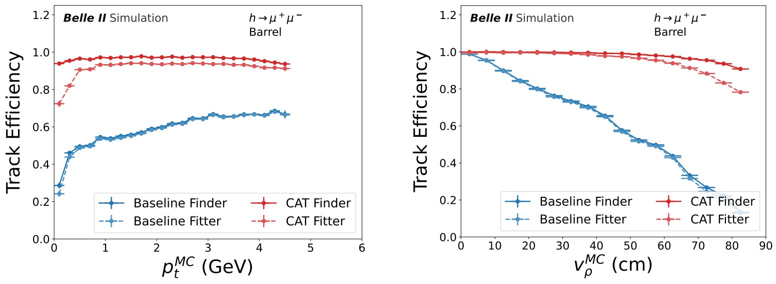


(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions. (d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

Figure C.17: Hit efficiency (top) and hit purity (bottom) vs. N_{seen} for all primary matched $p^{\pm}$ tracks. Left: distributions; right: medians.

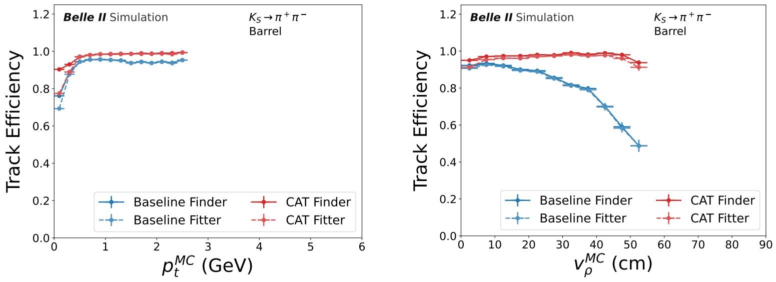
C.5 Track efficiency for $h \rightarrow \mu^+\mu^-$, $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$ and $K_S^0 \rightarrow \pi^+\pi^-$

The track finding efficiencies, and the combined track finding and track fitting efficiency for the *Baseline Finder* in comparison with the *CAT Finder* for $h \rightarrow \mu^+\mu^-$, $K_S^0 \rightarrow \pi^+\pi^-$, and $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$ events are shown in Fig. C.18.



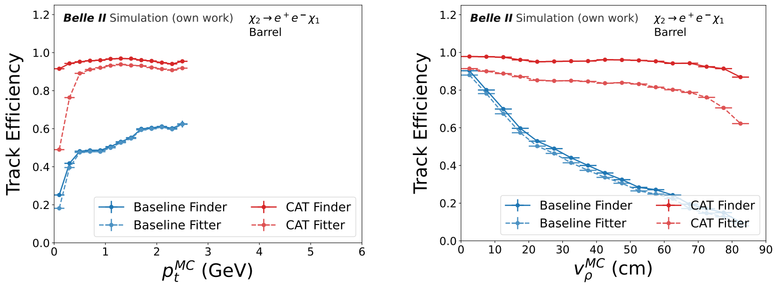
(a) $h \rightarrow \mu^+\mu^-$, p_t^{MC}.

(b) $h \rightarrow \mu^+\mu^-$, v_ρ^{MC}.



(c) $K_S^0 \rightarrow \pi^+\pi^-$, p_t^{MC}.

(d) $K_S^0 \rightarrow \pi^+\pi^-$, v_ρ^{MC}.



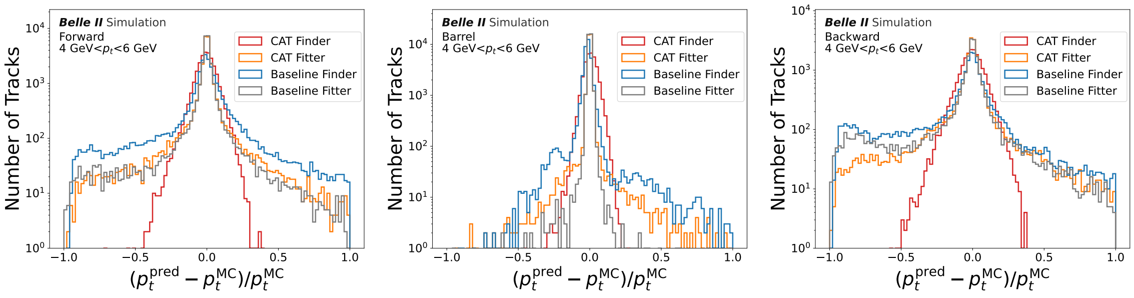
(e) $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$, p_t^{MC}.

(f) $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$, v_ρ^{MC}.

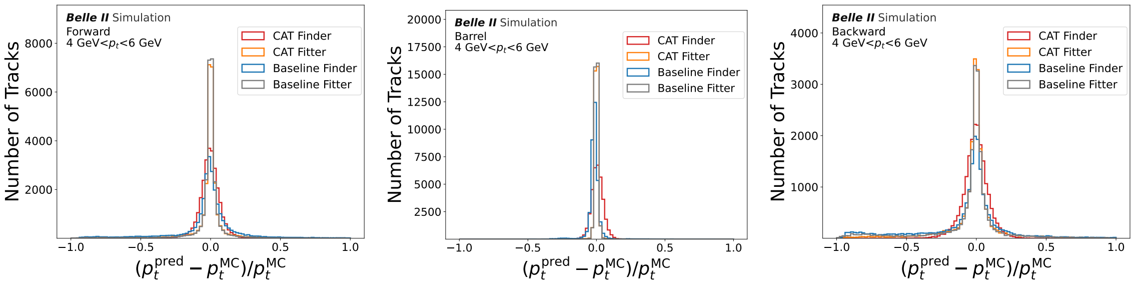
Figure C.18: Figures and text taken from [1]: Track finding (empty markers) and combined track finding and fitting efficiency (filled markers) for (top) displaced tracks in $h \rightarrow \mu^+\mu^-$ events, in (middle) $K_S^0 \rightarrow \pi^+\pi^-$ events, and for electrons in $e^+e^- \rightarrow h(\rightarrow \mu^+\mu^-)\chi_1\chi_2(\rightarrow e^+e^-\chi_1)$ events with *high data beam backgrounds*, as function of (left) the true simulated transverse momentum p_t^{MC}, and (right) the true simulated displacement v_ρ^{MC}.

C.6 High transverse momentum track resolution

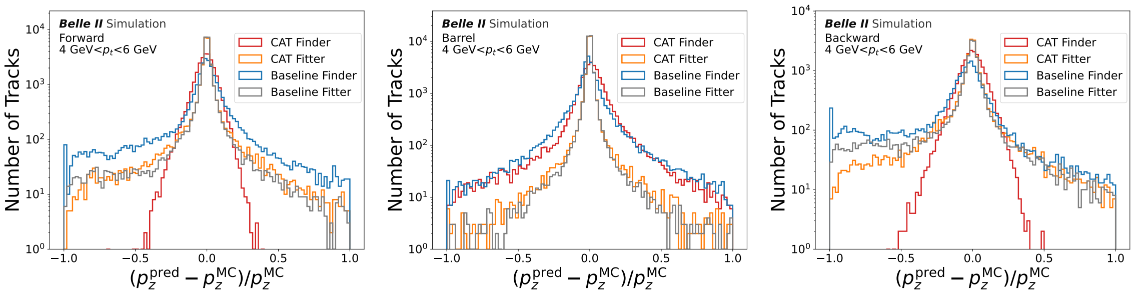
Fig. C.19 presents the relative transverse momentum resolution for tracks reconstructed and fitted by both the *CAT Finder* and the *Baseline Finder* (Fig. C.19a–Fig. C.19f), as well as the relative longitudinal momentum resolution in the transverse momentum range $4\ \text{GeV} < p_t < 6\ \text{GeV}$. This is illustrated in Fig. C.19c, where no significant tail is observed on the right side. Even for tracks with true momentum well above 6 GeV, the initial prediction from the *CAT Finder* provides a sufficiently accurate starting value for the subsequent track fitting algorithm. While the central region of the resolution distribution for the *CAT Finder* is broader compared to the *Baseline Finder*, the tails are considerably smaller.



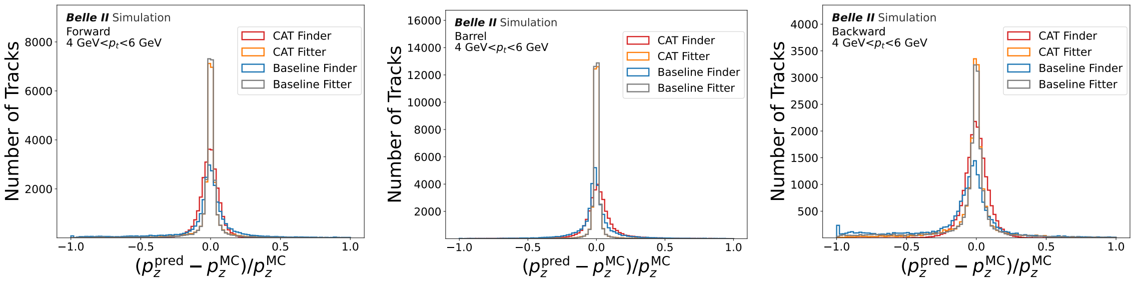
(a) Forward endcap. (b) Barrel. (c) Backward endcap.



(d) Forward endcap. (e) Barrel. (f) Backward endcap.



(g) Forward endcap. (h) Barrel. (i) Backward endcap.



(j) Forward endcap. (k) Barrel. (l) Backward endcap.

Figure C.19: Figures and text taken from [1]: Relative (Fig. C.19a-Fig. C.19f) transverse and (Fig. C.19g-Fig. C.19l) longitudinal momentum resolution as function of simulated transverse momentum p_t^{MC} for the intersecting prompt evaluation sample (category 1-3, see Table 3.2) in the (left) forward endcap, (center) barrel, and (right) backward endcap for tracks found (red) and fitted (orange) by both the *CAT Finder* and (blue and grey) the *Baseline Finder* for the high transverse momentum bin of $4\,\mathrm{GeV} < p_t < 6$ GeV.

C.7 Track helix parametrization resolution

The track parametrization follows a helix model and is evaluated at the point of closest approach (POCA) to the collision point (see Section 5.2). The distance between the POCA and the collision point in the transverse plane is denoted as d_0, and in the longitudinal direction as d_z. The angles of the transverse and longitudinal momentum at the POCA are denoted by ϕ and θ, respectively.

For matched tracks, the absolute residuals of these track parameters are defined, following Eq. (4.10), as

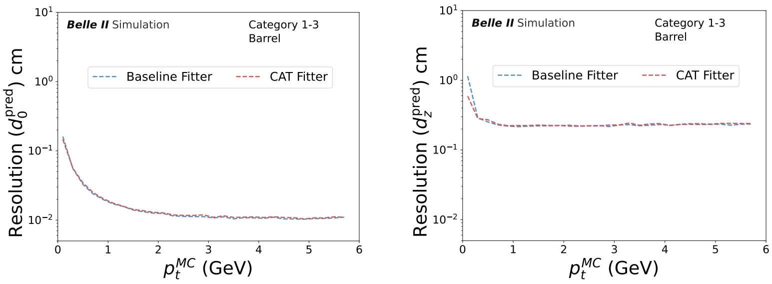
$$\eta(\phi, \theta, d_0, d_z) = (\phi, \theta, d_0, d_z)_{\text{rec}} - (\phi, \theta, d_0, d_z)_{\text{simulated}}. \tag{C.1}$$

The resolution of the track parameters, $r(\phi, \theta, d_0, d_z)$, is defined as the 68% coverage of the absolute residuals, following Eq. (4.11) :

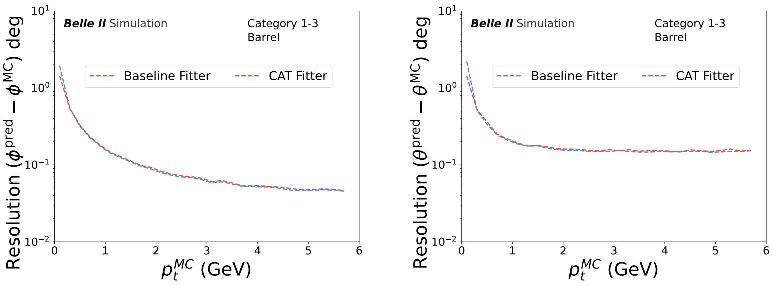
$$r(\phi, \theta, d_0, d_z) = P_{68\%}\Big(\big|\eta - P_{50\%}(\eta)\big|\Big), \tag{C.2}$$

where P_q denotes the q-th quantile of the distribution of η, and $P_{50\%}$ is the median [46].

The resulting resolutions of the fitting parameters as a function of the simulated transverse momentum p_t^{MC} are presented in Fig. C.20.



(a) d_0 Resolution. (b) d_z resolution.

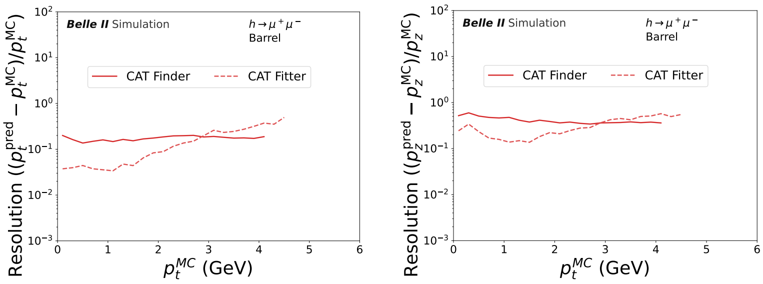


(c) ϕ resolution. (d) θ resolution.

Figure C.20: Figures and text taken from [1]: parameter resolution as function of simulated transverse momentum p_t^{MC} for the intersecting prompt evaluation sample (category 1-3, see Table 3.2) barrel.

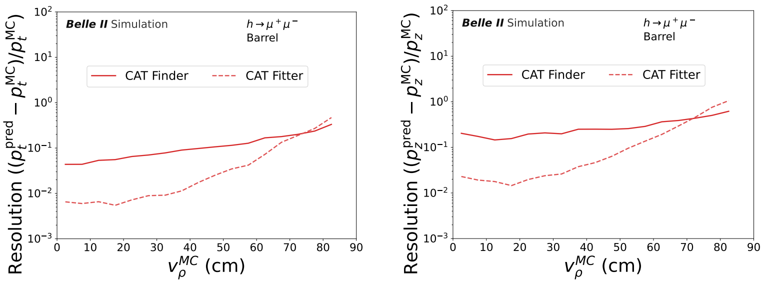
C.8 Track momentum resolution for additional *CAT Finder* samples in $h \to \mu^+\mu^-$ events

The relative momentum resolutions for displaced tracks from $h \to \mu^+\mu^-$ decays in the barrel for tracks only found by *CAT Finder* are shown in Fig. C.21.



(a) $\eta(p_t)$ Barrel.

(b) $\eta(p_z)$ Barrel.



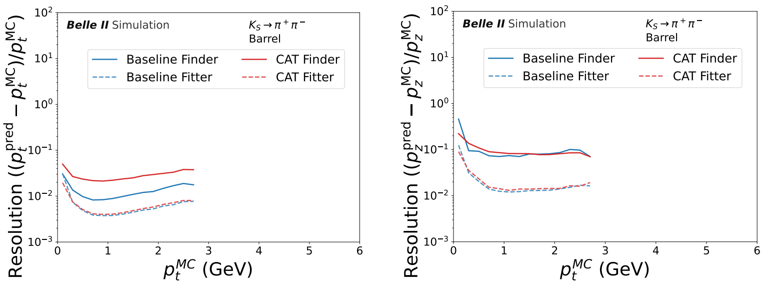
(c) $\eta(p_t)$ Barrel.

(d) $\eta(p_z)$ Barrel.

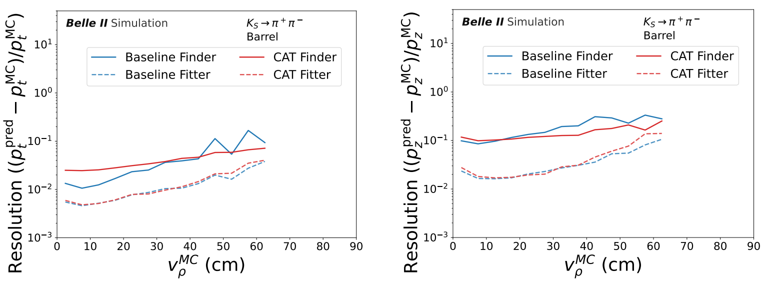
Figure C.21: Figures and text taken from [1]: Relative resolution of (first column) transverse and (second column) longitudinal momentum as function of (top row) simulated transverse momentum p_t^{MC} and (bottom row) simulated displacement v_ρ^{MC} for displaced tracks from $h \to \mu^+\mu^-$ decays in the barrel for tracks only found by *CAT Finder*.

C.9 Track momentum resolution in $K_S^0 \to \pi^+\pi^-$ events

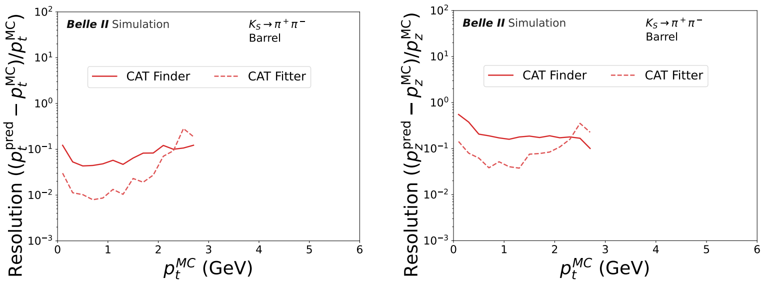
The relative momentum resolutions for displaced tracks from $K_S^0 \to \pi^+\pi^-$ decays in the barrel are shown in Fig. C.22.



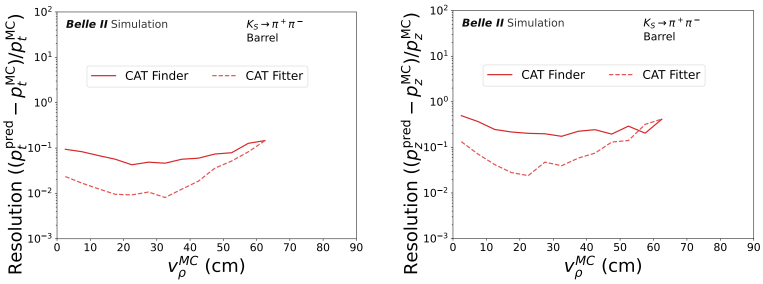
(a) $\eta(p_t)$ Barrel. (b) $\eta(p_z)$ Barrel.



(c) $\eta(p_t)$ Barrel. (d) $\eta(p_z)$ Barrel.



(e) $\eta(p_t)$ Barrel. (f) $\eta(p_z)$ Barrel.



(g) $\eta(p_t)$ Barrel. (h) $\eta(p_z)$ Barrel.

Figure C.22: Figures and text taken from [1]: Relative resolution of (first column) transverse and (second column) longitudinal momentum as function of (top row) simulated transverse momentum p_t^{MC} and (bottom row) simulated displacement v_ρ^{MC} for displaced tracks from $K_S^0 \to \pi^+\pi^-$ decays. Top (a-d) row shows the resolution for tracks found by both *CAT Finder* (red) and *Baseline Finder* (blue), and bottom (e-h) row for tracks only found by *CAT Finder*.

C.10 Hit efficiency for displaced electrons in $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$

The hit efficiency for electrons with a displaced starting position of at least 5 cm from the interaction point are given in Fig. C.23.

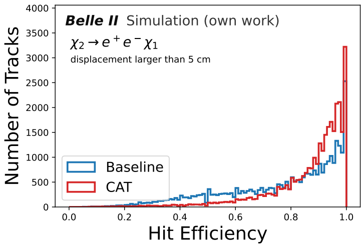
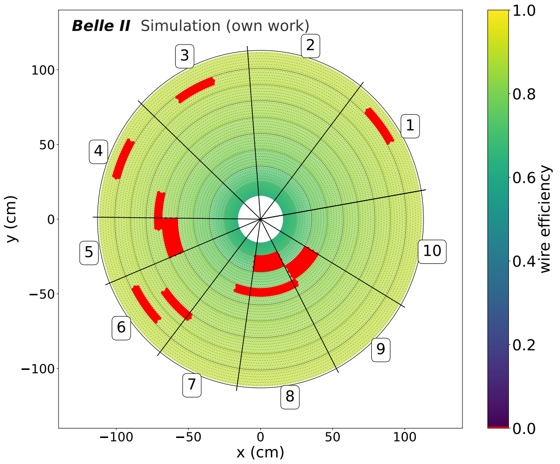


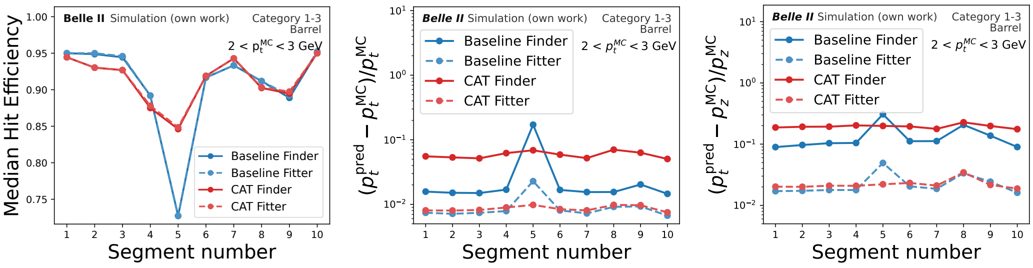
Figure C.23: Hit efficiency for electrons in $e^+e^- \to h(\to \mu^+\mu^-)\chi_1\chi_2(\to e^+e^-\chi_1)$ with a displacement $v_\rho > 5\,\mathrm{cm}$ for the intersecting sample between *CAT Finder* (red) and *Baseline Finder* (blue).

C.11 Hit efficiency and resolution for the intersecting sample for the robustness studies

The overall hit efficiency for the *Baseline Finder* for particles with 2 GeV< p_t^{MC} <3 GeV for the intersecting sample with *CAT Finder* is between 95% and 90%. For segment 5 with the axial hole, the hit efficiency drops to only 72%. For the *CAT Finder*, the hit efficiency also drops to 84%, but this is less steep than for the *Baseline Finder*. This drop in hit efficiency also reflects in the resolution of the transverse and longitudinal momentum, as for both the resolution drops for segment 5 for the *Baseline Finder* but not *CAT Finder*. Interesting to see now is the decrease in resolution for $\eta(p_z)$ both *CAT Finder* and *Baseline Finder* for segment 8 which includes the stereo hole in U2, while $\eta(p_t)$ stays the same. As the efficiency of the *Baseline Finder* for segment 9 is only at 33%, no large drop in resolution is seen as most tracks are not event found. The difference observed overall between *CAT Finder* and *Baseline Finder* is 0.001 worse $\eta(p_t)$ for the *CAT Finder* for the first 4 segments. This can be improved by fine-tuning the model either on a set of wire efficiencies and bad boards or directly on the corresponding wire efficiency map, as shown in [35].



(a) Wire efficiency map sectioned in ϕ segments.



(b) $\varepsilon_{\mathrm{hit}}$.

(c) $\eta(p_t)$.

(d) $\eta(p_z)$.

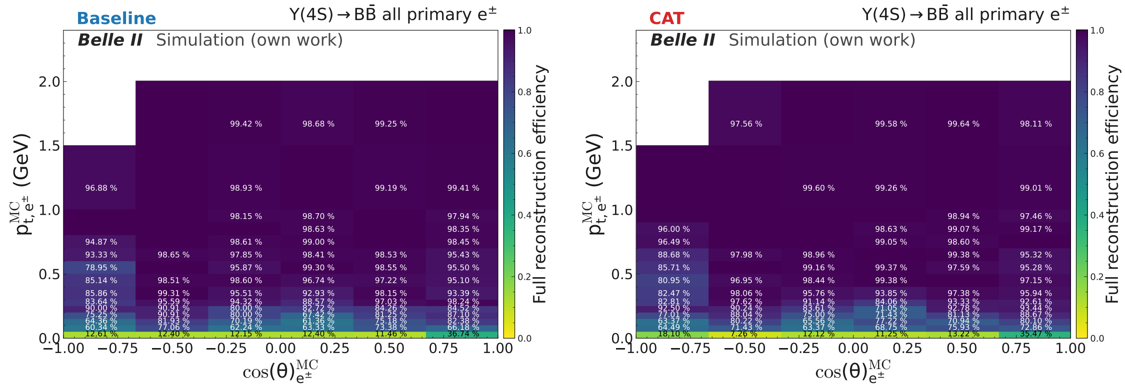
Figure C.24: The approximated wire efficiency map, described in detail in Section 3.1.2, sectioned in ϕ-segments in Fig. C.24a. For the intersecting sample for tracks with 2 GeV< p_t^{MC} <3 GeV, the hit efficiency $\varepsilon_{\mathrm{hit}}$, transverse and longitudinal resolution is shown at the bottom.

Appendix D

Full Reconstruction

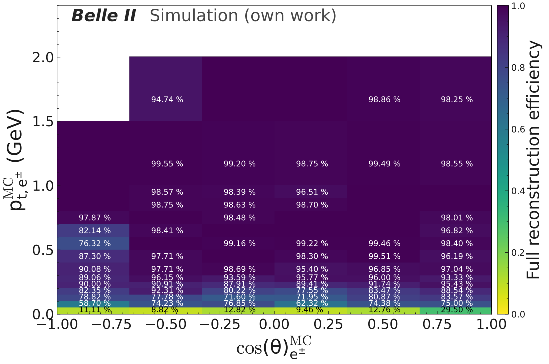
D.1 Prompt Tracks in $B\bar{B}$

The full reconstruction efficiency (see Eq. (4.9)) for the electrons (Fig. D.1), muons (Fig. D.3), kaons (Fig. D.5), and protons (Fig. D.7), as well as the hit efficiency depending on the number of seen particles in the event (see Fig. D.2 for electrons, Fig. D.4 for muons, Fig. D.6 for kaons, and Fig. D.8 for protons), is given in the following figures.

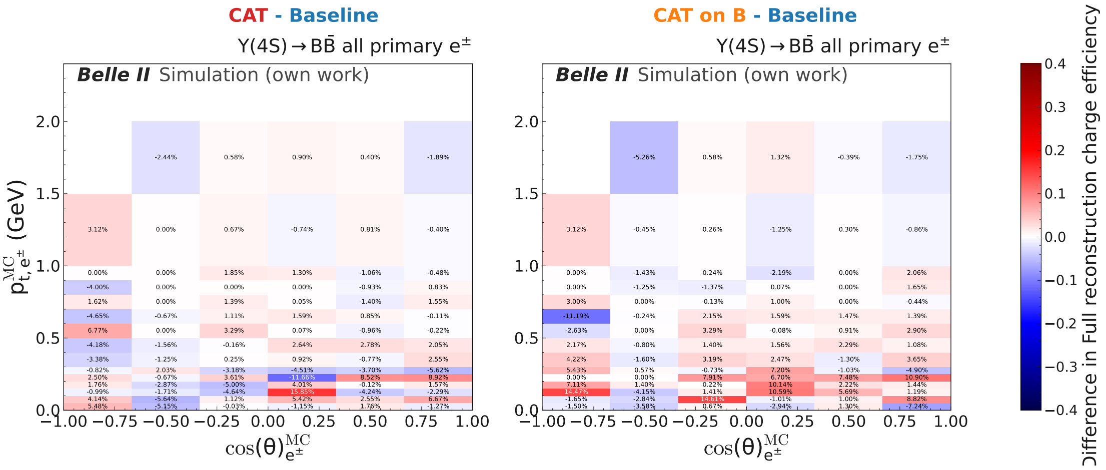


(a) *Baseline Finder*.

(b) *CAT Finder*.

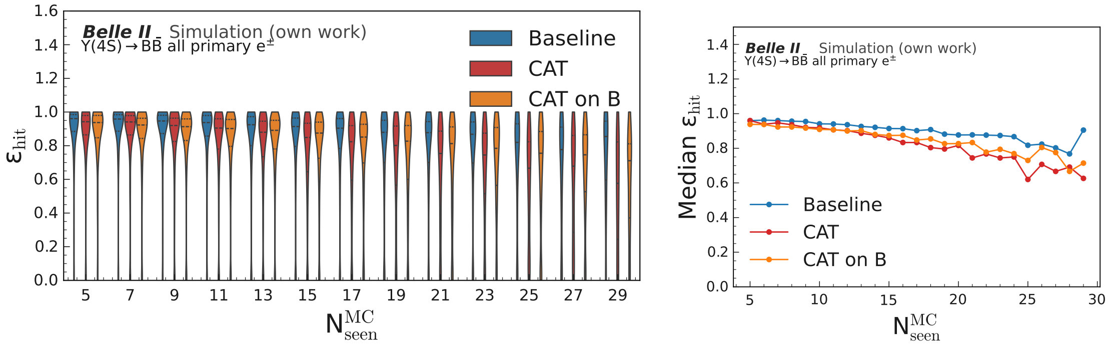


(c) *CAT B Finder*.



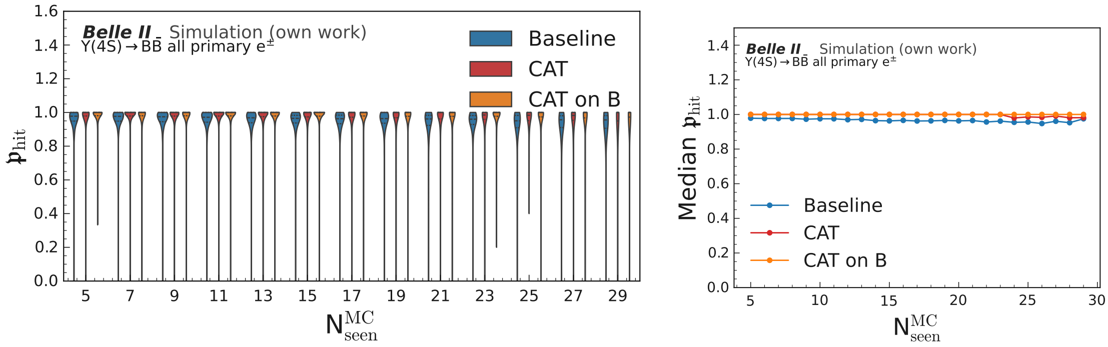
(d) Difference in the full reconstruction charge efficiency for the *CAT Finder* and *CAT B Finder* vs the *Baseline Finder*.

Figure D.1: The full reconstruction efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. D.1a), *CAT Finder* (Fig. D.1b), and *CAT B Finder* (Fig. D.1c). The difference between the two new tracking algorithms to the *Baseline Finder* is given as difference in full reconstruction efficiency in Fig. D.1d. The same plot but for the CDC-only track charge finding and fitting efficiency is given in Section C.4.



(a) Hit efficiency ε_{hit} distributions.

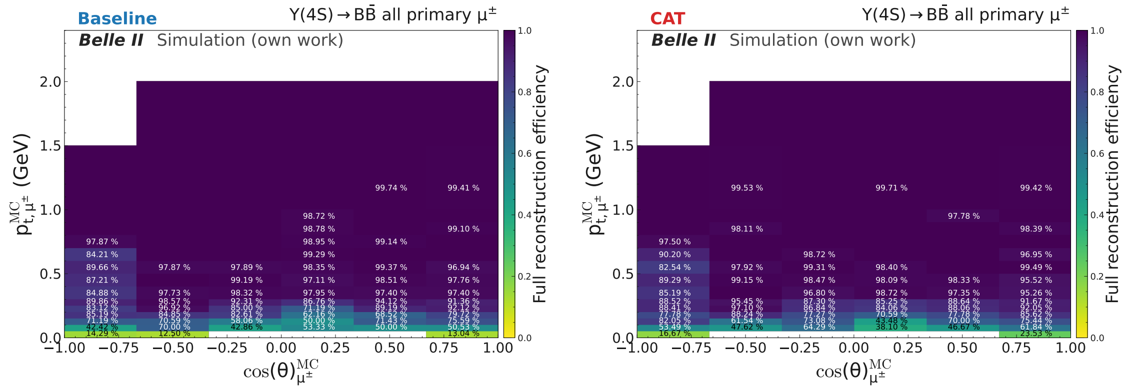
(b) Median hit efficiency ε_{hit}.



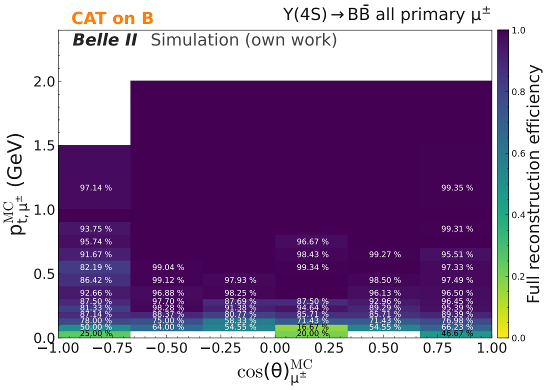
(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

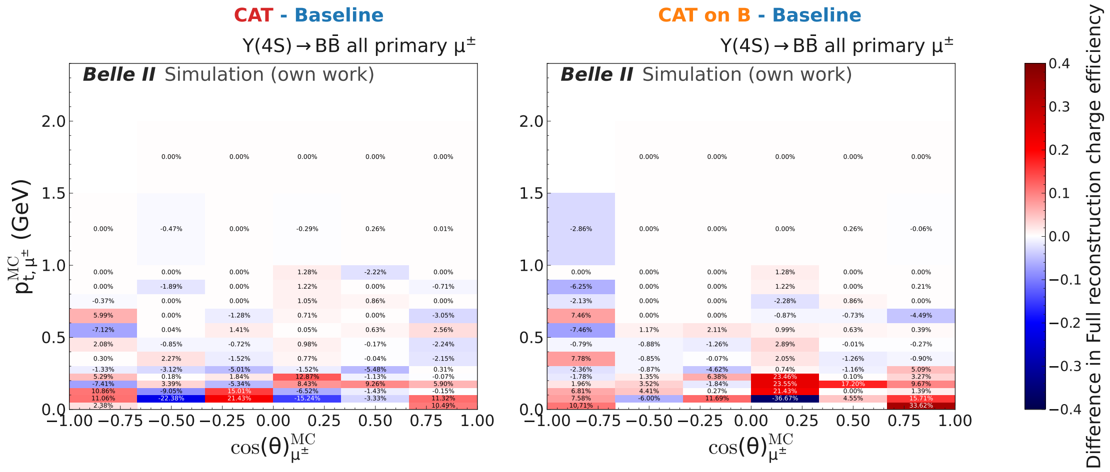
Figure D.2: Hit efficiency (top row) and hit purity (bottom row) over the number of particles seen in the CDC for all primary matched electrons. The left column shows the distributions for the *Baseline Finder* in blue, *CAT Finder* in red, and *CAT B Finder* in orange, while the right column shows the median for all three algorithms.



(a) *Baseline Finder*. (b) *CAT Finder*.

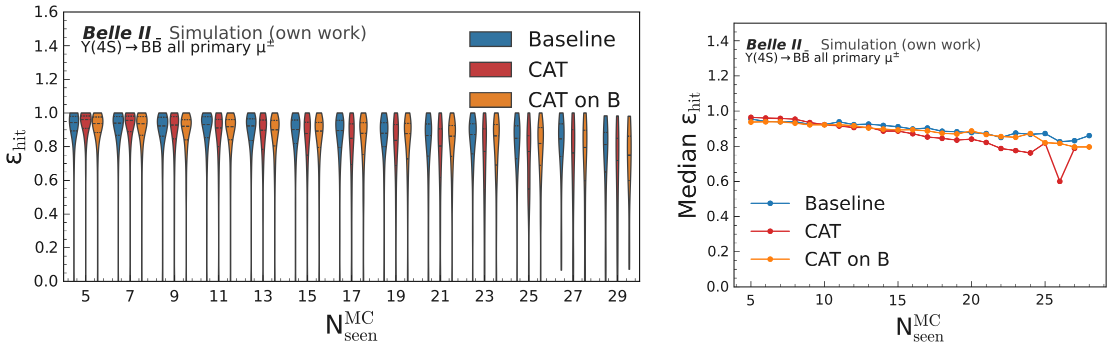


(c) *CAT B Finder*.



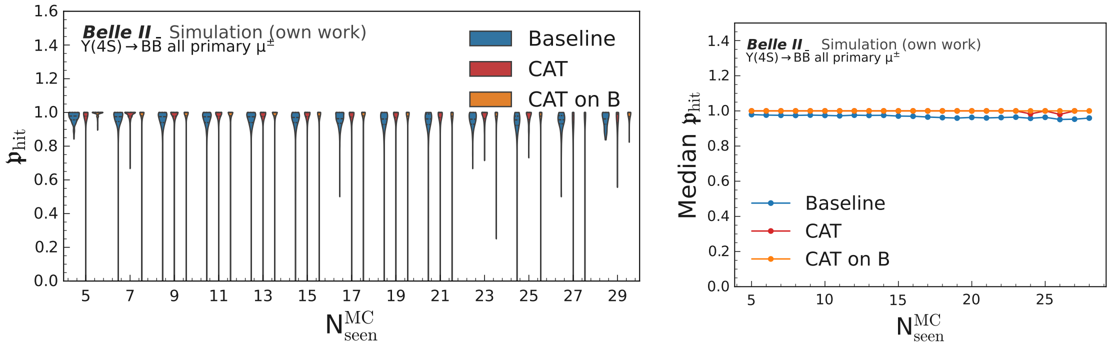
(d) Difference in the full reconstruction charge efficiency for the *CAT Finder* and *CAT B Finder* vs the *Baseline Finder*.

Figure D.3: The full reconstruction efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. D.1a), *CAT Finder* (Fig. D.1b), and *CAT B Finder* (Fig. D.1c). The difference between the two new tracking algorithms to the *Baseline Finder* is given as difference in full reconstruction efficiency in Fig. D.1d. The same plot but for the CDC-only track charge finding and fitting efficiency is given in Section C.4.



(a) Hit efficiency ε_{hit} distributions.

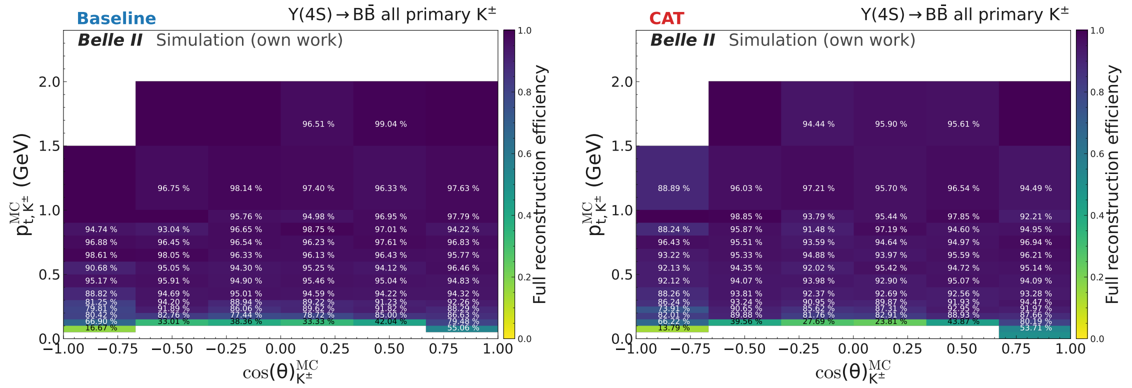
(b) Median hit efficiency ε_{hit}.



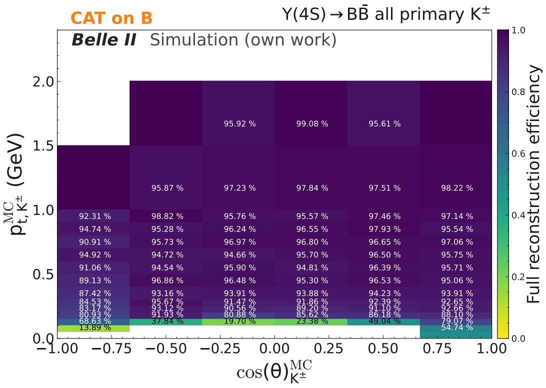
(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

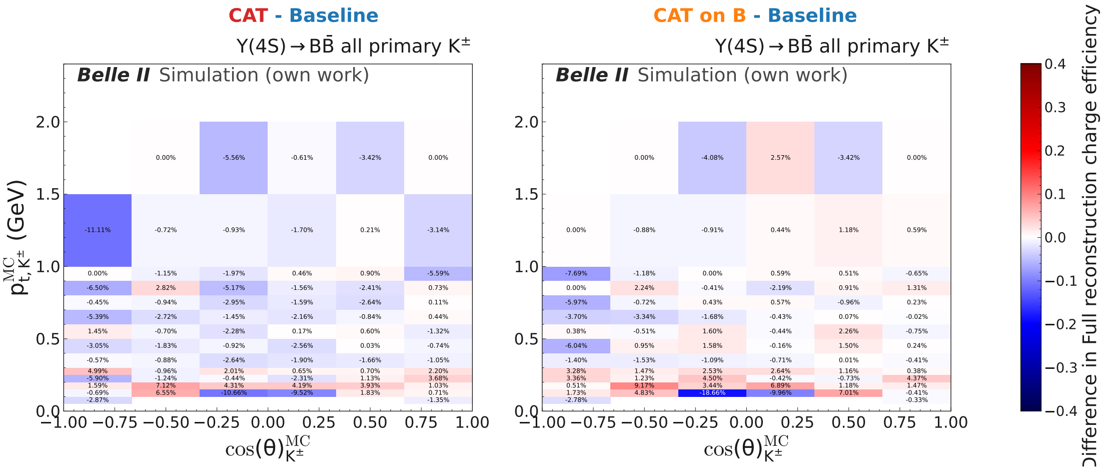
Figure D.4: Hit efficiency (top row) and hit purity (bottom row) over the number of particles seen in the CDC for all primary matched muons. The left column shows the distributions for the *Baseline Finder* in blue, *CAT Finder* in red, and *CAT B Finder* in orange, while the right column shows the median for all three algorithms.



(a) *Baseline Finder*. (b) *CAT Finder*.

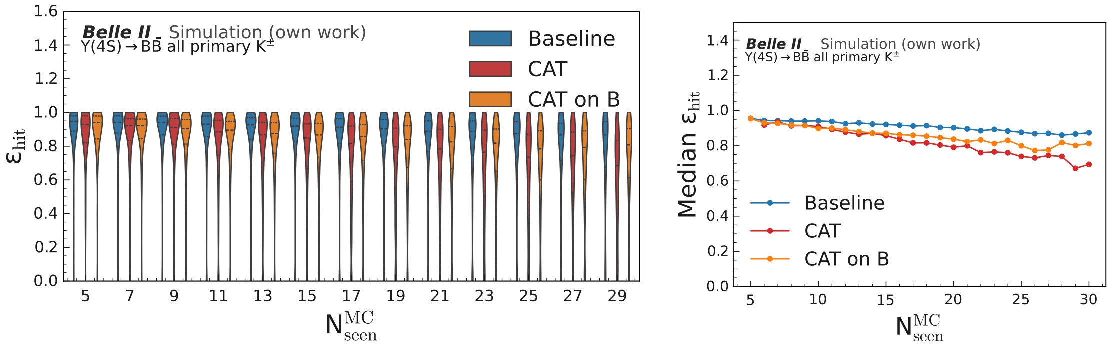


(c) *CAT B Finder*.



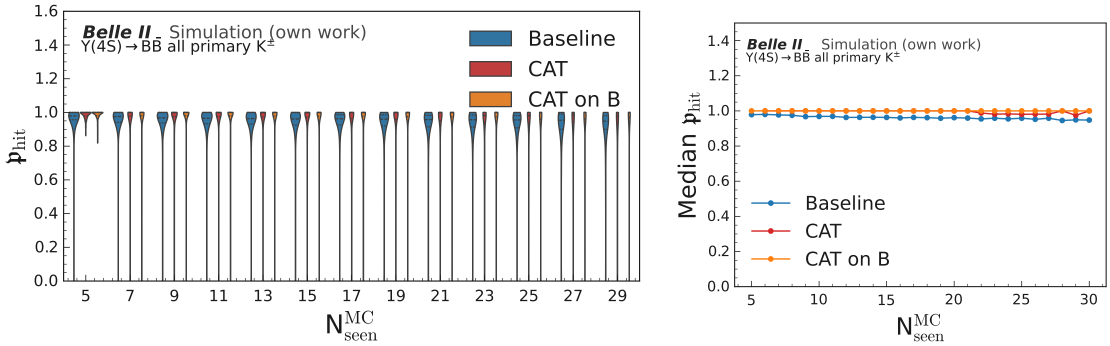
(d) Difference in the full reconstruction charge efficiency for the *CAT Finder* and *CAT B Finder* vs the *Baseline Finder*.

Figure D.5: The full reconstruction efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. D.5a), *CAT Finder* (Fig. D.5b), and *CAT B Finder* (Fig. D.5c). The difference between the two new tracking algorithms to the *Baseline Finder* is given as difference in full reconstruction efficiency in Fig. D.5d. The same plot but for the CDC-only track charge finding and fitting efficiency is given in Section C.4.



(a) Hit efficiency $\varepsilon_{\mathrm{hit}}$ distributions.

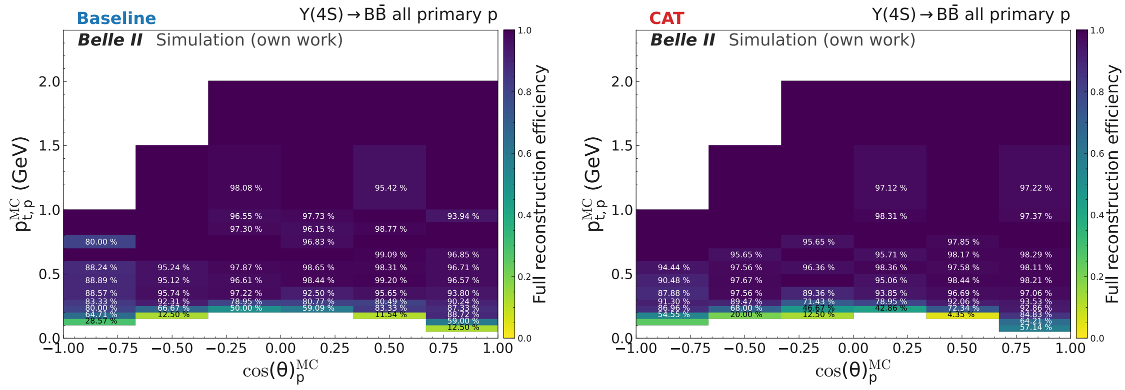
(b) Median hit efficiency $\varepsilon_{\mathrm{hit}}$.



(c) Hit purity $\mathfrak{p}_{\mathrm{hit}}$ distributions.

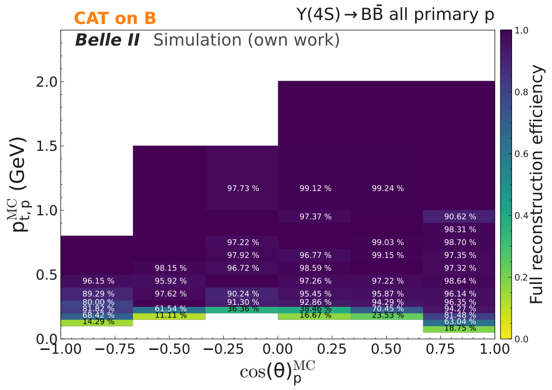
(d) Median hit purity $\mathfrak{p}_{\mathrm{hit}}$.

Figure D.6: Hit efficiency (top row) and hit purity (bottom row) over the number of particles seen in the CDC for all primary matched muons. The left column shows the distributions for the *Baseline Finder* in blue, *CAT Finder* in red, and *CAT B Finder* in orange, while the right column shows the median for all three algorithms.

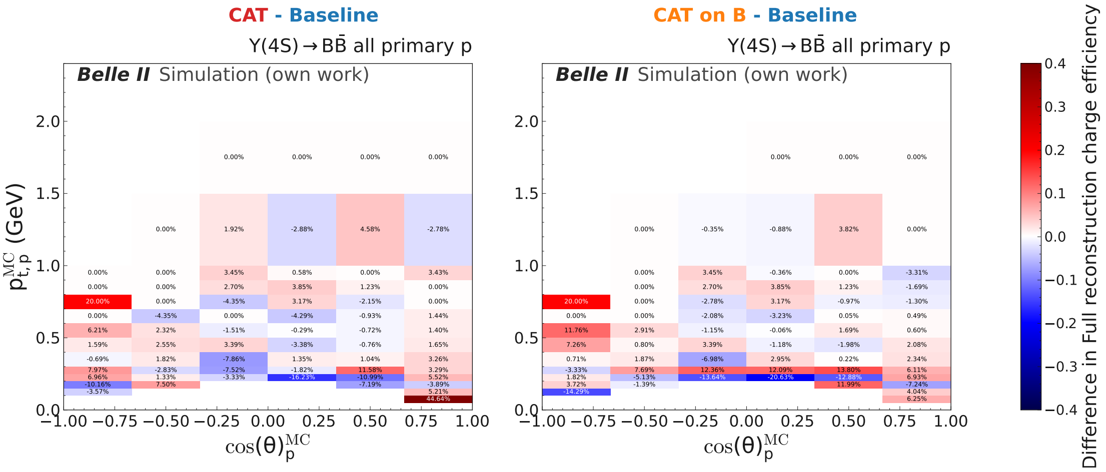


(a) *Baseline Finder*.

(b) *CAT Finder*.

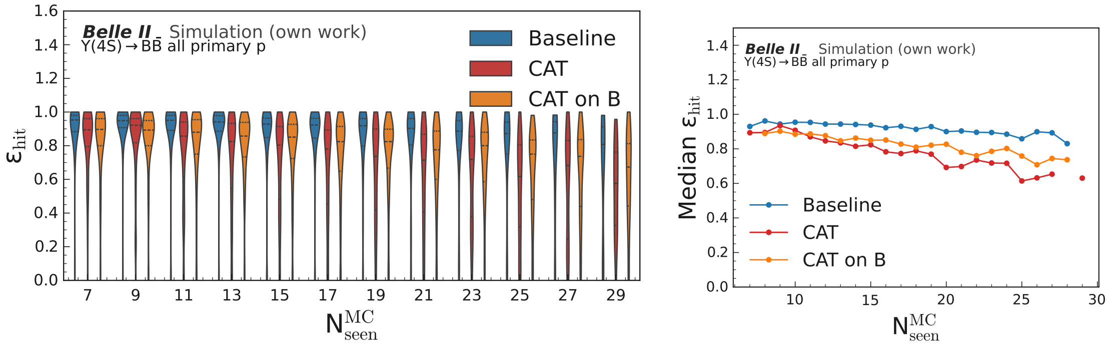


(c) *CAT B Finder*.



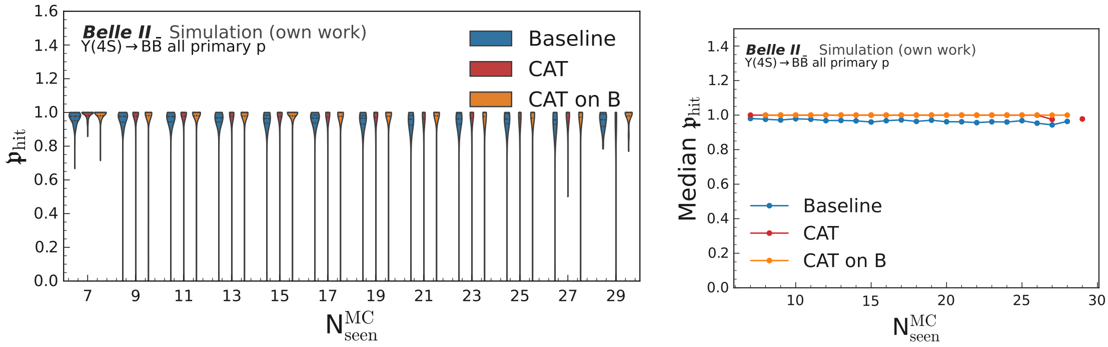
(d) Difference in the full reconstruction charge efficiency for the *CAT Finder* and *CAT B Finder* vs the *Baseline Finder*.

Figure D.7: The full reconstruction efficiency for the true transverse momentum p_t^{MC} over the polar angle for the *Baseline Finder* (Fig. D.7a), *CAT Finder* (Fig. D.7b), and *CAT B Finder* (Fig. D.7c). The difference between the two new tracking algorithms to the *Baseline Finder* is given as difference in full reconstruction efficiency in Fig. D.7d. The same plot but for the CDC-only track charge finding and fitting efficiency is given in Section C.4.



(a) Hit efficiency ε_{hit} distributions.

(b) Median hit efficiency ε_{hit}.



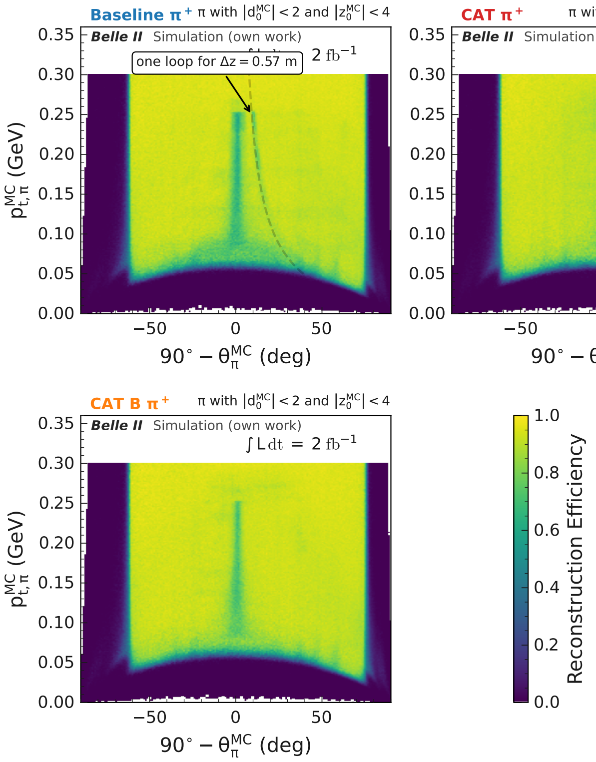
(c) Hit purity $\mathfrak{p}_{\text{hit}}$ distributions.

(d) Median hit purity $\mathfrak{p}_{\text{hit}}$.

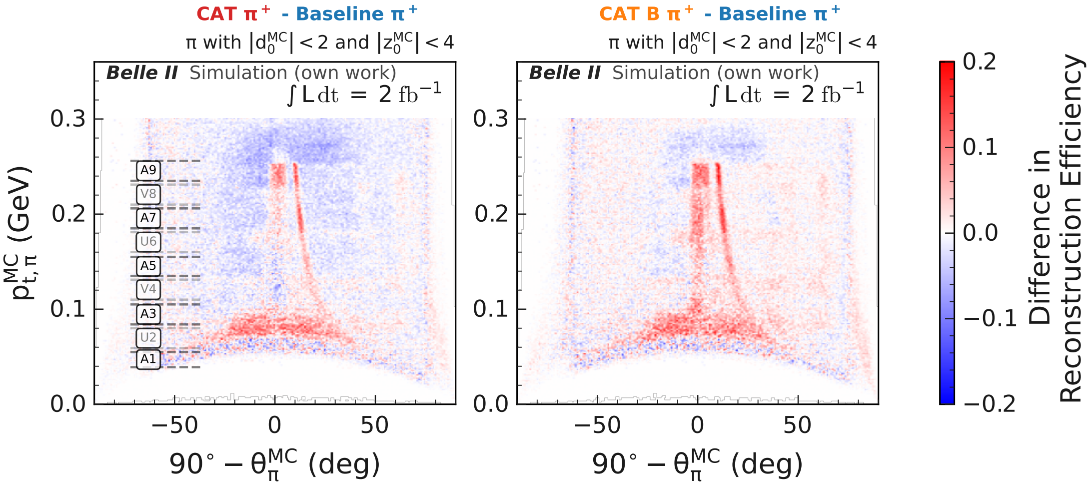
Figure D.8: Hit efficiency (top row) and hit purity (bottom row) over the number of particles seen in the CDC for all primary matched muons. The left column shows the distributions for the *Baseline Finder* in blue, *CAT Finder* in red, and *CAT B Finder* in orange, while the right column shows the median for all three algorithms.

D.2 Reconstruction efficiency per charge

The full reconstruction efficiency for positively charged pions is given in Fig. D.9 and for negatively charged pions in Fig. D.10.

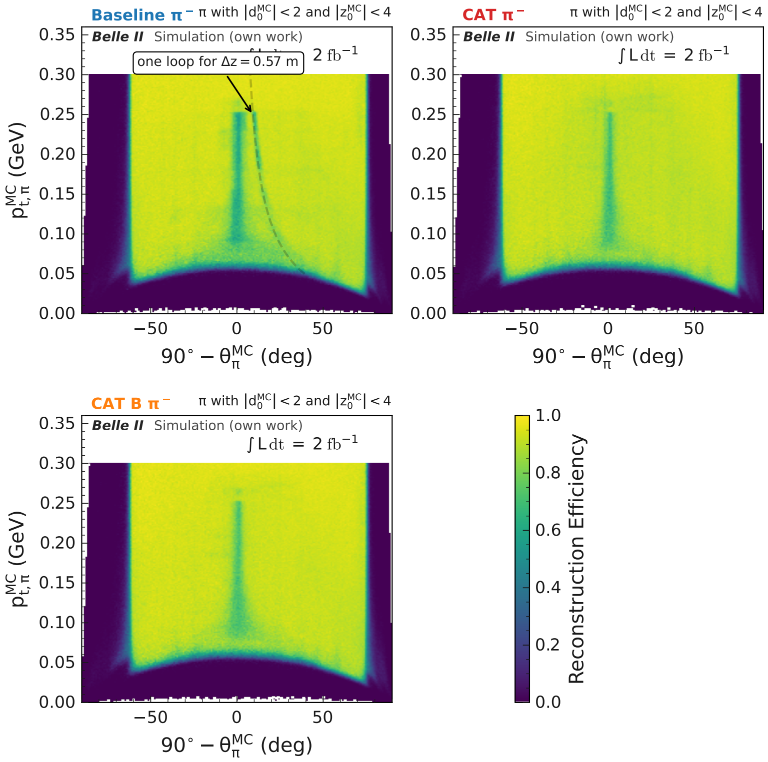


(a) Reconstruction efficiency for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom left).

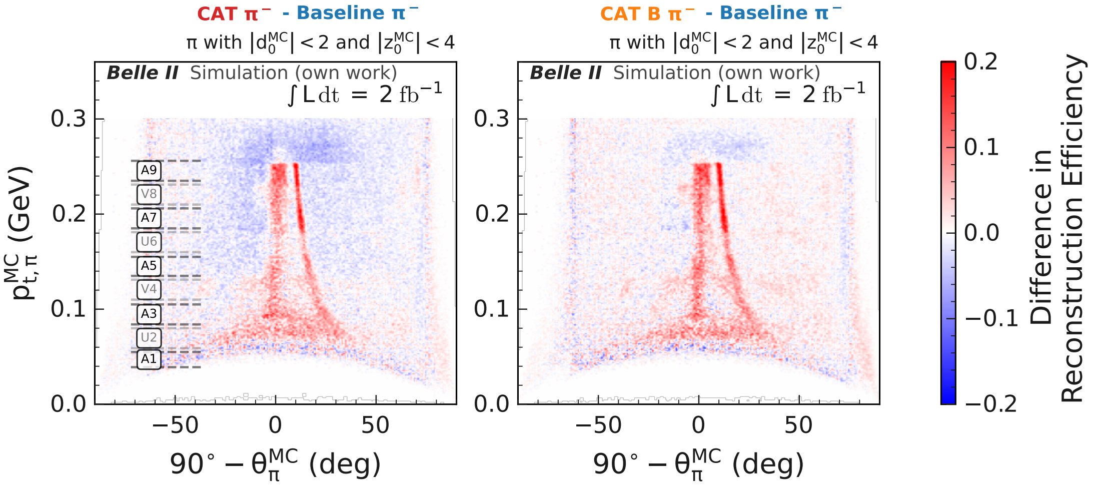


(b) Difference in reconstruction efficiency for *CAT Finder* and *Baseline Finder* (left) and *CAT B Finder* and *Baseline Finder* (right).

Figure D.9: Reconstruction efficiency of prompt pions with positive charge for all three track finding algorithms in full reconstruction. Fig. 7.6a shows the overall values, in dashed black the function for the particle traversing exactly one loop (Eq. (7.4)) for the respective transverse momentum and polar angle for the inner CDC boundary in forward direction of $z = 57\,\mathrm{cm}$.Fig. 7.6b shows the difference of detection asymmetry.



(a) Reconstruction efficiency for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom left).



(b) Difference in reconstruction efficiency for *CAT Finder* and *Baseline Finder* (left) and *CAT B Finder* and *Baseline Finder* (right).

Figure D.10: Reconstruction efficiency of prompt pions with negative charge for all three track finding algorithms in full reconstruction. Fig. 7.6a shows the overall values, in dashed black the function for the particle traversing exactly one loop (Eq. (7.4)) for the respective transverse momentum and polar angle for the inner CDC boundary in forward direction of $z = 57\,\mathrm{cm}$.Fig. 7.6b shows the difference of detection asymmetry.

D.3 Ionization energy loss in the CDC and SVD

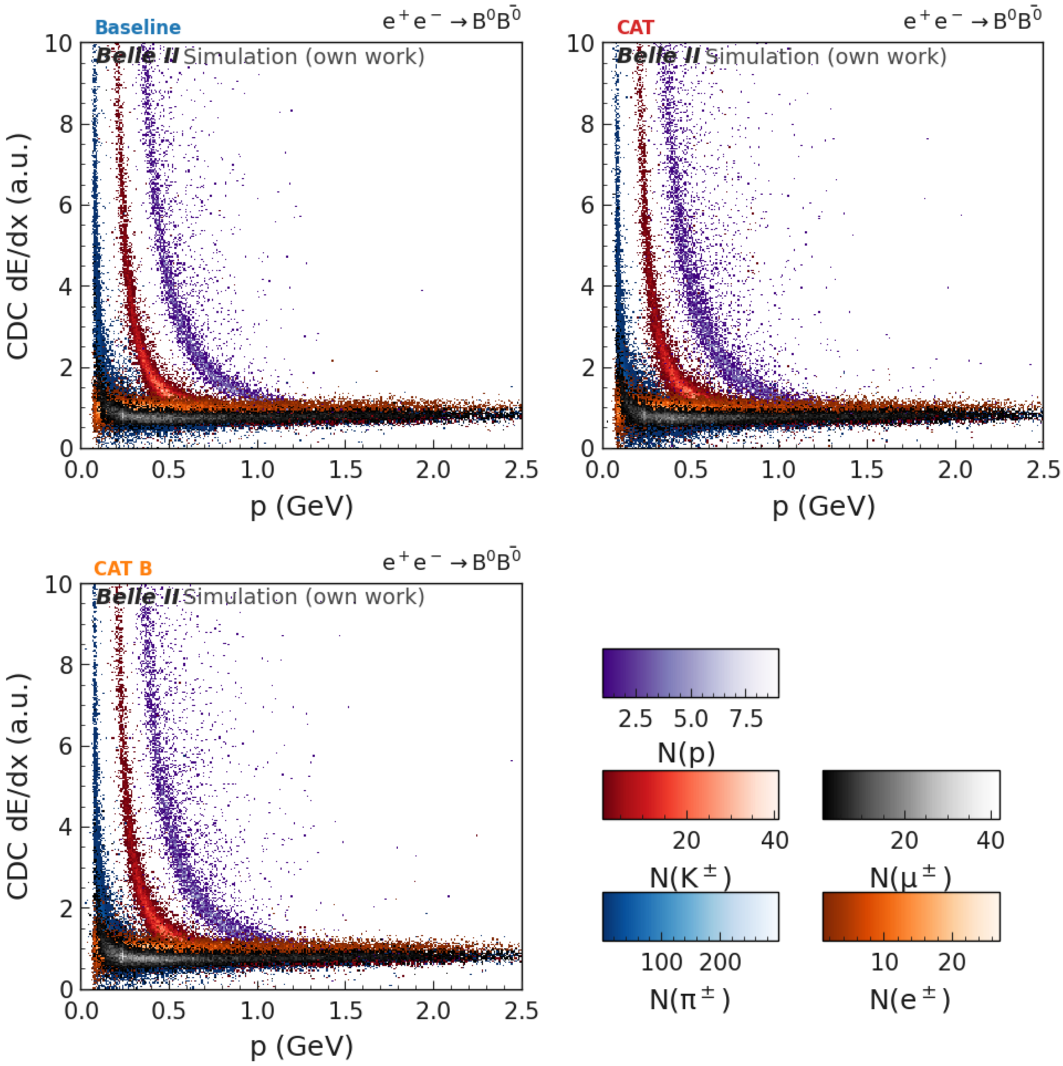


Figure D.11: CDC dE/dx over momentum for electrons (grey), muons (orange), pions (blue), Kaons (reds) and protons (purple) for the *Baseline Finder* (top left), *CAT Finder* (top right) and *CAT B Finder* (bottom left). The dE/dx values are normalized to the average energy loss for an electron at the Fermi plateau [105], as the absolute value is not necessary to distinguish between the different particle types.

A similar comparison is shown for the SVD dE/dx in Fig. D.12 with full distributions available in Fig. D.13. Compared to the CDC, the ionization energy loss curves in the SVD are less pronounced. Nevertheless, at low momentum, the three particle types are still clearly distinguishable, and the results are consistent across all three tracking algorithms. This confirms that the SVD extrapolation is functioning correctly and that a sufficient number of SVD clusters are being assigned to the reconstructed tracks.

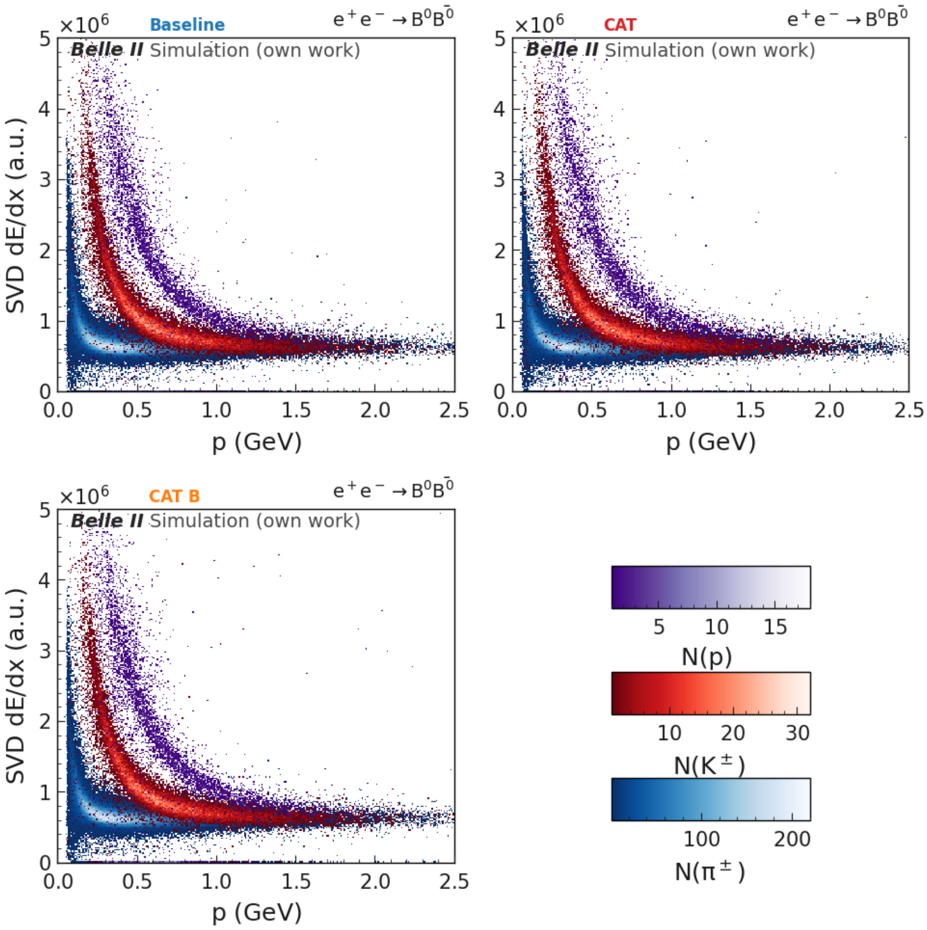


Figure D.12: SVD dE/dx over momentum for pions (blue), Kaons (reds) and protons (purple) for the *Baseline Finder* (top left), *CAT Finder* (top right) and *CAT B Finder* (bottom left). See Fig. D.11 for details.

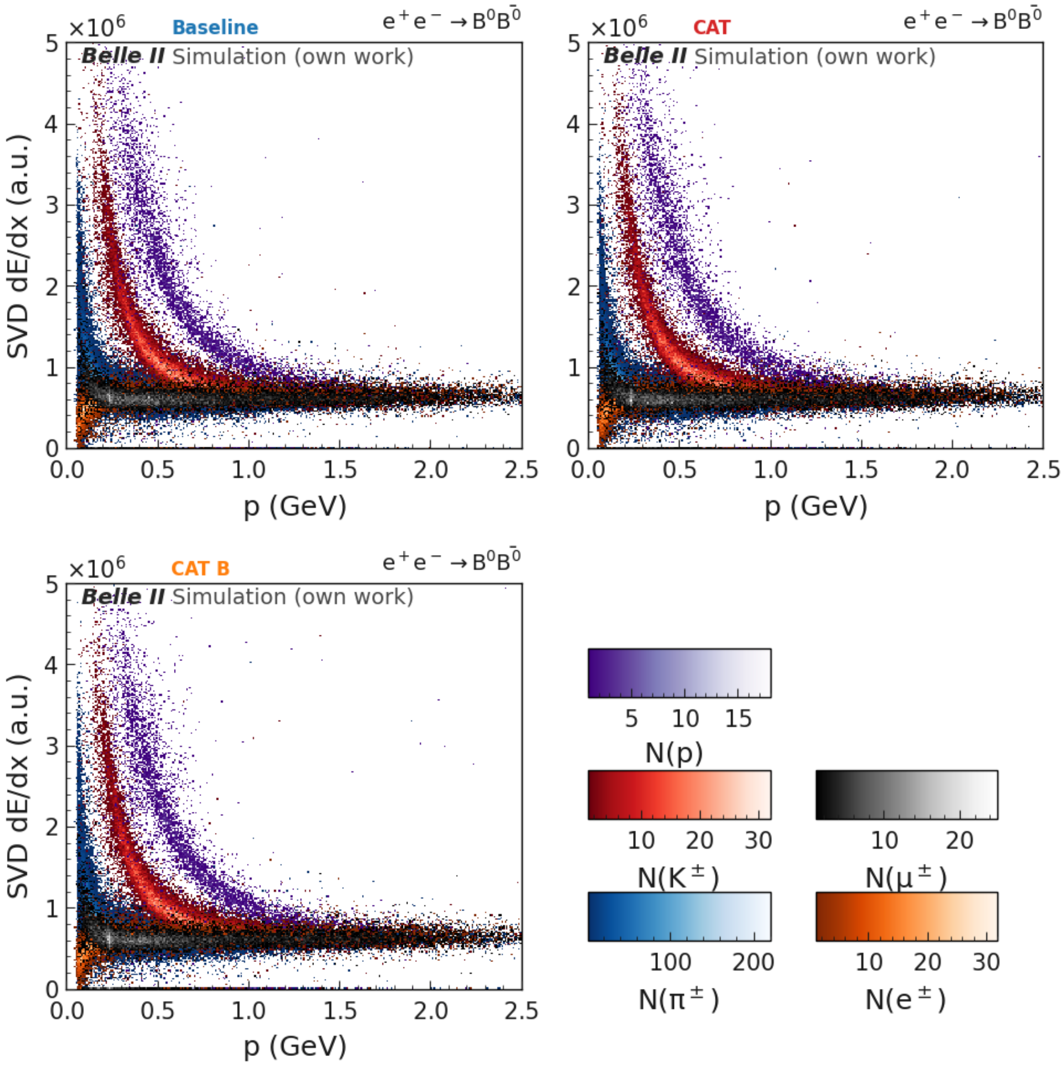
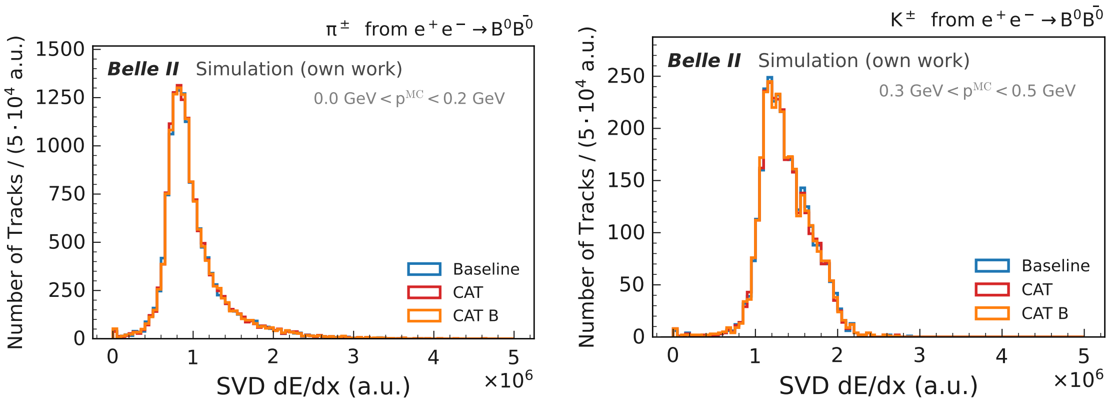
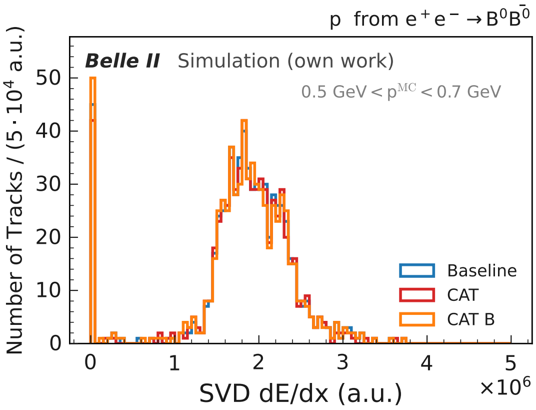


Figure D.13: SVD dE/dx over momentum for electrons (grey), muons (orange), pions (blue), Kaons (reds) and protons (purple) for the *Baseline Finder* (top left), *CAT Finder* (top right) and *CAT B Finder* (bottom left). See Fig. D.11 for details.



(a) $\pi^{\pm}$ dE/dx for $0\,\mathrm{GeV} < p < 0.2\,\mathrm{GeV}$.

(b) $K^{\pm}$ dE/dx for $0.3\,\mathrm{GeV} < p < 0.5\,\mathrm{GeV}$.



(c) $p/\bar{p}$ dE/dx for $0.5\,\mathrm{GeV} < p < 0.7\,\mathrm{GeV}$.

Figure D.14: SVD dE/dx distribution for a given true momentum range for pions (left), kaons (middle) and protons (right) for the *Baseline Finder* (blue), *CAT Finder* (red) and *CAT B Finder* (orange) for the intersecting sample between all three algorithms. See Fig. D.11 for details.

Appendix E

Validation on Data

E.1 Number of CDC hits in $J/\psi \to e^+e^-$

Figs. E.1 to E.4 show the number of CDC hits for the e^+ of the reconstructed J/ψ for runs from experiment 22 and experiment 26.

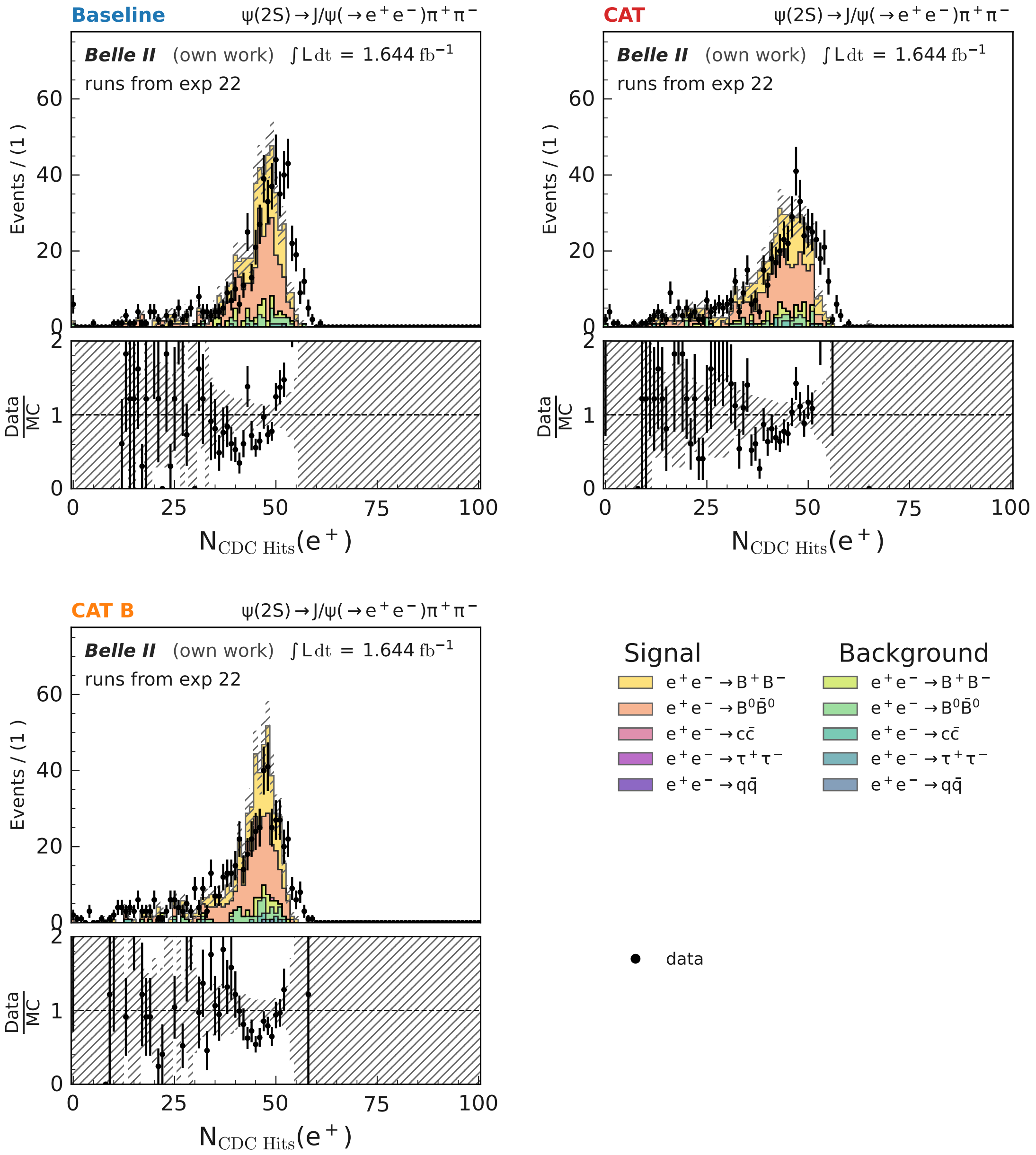


Figure E.1: Number of CDC hits for the e^+ of the reconstructed J/ψ mass from two opposite charged muons $M_{e^+e^-}$ in the signal region of 3.07 GeV to 3.12 GeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(e^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

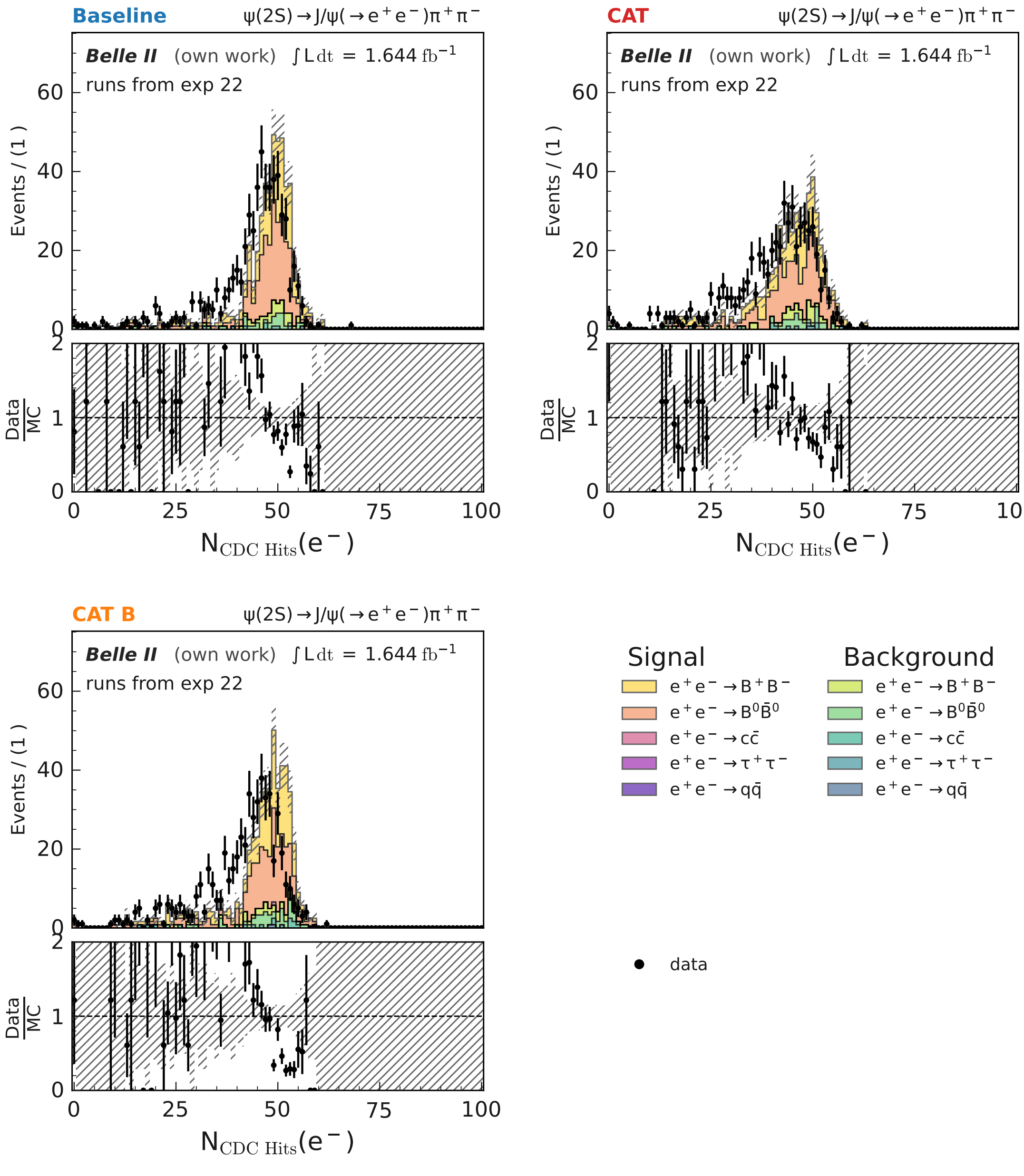


Figure E.2: Number of CDC hits for the e^- of the reconstructed J/ψ mass from two opposite charged muons $M_{e^+e^-}$ in the signal region of 3.07 GeV to 3.12 GeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(e^-)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

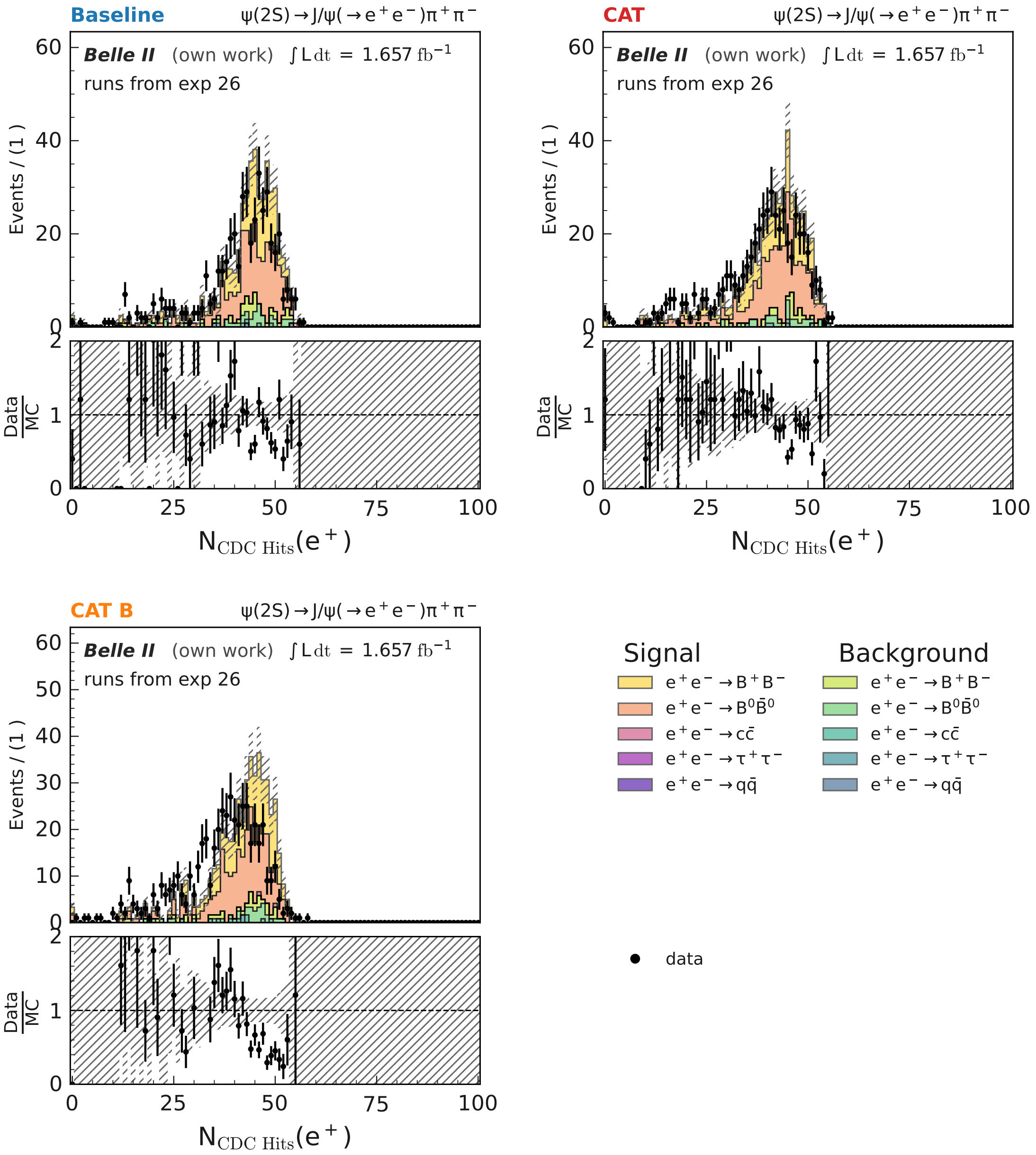


Figure E.3: Number of CDC hits for the e^+ of the reconstructed J/ψ mass from two opposite charged muons $M_{e^+e^-}$ in the signal region of 3.07 GeV to 3.12 GeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(e^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.657\,\mathrm{fb}^{-1}$.

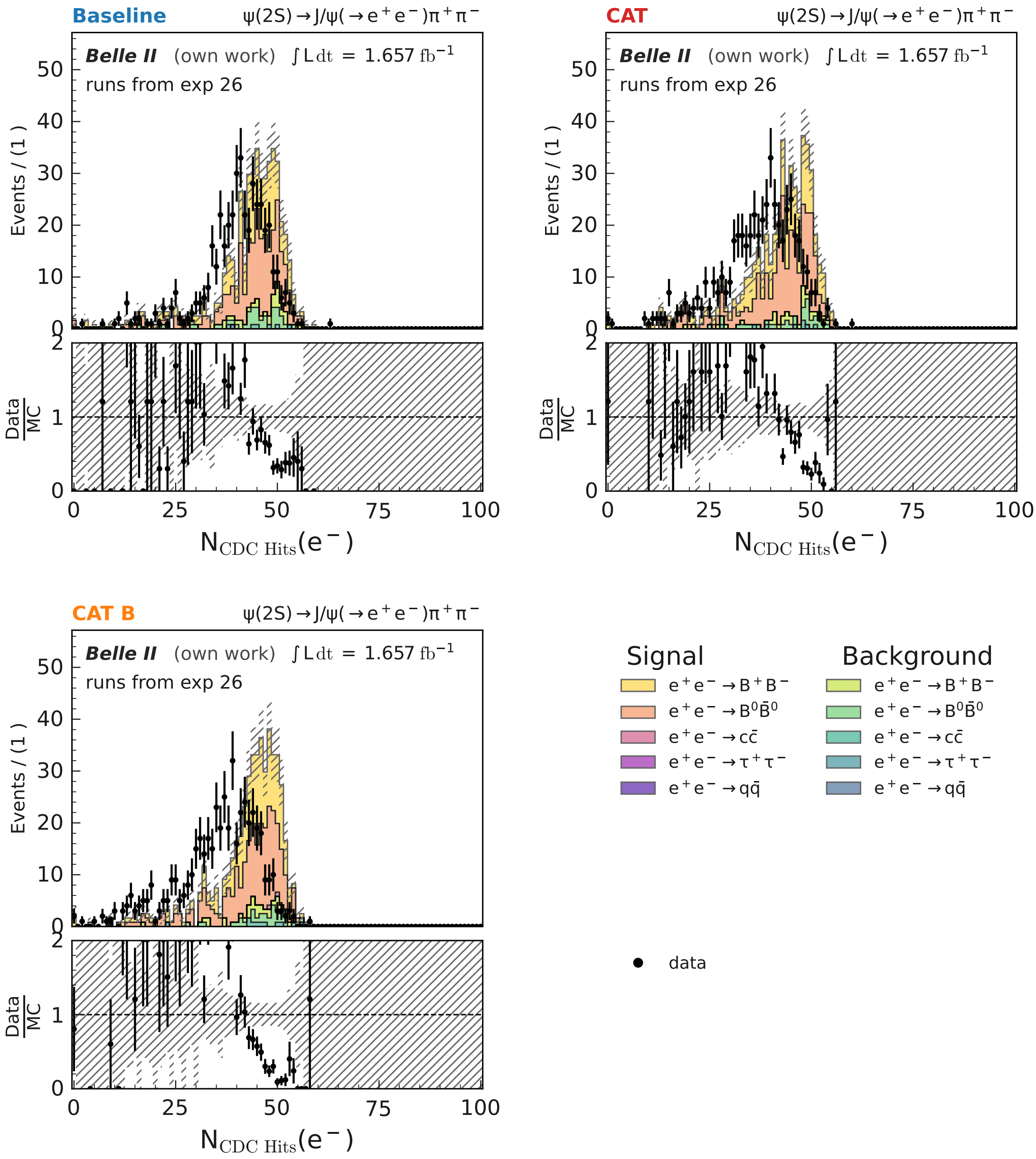


Figure E.4: Number of CDC hits for the e^- of the reconstructed J/ψ mass from two opposite charged muons $M_{e^+e^-}$ in the signal region of 3.07 GeV to 3.12 GeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(e^-)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed J/ψ. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.657\,\mathrm{fb}^{-1}$.

E.2 $D^{*+} \to D^0(\to K_s^0(\to \pi^+\pi^-)\pi^+\pi^-)\pi_S^+$

The transverse vertex distance $d_{\rho,K_S^0} = \sqrt{d_{x,K_S^0}^2 + d_{y,K_S^0}^2}$ from the interaction point of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV are given in Fig. E.5 and Fig. E.6 for measured data and simulation. The number of CDC hits of the π^+ and π^- of the K_S^0 are given in Figs. E.7 to E.10. The hit purity and hit efficiency on simulation only is given in Figs. E.11 and E.12

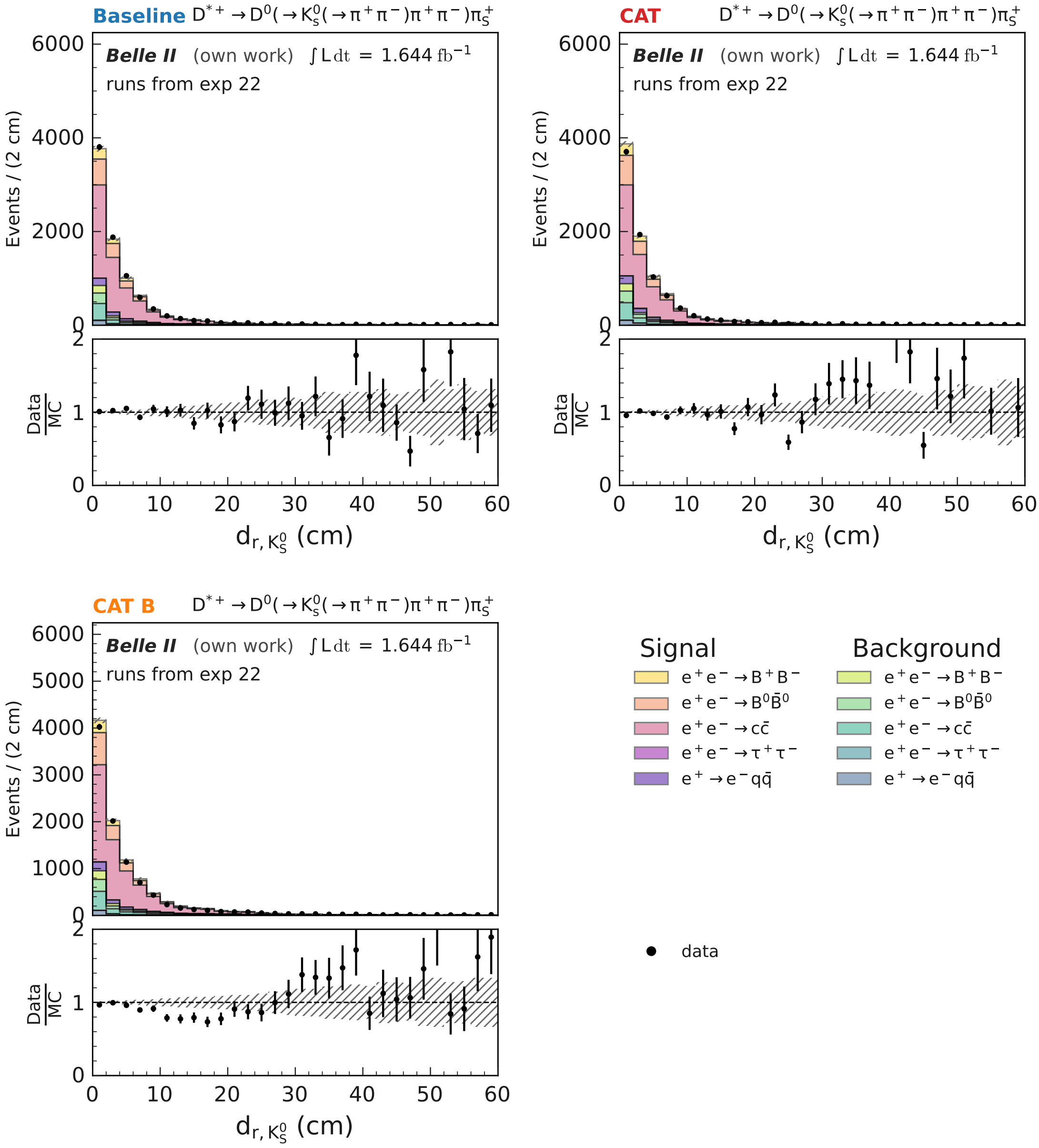


Figure E.5: Transverse vertex distance $d_{\rho,K_S^0} = \sqrt{d^2_{x,K_S^0} + d^2_{y,K_S^0}}$ from the interaction point of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *Baseline Finder* followed by *CAT Finder* (purple). The statistical uncertainties are correlated, as the same measured data is used for the different algorithms. Signal and background for the simulation refer to the correctly reconstructed K_S^0. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

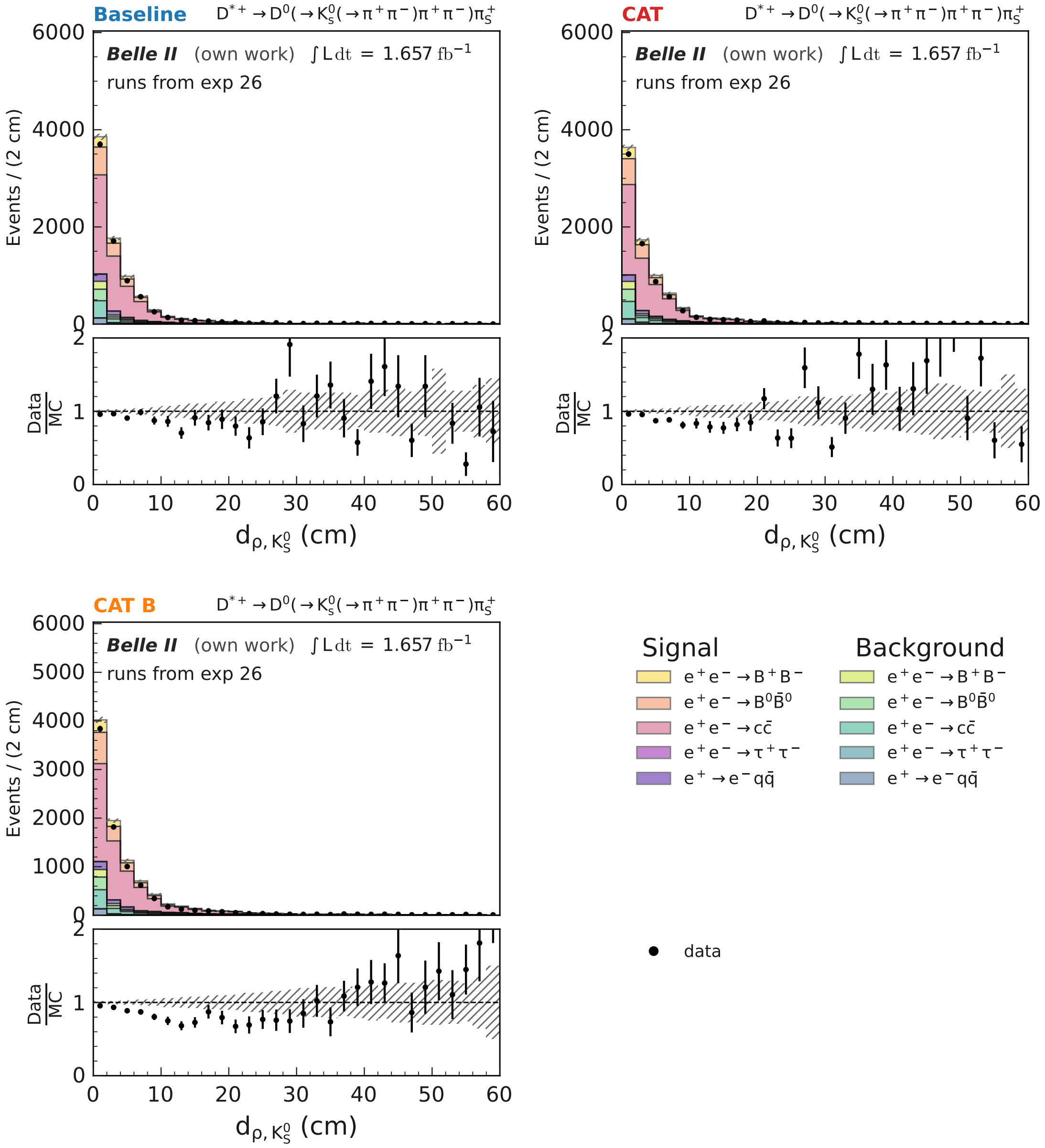


Figure E.6: Transverse vertex distance $d_{\rho,K_S^0} = \sqrt{d^2_{x,K_S^0} + d^2_{y,K_S^0}}$ from the interaction point of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for measured data, for the full reconstruction using the *Baseline Finder* (blue), *CAT Finder* (red) and *Baseline Finder* followed by *CAT Finder* (purple). The statistical uncertainties are correlated, as the same measured data is used for the different algorithms. Signal and background for the simulation refer to the correctly reconstructed K_S^0. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.657\,\text{fb}^{-1}$.

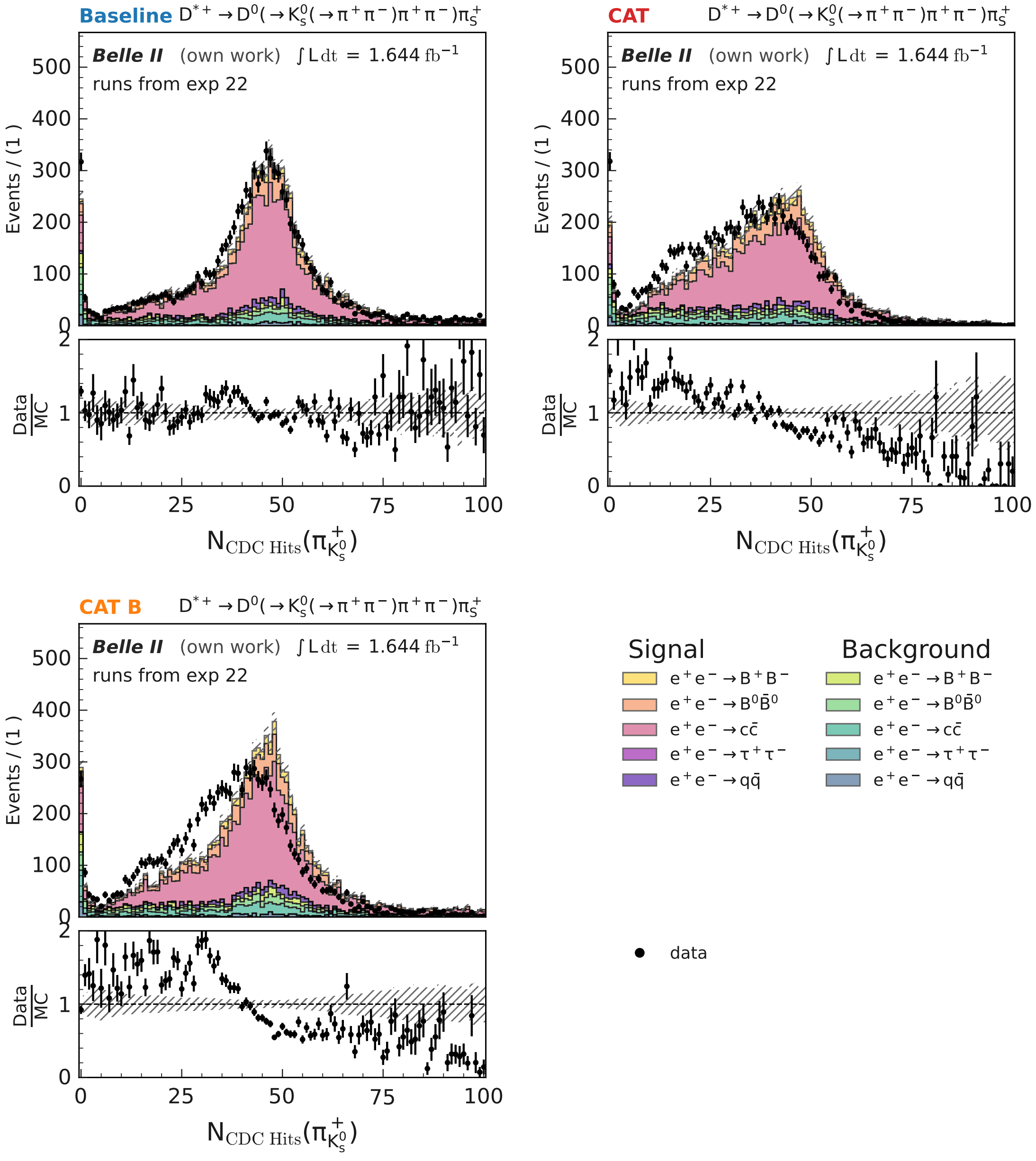


Figure E.7: Number of CDC hits for the π^+ of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(\pi^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed K_S^0 . The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

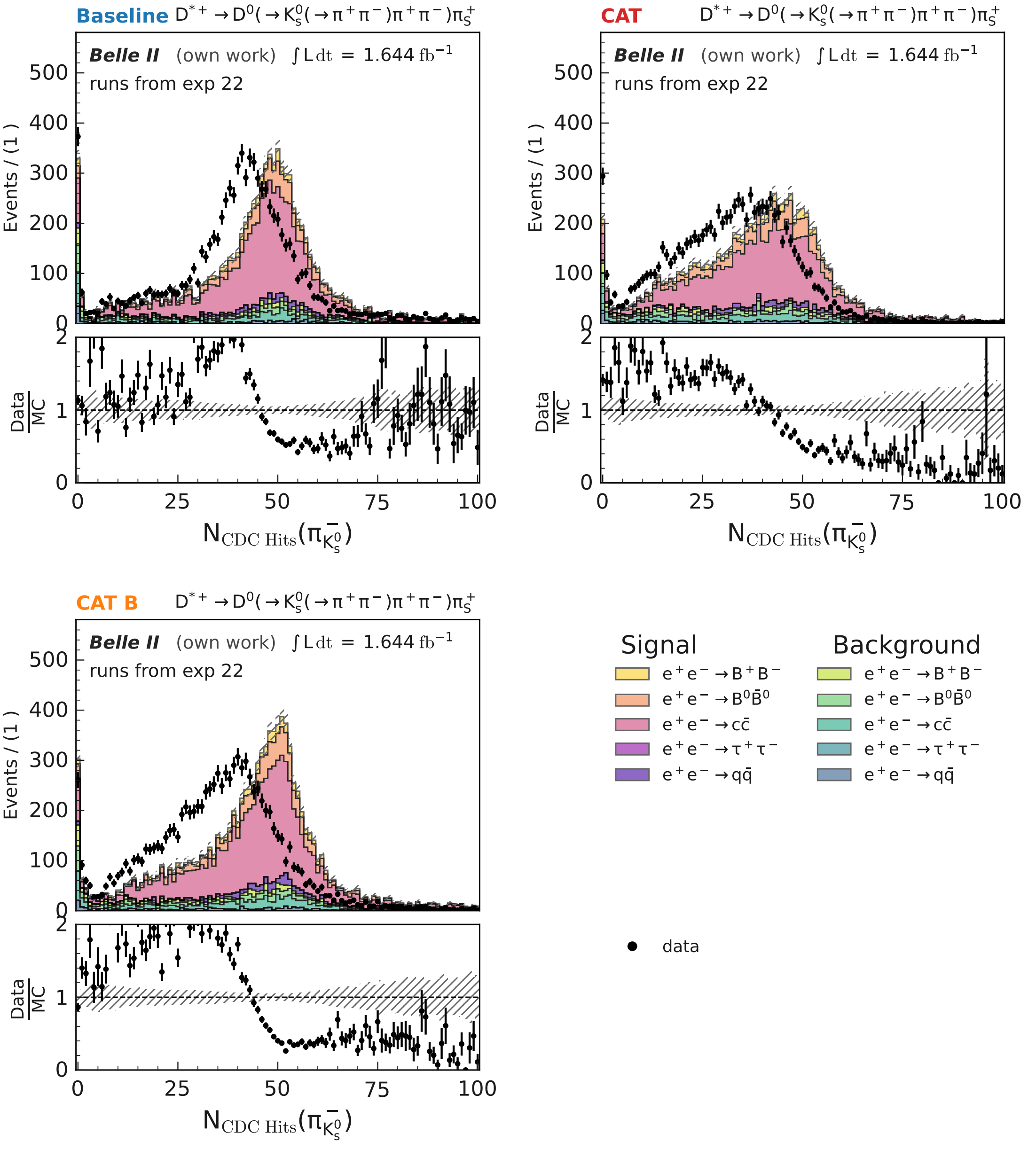


Figure E.8: Number of CDC hits for the π^- of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for both simulation and measured data for runs from experiment 22 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\mathrm{CDC\ Hits}}(\pi^-)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed K_S^0. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.644\,\mathrm{fb}^{-1}$.

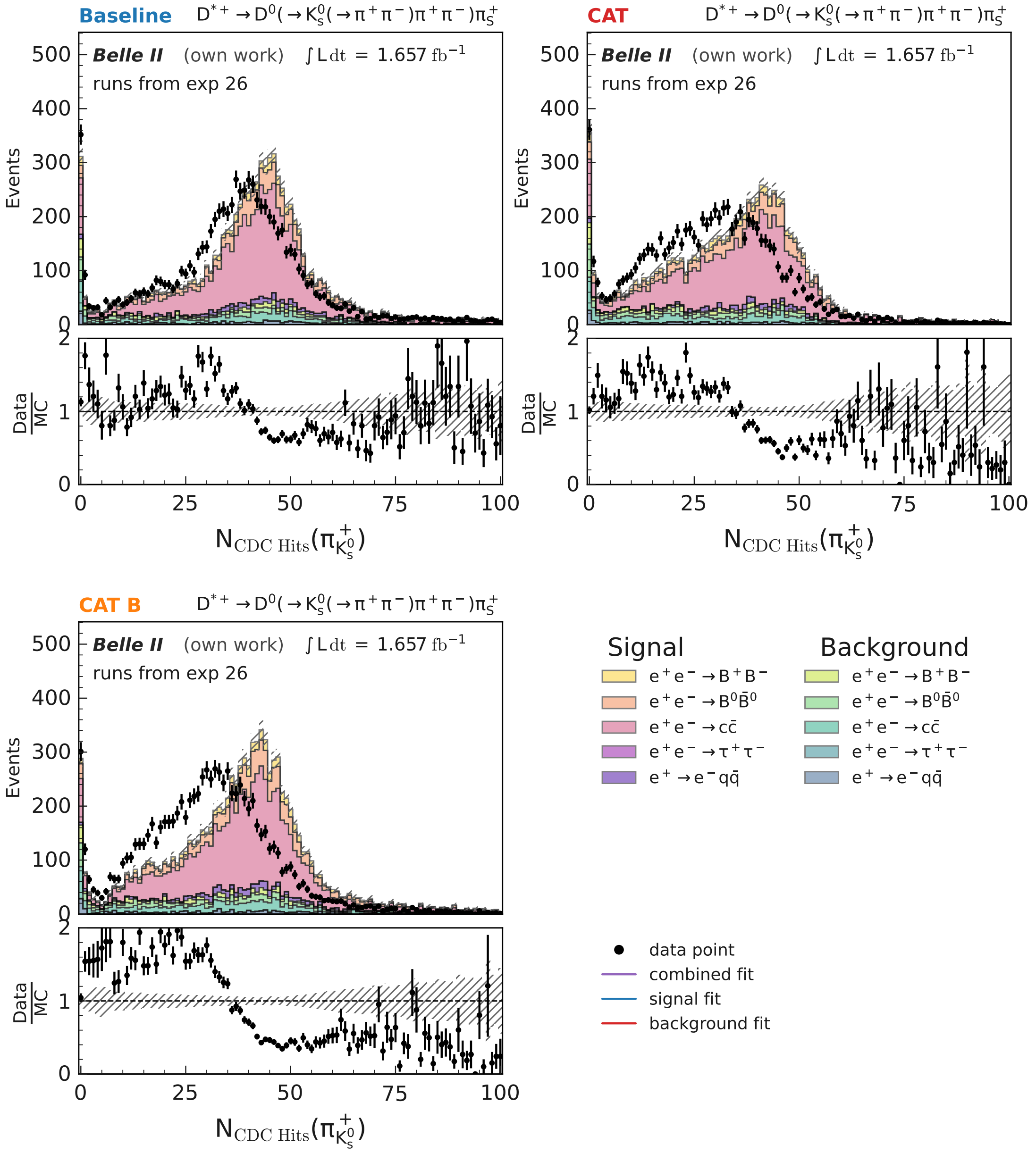


Figure E.9: Number of CDC hits for the π^+ of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\mathrm{CDC\ Hits}}(\pi^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed K_S^0 . The simulation is scaled to the integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1.657\,\mathrm{fb}^{-1}$.

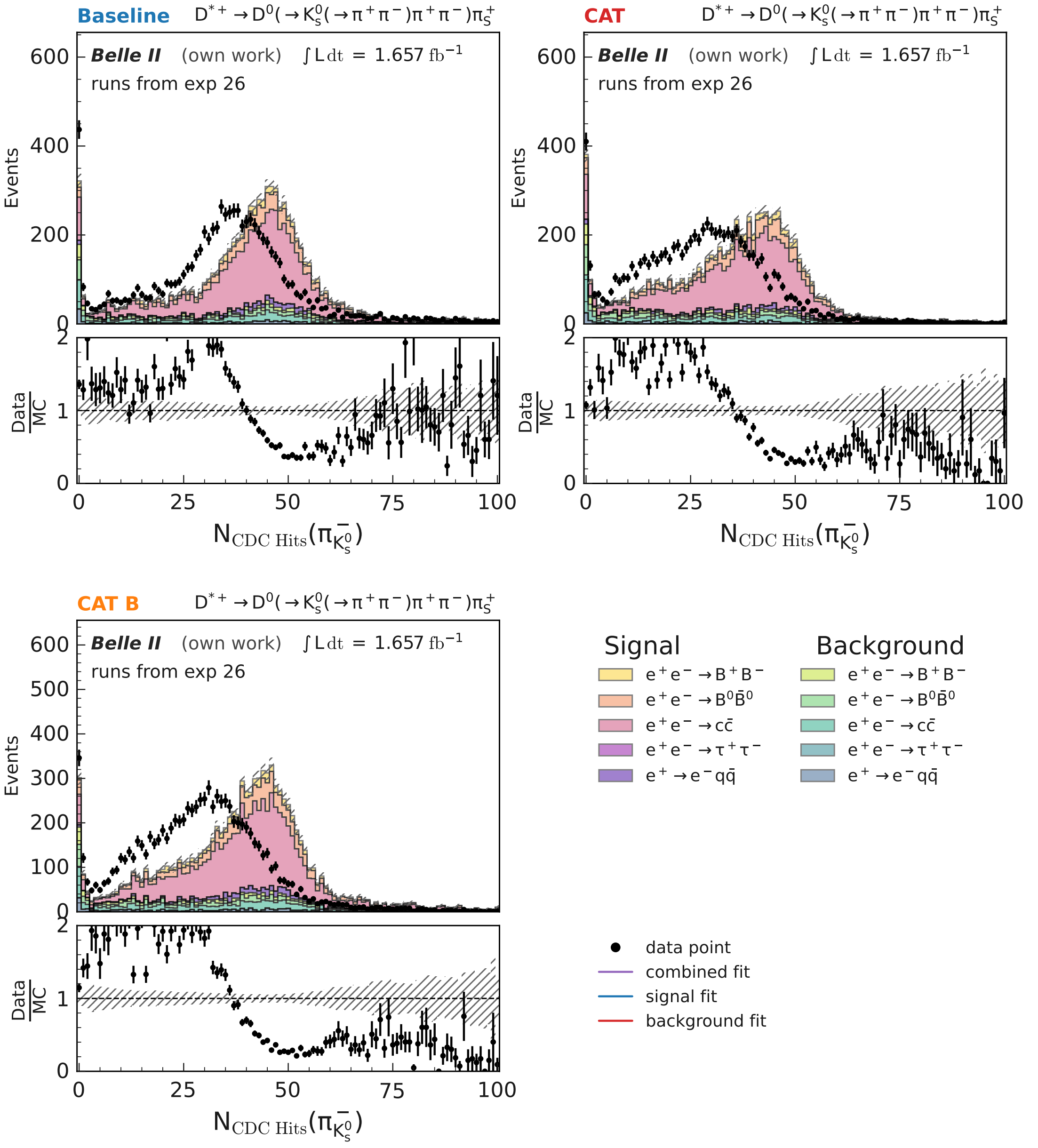


Figure E.10: Number of CDC hits for the π^- of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for both simulation and measured data for runs from experiment 26 (see Section 3.1.1 for details), including the comparison between simulation and measured data. The plots show $N_{\text{CDC Hits}}(\pi^+)$ using the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background for the simulation refer to the correctly reconstructed K_S^0. The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.657\,\text{fb}^{-1}$.

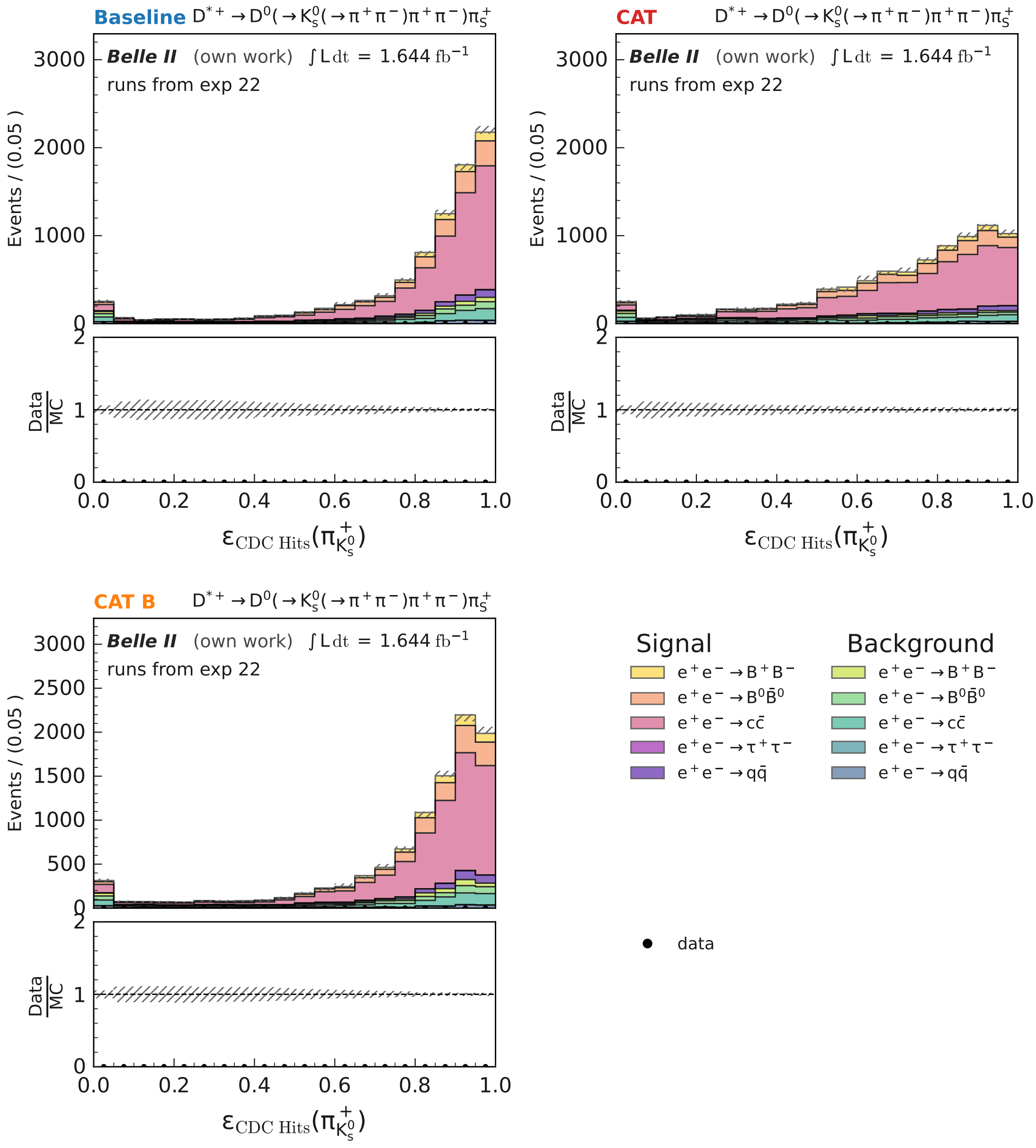


Figure E.11: CDC hit efficiency for the π^+ of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for simulation. The CDC hit efficiency is shown for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background refer to the correctly reconstructed K_S^0 . The simulation is scaled to the integrated luminosity of $\int \mathcal{L} dt = 1.644\,\mathrm{fb}^{-1}$.

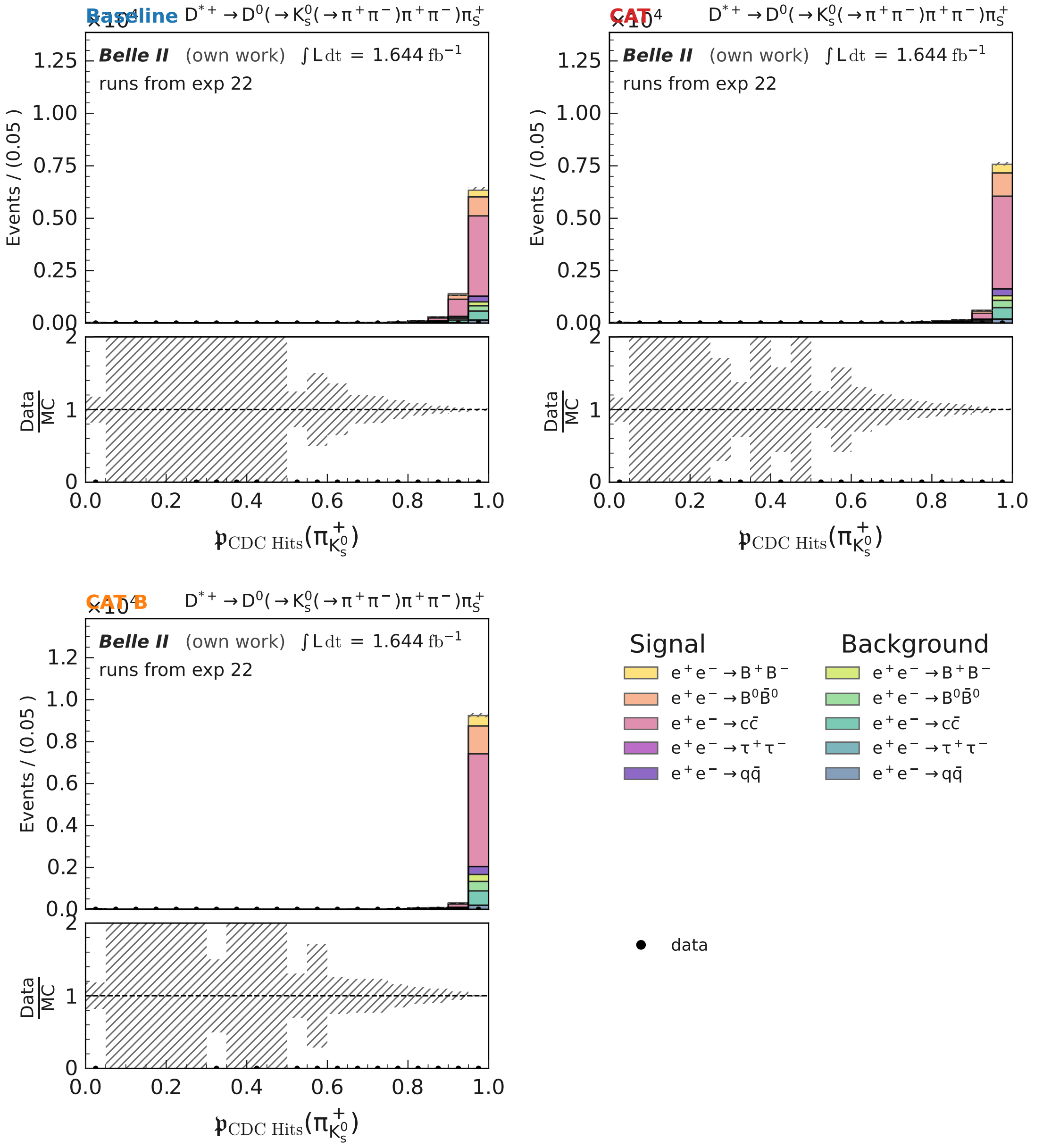


Figure E.12: CDC hit purity for the π^+ of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for simulation. The CDC hit purity is shown for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background refer to the correctly reconstructed K_S^0 . The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.644\,\mathrm{fb}^{-1}$.

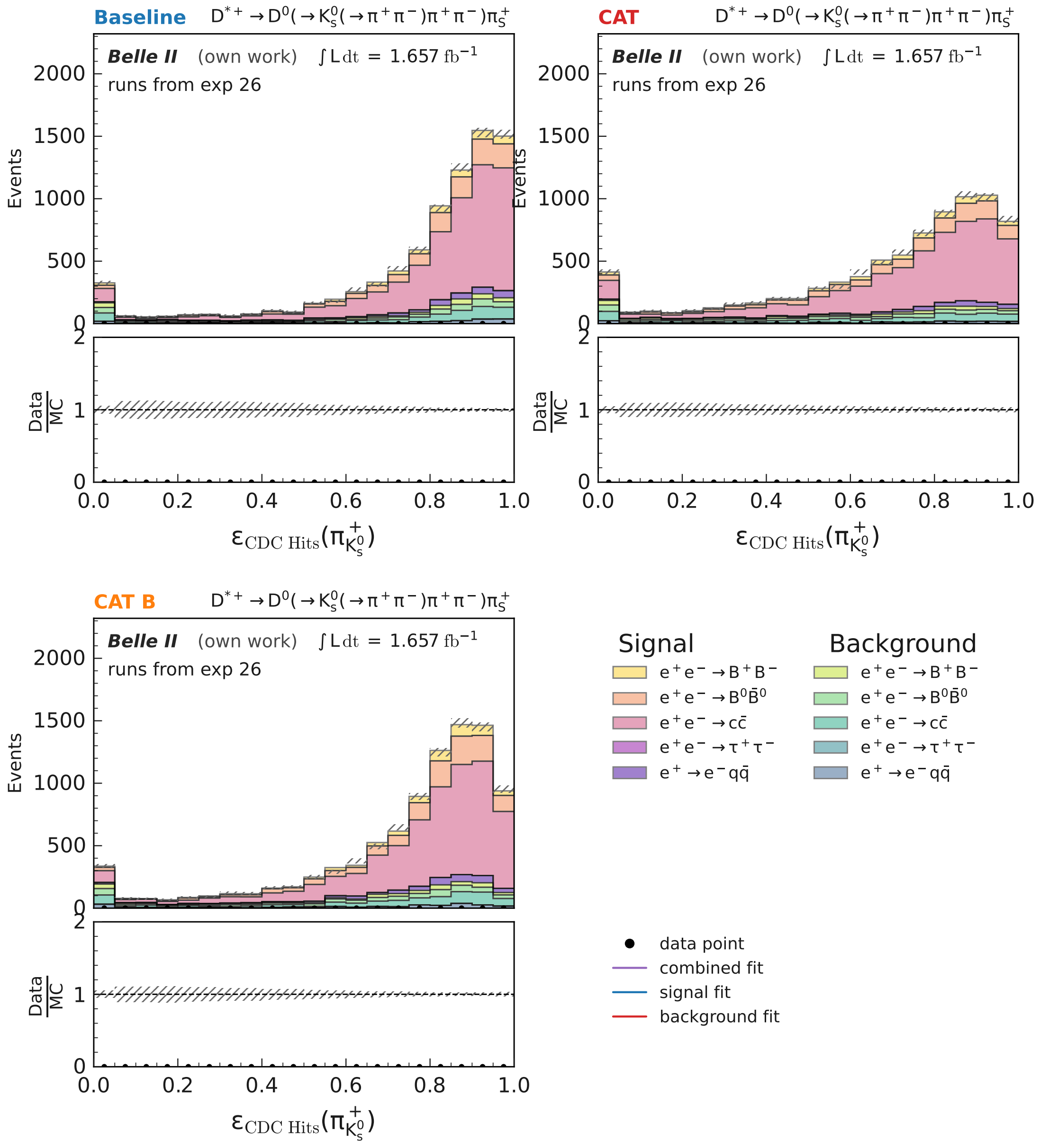


Figure E.13: CDC hit efficiency for the π^+ of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for simulation. The CDC hit efficiency is shown for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background refer to the correctly reconstructed K_S^0 . The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.657\,\text{fb}^{-1}$.

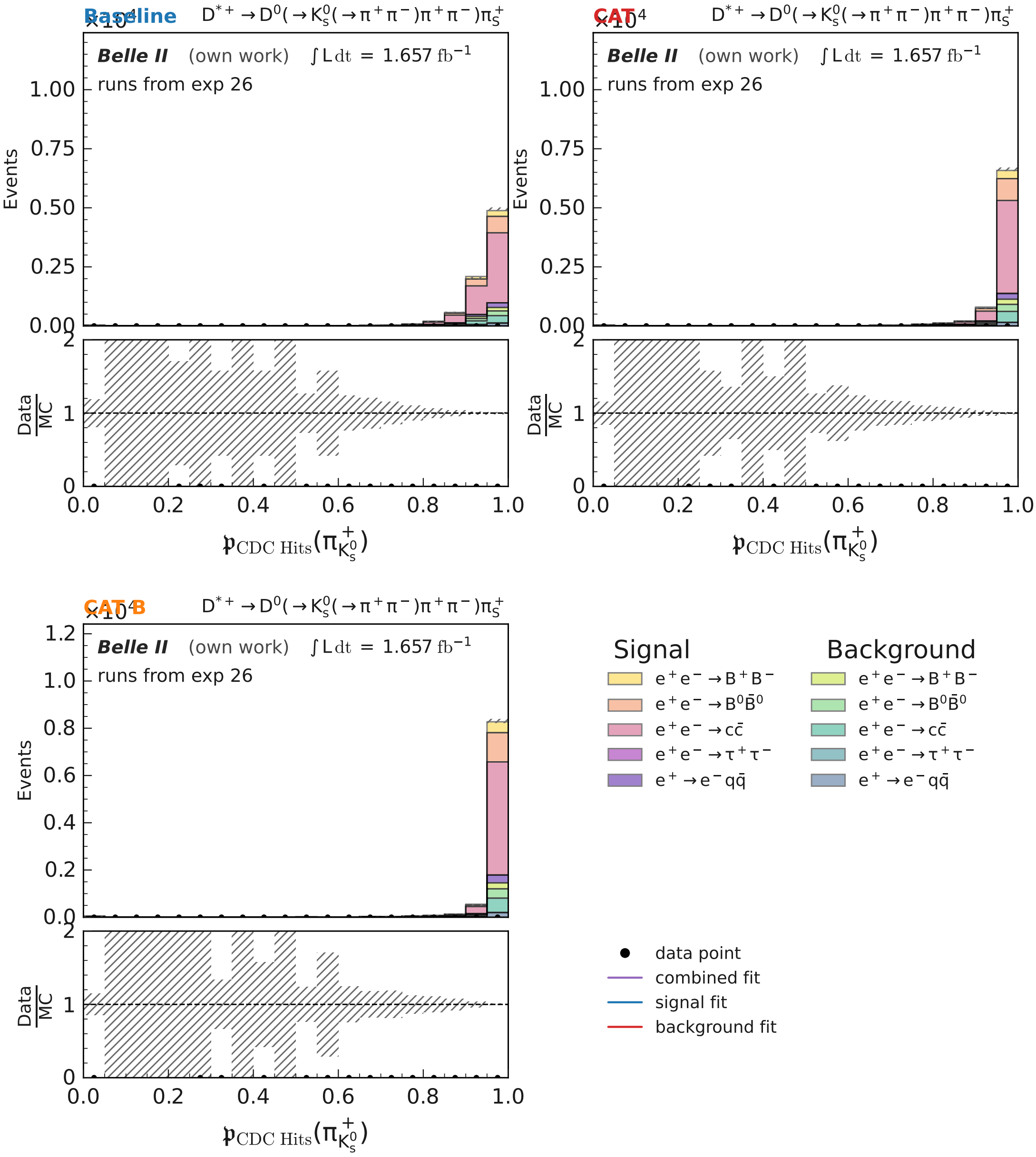


Figure E.14: CDC hit purity for the π^+ of the reconstructed K_S^0 in the signal region of 494.8 MeV to 500.8 MeV for simulation. The CDC hit purity is shown for the *Baseline Finder* (top left), *CAT Finder* (top right), and *CAT B Finder* (bottom) in the full reconstruction. Signal and background refer to the correctly reconstructed K_S^0 . The simulation is scaled to the integrated luminosity of $\int \mathcal{L}dt = 1.657\,\text{fb}^{-1}$.

E.3 Radiative $e^+e^- \to \mu^+\mu^-\gamma$

The number of CDC hits for different bins of transverse momentum for muons from radiative $e^+e^- \to \mu^+\mu^-\gamma$ events of runs from experiment 26 in Fig. E.15., following the selection in Table 3.5.

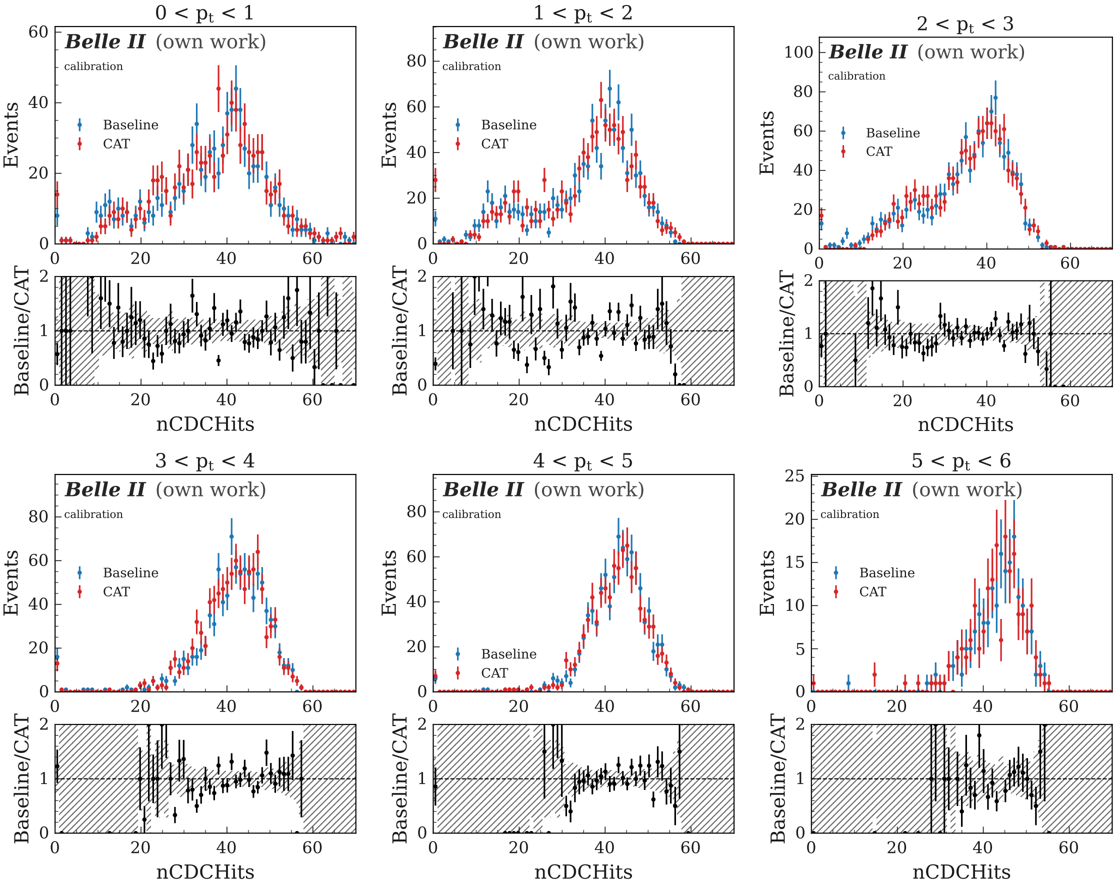


Figure E.15: Comparison between the number of CDC hits for the *CAT Finder* (red) and the *Baseline Finder* (blue) in bins of transverse momentum for muons from $e^+e^- \to \mu^+\mu^-\gamma$ for runs from experiment 26. The statistical uncertainties are correlated as the same measured data is evaluated for both algorithms.

E.4 Bhabha Resolution

E.4.1 Momentum Resolution

The momentum resolution for electrons from $e^+e^- \to e^+e^-$ following the selection in Section 8.3.2 is shown in Fig. E.16.

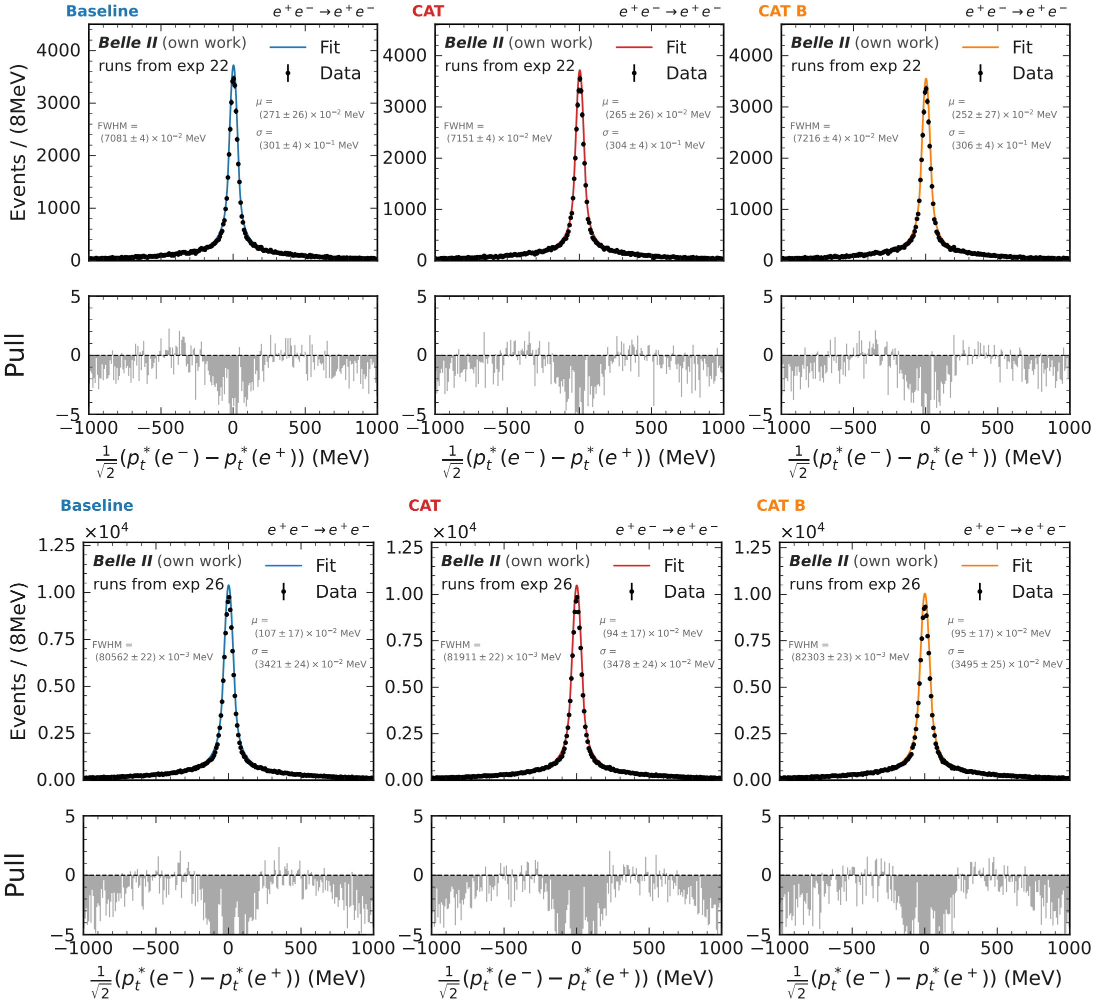


Figure E.16: The momentum resolution as the difference between the transverse momentum in the center-of-mass frame between the e^- and e^+ for the three algorithms *Baseline Finder* (left), *CAT Finder* (middle), and *CAT B Finder* (right) for experiment 22 (top) and experiment 26 (bottom).

E.4.2 PXD and SVD extrapolation confirmation using d_0 and z_0 resolution

The resolution of the impact parameters for electrons from $e^+e^- \to e^+e^-$ following the selection in Section 8.3.2 is shown in Fig. E.17 and Fig. E.18.

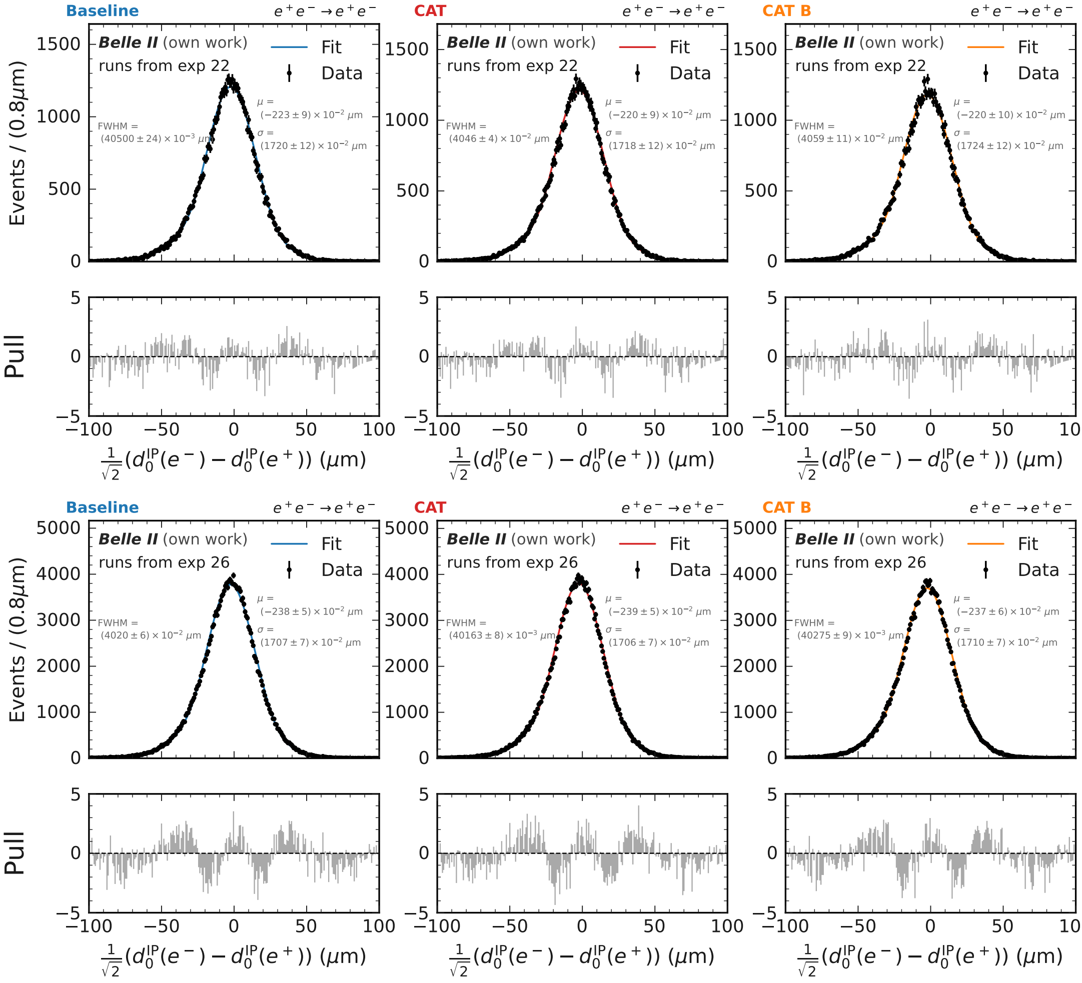


Figure E.17: The d_0 difference of the e^- and e^+ for the three algorithms *Baseline Finder* (left), *CAT Finder* (middle), and *CAT B Finder* (right) for experiment 22 (top) and experiment 26 (bottom).

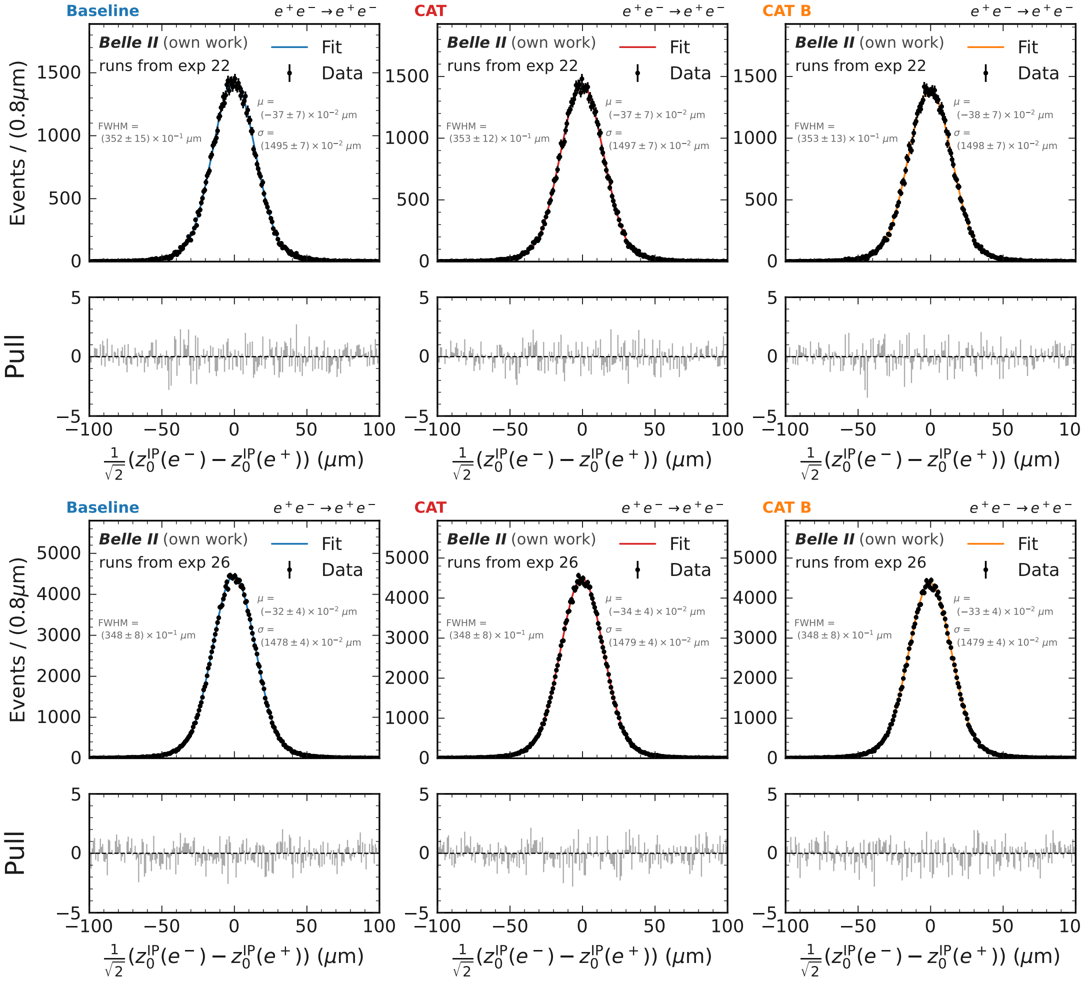


Figure E.18: The z_0 difference of the e^- and e^+ for the three algorithms *Baseline Finder* (left), *CAT Finder* (middle), and *CAT B Finder* (right) for experiment 22 (top) and experiment 26 (bottom).

Appendix F

Backgrounds for the sensitivity study

The backgrounds without the PID selection on the S candidate daughters given in Fig. F.1. The backgrounds used to calculate the expected limits in Section 9.4 are given in Fig. F.2, with the scaled *Baseline Finder* backgrounds used for the *CAT Finder*.

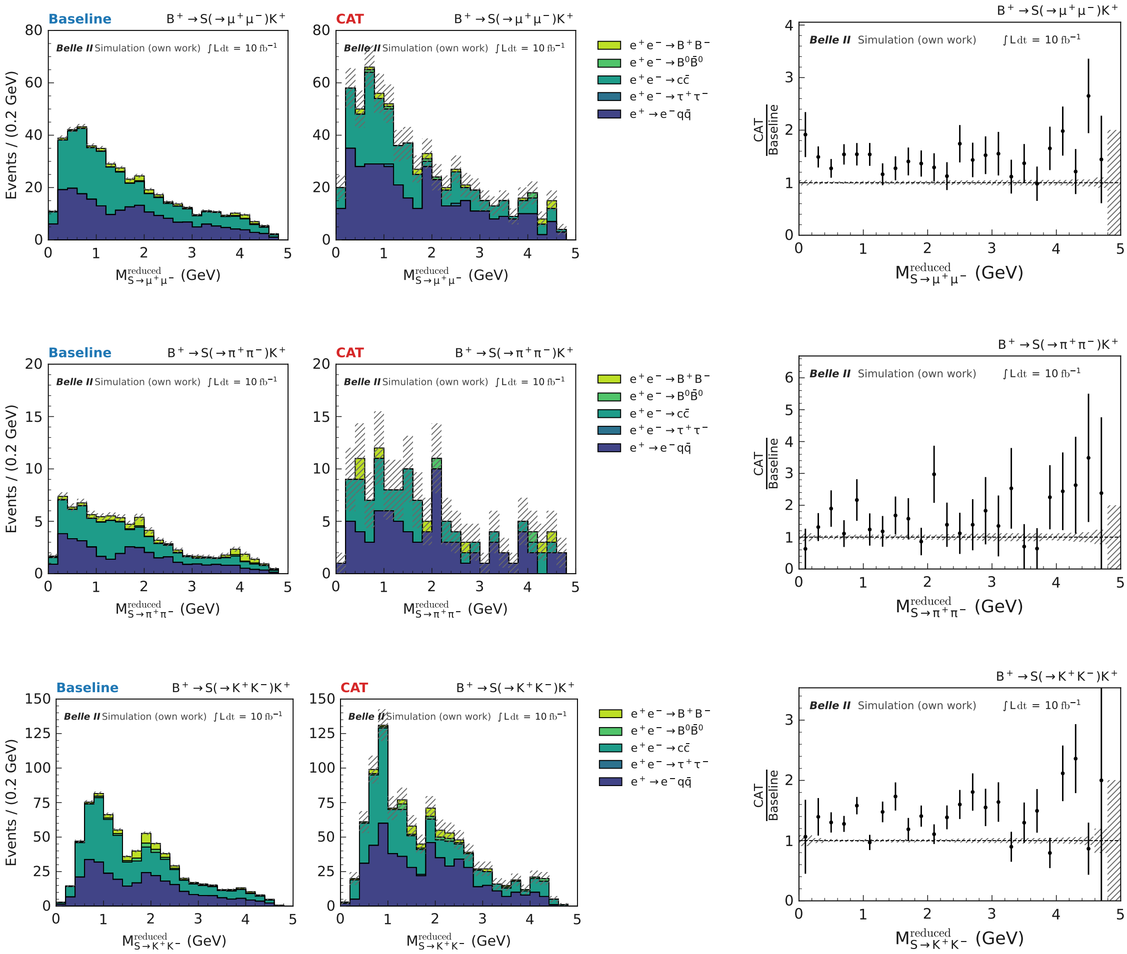


Figure F.1: Comparison for the backgrounds for the observable $M^{\text{reduced}}_{S\to x^+x^-}$ including the event selections in Table 9.1 except the PID requirement on the displaced tracks forming the S candidate for the *Baseline Finder* (left) scaled down from $\int Ldt = 500\,\mathrm{fb}^{-}1$ to $\int Ldt = 10\,\mathrm{fb}^{-}1$ of the *CAT Finder* (middle). On the right, the ratio of the *CAT Finder* and *Baseline Finder* is given.

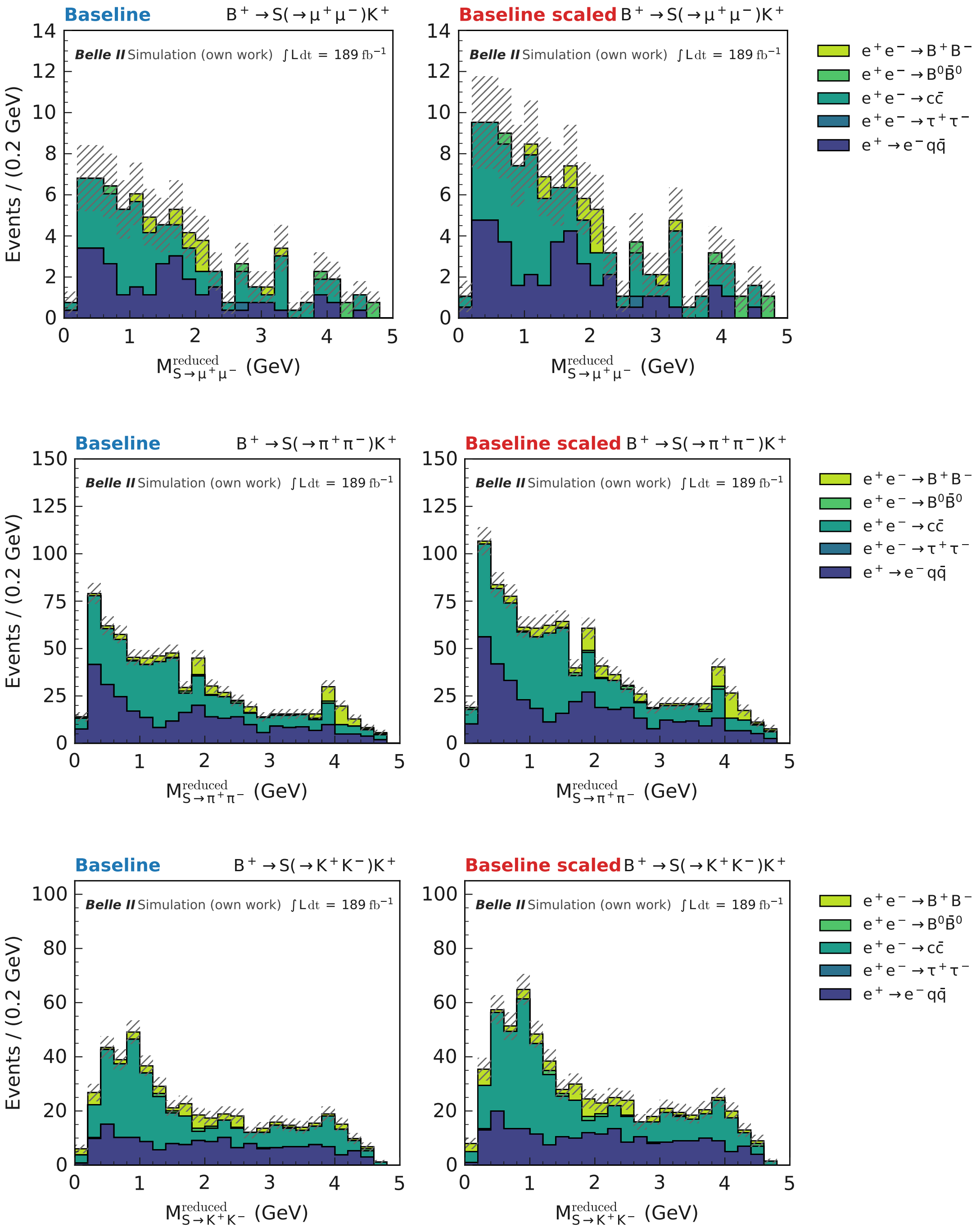


Figure F.2: Scaled backgrounds from the *Baseline Finder* scaled to $\int Ldt = 189\,\mathrm{fb}^{-1}$ for the observable $M^{\mathrm{reduced}}_{S\to x^+x^-}$ including all event selections in Table 9.1. These are used for the background estimation for the *Baseline Finder* (left) and for the *CAT Finder* (right), with the scaling factors of Table 9.3. Top shows $S \to \mu^+\mu^-$, middle $S \to \pi^+\pi^-$ and bottom $S \to K^+K^-$.

Glossary

BabaYaga@NLO A Monte Carlo generator for Bhabha scattering and photon pair production in electron-positron collisions. See [42, 134]. 25

Belle II A second-generation B factory and the successor of the Belle experiment. 1– 6, 8, 13– 15, 17, 23, 25, 28– 30, 35, 36, 39– 41, 46, 55, 57, 58, 66, 68, 84, 85, 87, 93, 99, 101, 109, 110, 141, 151, 159, 175– 177, 179, 180, 182, 185, 197, 198

EvtGen A Monte Carlo Generator suited for the decay of heavy flavour particles. See [38]. 20, 23, 25, 30

GEANT4 A toolkit for simulating the passage of particles through matter using a wide variety of phenomenological models. See [135]. 13, 60, 64, 180

GENFIT2 An experiment-independent framework for track reconstruction in particle and nuclear physics. See [51]. 76, 77, 79, 190

KKMC A Monte Carlo generator specifically for lepton and quark pair production at lepton colliders. See [40]. 25, 26

POCA The point of closest approach (POCA) is a tracks' signed distance to the z axis as determined by the tracking algorithm. 41

PYTHIA8 A general-purpose Monte Carlo generator used to describe hard and soft interactions, parton distributions, initial- and final-state parton showers, multiparton interactions, fragmentation and decay. See [39]. 25

SuperKEKB An upgrade of the KEKB electron-positron collider and the accelerator at which the Belle II experiment is located. 1, 3, 4, 8

Tauola A Monte Carlo generator suited for the decay of τ leptons. See [41]. 25

zfit A python-based model fitting library optimised for simple and direct manipulation of probability density functions. See [107]. 112

Acronyms

ADC Analog-to-Digital Converter. 9, 36, 41– 43, 56, 143, 180

ARICH Aerogel Ring-Imaging Cherenkov detector. 7, 100, 103– 105

basf2 Belle II Analysis Software Framework. 2, 13, 35, 56, 58, 103, 107, 109, 136, 143, 179, 184, 186, 190, 193

BDT Boosted Decision Tree. 38, 184

BNL Batch Norm Layer. 44– 46

CDC Central Drift Chamber. viii, 2, 3, 5, 6, 8, 10, 13– 15, 17– 21, 23, 24, 26– 28, 31, 32, 35– 37, 39, 40, 46, 56– 62, 66, 72, 82, 85, 87– 91, 94, 95, 97– 102, 107, 109– 111, 115, 118– 123, 128, 136– 142, 144, 147– 149, 151, 154– 156, 160, 163, 179, 180, 182, 183, 185, 186, 189– 191, 193, 194, 197, 198, 220, 221, 244– 251, 253– 255, 259– 264, 267– 275

CKF Combinatorial Kalman Filter. 35, 56, 58, 91, 99, 141, 180, 190

cms center-of-mass system. 3, 25, 29, 40, 147, 150, 154, 160

CNN convolutional neural network. 38, 39

CP Charge Parity. 93

DAF Deterministic Annealing Filter. 36

DAQ Data Acquisition. 8, 190

DM dark matter. 1, 28, 29, 159

DSCB Double Sided Crystal Ball. 112, 117, 127, 130, 134, 150, 151, 166, 169, 170

ECL Electromagnetic Calorimeter. 7, 8, 26– 28, 58, 60, 66, 99, 100, 104, 105, 148, 160

FPGA field-programmable gate array. 8

FSP Final State Particle. 23, 32, 68, 99, 100, 113

GNN graph neural network. 2, 13, 35, 39, 40, 44, 45, 47, 52, 87, 96, 107, 115, 143, 156, 159, 179, 181, 183, 186, 189– 192, 195

HEP high energy physics. 38, 39

HER High-Energy Ring. 3

HL-LHC High Luminosity LHC. 39

HLT High Level Trigger. 8, 15, 26– 28, 39, 58, 109, 110, 129, 147, 154, 186

IP Interaction Point. 4, 5, 20, 29, 36– 38, 41, 147, 148, 154

KLM K^0_L and muon detector. 7, 8, 26, 27, 58, 99, 100, 105, 106

L1 trigger Level 1 trigger. 8, 15

LER Low-Energy Ring. 3

LHC Large Hadron Collider. 39, 181

LL Linear Layer. 45, 46, 49, 53

MadGraph5 MadGraph5_aMC@NLO. *Glossary:* MadGraph5_aMC@NLO

ML machine learning. 38, 39, 185

MVA multivariate analysis. 58, 143

PDF probability density function. 112, 113, 165, 166, 170

PID particle identification. 58, 99, 100, 103, 107, 108, 111, 163– 165, 177, 182, 279, 280

POCA Point of closest approach. 41– 43, *Glossary:* POCA

PXD Pixel Detector. 5, 6, 8, 36, 58, 90, 91, 99, 151, 155, 190, 193

run I data taking from 2018-2022. 6, 15, 17– 19, 90, 159, 170, 180, 194

run II data taking starting from 2024. 6, 18, 19, 84, 180, 194

SM Standard Model of particle physics. 1, 29, 30, 159, 160, 165, 170, 175

SVD Silicon Vertex Detector. 5, 6, 8, 29, 56– 58, 60, 87, 89– 92, 99, 100, 102, 130, 131, 136, 141, 142, 144, 151, 155, 156, 159, 179, 190– 194, 255– 258

TDC Time-to-Digital Converter. 9, 36, 41– 43, 56, 143, 180

TOP Time-Of-Propagation counter. 6– 8, 60, 66, 67, 100, 103– 105, 182

TOT Time over Threshold. 9, 36, 41, 56, 180, 181