Input Feature Optimization for Track Reconstruction using Graph Neural Networks at Belle II

Jonathan Bauer

Bachelor Thesis

28th October 2024

Institute of Experimental Particle Physics (ETP)

Reviewer:	Prof. Dr. Torben Ferber
Second Reviewer:	Dr. Giacomo De Pietro
Advisor:	Lea Reuter

Editing time: 1st July 2024 – 28th October 2024

www.kit.edu

Optimierung der Eingangsparameter für die Spurdetektion mit Graph Neuralen Netzwerken bei Belle II

Jonathan Bauer

Bachelorarbeit

28. Oktober 2024

Institut für Experimentelle Teilchenphysik (ETP)

Referent: Prof. Dr. Torben Ferber
Korreferent: Dr. Giacomo De Pietro
Betreuerin: Lea Reuter

Bearbeitungszeit: 1. Juli 2024 – 28. Oktober 2024

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 28. Oktober 2024

. . . .
(Jonathan Bauer)

Disclaimer

The results of modern particle physics are a collaborative effort. The work in this thesis is built upon the work of Lea Reuter [1], who developed the Python code for the tracking software CDC AI Tracking (CAT) and provided the Python scripts to train and evaluate the models presented in this thesis. The Belle II collaboration [2] built and maintains the Belle II detector, which is the main subject of this thesis. They also developed and maintain the Belle II Analysis Software Framework (basf2) [3], which is used in this thesis for the simulated events. Those events for this thesis are simulated by Lea Reuter. The event displays shown in Chapter 4 are taken from Lea Reuter. The training of the models in Chapter 5 was performed on the TOpAS GPU cluster at the Scientific Computing Center (SCC) at KIT. The metrics used to evaluate the tracking performance in Chapter 5 are taken from the Belle II tracking group [4], where Lea Reuter provided the scripts to access the necessary information.

My contributions include the selection of the optimization methods for the input features and the implementation of those in the Python code. The training and evaluation of those models with the respective analysis are done by me. The plots and the tables, starting from Chapter 4, are created by me.

This thesis incorporates the use of Artificial Intelligence (AI) tools to help with grammatical or stylistic improvement of text creation. Grammarly[1] is utilized 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.

---

[1]Grammarly Inc. *Grammarly*. URL: https://www.grammarly.com (visited on Oct. 18, 2024).

Contents

List of Figures

List of Tables

1. Introduction

Physicists are always striving for higher precision in their measurements to get a deeper understanding of our universe and its fundamental rules. To measure those rules and get answers to the open questions of particle physics, large experiments are carried out. Among the leading experiments in the flavor and dark sector is the Belle II detector [2] at the SuperKEKB accelerator facility [6] in Tsukuba, Japan. Its goal is to investigate rare processes of the B-meson decay and new physics with unparalleled precision.

One of the main tasks of the Belle II detector is the reconstruction of particle trajectories to measure the particle kinematics at the production point, called Interaction Point (IP). This is mainly done in the Central Drift Chamber (CDC), where the particles ionize a gas along their trajectories, also called tracks. The energy deposition along those tracks is then measured by long wires in the CDC. An algorithm working with Legendre-transformation finds those tracks as a pattern recognition task [4]. To get a higher precision than the present algorithm for the CDC, a new algorithm CDC AI Tracking (CAT) [1] is built based on Graph Neural Networks (GNNs). Significant improvement could be made in finding so-called displaced tracks, which do not originate from the IP.

The main content of this work is to improve the CAT finder by input feature optimization. Different methods to improve the input features of Neural Networks (NNs) are discussed among researchers like feature selection, feature scaling, and feature criteria [7, 8]. With those methods, new models of the CAT finder are trained with the aim of increasing the performance of the models and time-savings of the training. The model with the best input feature configuration will then be compared with the current standard configuration of the CAT finder and the Legendre-based track finding algorithm.

This thesis is organized as follows. In Chapter 2, a short overview is given of the SuperKEKB accelerator, the Belle II detector, and the structure and readout electronics of the CDC. Chapter 3 describes the track finding in the CDC with the Legendre-based track finding algorithm and the CAT finder, as well as the data set with simulated events. An investigation of the input features with a visualization of those and the methods of input feature optimization will be given in Chapter 4. The main evaluation of the trained models and the analysis is given in Chapter 5. At last, a conclusion is drawn in Chapter 6.

2. The Belle II Experiment

2.1. The SuperKEKB Accelerator and the Belle II Detector

In this section the experimental setup is briefly described, based on [6, 2, 9]. For further information about the accelerator, see [6] and for the detector, see [2].

The SuperKEKB accelerator is a high-intensity asymmetric electron-positron-collider located in Tsukuba, Japan. It collides beams of $4\,\text{GeV}$ positrons and $7\,\text{GeV}$ electrons with the center-of-mass energy at the $\Upsilon(4S)$ resonance. This energy is over the threshold of the B-meson pair production $\Upsilon(4S) \to B\bar{B}$ where no fragmentation particles are produced. For this reason, the SuperKEKB collider is also called B-Factory. Because of the asymmetric beam energies, the particles are boosted in the lab system, which allows for precise measurements of lifetimes, mixing parameters and the time-dependent Charge Parity (CP) symmetry violation can be studied [6]. The instantaneous luminosity goal at SuperKEKB is set to $6.3 \times 10^{35}\,\text{cm}^{-2}\,\text{s}^{-1}$ [10], which is nearly 40 times higher than the achieved luminosity at the old accelerator KEKB [2].

The beams from the accelerator collide at the Interaction Point (IP) inside the Belle II detector. A schematic view of the detector can be seen in Fig. 2.1. Because of the asymmetric collision of the particles and the resulting boost in the center-of-mass frame, the direction of the electrons is considered forward and the other direction is called backward. The positive z axis is parallel to the solenoid axis. The x and y axis point away from the IP, where the x axis is horizontal and the y axis is vertical. The azimuthal angle is called ϕ and the polar angle θ. The cylindrical area around the z axis is called the barrel region and in the forward and backward directions are the endcaps [2]. The components of the detector are arranged like layers inside of the detector and are the following from innermost to outermost:

- **Beam pipe:** In the double-walled Beryllium beam pipe the beams are accelerated and collided in the IP with a crossing angle of 83 mrad. Because of the Nano-Beam scheme, the beam pipe can be smaller and the beam has a higher current than its predecessor [2].
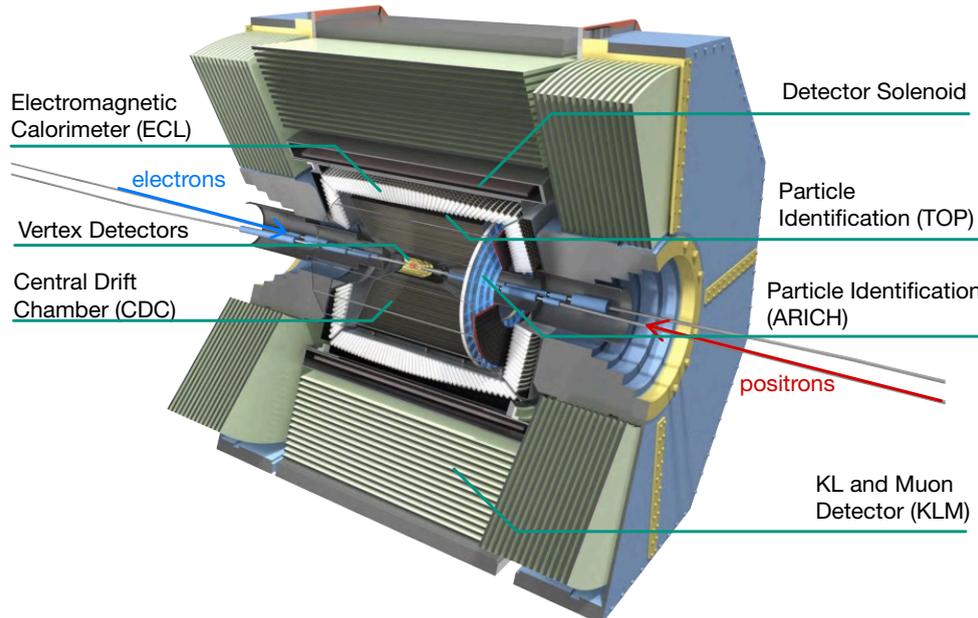
Figure 2.1.: Schematic overview of the Belle II detector showing the layers of different components to detect and measure particles. Adapted from [11].

- **Vertex detectors:** The innermost detector, the Pixel Detector (PXD), has the highest precision in determining the IP. The PXD consists of two layers of sensors, with radii of 14 mm and 22 mm [2].
  The Silicon Vertex Detector (SVD) is composed of four layers and has an acceptance range of $17° < \theta < 150°$ [2]. The main purpose of the SVD is to have high efficiency in reconstructing the vertices of $K_S^0 \to \pi^+\pi^-$ decays [9].

- **Central Drift Chamber (CDC):** The CDC is the central tracking device of the Belle II detector. This large volume drift chamber is filled with sense wires and a gas mixture of helium and ethane. It has the same acceptance rate as the SVD of $17° < \theta < 150°$ [2, 9]. The main purposes of the CDC are the determination of the particle trajectory and its momentum, the measurement of energy losses in the gas volume to identify the particle, and providing the trigger information on the charged particles [12]. This is the main detector for this thesis and therefore it is described in more detail in Sec. 2.2.

- **Particle identification system:** In the barrel region there is the Time-of-Propagation (TOP) counter. This Cherenkov detector consists of quartz bars and photo-detectors at the end. With two-dimensional information on the Cherenkov rings, the particles can be identified [9, 13].
  The Aerogel Ring-imaging Cherenkov (ARICH) detector is used in the forward endcap region. It consists of an aerogel radiator and an array of position-sensitive photon detectors. Both systems can separate pions and kaons in a high-impulse spectrum. [2]

- **Electromagnetic Calorimeter (ECL):** This calorimeter consists of CsI(Tl) scintillation crystals in the barrel and endcap region. It covers a polar angle region of $12.4° < \theta < 155.1°$ The main purposes are the detection and precise determination of the energy of photons and the electron identification [2, 9].

- **Detector Solenoid:** The superconducting solenoid provides a magnetic field of $1.5\,\text{T}$ with a $3.1\,\text{m}$ diameter. Because of this, the charged particles have a helical trajectory. With the radius of this trajectory and the charge of the particle, the impulse can be determined [2].

- **K_L and Muon Detector (KLM):** This detector consists of an alternating sandwich of iron plates and active detector elements. The iron plates serve as the magnetic flux return for the solenoid and provide more material, in which K_L^0 mesons can shower hadronically. The total angular coverage with the endcaps is $20° < \theta < 155°$. Its main purpose is to detect the long-living particles K_L^0 and muons [2, 9].

2.2. Central Drift Chamber

2.2.1. Geometry and Structure

Inside the volume of the CDC are 42 240 field wires and 14 336 sense wires spanned parallel to the magnetic field of the solenoid. Additionally, in the volume, there is a gas mixture of helium and ethane. When a charged particle flies through this gas, around its path electrons are produced that drift in the electrical field towards the sense wires, where the particle can be tracked. The inner and outer cylinder of the CDC consists of a thin carbon fiber reinforced plastic and two round aluminum endplates around the IP. The detector has a polar angular acceptance of $17° < \theta < 150°$. The radius of the inner cylinder is 160 mm and of the outer 1130 mm, the length of the CDC is 2417 mm. [2]. The structure can be seen as a cross-section in Figure 2.2.
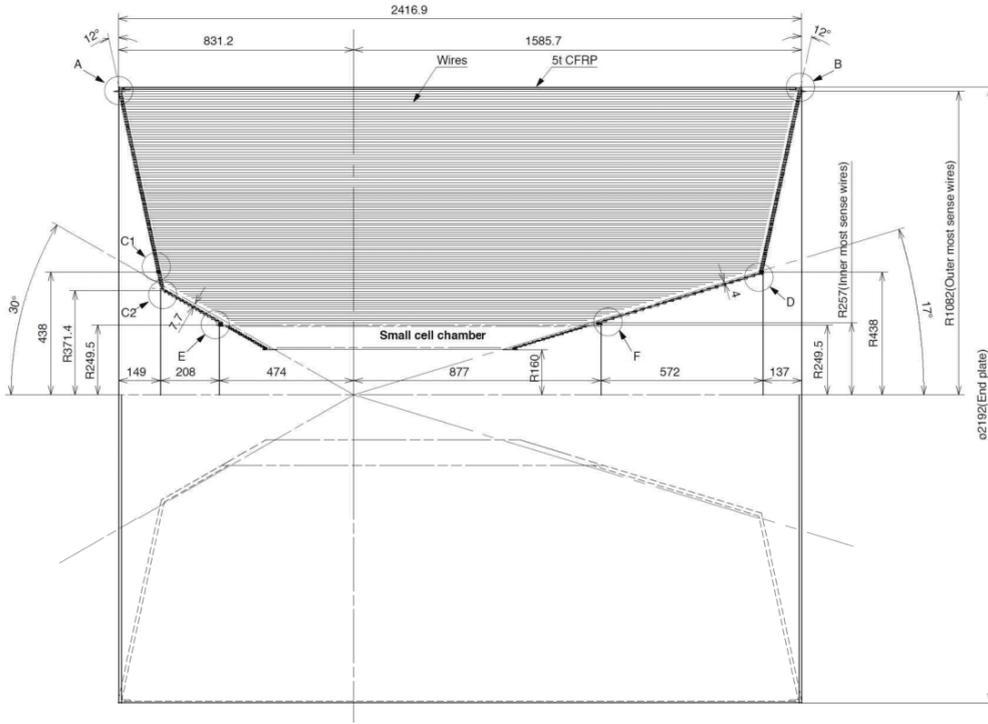
Figure 2.2.: Schematic structure of the CDC as a cross-section in the Belle II detector. Taken from [2].

The CDC has a cellular structure, where cells consist of eight field wires made out of aluminum and one sense wire made out of tungsten. Respectively eight field wires create an electrical field, with one sense wire in the middle. This can be seen in figure 2.3. When a charged particle flies through the gas inside the CDC, electrons are created in the ionization process around its path. The electrons then drift in the electrical field towards the sense wire. With the information of the time, the electrons need to drift to the sense wire, a circle with the distance between sense wire and particle can be drawn, a so called drift circle. The sense wire then gives a signal to the readout electronics, which allows the trajectory of the particle to be reconstructed [2, 12].
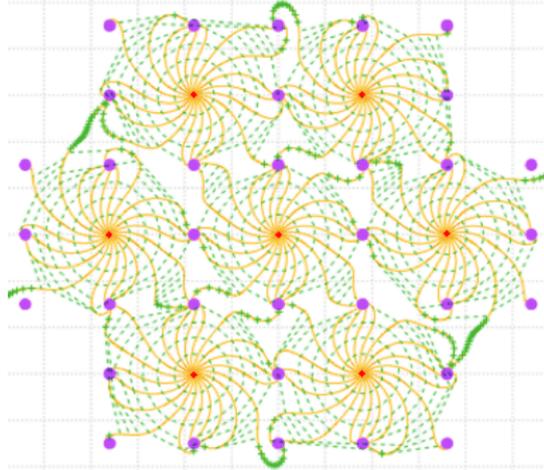
Figure 2.3.: Visualization of the electrical field created in one cell of the CDC. The electrical field (green dashed line) is created by the field wires (violet points) and the drift paths (yellow line) of the electrons point toward the sense wires (red points). Taken from [14].

The sense wires are arranged in 56 layers around the IP. Those layers are grouped into 9 superlayers, with 8 layers in the innermost superlayer and 6 layers in the remaining. To get better tracking in z-direction, parallel to the magnetic field, some of those superlayers are skewed by a small angle. The parallel superlayers are called axial (A), and the intervening superlayers, called stereo, alternate between positive (U) and negative (V) angles. The total sequence is AUAVAUAVA [2]. Detailed information about the superlayers can be seen in Tab. 2.1, where superlayer 1 is the innermost.

Table 2.1.: Configuration of the superlayers in the CDC. Taken from [2].

superlayer	type	layers	cells per layer	radius (mm)	stereo angle (mrad)
1	axial A	8	160	168.0 − 238.0	0
2	stereo U	6	160	257.0 − 348.0	45.4 − 45.8
3	axial A	6	192	356.2 − 455.7	0
4	stereo V	6	224	476.9 − 566.9	-55.3 − -64.3
5	axial A	6	256	584.1 − 674.1	0
6	stereo U	6	288	695.3 − 785.3	63.1 − 70.0
7	axial A	6	320	802.5 − 892.5	0
8	stereo V	6	352	913.7 − 1003.7	-68.5 − -74.0
9	axial A	6	284	1020.9 − 1111.4	0

2.2.2. Readout Electronics

The readout electronics of the CDC provide the digitized analog signal, the Analog-to-Digital Converter (ADC) count, the associated time information, the Time-to-Digital Converter (TDC) count and the Time-Over-Threshold (TOT) count to the Data Acquisition (DAQ) system. The ADC count refers to the deposited energy of a wire hit in the CDC. It is calculated as a truncated mean of a time window with a frequency of approximately 30 MHz. The TDC count is the time information of hits in the CDC and has a resolution of 1 ns. The TOT count gives the duration time of the signal over a specific threshold [15, 2].

The timer of the TDC is read out on a periodical external trigger, approximately every 30 ns. Subtracting this time difference between the last external trigger signal and the arrival of the electrons at the wire T_{count} by the known clock time T_{clock} gets the full measured time T_{meas}. The measured time T_{meas} is given by (see [16])

$$T_{\mathrm{meas}} = T_0 + T_{\mathrm{flight}} + T_{\mathrm{drift}} + T_{\mathrm{prop}} + T_{\mathrm{walk}}. \tag{2.1}$$

Here, T_0 is given by the difference between the external trigger signal and the collision, which was producing the particle. T_{flight} is the flight time of the particle until it reaches the drift cell, T_{drift} is the time the electron drifts from the ionization to the sense wire, and T_{prop} is the time, the electrical signal propagates through the wire to the readout electronics. When the current of the wire starts rising, the timer at the readout electronics gets started. This is the neglectable time T_{walk}. With the drift time T_{drift} and because of the constant drift velocity of the electrons, the closest distance of the particle and the sense wire can be determined, which can be visualized as a drift circle. With this, the accuracy of the track fitting gets higher [16]. A visualization of the time components is shown in Fig. 2.4.
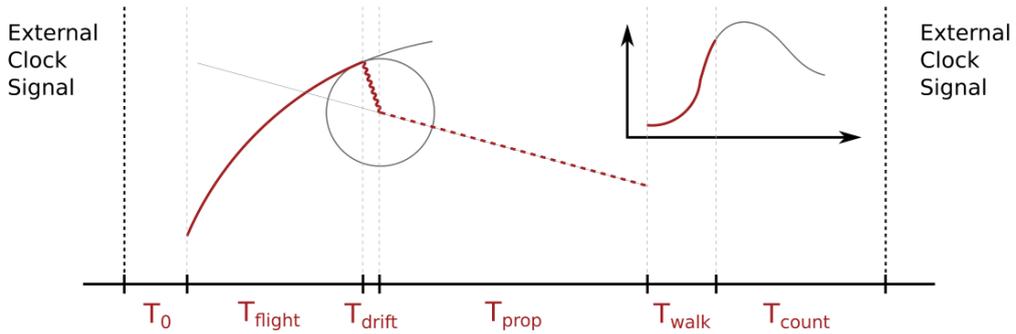
Figure 2.4.: Visualisation of the time components of a CDC wire hit. The particle, which was produced at T_0, travels to the cell (T_{flight}), where the electrons from the ionization drift to the sense wire (T_{drift}) and the electrical signal propagates through the wire (T_{prop}) to the readout electronics, which start a timer (T_{walk}). At the next external clock signal, the timer is read out (T_{count}). The lengths of the times on the x axis are not to scale. Taken from [16].

2.3. Beam Background

The beam-induced backgrounds, here from now on referred to as beam background, is one of the key challenges for tracking at Belle II. Because the luminosity of the SuperKEKB collider is increased by reducing the beam size and by higher beam currents, the background is higher than for the KEKB accelerator and is expected to get higher in the next ten years. The main sources of the beam background are the following, based on [10]:

- **Particle scattering:** The particles fall out of the stable orbit because of scattering with gas molecules or Coulomb scattering in the beam bunch (Touschek effect) and reach the wall of the beam pipe. Here they create an electromagnetic (EM) shower, which is then detected in the detector as background.

- **Colliding beams:** Some collision processes, which are not of high interest for physics measurements but have a high cross-section, create a radiative background. Those are for example radiative Bhabha scattering and two photon processes.

- **Beam injection:** When new bunches are injected, because of the limited lifetime of the beam, the particle loss gets higher during the first rounds, until the beam is stable again.

Other sources for background in the CDC can be the interference from different channels, because of spatially close sense wires [17] and cosmic rays, which fly through the detector. When being studied by Monte Carlo (MC) simulations, a background rate of 150 kHz/wire in the CDC is considered acceptable [18].

3. Track Finding in the Central Drift Chamber

The Belle II tracking software consists of independent and interchangeable data processing modules, which perform the corresponding task. The information from all modules gets integrated into one final set of tracks that are available for physics analyses. Different algorithms are used for the different tracking detectors [4]. The combination of the reconstructed tracks from each of the tracking detectors is done by a Combinatorial Kalman Filter (CKF) [16]. After this, the track fitting algorithm performs a track fit. Since the main focus of this thesis is on the CDC, only the CDC track finding algorithm will be explained here. For further information, see [4].

3.1. Legendre-based Track Finder

The trajectories of the charged particles in the CDC can be approximated as circles in the xy plane. The reason for this is the magnetic field in the z direction from the solenoid and the Lorentz force (see Sec. 2.2.1). Since the axial wires are parallel to the z axis, those can provide the correct xy information of the trajectory for the first step of track finding. It is assumed that all charged particles come from the IP in some approximation, which makes a common point of all tracks. The wire hits in the CDC can be visualized as drift circles around the sense wire, where the trajectory of the particle has to be a tangent (see Sec. 2.2.2) [12].

The Legendre-based track finder, in the following called the *Baseline* finder, uses the characteristics of the trajectories to find and fit them with high precision. Because of the circular trajectories, they first have to be transformed into straight lines. This is done by conformal mapping, which transforms circles in the real space to straight lines in the conformal space. The problem then is to only find a tangent common of all circles, with known radius and position, which is done by the Legendre transformation. After the Legendre transformation, the drift circle can be represented as a concave r_+ and convex r_- part with:

$$r_\pm = x_0 \cos\theta + y_0 \sin\theta \pm R_{\mathrm{dr}} \tag{3.1}$$

Here (x_0, y_0) represents the center of the drift circle in the conformal space and R_{dr} its radius (which is the same as in the real space). The parameters of the Legendre space are r and θ. The common tangents of the drift circles in the conformal space correspond to

intersections of the sinograms in the Legendre space [19, 12]. A representation of a common tangent of drift circles in the Legendre space can be seen in Fig. 3.1.
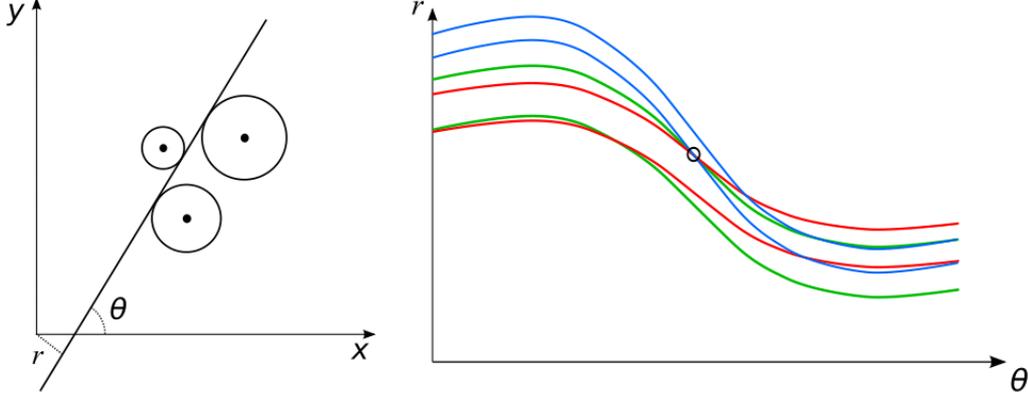
Figure 3.1.: Visualization of a common tangent of drift circles (on the left) represented in the Legendre space as the intersection of sinograms for the Legendre-based track finding in the CDC (on the right). Here, x and y are the coordinates of the conformal space, and r and θ are the Legendre parameters. Taken from [12].

To determine the Legendre parameters, the most populated regions in the Legendre space must be found. This is done by a two-dimensional binary search algorithm. This algorithm works recursive, by dividing the r-θ space into four equally sized bins and selecting the most populated, until convergence. The stopping of the binary search is determined by a resolution parameter, which is dependent on r [12, 4].

In the next step, the hits from the stereo layers are added. Those hits contain the z information because of the skewed angle of the stereo layers. The same algorithm as described above can be used with the conformal mapping of the ρ-ϕ space, where $\rho^2 = x^2 + y^2$. The trajectory is now straight in the s-z space, described by the equation

$$z_0 = z_{\text{rec}} - \tan \lambda \cdot s_{\text{rec}}. \tag{3.2}$$

The reconstructed z coordinate z_{rec} and s coordinate s_{rec} are transformed in the Legendre space (z_0, λ). The track parameters are determined by the same method. After the two-dimensional track finding and the last step of adding the stereo wires, the tracks are found and the momentum is determined [4].

3.2. Track Reconstruction with Graph Neural Networks

The *Baseline* finder is facing higher difficulties because of the low efficiency in finding tracks, that do not originate from the IP (see sec. 3.1) and because of the higher background conditions in the CDC expected in the future (see sec. 2.3). This is the reason a finder based on Artificial Intelligence (AI) is being developed [1].

3.2.1. CAT Finder

The new track finding algorithm CDC AI Tracking (CAT) uses Graph Neural Networks (GNNs) to detect and fit the trajectories of charged particles in the CDC. The GNN architecture allows the utilization of the variable input size and non-uniform arrangement of the input wires. This algorithm aims to predict the number of tracks, the three-momentum, the starting point of the track, the charge, and the hits associated with the number of tracks [1].

Each node in the Graph corresponds to one wire hit. The standard input features are the x and y position of the wire, the layer information (see Sec. 2.2.1), and the ADC and TDC count (see Sec. 2.2.2). An overview of all used input features with their names can be seen in the following:

- `middle_x`: x position of the wire, taken at the z middle of the wire.

- `middle_y`: y position of the wire, taken at the z middle of the wire.

- `superlayer`: Number of the superlayer in which the wire is, counting from 1 outwards to 9.

- `layer`: Number of the layer inside the superlayer, counting from 1 outwards to 8 for the first and 6 for the remaining superlayers.

- `clayer`: Total number of the layer, counting from 1 outwards to 56.

- `adc`: Signal amplitude, also called ADC count, which is proportional to the energy deposition of the particle.

- `tdc`: Signal length, also called TDC count, which is the time difference between the signal time and an external clock time.

The input features of the detector hits are first scaled to be in the same range and then put in an input matrix. This runs through the initial Linear Layer (LL) [1]. The GravNet block consists of a GravNet layer, which is responsible for building the graph and for message passing between the nodes [1, 20]. After this comes a final LL. The output of the first LL is used for the GravNet block and gets concatenated with the final layer [1]. A visualization of the architecture of the CAT finder can be seen in Fig. 3.2.
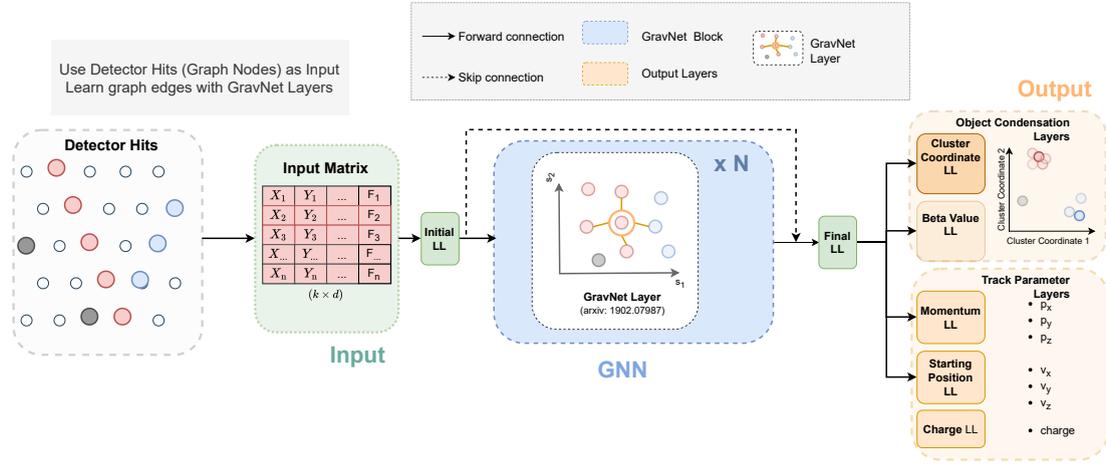
Figure 3.2.: Visualization of the architecture of the CAT finder using GNNs. Taken with permission of Lea Reuter [1].

3.2.2. Simulated Events

The data set used for training and evaluation in this thesis are simulated Belle II events. The MC simulation is done by the software tool GEANT4 [21]. Here, the full detector geometry and all interactions of final state particles with the detector material can be simulated. The software tool provides the simulation of a detector response to create digitized detector hits. The simulated events are reconstructed and analyzed with the Belle II Analysis Software Framework (basf2) [22, 3].

The particles in the simulation are created without bias toward specific event signatures. The simulated, charged particles are muons, protons, electrons, and pions with their respective antiparticle. All events include 1 to 11 particles and the charge is selected randomly. The samples include displaced events, with a starting point not in the IP, and displaced angled events with a new momentum direction. They also include vertex samples, with a displaced starting point along the original momentum direction, where two particles with opposite charge are created, with a new, random momentum direction [1]. Those topologies of the data set can be seen in Fig. 3.3.

In addition to the simulated signal is an overlay of beam background in the events (see Sec. 2.3). This is done by simulated beam background corresponding to the collider conditions in 2021 [10], in the following called *low background* and by data recorded by Belle II in the last data-taking period during Run 1, in the following called *high background*. In Fig. 3.4 a comparison between *low* and *high background* can be seen with two typical event displays. The x and y information of the wire hits are plotted at $z = 0$. The wire hits are visualized as dots over a schematic cross-section of the CDC. Because of the `mc_id` parameter in the data set, the differentiation of each track from a particle and the background is possible. The grey dots are marked as background and the colored dots each belong to a particle track.
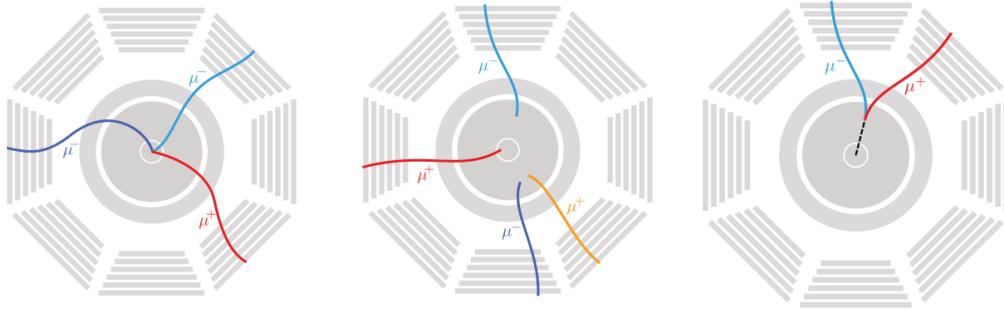
Figure 3.3.: Topologies of the data set used for training and validating the CAT finder. Particles and anti-particles are created in the IP (on the left), with displaced starting points and displaced angles (in the middle), and with displaced vertices, where two particles with opposite charges are created (on the right). Taken with permission of Lea Reuter [1].
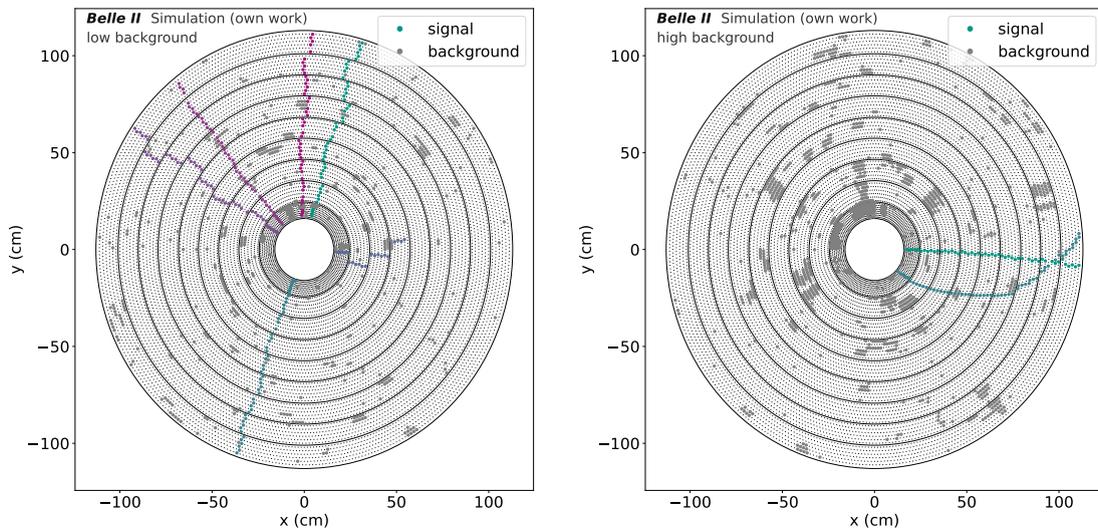
Figure 3.4.: Visualization of a simulated event display at *low* (left) and *high background* (right). The grey dots are marked as background and the colored points each belong to a track of a simulated particle.

4. Investigation of the Input Features

4.1. Visualization of the Input Features

In the following, a visualization of the input features from the CDC will be given. The data sets used for the studies of this thesis are simulated events, as discussed in Sec. 3.2.2. The visualization leads to a better understanding of the input features, used for the CAT finder and therefore will help with the upcoming optimization of the input features. Comparable and extended studies can be found in [23].

4.1.1. Comparison between the Position Features

As discussed in Sec 3.2.1, the standard position parameters of the CAT finder are `middle_x` and `middle_y` at the z middle of the wire from $z = 0\,\text{cm}$ up to $z = 40\,\text{cm}$. The other option would be the x and y position at $z = 0$, called `x` and `y`. Those parameters differ, because of the non-symmetric structure of the CDC in the xz plane and because of the skewed stereo layers (see Sec. 2.2.1). A visualization of the different position features for the CAT finder can be seen in Fig. 4.1.

Fig. 4.2 shows a comparison of the two position parameters for different angles. In the event display is a selection of tracks with specific angles θ. Those angles are calculated from the momentum information of the simulation. The position parameters are projected on the xy plane for a better comparison. The differences can then be seen at the stereo layers because of their deviating position at $z = 0$ and the middle of the wire.

The signal at $z = 0$ is in one line in the xy plane at an angle of $\theta = 90°$ and the signal at the middle of the wire is slightly shifted at the stereo layers. At an angle of about $\theta \approx 60°$ is the signal at the middle of the wire in one line and the signal at $z = 0$ is shifted at the stereo layers. In the forward direction with small angles, the shift in the stereo layers is higher at the signal at $z = 0$ as at the middle of the wire (see $\theta = 37.5°$). In the backward direction the other way around (see $\theta = 122.9°$). The shift of the track at the stereo layers gives information about the z position of the track, where a higher shift means a higher z value. Because of the asymmetric beam energies, most tracks are in the forward direction. This means, that the shift in the forward direction is important than in the backward direction. With this, the position at $z = 0$ can give better resolution in the z direction.
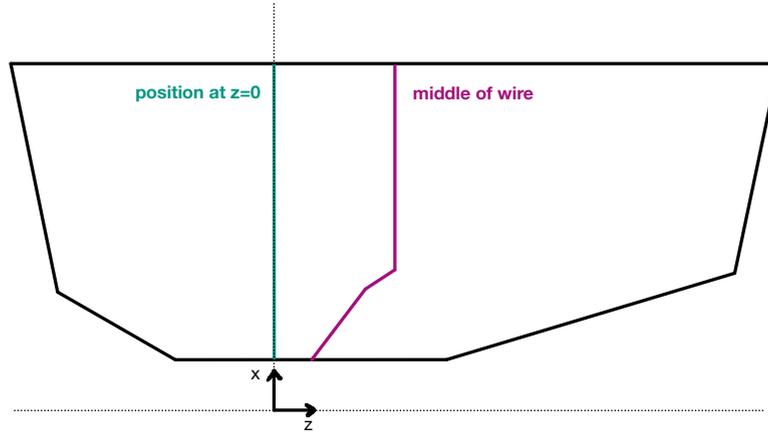
Figure 4.1.: Visualization of the position features for the CAT finder as a cross-section of the CDC in the xz plane. One position feature is taken at $z = 0$ and the other at the z middle of the wire. It can be seen that the z middle of the wire is not in one line, but shifted for the lower layers. Dimension of the CDC from [2].
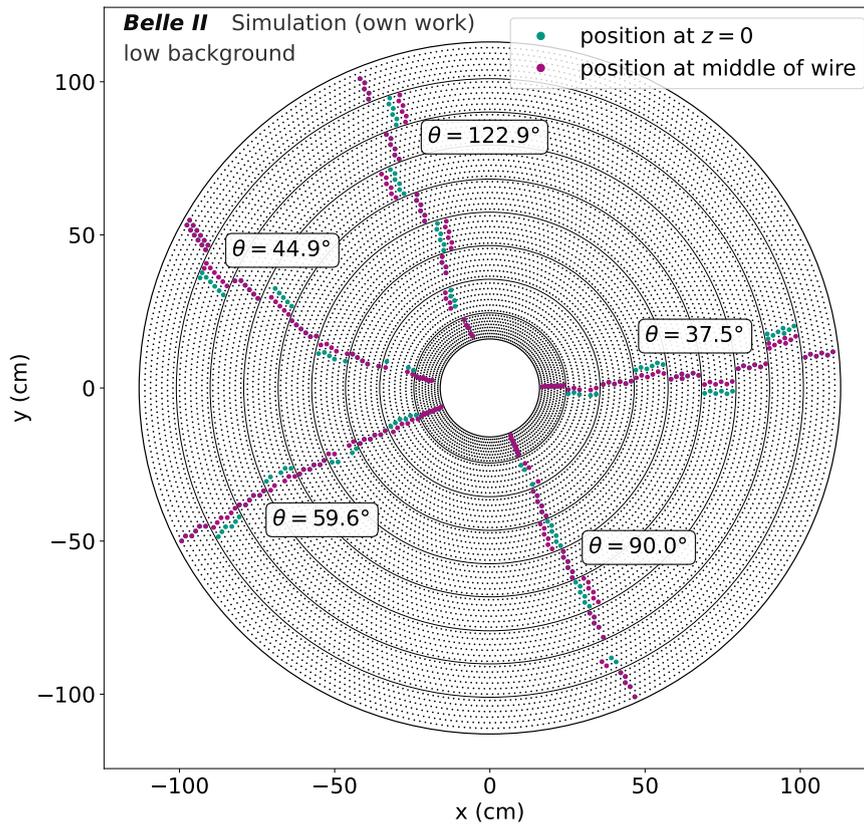
Figure 4.2.: Event display of selected tracks taken at low background with specific θ angles to compare the two different position parameters. One position parameter is taken at the z middle of the wire and the other at $z = 0$. The position of the wires shown in the visualization of the event display in the background is taken at $z = 0$.

4.1.2. Distribution of the Input Features

In the following, the distribution of the ADC count, TDC count, and TOT count (see Sec. 2.2.2) is shown for muons, electrons, and protons. Those can then be compared to the distributions found in low or high background, regarding the data set. The high background is of more interest because it is taken from real data and not simulated, as the low background. The low background is of special importance in this thesis because the training of the models is done with the low background.

Fig. 4.3 to 4.8 show the distributions for the low or high background hits as well as the three signal particle hits for the ADC, TDC, and TOT count. Due to the different amounts of background and signal hits, the histograms are normalized. An overview of the medians and means of the ADC, TDC, and TOT count per particle can be seen in Tab. 4.1.

Table 4.1.: Median and mean values of the input features for muons, electrons, and protons.

input feature	bg type	muon		electron		proton	
		median	mean	median	mean	median	mean
ADC count	low	63	87.71	73	104.14	67	133.67
	high	67	98.86	78	116.73	75	157.38
TDC count	low	4802	4765.0	4801	4762.6	4792	4749.3
	high	4812	4767.6	4813	4767.7	4803	4755.4
TOT count	low	7	6.627	7	6.581	7	6.545
	high	7	6.614	6	6.572	7	6.592

The distribution of the ADC count in dependency of the particle type compared to the low background can be seen in Fig. 4.3 and for the high background in Fig. 4.4. The signal has a distinct peak at around 60 ADC counts. The muons have the lowest median and electrons the highest. All particles have a long tail to high values. The limit of the x axis is set to 300 ADC counts, but the highest values of the ADC go up to $50\,000$ ADC counts, which can be neglected. The y axis with the relative events is logarithmic.

Most of the ADC counts that belong to the background are very small, between 0 ADC counts and 3 ADC counts. This maximum goes for the low background up to 1.6×10^{-2} relative events and for the high background to 9.5×10^{-2} relative events. The background has a second smaller maximum at around 150 ADC counts and also a long tail towards high values. Because of the different shapes of the distribution of the signal and background hits, the ADC count can be used to distinguish signal and background and even give a hint about the particle type.
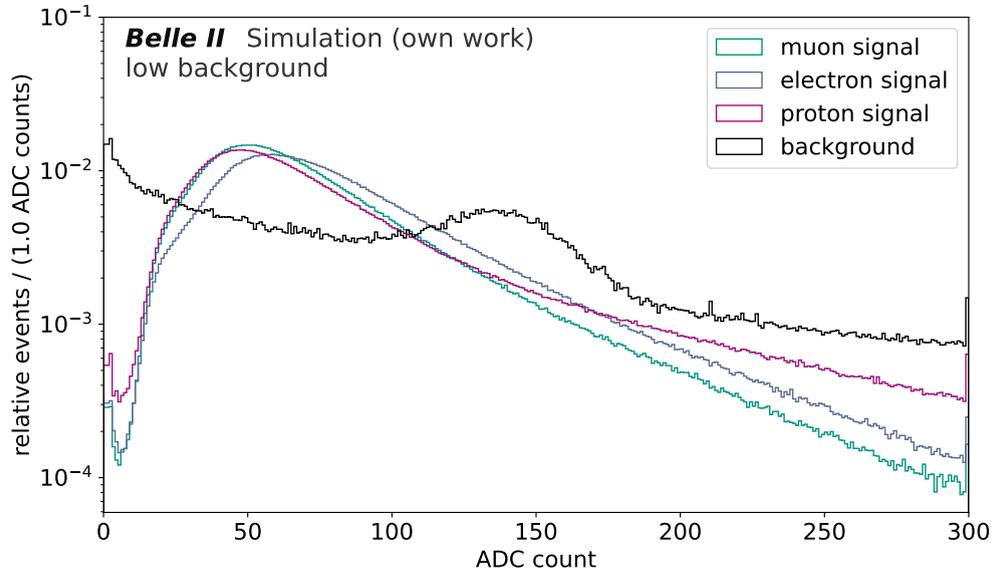
Figure 4.3.: Distribution of the ADC count for muons, electrons, and protons compared with the low background from the simulated data set.
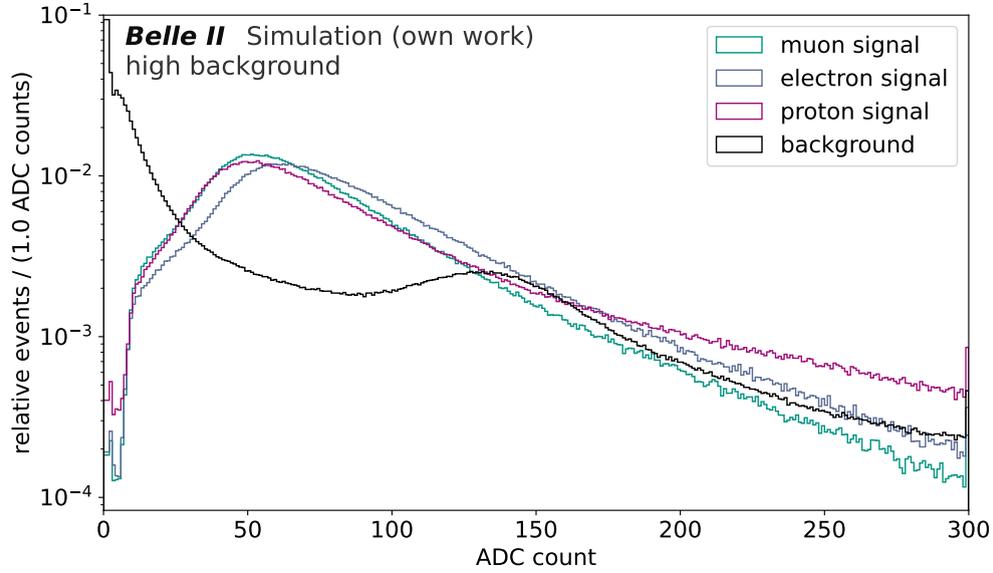
Figure 4.4.: Distribution of the ADC count for muons, electrons, and protons compared with the high background from the simulated data set.

The distribution of the TDC count in dependency of the particle type compared to the low background can be seen in Fig. 4.5 and for the high background in Fig. 4.6. The signal has a distinct maximum at around 4880 TDC counts. The distribution of muons and electrons behaves very similarly and protons have slightly smaller values. All particles have a tail to smaller values down to 4200 TDC counts. There are no signal hits over 4930 TDC counts.

The TDC count of the background is more evenly distributed with a small peak at 5000 TDC counts. The peak at the low background is more significant than at the high background. There are even values over 5000 TDC counts. This concludes that the TDC count can also be used to distinguish signal and background.
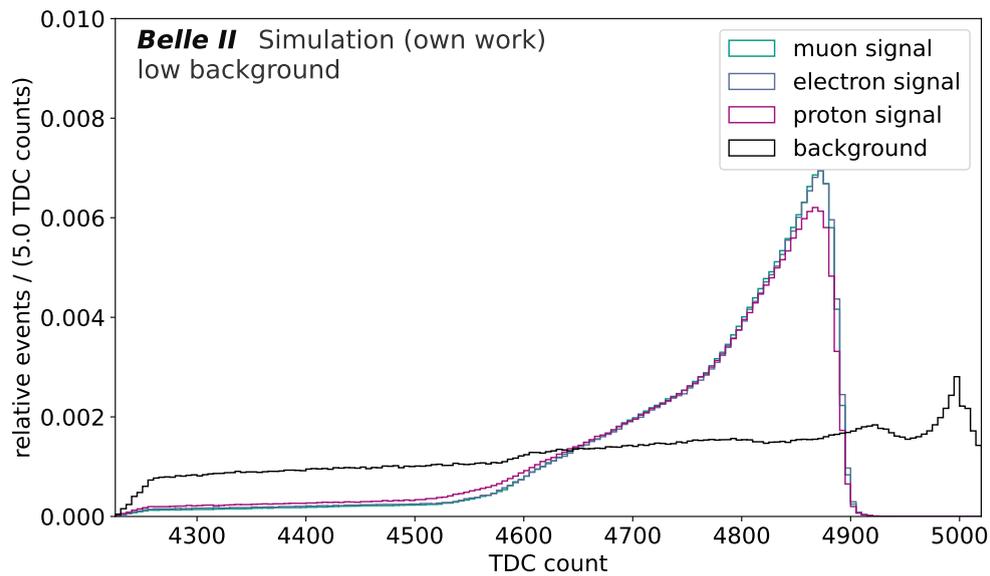
Figure 4.5.: Distribution of the TDC count for muons, electrons, and protons compared with the low background from the simulated data set.
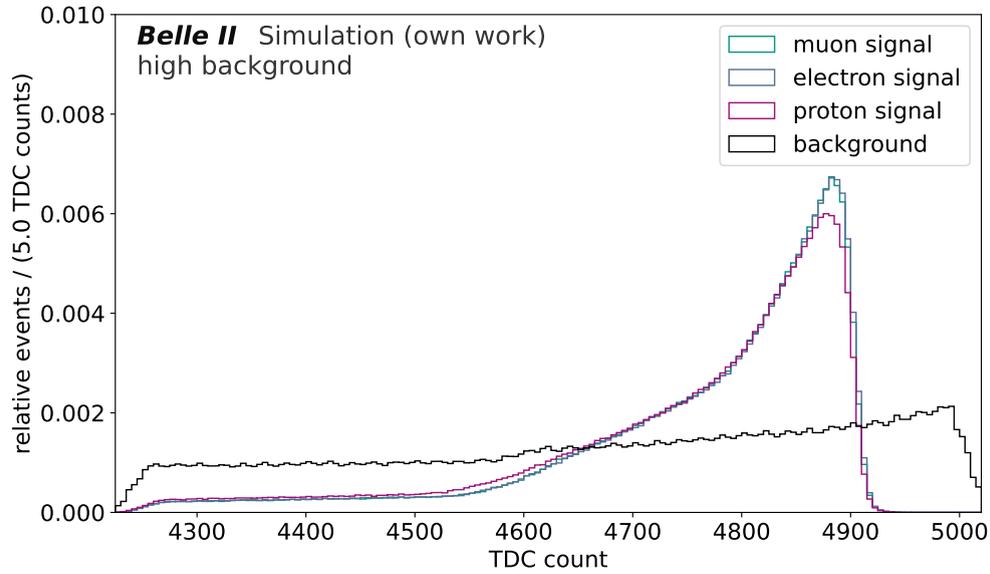
Figure 4.6.: Distribution of the TDC count for muons, electrons, and protons compared with the high background from the simulated data set.

The distribution of the TOT count in dependency of the particle type compared to the low background can be seen in Fig. 4.7 and for the high background in Fig. 4.8. The signal has a maximum of around 7 TOT counts. The distribution of muons and electrons is still quite similar and protons have slightly lower values. There are very few signal hits with under 3 TOT counts. The distribution of the signal hits has a tail to high values. The x axis is cut off at 17 TOT counts, but the highest values go up to 30 TOT counts.

The background hits have smaller TOT counts with a peak at 4 TOT counts for the low background and 2 TOT counts for the high background. The distributions of both background conditions have also a tail of up to 30 TOT counts. Although the TOT count has very small values and the values are less distributed, the TOT count still shows different distributions for signal and background and can therefore also be used to distinguish both.
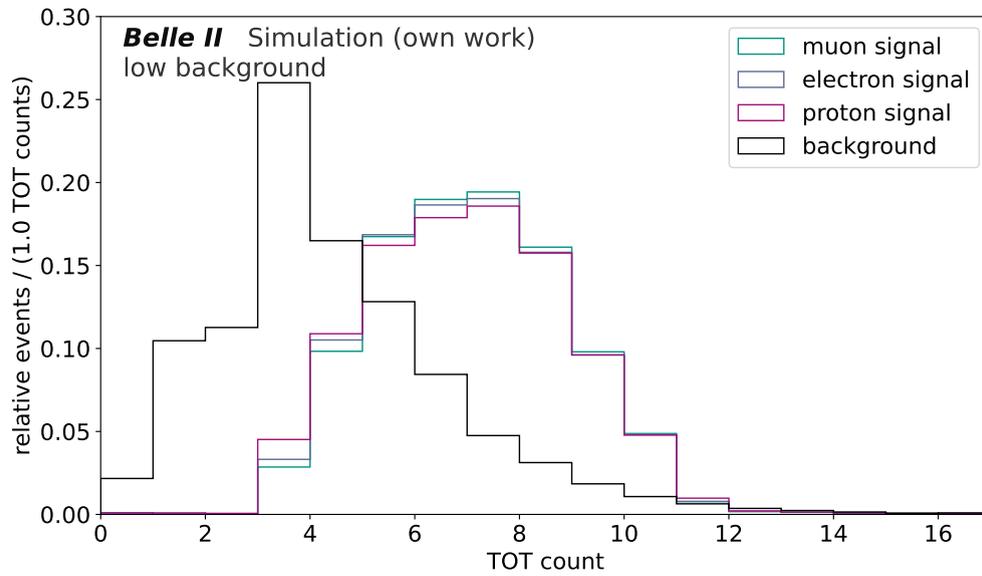
Figure 4.7.: Distribution of the TDC count for muons, electrons, and protons compared with the low background from the simulated data set.
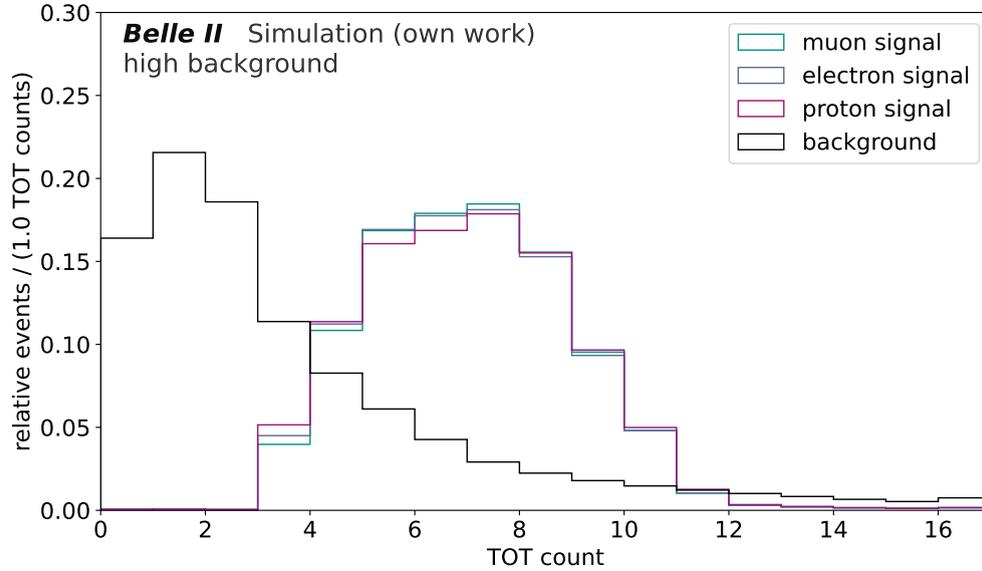
Figure 4.8.: Distribution of the TDC count for muons, electrons, and protons compared with the high background from the simulated data set.

4.1.3. Areas of the Excess of Signal and Background Hits

For a deeper investigation of the input features, the areas of the excess of signal over background hits and contrariwise are visualized in this section with two-dimensional histograms. At first, the TDC count is plotted over the ADC count, and then the TOT count is plotted over the ADC count. For signal hits, muons are chosen, because the evaluation of the models will also be taken with muons. The same plots for electron and proton hits can be seen in Sec. A.3.

The value of each bin n_{bin} in the histogram is given by the number of signal hits n_{signal} and the number of background hits $n_{\text{background}}$ per bin with the following formula:

$$n_{\text{bin}} = \frac{n_{\text{signal}} - n_{\text{background}}}{n_{\text{signal}} + n_{\text{background}}}. \tag{4.1}$$

A value of 1 (green) means a total excess of signal hits in this bin and -1 (violet) means a total excess of background hits. A value of 0 (white) is the result of a balance between the number of signal hits and background hits in one bin or an empty bin. Those conditions can therefore not be distinguished. The total amount of hits per bin can also not be determined by those plots and can be seen in Sec. A.2.

The two-dimensional histogram with the TDC count over the ADC count, with a comparison of the signal hits from a muon and the low background hits can be seen in Fig. 4.9 and in Fig. 4.10 for the high background. The highest bin at the y axis ends at 5020 TDC counts. A clear excess of background over signal hits (violet) can be seen at values under 20 ADC counts and over 4900 TDC counts. The area of the excess of signal hits (green) is bigger for the low background than for the high background. This is because about four times more background hits are measured per event and therefore the bins are filled with more background. At the low background, the green area is divided into two parts horizontally. This can be explained by the second maxima of the background hits at around 150 ADC counts. For higher values of the ADC count and lower values of the TDC count, there are fewer hits in total, but for both background conditions, there is an excess of background over signal hits.

The two-dimensional histogram with the TOT count over the ADC count can be seen in Fig. 4.11 at the low background and in Fig. 4.12 at the high background. The bins of the y axis end at 14 TOT counts and at the x axis at 300 ADC counts. An excess of background hits can be seen especially under 4 TOT counts and over 20 ADC counts. For low background, there are no hits with an ADC count which is less than double the value of the TOT count. For the high background, this is only true for signal hits. The area of the excess of signal hits is still bigger at the low background than at the high background. The most signal hits are between 8 and 11 TOT counts and for low ADC counts. For high TOT and ADC counts there are fewer hits, with a balance of signal and background hits at low background and an excess of background hits at high background.

This shows that the distinct areas of the excess of signal hits can be used to find criteria, where background hits are filtered out. But for higher background conditions the area of the excess of signal hits is getting smaller, which makes the challenge more difficult to distinguish between signal and background hits.
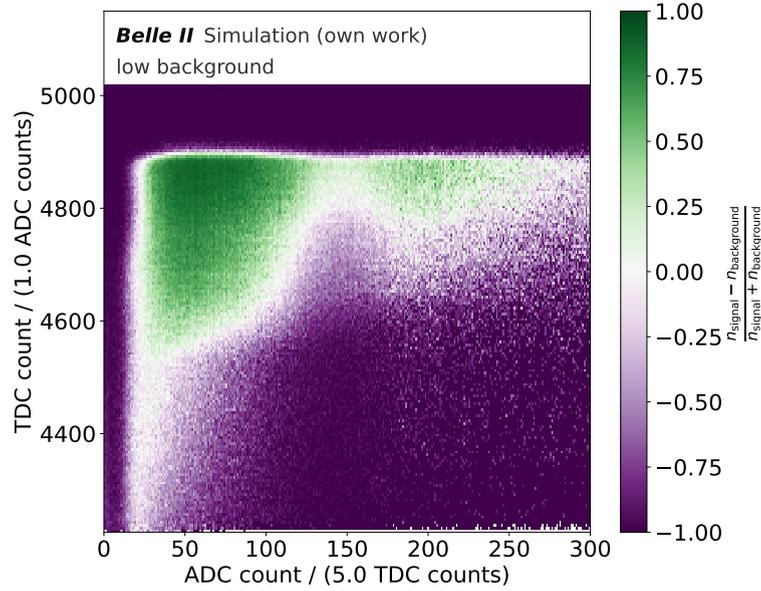
Figure 4.9.: Comparison between the signal hits from a muon and low background hits from the simulated data set in a two-dimensional histogram with the TDC count over the ADC count.
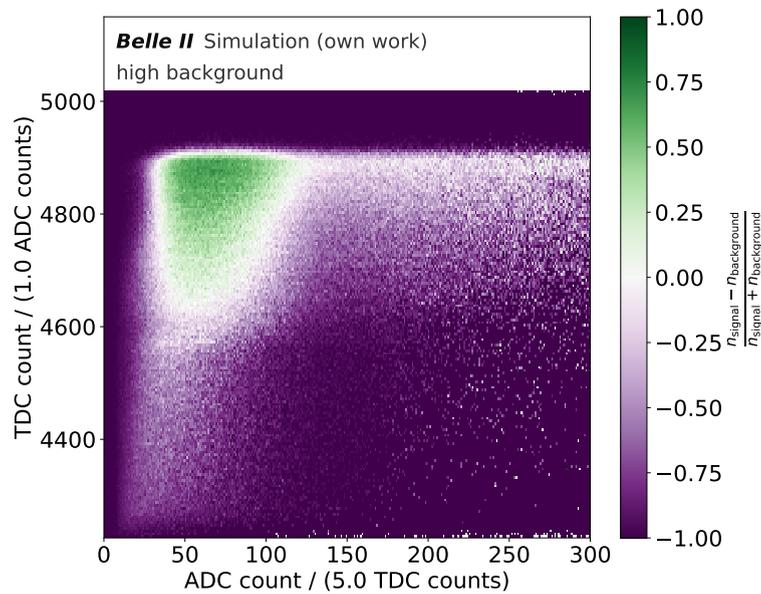
Figure 4.10.: Comparison between the signal hits from a muon and high background hits from the simulated data set in a two-dimensional histogram with the TDC count over the ADC count.
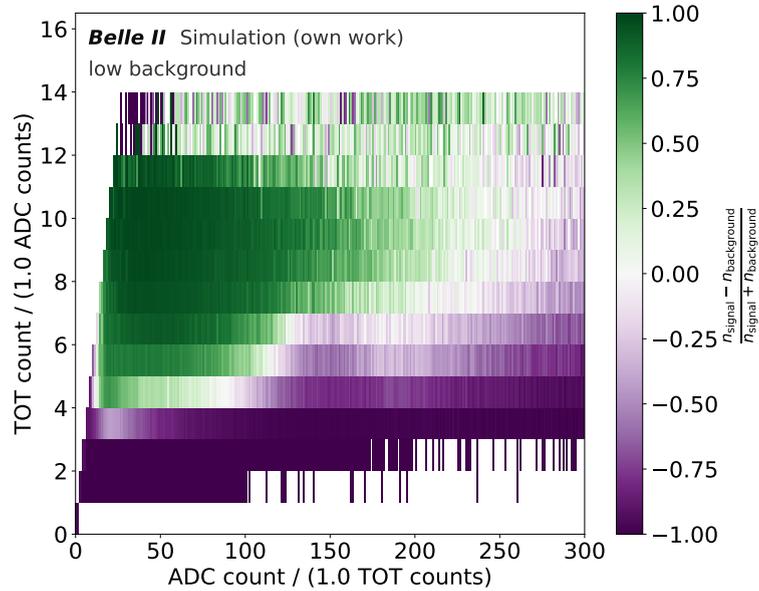
Figure 4.11.: Comparison between the signal hits from a muon and low background hits from the simulated data set in a two-dimensional histogram with the TOT count over the ADC count.
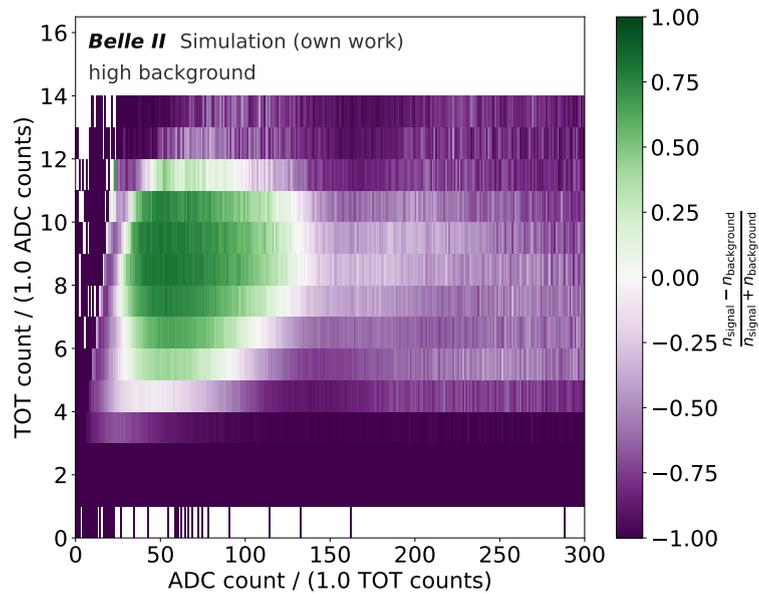
Figure 4.12.: Comparison between the signal hits from a muon and high background hits from the simulated data set in a two-dimensional histogram with the TOT count over the ADC count.

4.2. Methods of Input Feature Optimization

There are three methods of input feature optimization, which will be examined in this thesis. The first one is the *input feature selection*, the second is the *input features scaling*, and the third is the *input feature criteria*. The standard methods of input feature optimization for the CAT finder will be shortly described here and new methods will be proposed. The impact of those modifications will then be evaluated in the next Chapter.

4.2.1. Input Feature Selection

The *input feature selection* of the standard CAT finder can be seen in Sec. 3.2.1. There are different options to extend and exchange the standard features. The position features `middle_x` and `middle_y` can be exchanged by the features at $z = 0$, called `x` and `y`. The visualization of those input features in Sec. 4.1.1, shows that this can lead to a better z resolution.

Another option is to use the calculated features given by basf2. Those are the `charge_deposit`, calculated from the ADC count, and the `drift_time` and `drift_length`, calculated from the TDC count. Those features can hold additional information about the IP, the collision time, the electrical field in the cells of the CDC, and the charge of the particle.

To add additional information about the wire configuration of the CDC (see Sec. 2.2.1), the `angle` feature could be implemented. This feature then holds information about the skewed angle of the stereo wires. Since all layers in the stereo wires have a different angle, the information from the `clayer` feature can be translated to the angle via a matrix with all angles of the respective layers. The angle of each layer can be seen in [24].

4.2.2. Input Feature Scaling

The *input feature scaling* is important for the efficiency of the Neural Network (NN). The scaling of the input features puts them in a better form for the NN to use, which makes the training of the NN faster. There are multiple methods of *input feature scaling* like the statistical normalization or the min-max-normalization. The first one scales the features so that the mean of the feature is 0 and the standard deviation is 1. The advantage of this method is, that it reduces the effects of outliers in the data set. The second method rescales the features linearly in a new range, from 0 to 1 or from -1 to 1. The advantage of this method is that all relationships are preserved in the data set [7, 25].

For the CAT finder, a version of the min-max-normalization is chosen. To scale the value x of an input feature, two parameters a and b must be chosen. The scaled value x_{scaled} can then be calculated by

$$x_{\text{scaled}} = \frac{x}{a} + b. \tag{4.2}$$

In addition to this method, the features `adc` and `charge_deposit` are clipped at high values. This is to preserve the information at the lower values while removing the very high outliers of those features. The `adc` feature is clipped at 600 and the `charge_deposit` at 0. The scaling of all features in the CAT finder can be seen in Tab. 4.2.

Table 4.2.: Scaling of the input features in the CAT finder.

Input feature	unscaled range	scaled range	a	b
position features	-111 to 111	-1.11 to 1.11	100	0
superlayer	0 to 8	0 to 0.8	10	0
layer	0 to 7	0 to 0.7	10	0
clayer	0 to 55	0 to 0.98	56	0
adc	0 to 600 (clipped)	0 to 1	600	0
charge_deposit	0 to 0 (clipped)	0 to 1	0	0
tdc	4100 to 5100	0 to 0.91	1100	4100
drift_time	-130 to 710	-0.19 to 1.01	700	0
drift_length	-0.4 to 0.83	-0.4 to 0.83	1	0
tot	0 to 30	0 to 2.72	11	0

Options for a different *input feature scaling* in comparison to the standard CAT finder could be a logarithmic scaling [7] of the `adc` and `charge_deposit` features instead of the clipping at 600 ADC counts. With this, while keeping all the data, the more important low values are centered and the outliers have less weight. A visualization of the clipped and logarithmic `adc` feature can be seen in Fig. 4.13. It can be seen that for the clipping at 600 ADC counts, a lot of values are in the last bin, which goes over 25 relative events, while the values of the logarithmic scaling are more centered and have, after a division by 10, a range from 0 to 0.9.
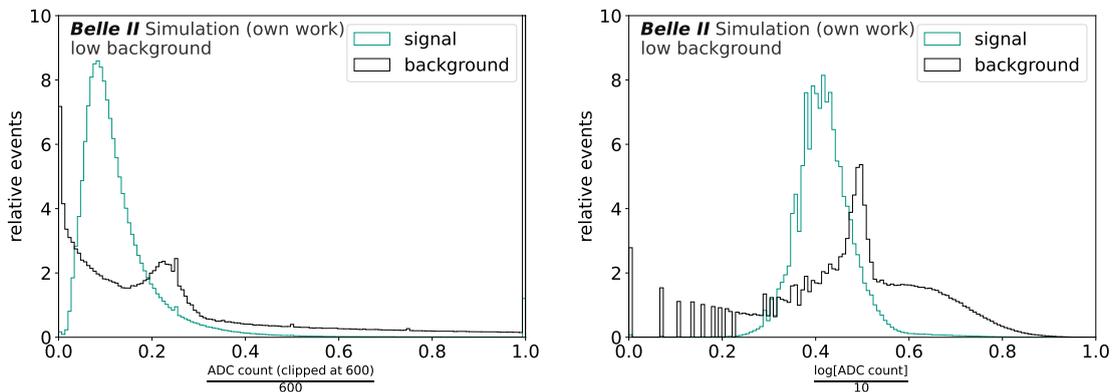
Figure 4.13.: Scaling of the `adc` feature in the CAT finder in comparison with the muon signal and low background. The standard scaling with a clipping at 600 ADC counts (left) is compared to a logarithmic scaling (right).

Another feature, whose scaling can be optimized is the `tot` feature. Because of the outliers up to 30 TOT counts, the range after scaling still goes up to 2.72. This could be fixed by applying a clipping at values of 11 TOT counts. With this, the range would be exactly between 0 and 1, while still keeping the outliers as the value 1. A visualization of the normal range and the clipped range of the `tot` feature can be seen in Fig. 4.14.
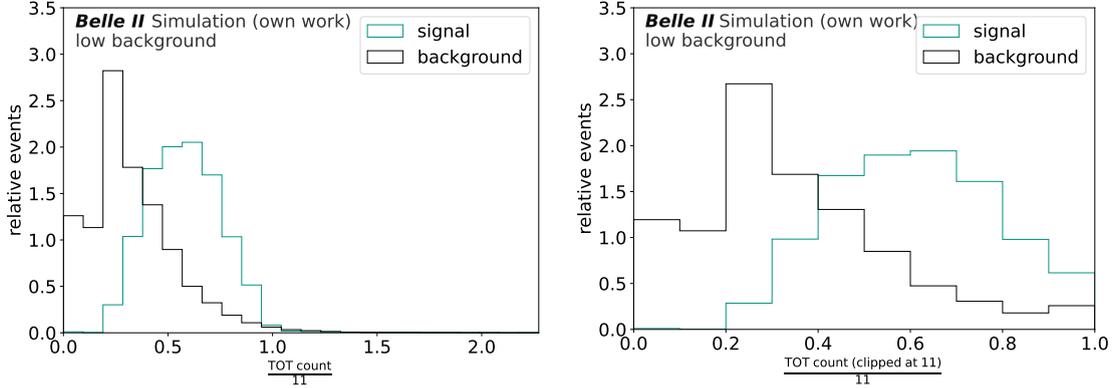
Figure 4.14.: Visualization of the scaling of the `tot` feature in the CAT finder in comparison with the muon signal and low background. The standard scaling (left) is compared to a clipping at 11 TOT counts (right).

4.2.3. Input Feature Criteria

As in Sec. 4.1.2 and 4.1.3 discussed, the distribution of the input features differs in dependency if they belong to a signal or a background hit. Because of this, *input feature criteria* can be found, to reduce the amount of background hits without losing many signal hits. The low values of the `tot` and `adc` features are mainly background hits, while the signal hits are at higher values. This can be used to find criteria, with a lower limit for those features. This would remove background hits while keeping most of the signal hits. The standard configuration of the CAT finder has no criteria for the input features. In the following, the `adc` and `tot` criteria will be discussed, but the same principles apply to other features like the `charge_deposit` (see Sec A.4). For the `tdc` feature no criteria will be used, because of the importance of this feature on the correct event timing.

The ratio of removed background hits $\epsilon_{\text{background}}$ is then given by

$$\epsilon_{\text{background}} = \frac{n_{\text{hits}}(\text{removed background})}{n_{\text{hits}}(\text{background})}. \tag{4.3}$$

Here, $n_{\text{hits}}(\text{removed background})$ is the amount of removed background hits, and the amount of all background hits without feature criteria is $n_{\text{hits}}(\text{background})$. The amount of removed signal hits ϵ_{signal} is given by

$$\epsilon_{\text{signal}} = \frac{n_{\text{hits}}(\text{removed signal})}{n_{\text{hits}}(\text{signal})}. \tag{4.4}$$

The amount of removed signal hits is $n_{\text{hits}}(\text{removed signal})$ and the total amount of signal hits without feature criteria is $n_{\text{hits}}(\text{signal})$. The ratio of removed signal hits is chosen to not be higher than one percent.

An overview of different lower limit criteria for the `adc` and `tot` features at low background can be seen in Tab. 4.3. The feature criteria gives a lower limit for either the `adc` or the

`tot` feature, to remove all hits lower than the limit. As expected, the ratio of removed background hits is always higher than for the signal hits. The ratio of removed muon hits is for most limits the lowest, the ratio of removed electron hits is in general higher. For the proton hits, the ratio of removed hits is up to two times higher than for the other particles.

The highest ratio of removed background is at the criteria ≥ 4 TOT counts, with around 50% removed background. But the ratio of removed signal for this criteria is also between 3% for muons and 5% for protons, which is over the chosen threshold of how much signal can be removed. The second highest amount of removed background is at ≥ 3 TOT counts, with 22% of removed background and around a tenth of percent removed signal. These criteria can therefore be considered the best because the ratio of removed backgrounds is the highest while the ratio of removed signal hits is under the threshold.

The ADC criteria seem to be less effective on the removed background than the TOT criteria. For example, the criteria ≥ 15 ADC counts and ≥ 2 TOT counts have about the same ratio of removed background hits, but the ratio of removed signal hits is about five times higher for the criteria on the ADC count. A combination of a lower limit for the `adc` feature and the `tot` feature would not be effective, because less background would be removed, while the ratio of removed signal hits would be the same as with only one limit (see Sec. A.4).

Table 4.3.: Ratio of removed signal $\epsilon_{\mathrm{signal}}$ and background hits $\epsilon_{\mathrm{background}}$ for different feature criteria with lower limits for the `adc` and `tot` features for a muon, electron, and proton signal at the low background.

feature criteria	ϵ_{muon} (%)	$\epsilon_{\mathrm{electron}}$ (%)	$\epsilon_{\mathrm{proton}}$ (%)	$\epsilon_{\mathrm{background}}$ (%)
≥ 6 ADC counts	0.125	0.141	0.255	5.568
≥ 8 ADC counts	0.155	0.171	0.320	6.949
≥ 10 ADC counts	0.195	0.212	0.398	8.194
≥ 15 ADC counts	0.526	0.513	0.821	10.894
≥ 20 ADC counts	1.734	1.436	2.059	13.431
≥ 1 TOT count	0.057	0.060	0.100	2.052
≥ 2 TOT counts	0.101	0.111	0.192	11.940
≥ 3 TOT counts	0.126	0.142	0.255	22.672
≥ 4 TOT counts	2.980	3.459	4.770	49.402

The amount of removed signal and background hits for high background can be seen in Tab. 4.4. While the amount of removed signal hits is not significantly higher than at the low background, the removed background hits are much larger. For example at a lower limit of 3 TOT counts, with a loss of around 0.1 % of signal hits for all particles, more than 50 % of background hits are removed. With a lower limit of 4 TOT counts, even 67 % of background hits can be removed. The ratio of removed signal hits is about twice as large for the ADC criteria at high background than for the low background, but the ratio of removed background hits is about four times higher.

Table 4.4.: Ratio of removed signal ϵ_{signal} and background hits $\epsilon_{\text{background}}$ for different feature criteria with lower limits for the `adc` and `tot` features for a muon, electron, and proton signal at the high background.

feature criteria	ϵ_{muon} (%)	$\epsilon_{\text{electron}}$ (%)	ϵ_{proton} (%)	$\epsilon_{\text{background}}$ (%)
≥ 6 ADC counts	0.095	0.104	0.214	28.727
≥ 8 ADC counts	0.159	0.172	0.305	33.777
≥ 10 ADC counts	0.383	0.378	0.514	37.885
≥ 15 ADC counts	1.580	1.273	1.525	44.748
≥ 20 ADC counts	3.226	2.444	2.919	48.947
≥ 1 TOT count	0.036	0.037	0.073	16.365
≥ 2 TOT counts	0.070	0.078	0.151	37.734
≥ 3 TOT counts	0.096	0.104	0.215	56.078
≥ 4 TOT counts	4.063	4.598	5.357	67.342

The criteria will be applied to the models in the next Chapter and the best criteria will be examined by the evaluation of the models. All cuts of the low background, where the ratio of removed signal hits is not over the chosen threshold of 1% will be tested.

5. Evaluation and Analysis

5.1. Evaluation Metrics

To measure the performance of the chosen configurations for the CAT finder, the tracking metrics from the Belle II tracking group [4] are reproduced in the following section. At first, the *hit efficiency* and *hit purity* are determined, to give criteria for the tracks of the simulated particles. Then the metrics for the performance of the track finding are defined.

The *hit efficiency* ε_{hit} of one track is given by the ratio of hits matched to a simulated particle and included in the found track $n_{\text{hits}}(\text{matched and in track})$ divided by the number of all matched hits $n_{\text{hits}}(\text{matched})$, as

$$\varepsilon_{\text{hit}} = \frac{n_{\text{hits}}(\text{matched and in track})}{n_{\text{hits}}(\text{matched})}. \tag{5.1}$$

The *hit purity* $\mathfrak{p}_{\text{hit}}$ is the ratio of the number of hits matched to a particle and included in a found track $n_{\text{hits}}(\text{matched and in track})$ divided by the number of all hits in the track $n_{\text{hits}}(\text{in track})$, as

$$\mathfrak{p}_{\text{hit}} = \frac{n_{\text{hits}}(\text{matched and in track})}{n_{\text{hits}}(\text{in track})}. \tag{5.2}$$

For those parameters and the number of hits related to this track $n_{\text{hits}}(\text{in track})$, the following criteria are chosen, to define, if a found track is related to a simulated particle: $\varepsilon_{\text{hit}} > 0.005$, $\mathfrak{p}_{\text{hit}} > 0.66$, and $n_{\text{hits}}(\text{in track}) \geq 7$. If the criteria for the *hit efficiency* or *hit purity* are not achieved, the track is called a *fake track*. The number of *fake tracks* $n_{\text{fake track}}$ divided by the number of all found tracks n_{track}, is called *fake rate* $\mathfrak{r}_{\text{fake}}$ and given by

$$\mathfrak{r}_{\text{fake}} = \frac{n_{\text{fake track}}}{n_{\text{track}}}. \tag{5.3}$$

The track with the highest *hit purity* is chosen if more than one found track can be related to the same simulated particle. Those tracks are then called *clone track*. The *clone rate* $\mathfrak{r}_{\text{clone}}$ is given by the ratio of the number of *clone tracks* $n_{\text{clone track}}$ divided by the number of all tracks related to this particle $n_{\text{tracks}}(\text{related to particle})$, as

$$\mathfrak{r}_{\text{clone}} = \frac{n_{\text{clone tracks}}}{n_{\text{track}}(\text{related to particle})}. \tag{5.4}$$

The ratio of the number of matched tracks $n_{\text{track}}(\text{matched})$ divided by the number of simulated particles that are matched to at least one hit $n_{\text{simulated}}$ is called *track efficiency* $\varepsilon_{\text{track}}$ and is given by

$$\varepsilon_{\text{track}} = \frac{n_{\text{track}}(\text{matched})}{n_{\text{simulated}}}. \tag{5.5}$$

The *track charge efficiency* $\varepsilon_{\text{charge}}$ is given by the ratio of the number of matched tracks with the correct reconstructed charge $n_{\text{track}}(\text{matched and corr. charge})$ divided by the number of simulated particles that are matched to at least one hit $n_{\text{simulated}}$, as

$$\varepsilon_{\text{charge}} = \frac{n_{\text{track}}(\text{matched and corr. charge})}{n_{\text{simulated}}}. \tag{5.6}$$

The *wrong charge rate* $\mathfrak{r}_{\text{wrong ch.}}$ is defined as the number of matched tracks with the wrong charge $n_{\text{track}}(\text{matched and wrong charge})$ divided by the number of all tracks, that are matched to a particle $n_{\text{track}}(\text{matched})$, as

$$\mathfrak{r}_{\text{wrong ch.}} = \frac{n_{\text{track}}(\text{matched and wrong charge})}{n_{\text{track}}(\text{matched})}. \tag{5.7}$$

For all metrics, a different value for the track finding and track fitting can be determined. This is because tracks can be found, but then lost in the fitting step. The following analysis will therefore differentiate between CAT finder and CAT fitter.

5.2. Training and Evaluation of the Input Feature Configurations

In this section, the evaluation of the chosen input feature configurations for the CAT finder and fitter will be presented. After the selection of the input features, a new scaling will be tested and at last, new feature criteria will be applied. The training of the chosen configurations is done with a smaller data set with 10 % of the data compared to the data set for the standard CAT finder.

Since the *charge efficiency* $\varepsilon_{\text{charge}}$ is considered the most important metric, the performance of the following models will be ranked by this metric. A *charge efficiency* of 100% is considered a perfect performance and means, that all tracks are matched to a simulated particle with the correct charge. All the following tables with rankings of the presented models, are sorted by the *charge efficiency*, from the highest to the lowest. When a new model beats the old one in this metric, it is considered the best model and the new standard for the following studies. The goal of this section is to find the best configuration of input features for the CAT finder and fitter.

The evaluation software from the Belle II tracking group determines the indicated errors. After training and evaluating the same model several times, the uncertainties of the training are measured as $\pm 0.2\%$ of the *charge efficiency*. An improvement of a model of around 1 % of the *charge efficiency* or more can therefore be considered significant. Unless otherwise stated, the training of all models is done at the low background with all particles, and the evaluations are done with muons. The evaluation was also done only in the barrel region, because of the importance of this region and the high performance of the baseline in this region. Only a selection of models and rankings is shown here. A total ranking of all configurations can be seen in Sec. B.2.

5.2.1. Selection of the Position Features

The selection of the input features starts with the standard CAT configuration (see Sec. 3.2.1). The standard position features are `middle_x` and `middle_y` but can be exchanged by the `x` and `y` features taken at $z = 0$. An overview of the models for the position feature selection can be seen in Tab. 5.1.

Table 5.1.: Feature selection of the position features.

model name	position features
middle-of-wire CAT	`middle_x` and `middle_y`
z = 0 CAT	`x` and `y`

The evaluation of those models can be seen in Fig. 5.2. As expected, the model $z = 0$ *CAT* with the position features at $z = 0$ has a higher efficiency in finding and fitting the tracks than the model with the position at the middle of the wire (*middle-of-wire CAT*). The *charge efficiency* and *track efficiency* are significantly higher for the $z = 0$ *CAT* finder and fitter. The *fake, clone,* and *wrong charge rate* are so small, that they can be neglected. In the following, the features `x` and `y` are used as the position features.

Table 5.2.: Evaluation with muons of the models for the selection of the position features with a model with the standard configuration (*middle-of-wire CAT*) and the $z = 0$ *CAT* model. Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$ (%)	$\varepsilon_{\text{track}}$ (%)	$\mathfrak{r}_{\text{fake}}$ (%)	$\mathfrak{r}_{\text{clone}}$ (%)	$\mathfrak{r}_{\text{wrong ch.}}$ (%)
$z = 0$ *CAT* finder	$95.27^{+0.06}_{-0.06}$	$96.77^{+0.05}_{-0.05}$	$1.72^{+0.04}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.55^{+0.04}_{-0.04}$
middle-of-wire CAT finder	$93.92^{+0.07}_{-0.07}$	$95.19^{+0.06}_{-0.06}$	$2.03^{+0.04}_{-0.04}$	$0.08^{+0.01}_{-0.01}$	$1.33^{+0.03}_{-0.03}$
$z = 0$ *CAT* fitter	$92.08^{+0.08}_{-0.08}$	$92.65^{+0.07}_{-0.07}$	$0.51^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.61^{+0.02}_{-0.02}$
middle-of-wire CAT fitter	$91.02^{+0.08}_{-0.08}$	$91.47^{+0.08}_{-0.08}$	$0.81^{+0.03}_{-0.03}$	0.02	$0.49^{+0.02}_{-0.02}$

5.2.2. Selection of the Charge and Time Features

The models for the selection of the charge and time features can be seen in Tab. 5.3. The *adc-tdc-tot CAT* model has the additional `tot` feature in comparison to the *no tot CAT* model, with the standard charge and time feature selection. In a model without `tdc` (*no tdc CAT*) the importance of this feature is examined. The `adc` and `tdc` features are exchanged by the calibrated features `charge_deposit`, `drift_time`, and `drift_length` in different configurations.

Table 5.3.: Feature selection of the charge and time features.

model name	charge features	time features	
adc-tdc-tot CAT	adc	tdc	tot
no tdc CAT	adc		tot
no tot CAT	adc	tdc	
drift CAT	adc	drift_time and drift_length	tot
charge CAT	charge_deposit	tdc	tot
charge-drift CAT	charge_deposit	drift_time and drift_length	tot

The evaluation of the models can be seen in Tab. 5.4. As expected, the addition of the `tot` feature in comparison to the standard charge and time features, `adc` and `tdc`, has a better *charge efficiency* for the CAT finder and fitter. The *adc-tdc-tot CAT* is overall the model with the highest performance for this selection. The models without `tot` (*no tot CAT*) and without `tdc` (*no tdc CAT*) have a significantly lower *charge efficiency*, with the *no tdc CAT* having the lowest performance. Especially the *no tdc CAT* has a higher *fake rate* than all other models. The other *fake*, *clone*, and *wrong charge rates* are so low that they can be neglected. This shows the importance of the time features `tdc` and `tot` and that they each give additional important information for tracking.

The *charge efficiency* of the *drift-charge CAT* finder is not significantly lower than the *adc-tdc-tot CAT* finder, but for the associated fitters the first one performs significantly lower than the second. A reason for this could be, that the calculated features, especially

the `drift_time` and `drift_length` are not calculated precisely enough for low multiplicity events to give the correct parameters for the track fitting. This could then be the reason, the *charge CAT* fitter, which does not use those features, performs slightly better than the *drift-charge CAT* fitter, while the associated finders have no significant difference. The models with a mixture of calculated and normal features, the *charge* and *drift CAT* have overall a lower performance than the models with only one type of feature. This could mean, that the normal features and the calculated features are not compatible. The *drift CAT* has a very low performance, which could come from the lower-performing calculated time features and the incompatibility of its features. Therefore in the following the `adc`, `tdc`, and `tot` features are the charge and time features. Those features are also used by basf2 for background filters.

Table 5.4.: Evaluation with muons of the models for the selection of the charge and time features (configuration see Tab. 5.3). Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$ (%)	$\varepsilon_{\text{track}}$ (%)	$\mathfrak{r}_{\text{fake}}$ (%)	$\mathfrak{r}_{\text{clone}}$ (%)	$\mathfrak{r}_{\text{wrong ch.}}$ (%)
adc-tdc-tot CAT finder	$96.29^{+0.05}_{-0.05}$	$97.62^{+0.04}_{-0.04}$	$1.44^{+0.03}_{-0.03}$	$0.19^{+0.01}_{-0.01}$	$1.36^{+0.03}_{-0.03}$
drift-charge CAT finder	$96.08^{+0.09}_{-0.09}$	$97.69^{+0.07}_{-0.07}$	$1.75^{+0.06}_{-0.06}$	$0.22^{+0.02}_{-0.02}$	$1.65^{+0.06}_{-0.06}$
charge CAT finder	$96.01^{+0.06}_{-0.05}$	$97.37^{+0.05}_{-0.04}$	$1.59^{+0.03}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$1.40^{+0.03}_{-0.03}$
no tot CAT finder	$95.27^{+0.06}_{-0.06}$	$96.77^{+0.05}_{-0.05}$	$1.72^{+0.04}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.55^{+0.04}_{-0.04}$
drift CAT finder	$95.05^{+0.06}_{-0.06}$	$96.47^{+0.05}_{-0.05}$	$1.71^{+0.04}_{-0.04}$	$0.12^{+0.01}_{-0.01}$	$1.47^{+0.03}_{-0.03}$
no tdc CAT finder	$94.64^{+0.06}_{-0.06}$	$95.94^{+0.06}_{-0.06}$	$3.63^{+0.05}_{-0.05}$	$0.14^{+0.01}_{-0.01}$	$1.36^{+0.03}_{-0.03}$
adc-tdc-tot CAT fitter	$93.2^{+0.07}_{-0.07}$	$93.62^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.44^{+0.02}_{-0.02}$
charge CAT fitter	$92.87^{+0.07}_{-0.07}$	$93.3^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.46^{+0.02}_{-0.02}$
drift-charge CAT fitter	$92.3^{+0.1}_{-0.1}$	$92.7^{+0.1}_{-0.1}$	$0.49^{+0.03}_{-0.03}$	$0.07^{+0.01}_{-0.01}$	$0.5^{+0.03}_{-0.03}$
no tot CAT fitter	$92.08^{+0.08}_{-0.08}$	$92.65^{+0.07}_{-0.07}$	$0.51^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.61^{+0.02}_{-0.02}$
drift CAT fitter	$91.91^{+0.08}_{-0.08}$	$92.33^{+0.08}_{-0.07}$	$0.56^{+0.02}_{-0.02}$	0.02	$0.46^{+0.02}_{-0.02}$
no tdc CAT fitter	$91.46^{+0.08}_{-0.08}$	$92.06^{+0.08}_{-0.08}$	$1.24^{+0.03}_{-0.03}$	$0.05^{+0.01}_{-0.01}$	$0.65^{+0.02}_{-0.02}$

5.2.3. Selection of the Layer Features

The models for the selection of the layer features can be seen in Tab. 5.5. In comparison to the standard wire feature selection (*no angle CAT*), the new feature `angle` is added in the *angle CAT* model. As scaling of this new feature, the parameter a is chosen as the highest value of the angles in mrad, $a = 79$ so that the scaled range is between -1 and 1. The importance of the `clayer` feature after adding the `angle` is tested with the *no clayer CAT* model and the importance of the `layer` feature with the *no layer CAT* model.

In Tab. 5.6 the evaluation of the models for the selection of the layer feature can be seen. Adding the `angle` feature in the *angle CAT* does not have the expected higher performance than without angle (*no angle CAT*). Both models have no significant differences in the

Table 5.5.: Feature selection of the layer features with the new `angle` feature.

model name	layer features			
angle CAT	superlayer	layer	clayer	angle
no angle CAT	superlayer	layer	clayer	
no clayer CAT	superlayer	layer		angle
no layer CAT	superlayer		clayer	angle

metrics for the finder and fitter. The reason could be, that the CAT finder gets enough information about the structure of the CDC from the other layer features.

The importance of the `clayer` and `layer` feature can be confirmed with the lower performance of the *no clayer CAT* and *no layer CAT* finder and fitter. But it can be seen, that the difference of those models to the full feature selection is not as high as for other features when they are left out.

Table 5.6.: Evaluation with muons of the models for the selection of the layer features (configuration see Tab. 5.5). Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$ (%)	$\varepsilon_{\text{track}}$ (%)	$\mathfrak{r}_{\text{fake}}$ (%)	$\mathfrak{r}_{\text{clone}}$ (%)	$\mathfrak{r}_{\text{wrong ch.}}$ (%)
no angle CAT finder	$96.29^{+0.05}_{-0.05}$	$97.62^{+0.04}_{-0.04}$	$1.44^{+0.03}_{-0.03}$	$0.19^{+0.01}_{-0.01}$	$1.36^{+0.03}_{-0.03}$
angle CAT finder	$96.15^{+0.05}_{-0.05}$	$97.46^{+0.04}_{-0.04}$	$1.55^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$1.34^{+0.03}_{-0.03}$
no clayer CAT finder	$95.35^{+0.06}_{-0.06}$	$96.83^{+0.05}_{-0.05}$	$1.68^{+0.04}_{-0.04}$	$0.16^{+0.01}_{-0.01}$	$1.53^{+0.03}_{-0.04}$
no layer CAT finder	$95.24^{+0.06}_{-0.06}$	$96.86^{+0.05}_{-0.05}$	$1.65^{+0.03}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.66^{+0.04}_{-0.04}$
angle CAT fitter	$93.26^{+0.07}_{-0.07}$	$93.78^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.55^{+0.02}_{-0.02}$
no angle CAT fitter	$93.2^{+0.07}_{-0.07}$	$93.62^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.44^{+0.02}_{-0.02}$
no clayer CAT fitter	$92.07^{+0.08}_{-0.08}$	$92.44^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.04^{+0.01}_{-0.01}$	$0.41^{+0.02}_{-0.02}$
no layer CAT fitter	$91.97^{+0.08}_{-0.08}$	$92.4^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.47^{+0.02}_{-0.02}$

The features of the *angle CAT* model are considered the best feature selection in the following studies. Those are `x` and `y` as the position features, `adc`, `tdc`, and `tot` as the charge and time features, and `superlayer`, `layer`, `clayer`, and `angle` as layer features.

5.2.4. Scaling of the Input Features

The model with the best feature selection and the standard scaling (*std. scaling CAT*) will be compared to the models with the new scaling. Those are a model with clipping of the `tot` feature at 11 TOT counts (*tot clip CAT*) and a model with logarithmic scaling of the `adc` feature (*adc log CAT*).

The evaluation of those models can be seen in Tab. 5.7. The new scaling methods do not have the expected performance advantages. The models have no significant difference in the *charge efficiency*. The standard scaling *std. scaling CAT* has a slightly better performance, especially for the fitter. The reason for this could be that most of the values of the `adc` feature are below 600 ADC counts and not affected by the clipping. The logarithmic scaling just shifts the values higher, but also does not contain the relationships in the data. The clipping of the `tot` feature seems to have no significant effect because most of the values are below the clipping and therefore with no change. With those results, the standard scaling is still considered the best scaling for the following models.

Table 5.7.: Evaluation with muons of the models with the input feature scaling (scaling methods see Sec. 5.2.4). Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$ (%)	$\varepsilon_{\text{track}}$ (%)	$\mathfrak{r}_{\text{fake}}$ (%)	$\mathfrak{r}_{\text{clone}}$ (%)	$\mathfrak{r}_{\text{wrong ch.}}$ (%)
std. scaling CAT finder	$96.15^{+0.05}_{-0.05}$	$97.46^{+0.04}_{-0.04}$	$1.55^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$1.34^{+0.03}_{-0.03}$
tot clip CAT finder	$96.03^{+0.06}_{-0.05}$	$97.54^{+0.04}_{-0.04}$	$1.61^{+0.03}_{-0.03}$	$0.24^{+0.01}_{-0.01}$	$1.55^{+0.03}_{-0.04}$
adc log CAT finder	$96.01^{+0.06}_{-0.05}$	$97.37^{+0.05}_{-0.04}$	$1.61^{+0.03}_{-0.03}$	$0.16^{+0.01}_{-0.01}$	$1.4^{+0.03}_{-0.03}$
std. scaling CAT fitter	$93.26^{+0.07}_{-0.07}$	$93.78^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.55^{+0.02}_{-0.02}$
adc log CAT fitter	$93.02^{+0.07}_{-0.07}$	$93.53^{+0.07}_{-0.07}$	$0.58^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.56^{+0.02}_{-0.02}$
tot clip CAT fitter	$92.6^{+0.07}_{-0.07}$	$93.11^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.11^{+0.01}_{-0.01}$	$0.54^{+0.02}_{-0.02}$

5.2.5. Criteria for the Input Features

Based on the best feature selection model (*no criteria CAT*), the feature criteria with a lower limit for the `tot` and `adc` feature will be evaluated here. The model name gives the lower limit, with the number being the lowest value of the feature, which is kept in the data set. For example at the *tot1 CAT* model, all values of `tot` smaller than 1 TOT count are removed.

The evaluation of the models with different criteria for the `adc` and `tot` feature can be seen in Tab. 5.8. The model *no criteria CAT* has the highest performance, but the difference between the models with lower criteria and the model without criteria is very small. This could mean, that the model without criteria learns on its own, how to distinguish between signal and background with the given input features and there is no need to remove hits, for a higher-performing model. The worst performing models are the ones with the lower limit of `tot` at 1 and `adc` at 10. The reason for this could be, that the amount of background is too small to affect the model, but the are still removed signal hits, which the model misses.

Table 5.8.: Evaluation with muons of the models with the feature criteria for the `adc` and `tot` feature (used criteria see Sec. 5.2.5). Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$ (%)	$\varepsilon_{\text{track}}$ (%)	$\mathfrak{r}_{\text{fake}}$ (%)	$\mathfrak{r}_{\text{clone}}$ (%)	$\mathfrak{r}_{\text{wrong ch.}}$ (%)
no criteria CAT finder	$96.15^{+0.05}_{-0.05}$	$97.46^{+0.04}_{-0.04}$	$1.55^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$1.34^{+0.03}_{-0.03}$
adc8 CAT finder	$95.99^{+0.06}_{-0.05}$	$97.27^{+0.05}_{-0.05}$	$1.45^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$1.32^{+0.03}_{-0.03}$
adc15 CAT finder	$95.99^{+0.06}_{-0.06}$	$97.29^{+0.05}_{-0.05}$	$1.57^{+0.04}_{-0.04}$	$0.17^{+0.01}_{-0.01}$	$1.34^{+0.04}_{-0.04}$
tot3 CAT finder	$95.95^{+0.06}_{-0.06}$	$97.41^{+0.04}_{-0.04}$	$1.37^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$1.5^{+0.03}_{-0.03}$
adc6 CAT finder	$95.93^{+0.06}_{-0.06}$	$97.3^{+0.05}_{-0.05}$	$1.53^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$1.41^{+0.03}_{-0.03}$
tot2 CAT finder	$95.86^{+0.06}_{-0.06}$	$97.37^{+0.05}_{-0.04}$	$1.68^{+0.04}_{-0.04}$	$0.18^{+0.01}_{-0.01}$	$1.56^{+0.04}_{-0.04}$
adc10 CAT finder	$95.29^{+0.07}_{-0.07}$	$96.73^{+0.06}_{-0.06}$	$1.49^{+0.04}_{-0.04}$	$0.16^{+0.01}_{-0.01}$	$1.49^{+0.04}_{-0.04}$
tot1 CAT finder	$95.29^{+0.06}_{-0.06}$	$96.4^{+0.05}_{-0.05}$	$1.63^{+0.03}_{-0.04}$	$0.11^{+0.01}_{-0.01}$	$1.16^{+0.03}_{-0.03}$
no criteria CAT fitter	$93.26^{+0.07}_{-0.07}$	$93.78^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.55^{+0.02}_{-0.02}$
adc15 CAT fitter	$93.16^{+0.08}_{-0.08}$	$93.6^{+0.08}_{-0.08}$	$0.5^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.46^{+0.02}_{-0.02}$
tot3 CAT fitter	$92.94^{+0.07}_{-0.07}$	$93.36^{+0.07}_{-0.07}$	$0.5^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.44^{+0.02}_{-0.02}$
adc6 CAT fitter	$92.89^{+0.07}_{-0.07}$	$93.34^{+0.07}_{-0.07}$	$0.49^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.47^{+0.02}_{-0.02}$
tot2 CAT fitter	$92.69^{+0.07}_{-0.07}$	$93.13^{+0.07}_{-0.07}$	$0.57^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.47^{+0.02}_{-0.02}$
adc8 CAT fitter	$92.5^{+0.07}_{-0.07}$	$92.88^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.42^{+0.02}_{-0.02}$
adc10 CAT fitter	$92.09^{+0.09}_{-0.09}$	$92.57^{+0.09}_{-0.09}$	$0.57^{+0.02}_{-0.03}$	$0.07^{+0.01}_{-0.01}$	$0.52^{+0.02}_{-0.02}$
tot1 CAT fitter	$92.01^{+0.08}_{-0.08}$	$92.42^{+0.07}_{-0.07}$	$0.5^{+0.02}_{-0.02}$	0.02	$0.44^{+0.02}_{-0.02}$

In Tab. 5.9 the evaluation for the same models with protons can be seen. This evaluation is of special interest because the removed signal hits are higher for protons than for muons and electrons. The model without criteria and the models with the highest criteria, *tot3 CAT* and *adc15 CAT* have no significant difference. Only the *adc15 CAT* model is noticeable with a slightly higher performance of the fitter. The lower criteria model, *tot1 CAT*, *tot2 CAT*, *adc6 CAT*, *adc8 CAT*, and *adc10 CAT* have the lowest performance.

An important advantage of the criteria for the input features is the faster time for the training of the models, when there are fewer hits in the data set. The most important metric is the duration per iteration, which ranks the models according to their time savings. The number of epochs can vary depending on the training, and therefore the total duration, which is dependent on the total number of epochs, can also vary strongly. However the duration per iteration varies only about $\pm 1\,\text{s}$, which was tested with the same model trained several times. The time saving is compared to the ratio of removed hits ϵ_{hits}, which is given by the total amount of removed hits by the criteria divided by the total amount of hits.

The time savings for the different models can be seen in Tab. 5.10. As expected, there is a significant time saving for the models after applying the feature criteria. For the model with the highest amount of removed hits, the *tot3 CAT*, the duration per iteration is nearly half as large as for the model with no criteria. This model is followed by all the other models with a criteria of the `tot` feature. The models with criteria of the `adc` feature have

Table 5.9.: Evaluation with protons of the models with the feature criteria for the `adc` and `tot` feature (used criteria see Sec. 5.2.5). Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$ (%)	$\varepsilon_{\text{track}}$ (%)	$\mathfrak{r}_{\text{fake}}$ (%)	$\mathfrak{r}_{\text{clone}}$ (%)	$\mathfrak{r}_{\text{wrong ch.}}$ (%)
tot3 CAT finder	$93.53^{+0.08}_{-0.08}$	$94.24^{+0.08}_{-0.08}$	$3.3^{+0.05}_{-0.05}$	$0.47^{+0.02}_{-0.02}$	$0.75^{+0.03}_{-0.03}$
adc15 CAT finder	$93.41^{+0.09}_{-0.09}$	$94.22^{+0.09}_{-0.08}$	$3.3^{+0.05}_{-0.05}$	$0.47^{+0.02}_{-0.02}$	$0.86^{+0.03}_{-0.04}$
no criteria CAT finder	$93.26^{+0.08}_{-0.08}$	$94.25^{+0.08}_{-0.08}$	$3.79^{+0.05}_{-0.05}$	$0.62^{+0.02}_{-0.02}$	$1.05^{+0.03}_{-0.04}$
adc6 CAT finder	$93.25^{+0.08}_{-0.08}$	$94.17^{+0.08}_{-0.08}$	$3.67^{+0.05}_{-0.05}$	$0.5^{+0.02}_{-0.02}$	$0.97^{+0.03}_{-0.03}$
adc8 CAT finder	$93.07^{+0.08}_{-0.08}$	$94.1^{+0.08}_{-0.08}$	$3.67^{+0.05}_{-0.05}$	$0.52^{+0.02}_{-0.02}$	$1.09^{+0.04}_{-0.04}$
adc10 CAT finder	$92.93^{+0.09}_{-0.09}$	$93.87^{+0.09}_{-0.09}$	$3.43^{+0.05}_{-0.05}$	$0.53^{+0.02}_{-0.02}$	$1.0^{+0.04}_{-0.04}$
tot2 CAT finder	$92.93^{+0.09}_{-0.08}$	$93.95^{+0.08}_{-0.08}$	$3.92^{+0.05}_{-0.05}$	$0.46^{+0.02}_{-0.02}$	$1.09^{+0.04}_{-0.04}$
tot1 CAT finder	$92.09^{+0.09}_{-0.09}$	$92.9^{+0.09}_{-0.08}$	$3.91^{+0.05}_{-0.05}$	$0.44^{+0.02}_{-0.02}$	$0.88^{+0.03}_{-0.03}$
adc15 CAT fitter	$91.3^{+0.1}_{-0.1}$	$92.06^{+0.1}_{-0.1}$	$1.08^{+0.03}_{-0.03}$	$0.25^{+0.01}_{-0.02}$	$0.85^{+0.03}_{-0.04}$
tot3 CAT fitter	$90.65^{+0.1}_{-0.1}$	$91.48^{+0.09}_{-0.09}$	$1.07^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$0.9^{+0.03}_{-0.03}$
no criteria CAT fitter	$90.5^{+0.1}_{-0.1}$	$91.43^{+0.09}_{-0.09}$	$1.22^{+0.03}_{-0.03}$	$0.3^{+0.01}_{-0.02}$	$1.02^{+0.03}_{-0.04}$
adc6 CAT fitter	$90.23^{+0.1}_{-0.1}$	$91.09^{+0.09}_{-0.09}$	$1.13^{+0.03}_{-0.03}$	$0.24^{+0.01}_{-0.01}$	$0.95^{+0.03}_{-0.03}$
adc10 CAT fitter	$90.2^{+0.1}_{-0.1}$	$91.2^{+0.1}_{-0.1}$	$1.2^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$1.04^{+0.04}_{-0.04}$
tot2 CAT fitter	$90.2^{+0.1}_{-0.1}$	$91.09^{+0.09}_{-0.09}$	$1.34^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$0.98^{+0.03}_{-0.03}$
adc8 CAT fitter	$90.13^{+0.1}_{-0.1}$	$91.05^{+0.09}_{-0.09}$	$1.23^{+0.03}_{-0.03}$	$0.23^{+0.01}_{-0.01}$	$1.02^{+0.03}_{-0.04}$
tot1 CAT fitter	$89.5^{+0.1}_{-0.1}$	$90.37^{+0.1}_{-0.1}$	$1.19^{+0.03}_{-0.03}$	$0.21^{+0.01}_{-0.01}$	$0.96^{+0.03}_{-0.03}$

a higher duration per iteration, although the ratio of removed hits is not smaller than for the models with the criteria of the `tot` feature. The models *tot2 CAT* and *adc15 CAT* have roughly the same ratio of removed hits, but the duration per iteration is significantly smaller for the first model. A reason for this could be, that the amount of removed signal hits is about 5 times higher for the criteria of the `adc` feature than for the `tot` feature, while the amount of background is about the same. This would mean, that the model *adc15 CAT* has fewer signal hits, to find the tracks, which makes it last longer in the training.

In conclusion, it is not easy to pick the best input feature criteria after this evaluation. For the performance of the model, removing hits is not necessary, because the model learns on its own to distinguish signal and background from the input features. When the time saving of the training is of high importance, then the model *tot3 CAT* is the best one. This model has roughly the same performance as a model without criteria, but the duration of the training is significantly quicker.

As already discussed, the criteria at high background are expected to have a higher impact, because of the higher ratio of removed background hits in comparison to the removed signal hits. A training of such models is not done in the context of this work, because the model size and the size of the data set would be too small for training at the high background. Higher background conditions would require more time for training to deal with the more complicated events, with more background hits than signal hits.

Table 5.10.: Time savings of the models with feature criteria. Sorted by the duration/iteration with the total duration, last epoch after early stopping and the ratio of removed hits ϵ_{hits}.

model name	duration/iteration (s)	total duration	last epoch	ϵ_{hits} (%)
tot3 CAT	75.03	15 h 46 min 40 s	757	12.79
tot2 CAT	102.41	23 h 4 min 14 s	811	6.75
tot1 CAT	123.09	22 h 48 min 19 s	667	1.18
adc15 CAT	124.77	23 h 58 min 57 s	692	6.35
adc10 CAT	127.99	35 h 33 min 14 s	1000	4.69
adc8 CAT	130.87	29 h 2 min 45 s	799	3.97
adc6 CAT	133.06	29 h 9 min 43 s	789	3.18
no criteria CAT	139.17	29 h 15 min 48 s	757	0.00

5.3. Comparison of the Best Configuration CAT to the Baseline

In this section, a comparison of the model with the best configuration will be compared with the baseline and the standard CAT finder. The model was trained with the same sized data set as the standard CAT finder. The used feature selections are `x` and `y` as the position features, `adc`, `tdc`, and `tot` as the charge and time features, and `superlayer`, `layer`, `clayer`, and `angle` as layer features. The criteria for the `tot` feature of ≥ 3 TOT counts was chosen. The training duration was over 12 days, which is more than 6 times higher than for the other models.

The evaluation of those models with muons can be seen in Tab. 5.11. It can be seen that the model with the best configuration (*best config. CAT*) has a significantly higher performance than the *Baseline*. This proves the fact, that a tracking software based on GNNs is more effective in finding and fitting tracks than the *Baseline* based on Legendre-transformations. The standard model (*std. CAT*) has a slightly higher performance than the best configuration model, which is in contrast to the previous studies. The reason for this could be, that the model is bigger with more GravNet blocks. The *std. CAT* model has about double the amount of total parameters than the *best config. CAT* model. For a correct comparison, the same model size should be compared.

Table 5.11.: Evaluation with muons of the best configuration model *best config. CAT* in comparison to the *Baseline* and the standard CAT model *std. CAT*. Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\mathfrak{r}_{\text{wrong ch.}}$
std. CAT finder	$99.2^{+0.03}_{-0.02}$	$99.54^{+0.02}_{-0.02}$	$0.93^{+0.03}_{-0.03}$	$0.03_{-0.01}$	$0.34^{+0.02}_{-0.02}$
best config. CAT finder	$98.99^{+0.03}_{-0.03}$	$99.41^{+0.02}_{-0.02}$	$0.73^{+0.02}_{-0.02}$	0.02	$0.43^{+0.02}_{-0.02}$
Baseline finder	$96.78^{+0.05}_{-0.05}$	$98.63^{+0.03}_{-0.03}$	$0.71^{+0.02}_{-0.02}$	$0.04^{+0.01}_{-0.01}$	$1.87^{+0.04}_{-0.04}$
std. CAT fitter	$97.43^{+0.04}_{-0.04}$	$97.55^{+0.04}_{-0.04}$	$0.41^{+0.02}_{-0.02}$	0.01	$0.12^{+0.01}_{-0.01}$
best config. CAT fitter	$97.19^{+0.05}_{-0.05}$	$97.32^{+0.05}_{-0.05}$	$0.23^{+0.01}_{-0.01}$	0.01	$0.13^{+0.01}_{-0.01}$
Baseline fitter	$96.1^{+0.06}_{-0.05}$	$97.37^{+0.05}_{-0.05}$	$0.51^{+0.02}_{-0.02}$	0.02	$1.3^{+0.03}_{-0.03}$

6. Conclusion and Outlook

The goal of this thesis was to improve the performance and training time of the track finding algorithm CAT for the CDC in the Belle II detector. For this, different methods of input feature optimization were applied to models of the CAT finder to get the model with the best configuration.

At first, the input features were investigated by visualizing them. By comparing the two position features, it could be seen that the x and y features, with the position taken at $z = 0$ contain better position information than the features taken at the z middle of the wire, `middle_x` and `middle_y`. In a comparison of the input features of signal and background hits, it could be seen, that the features `adc`, `tdc`, and `tot`, taken from the readout electronics of the CDC, can be used to distinguish signal and background hits. In all of those features are small areas of the excess of signal hits over background hits, which can be used to apply criteria to remove background hits and could even help in distinguishing the particle types.

In the main part, the models with the different configurations were trained at the low background, with 1 to 11 particles per event, including displaced, displaced angled, and displaced vertex events. Starting from the standard configuration model with a charge efficiency evaluated with muons of $\varepsilon_{\mathrm{charge}} = 93.9\%$ for the finder and $\varepsilon_{\mathrm{charge}} = 91.0\%$ for the fitter, the first improvement could be made by switching the position features to x and y. An improvement of over one percentage point for the finder and fitter could be made, which is clearly over the uncertainties of the training for similar models and therefore can be considered significant. In the selection of the charge and time features, by adding the `tot` feature, another improvement of over one percentage point could be made. The addition of the new `angle` feature did not have the expected improvements for the CAT finder and fitter but had roughly the same performance. The model with the position features x and y, charge and time features `adc`, `tdc`, and `tot` and layer features `superlayer`, `clayer`, `layer`, and `angle` is now considered the model with the best feature selection. The charge efficiency of the finder of this model is $\varepsilon_{\mathrm{charge}} = 96.2\%$ and of the fitter $\varepsilon_{\mathrm{charge}} = 93.26\%$, which is an improvement of more than two percentage points over the standard configuration.

The new scaling methods that were applied to the models, namely the logarithmic scaling of the `adc` feature and the clipping of the `tot` feature at 11 TOT counts, were not effective. The efficiency of the models with the new scaling was around a tenth of a percentage point less than that of the model with the standard scaling.

At last, different models with criteria of the input features, with lower limits of the `adc` and the `tot` feature were trained and evaluated. The criteria at ≥ 15 ADC counts and ≥ 3 TOT counts had slightly the same charge efficiency as the model with no criteria, the other models had a lower performance. The time savings on the other hand were significant, with a duration per iteration of around $75\,\text{s}$ for the criteria ≥ 3 TOT counts compared to around $139\,\text{s}$ for the model without criteria. This time-saving of around 46% is very significant for the training of the CAT models.

A model was trained with the same sized data set as the standard CAT finder and the best configuration, which has the best feature selection and a feature criteria of ≥ 3 TOT counts. A comparison was made with this best configuration model, the baseline, and the standard CAT finder. It could be seen that the performance is significantly higher for the best configuration CAT model than for the baseline, with a difference of over two percentage points for the model with the best configuration over the baseline finder and over one percentage point for the baseline fitter. The standard CAT model had the highest performance by a slight margin over the best configuration model for the finder and fitter. This could be explained by the larger model size of the standard CAT model with more layers and parameters.

In conclusion, the improvement of the CAT finder by input feature optimization was successful. Especially by the input feature selection, the performance was improved significantly. The criteria of the `tot` feature resulted in significant time savings for the training of the model without losing performance. This shows, that studies on input feature optimization can improve Neural Networks significantly.

One major limitation of this study is the size of the models and the used data set. For Neural Networks, those are very important for the performance of the Network. In the end, it could be seen that even after finding the best configuration, the larger model has still a higher efficiency. The other limitation was the so-called low background, with which the models were trained. Those background hits are simulated as opposed to the high background presented in this study, which was taken from the detector.

Based on those studies, in the future, the best configuration should be tested with larger models. Those could then be expanded to data sets with the high background taken from the detector. At the high background, one should also study the effects of the feature criteria, because of their higher ratio of removed background hits.

The CDC of the Belle II detector is facing increasing background conditions and finding particles is getting more and more difficult. This study can help improve the track finding in the future with higher-performing models and a faster training time.

Bibliography

[1] L. Reuter et al. *Graph Neural Network based Track Finding in the Central Drift Chamber at Belle II*. 2023. URL: https://indico.jlab.org/event/459/contributions/11761/ (visited on Oct. 15, 2024).

[2] T. Abe et al. (Belle II Collaboration). *Belle II Technical Design Report*. 2010. arXiv: 1011.0352 [physics.ins-det]. URL: https://arxiv.org/abs/1011.0352.

[3] Belle II Collaboration. *Belle II Analysis Software Framework (basf2)*. Version release-06-00-09. Aug. 2022. DOI: 10.5281/zenodo.6949513. URL: https://doi.org/10.5281/zenodo.6949513.

[4] Bertacchi et al. (Belle II Tracking Group). "Track finding at Belle II". In: *Computer Physics Communications* 259 (Feb. 2021), p. 107610. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2020.107610. URL: http://dx.doi.org/10.1016/j.cpc.2020.107610.

[5] Grammarly Inc. *Grammarly*. URL: https://www.grammarly.com (visited on Oct. 18, 2024).

[6] K. Akai, K. Furukawa, and H. Koiso (SuperKEKB accelerator team). "SuperKEKB collider". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (2018). Advances in Instrumentation and Experimental Methods (Special Issue in Honour of Kai Siegbahn), pp. 188–199. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2018.08.017. URL: https://www.sciencedirect.com/science/article/pii/S0168900218309616.

[7] L. Obaid et al. "Effect of feature optimization on performance of machine learning models for predicting traffic incident duration". In: *Engineering Applications of Artificial Intelligence* 131 (2024), p. 107845. ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2024.107845. URL: https://www.sciencedirect.com/science/article/pii/S0952197624000034.

[8] M. J. Willatt, F. Musil, and M. Ceriotti. "Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements". In: *Phys. Chem. Chem. Phys.* 20 (47 2018), pp. 29661–29668. DOI: 10.1039/C8CP05921G. URL: http://dx.doi.org/10.1039/C8CP05921G.

[9] E. Kou et al. (Belle II Collaboration). "The Belle II Physics Book". In: *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019). ISSN: 2050-3911. DOI: 10.1093/ptep/ptz106. URL: http://dx.doi.org/10.1093/ptep/ptz106.

[10] A. Natochii et al. (Belle II Collaboration). *Beam background expectations for Belle II at SuperKEKB*. 2022. arXiv: 2203.05731 [hep-ex]. URL: https://arxiv.org/abs/2203.05731.

[11] Belle II Software Group. *Belle II Software Documentation*. URL: https://software.belle2.org/development/sphinx/online_book/fundamentals/02-datataking.html (visited on Oct. 10, 2024).

[12] V. Trusov. "Development of Pattern Recognition Algorithms for the Central Drift Chamber of the Belle II Detector". PhD thesis. Karlsruhe Institute of Technology (KIT), 2016.

[13] K. Nishimura. "The time-of-propagation counter for Belle II". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 639.1 (May 2011), pp. 177–180. ISSN: 0168-9002. DOI: 10.1016/j.nima.2010.09.164. URL: http://dx.doi.org/10.1016/j.nima.2010.09.164.

[14] KEK HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION. *New electronics tested for Belle II central drift chamber*. 2010. URL: https://www2.kek.jp/proffice/archives/feature/2010/BelleIICDCDesign.html (visited on Sept. 30, 2024).

[15] N. Taniguchi et al. "All-in-one readout electronics for the Belle-II Central Drift Chamber". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 732 (2013). Vienna Conference on Instrumentation 2013, pp. 540–542. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2013.06.096. URL: https://www.sciencedirect.com/science/article/pii/S0168900213009509.

[16] N. Braun. "Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment". PhD thesis. Karlsruhe: Karlsruhe Institute of Technology (KIT), 2018. URL: https://doi.org/10.5445/IR/1000089317.

[17] S. Shimazaki et al. "Front-end electronics of the Belle II drift chamber". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 735 (2014), pp. 193–197. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2013.09.050. URL: https://www.sciencedirect.com/science/article/pii/S0168900213012849.

[18] A. Natochii et al. (Belle II Collaboration). "Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1055 (Oct. 2023), p. 168550. ISSN: 0168-9002. DOI: 10.1016/j.nima.2023.168550. URL: http://dx.doi.org/10.1016/j.nima.2023.168550.

[19] T. Alexopoulos et al. "Implementation of the Legendre Transform for track segment reconstruction in drift tube chambers". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 592.3 (2008), pp. 456–462. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2008.04.038. URL: https://www.sciencedirect.com/science/article/pii/S0168900208005780.

[20] S. R. Qasim et al. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". In: *The European Physical Journal C* 79.7 (July 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7113-9. URL: http://dx.doi.org/10.1140/epjc/s10052-019-7113-9.

[21] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[22] T. Kuhr et al. "The Belle II Core Software: Belle II Framework Software Group". In: *Computing and Software for Big Science* 3.1 (Nov. 2018). ISSN: 2510-2044. DOI: 10.1007/s41781-018-0017-9. URL: http://dx.doi.org/10.1007/s41781-018-0017-9.

[23] P. Dorwarth. "Graph-Building and Input Feature Analysis for Edge Classification in the Central Drift Chamber at Belle II". MA thesis. Karlsruhe Institute of Technology (KIT), 2023.

[24] N. Taniguchi. "CDC wire configuration". In: *KEK* (July 2016).

[25] T. Jayalakshmi and A. Santhakumaran. "Statistical normalization and back propagation for classification". In: *International Journal Computer Theory Engineering (IJCTE)* 3 (Jan. 2011), pp. 89–93.

A. Appendix: Investigation of the Input Features

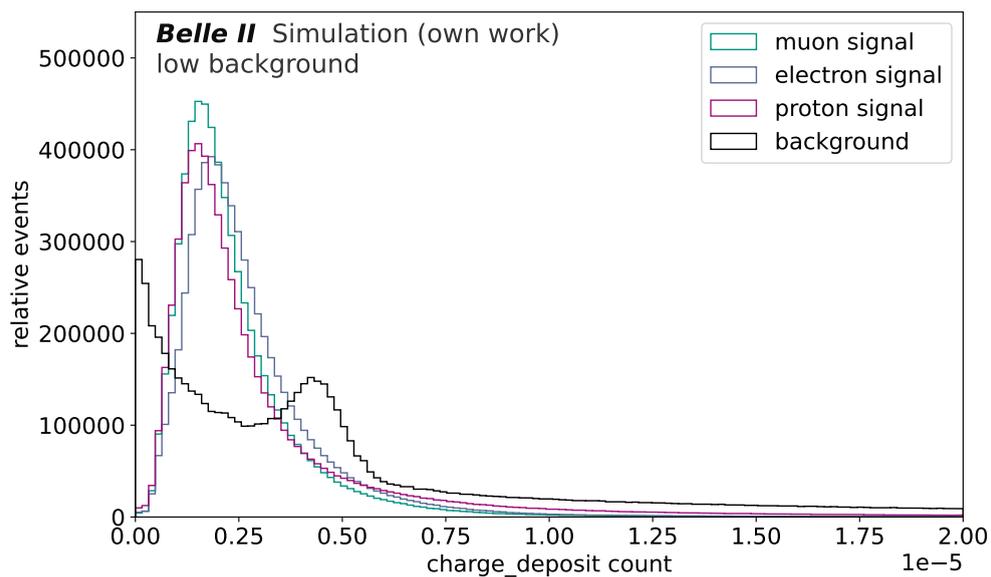## A.1. Distribution of the Input Features

Figure A.1.: Distribution of the charge deposit for muons, electrons, and protons compared with the low background from a simulated data set.
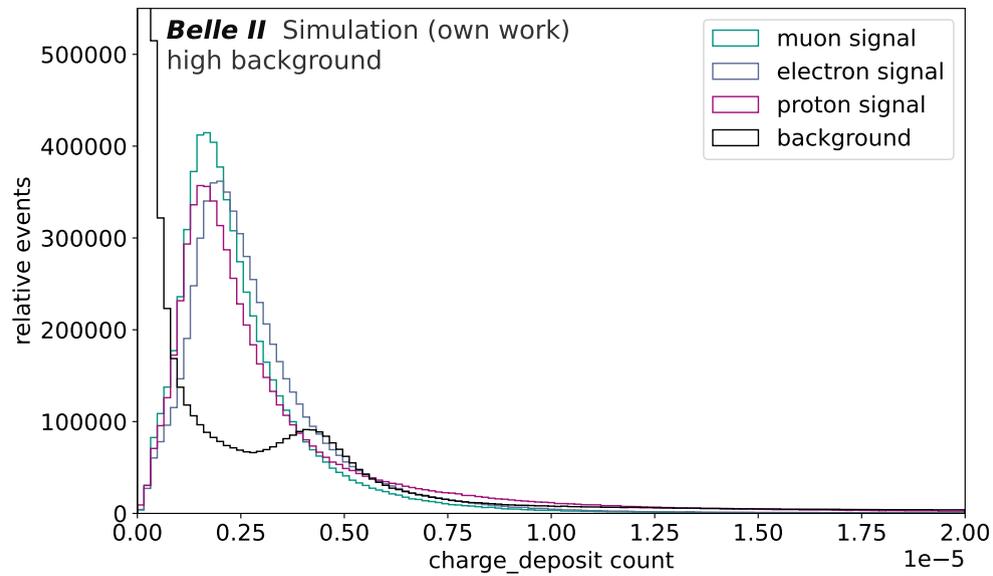
Figure A.2.: Distribution of the charge deposit for muons, electrons, and protons compared with the high background from a simulated data set.

Figure A.3.: Distribution of the drift time for muons, electrons, and protons compared with the low background from a simulated data set.
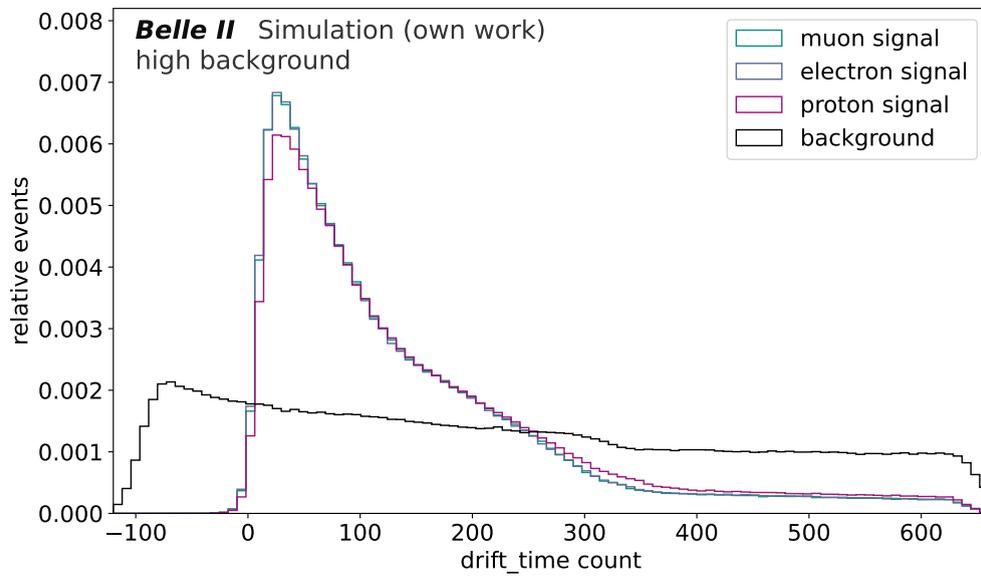
Figure A.4.: Distribution of the drift time for muons, electrons, and protons compared with the high background from a simulated data set.
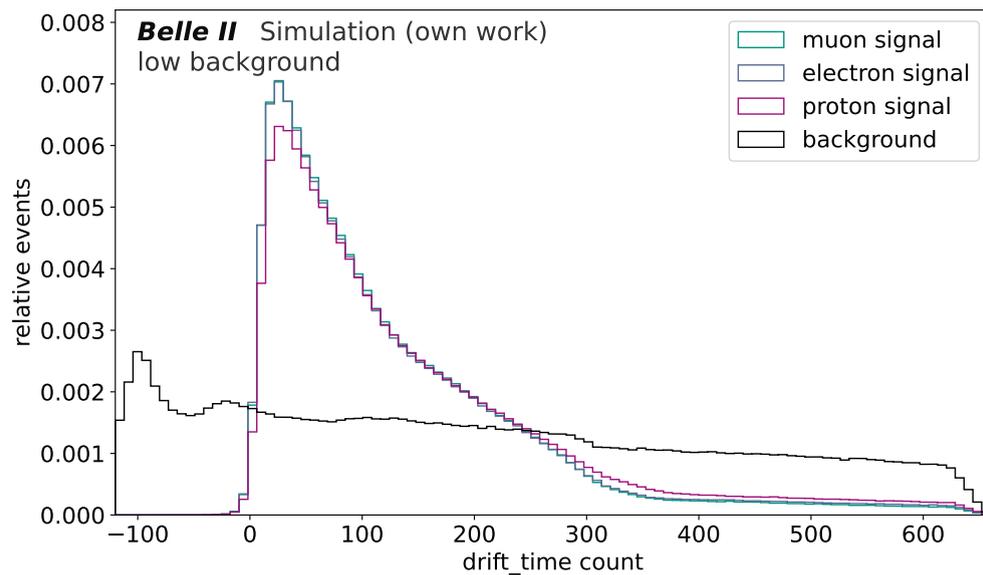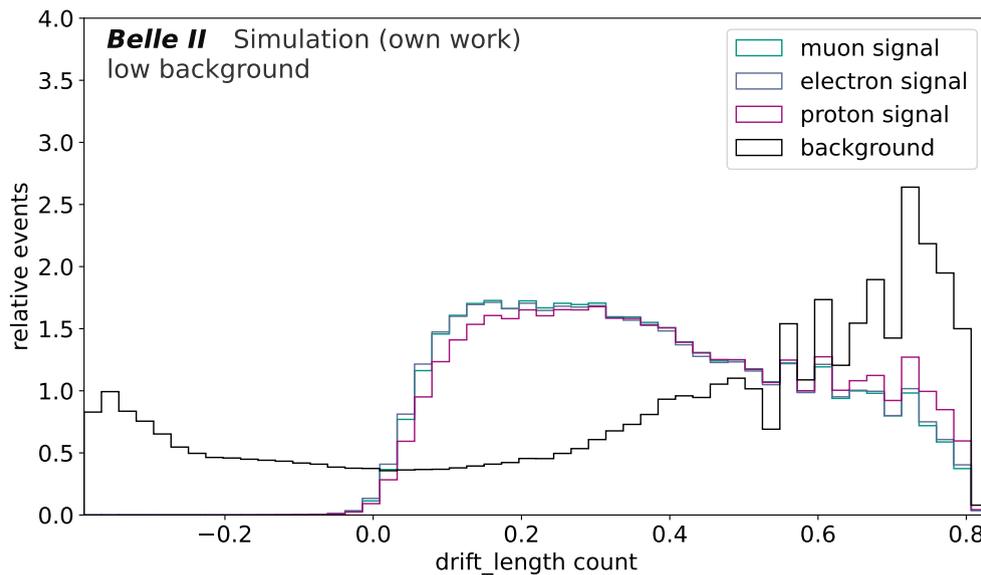

Figure A.5.: Distribution of the drift length for muons, electrons, and protons compared with the low background from a simulated data set.
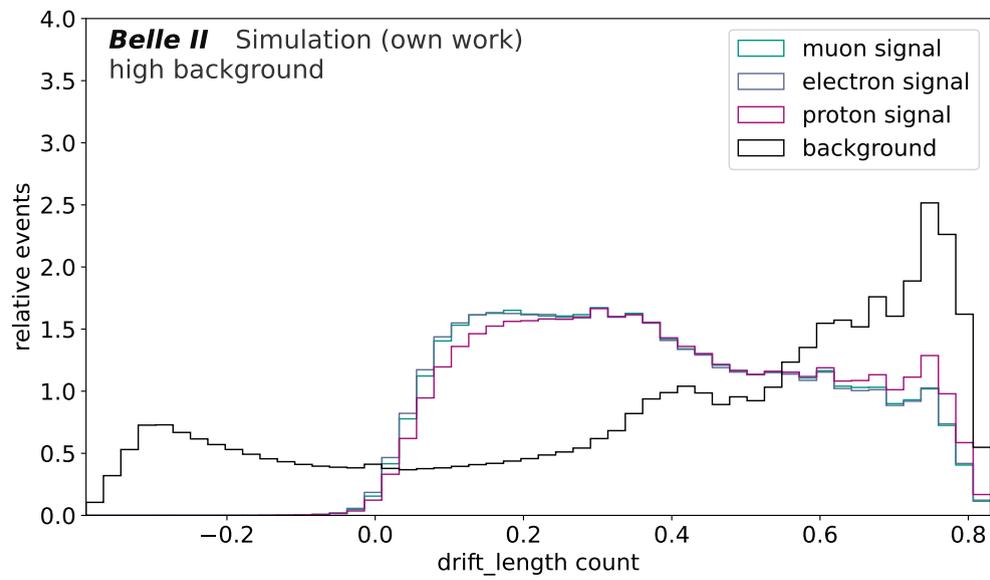
Figure A.6.: Distribution of the drift length for muons, electrons, and protons compared
with the high background from a simulated data set.

A.2. Correlation between the Input Features

Figure A.7.: Correlation between the ADC count and TDC count for a muon signal at the low background from a simulated data set.

Figure A.8.: Correlation between the ADC count and TDC count for an electron signal at the low background from a simulated data set.

Figure A.9.: Correlation between the ADC count and TDC count for a proton signal at the low background from a simulated data set.

Figure A.10.: Correlation between the ADC count and TDC count for the low background from a simulated data set.

Figure A.11.: Correlation between the ADC count and TDC count for a muon signal at the high background from a simulated data set.
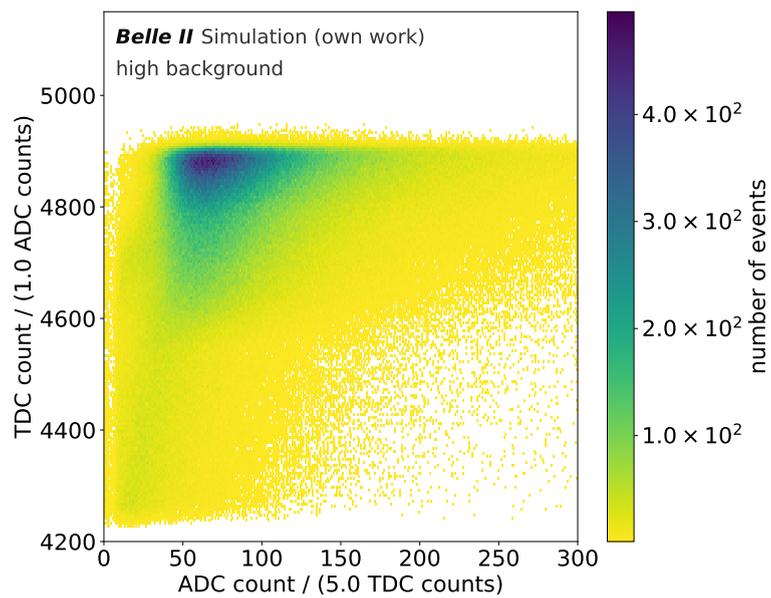
Figure A.12.: Correlation between the ADC count and TDC count for an electron signal at the high background from a simulated data set.
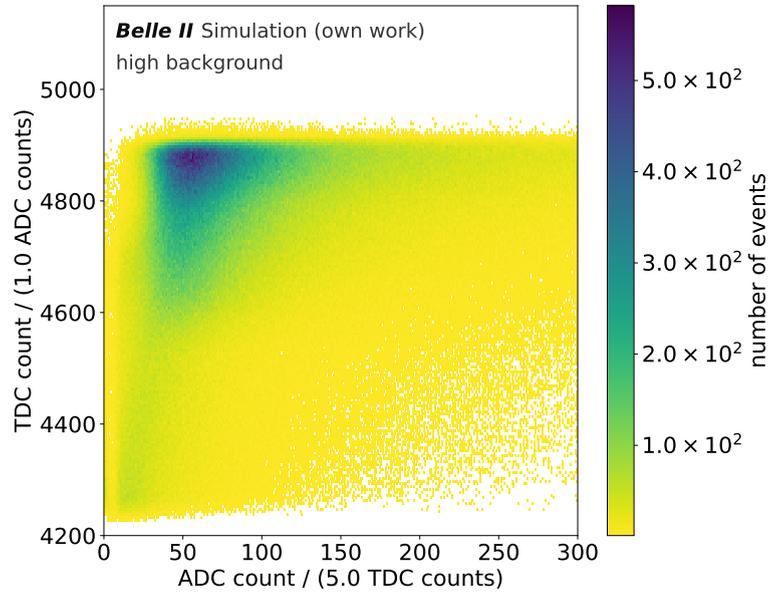
Figure A.13.: Correlation between the ADC count and TDC count for a proton signal at the high background from a simulated data set.
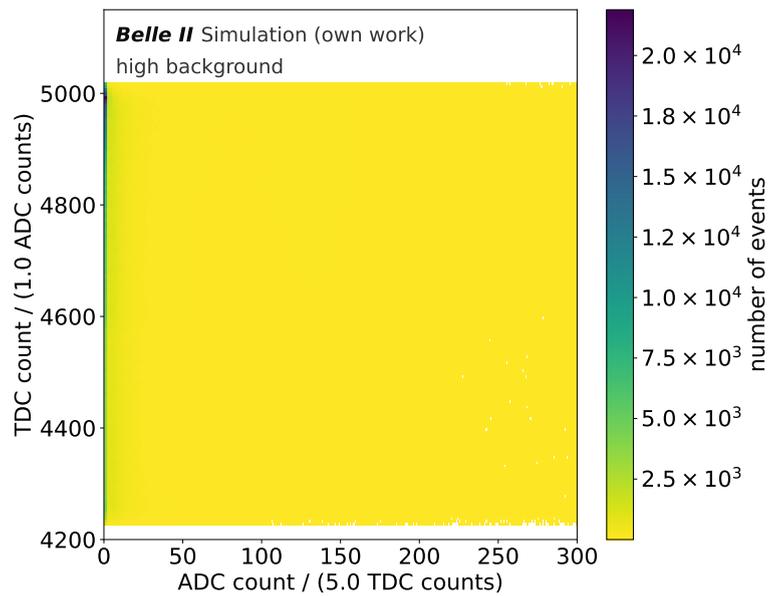
Figure A.14.: Correlation between the ADC count and TDC count for the high background from a simulated data set.
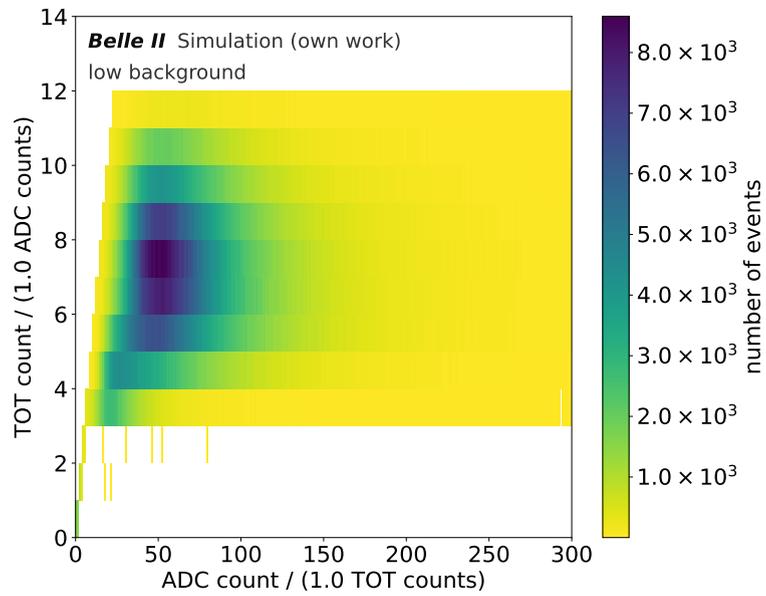
Figure A.15.: Correlation between the ADC count and TOT count for a muon signal at the low background from a simulated data set.
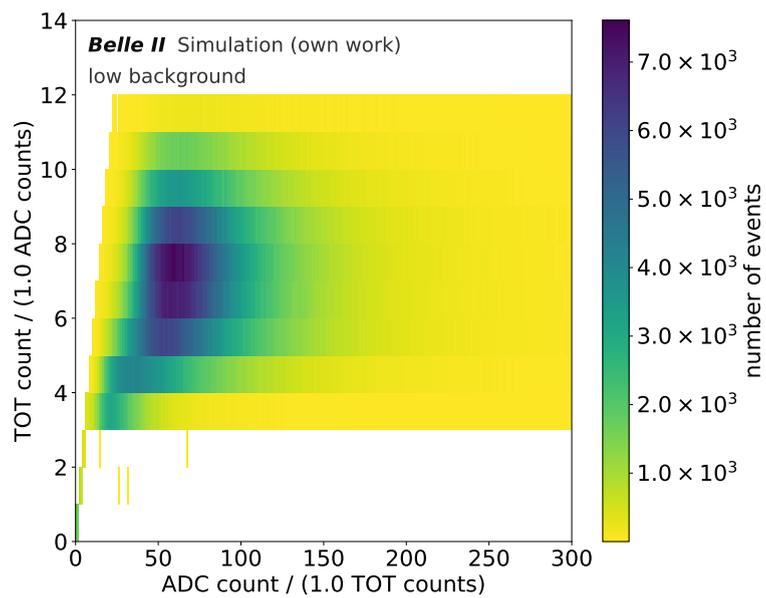
Figure A.16.: Correlation between the ADC count and TOT count for an electron signal at the low background from a simulated data set.
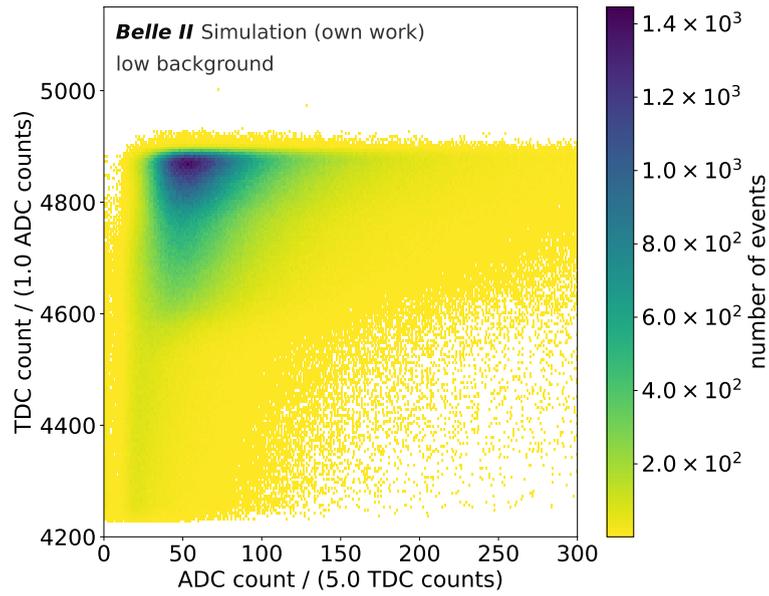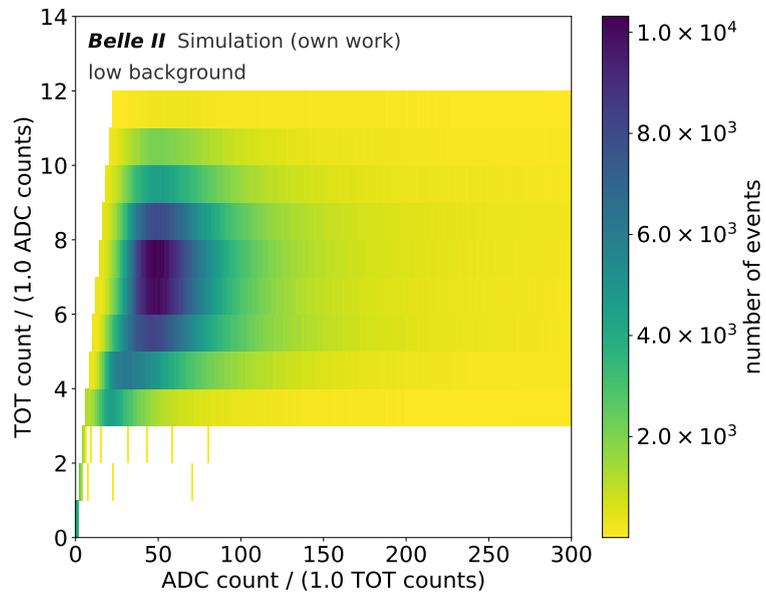
Figure A.17.: Correlation between the ADC count and TOT count for a proton signal at the low background from a simulated data set.
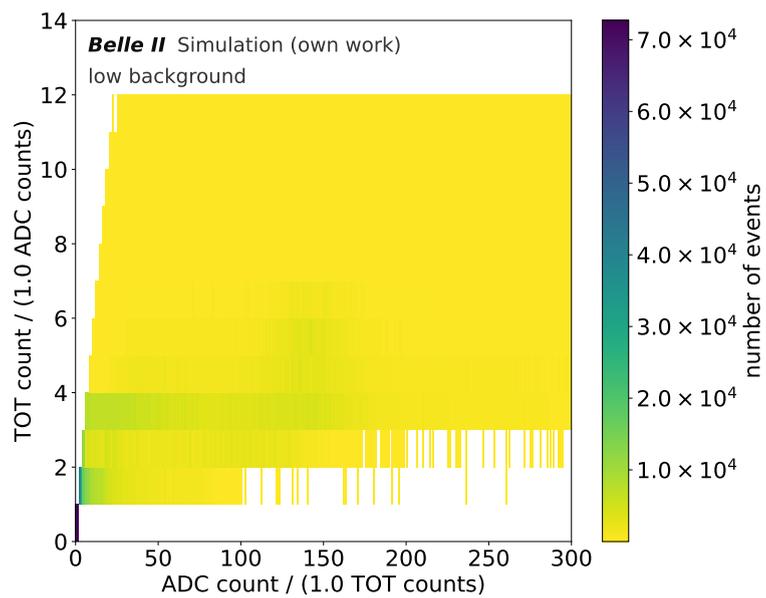
Figure A.18.: Correlation between the ADC count and TOT count for the low background from a simulated data set.
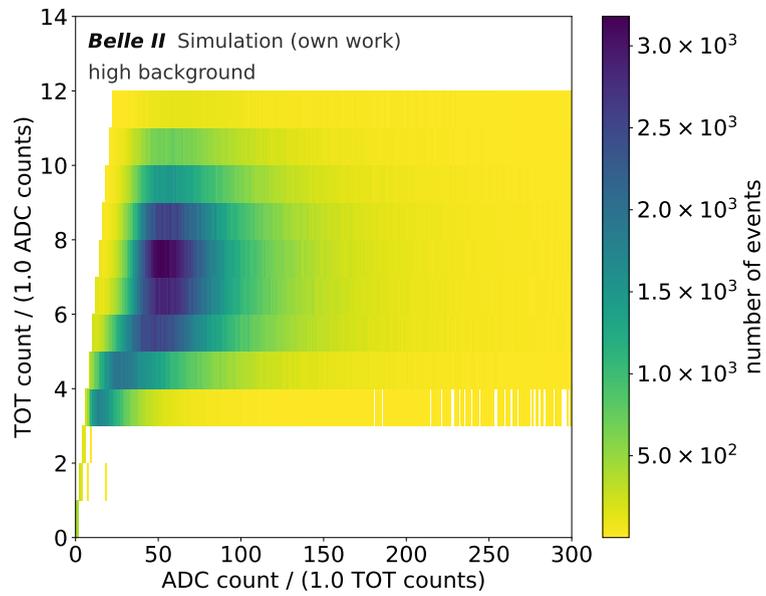
Figure A.19.: Correlation between the ADC count and TOT count for a muon signal at the high background from a simulated data set.
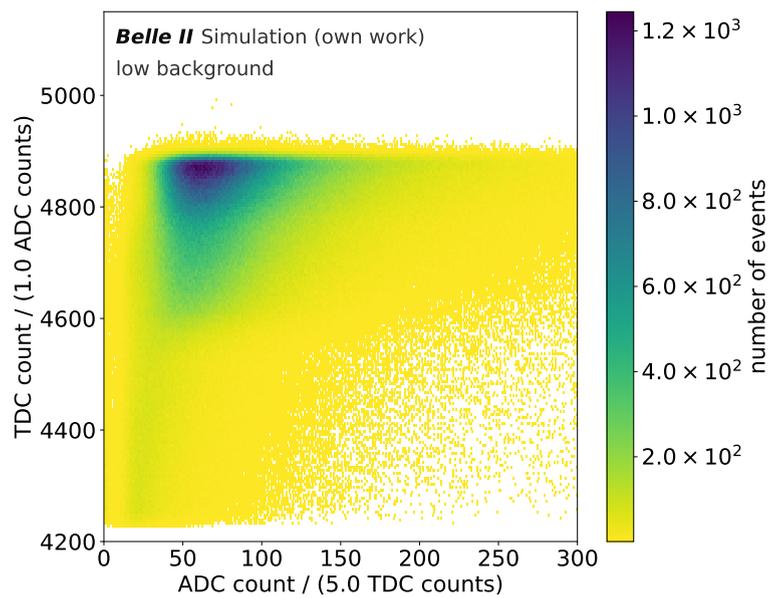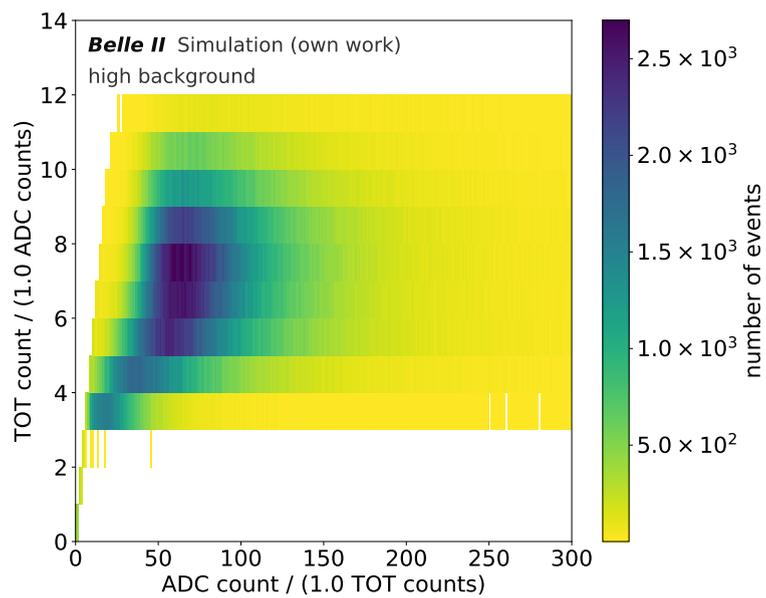

Figure A.20.: Correlation between the ADC count and TOT count for an electron signal at the high background from a simulated data set.
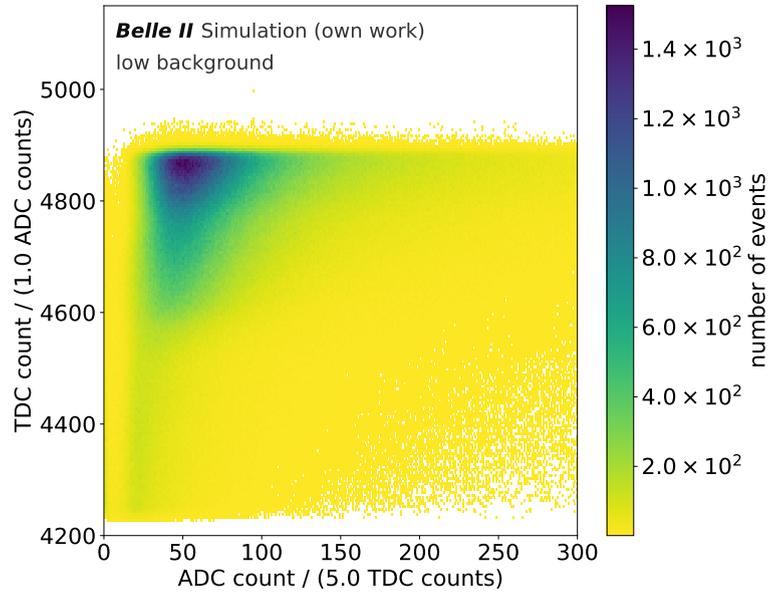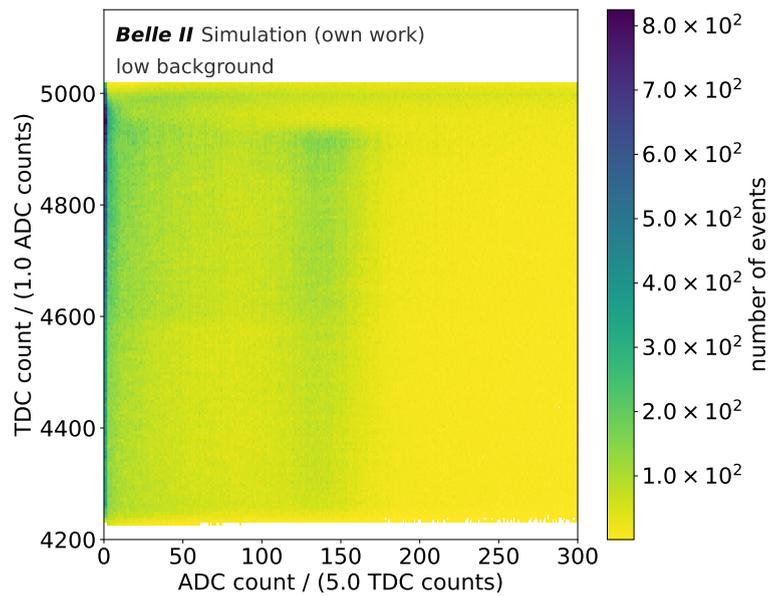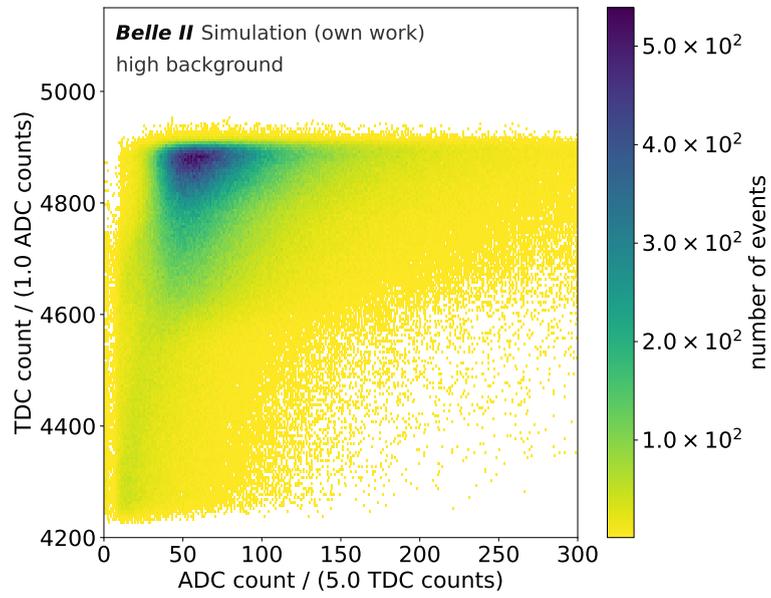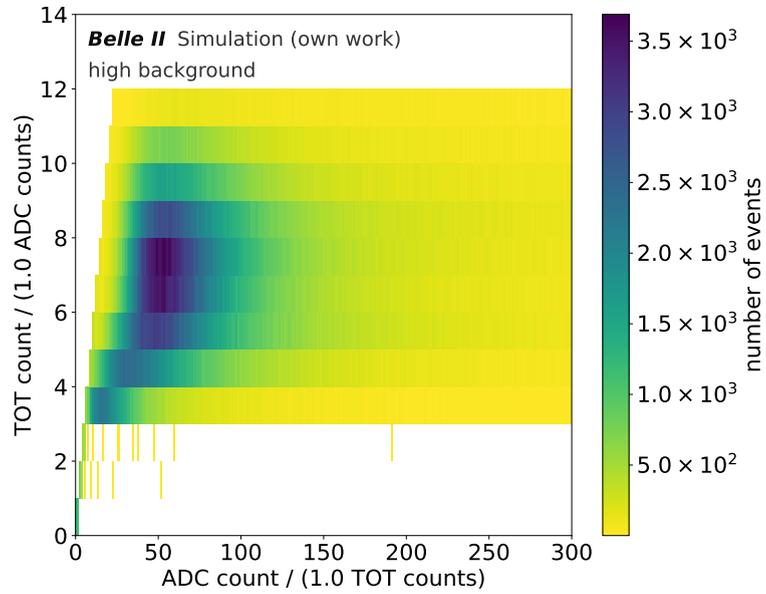
Figure A.21.: Correlation between the ADC count and TOT count for a proton signal at the high background from a simulated data set.
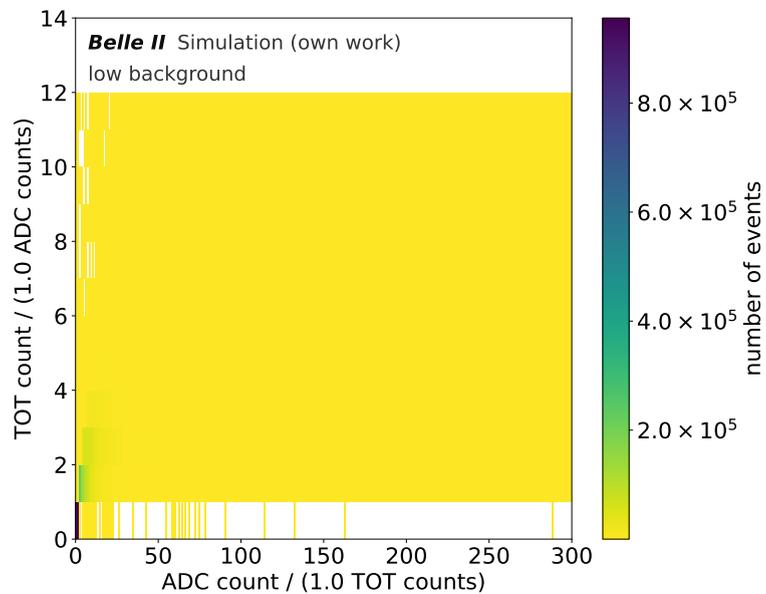
Figure A.22.: Correlation between the ADC count and TOT count for the high background from a simulated data set.

## A.3. Areas of the Excess of Signal and Background Hits
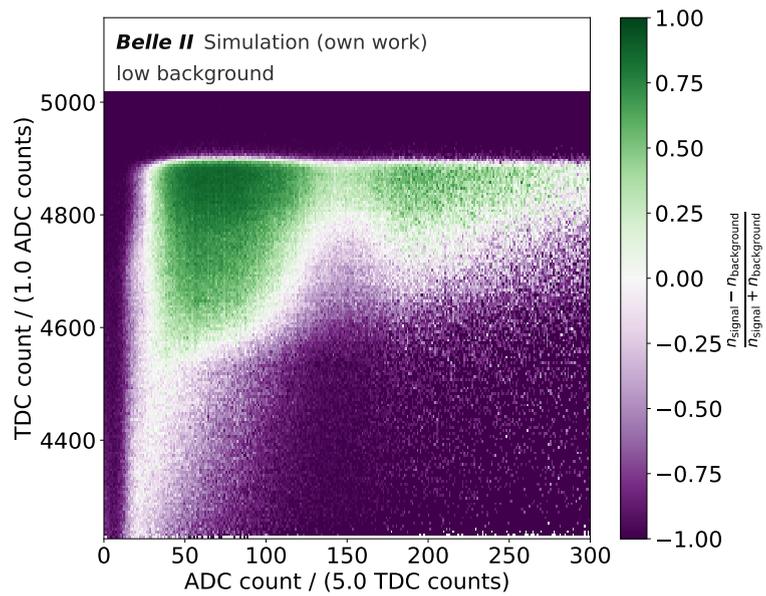
Figure A.23.: Comparison between the signal hits from an electron and low background hits from the simulated data set in a two-dimensional histogram with the TDC count over the ADC count.
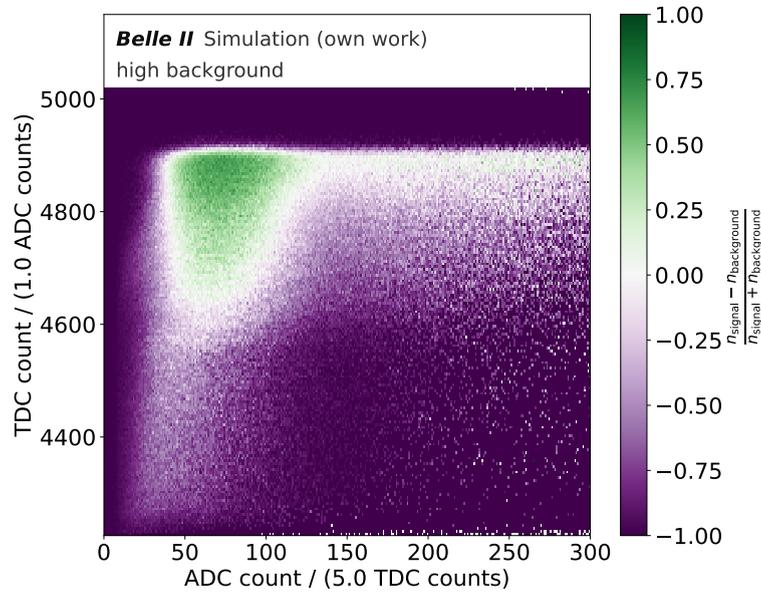
Figure A.24.: Comparison between the signal hits from an electron and high background hits from the simulated data set in a two-dimensional histogram with the TDC count over the ADC count.
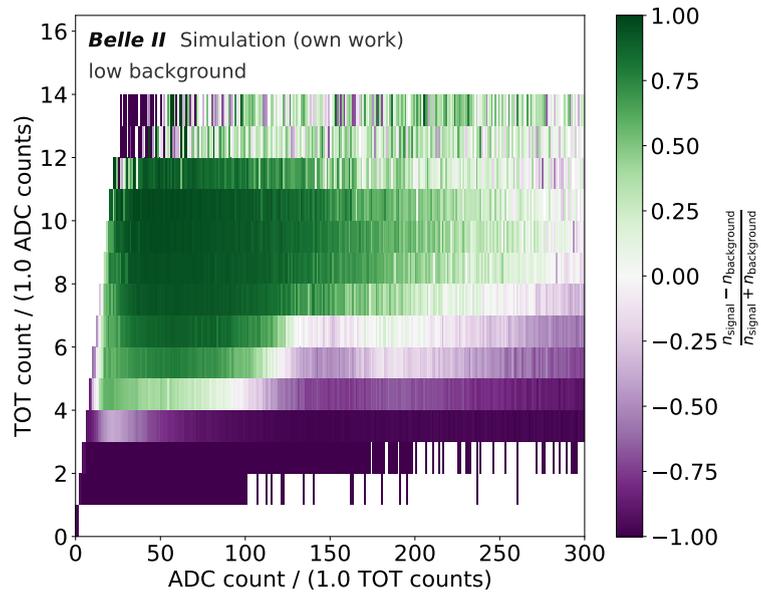
Figure A.25.: Comparison between the signal hits from an electron and low background hits from the simulated data set in a two-dimensional histogram with the TOT count over the ADC count.
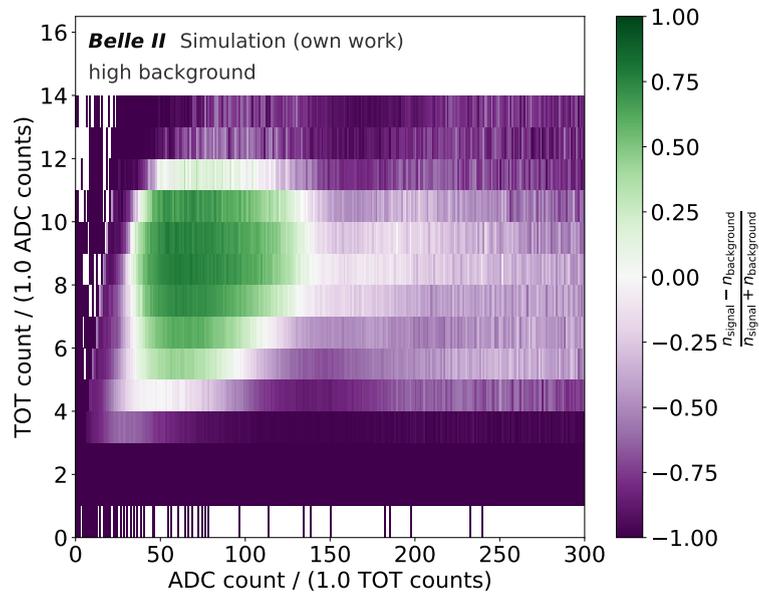
Figure A.26.: Comparison between the signal hits from an electron and high background hits from the simulated data set in a two-dimensional histogram with the TOT count over the ADC count.
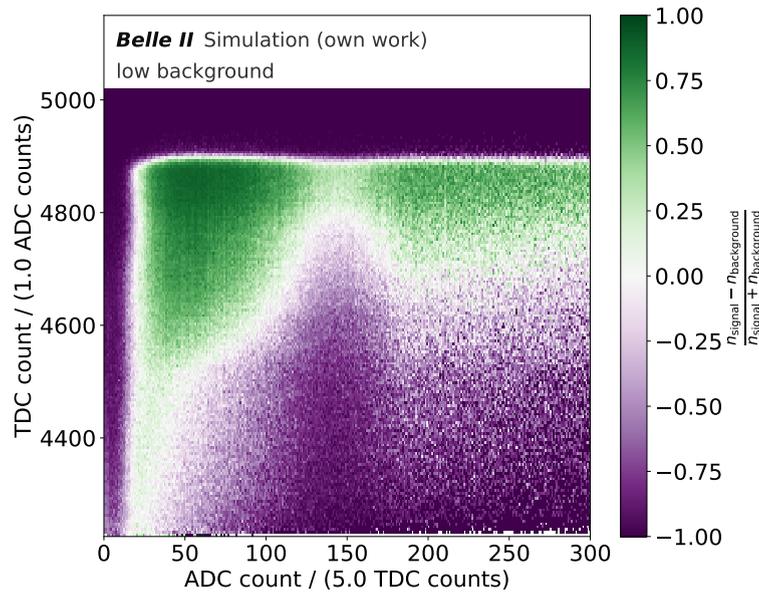
Figure A.27.: Comparison between the signal hits from a proton and low background hits from the simulated data set in a two-dimensional histogram with the TDC count over the ADC count.
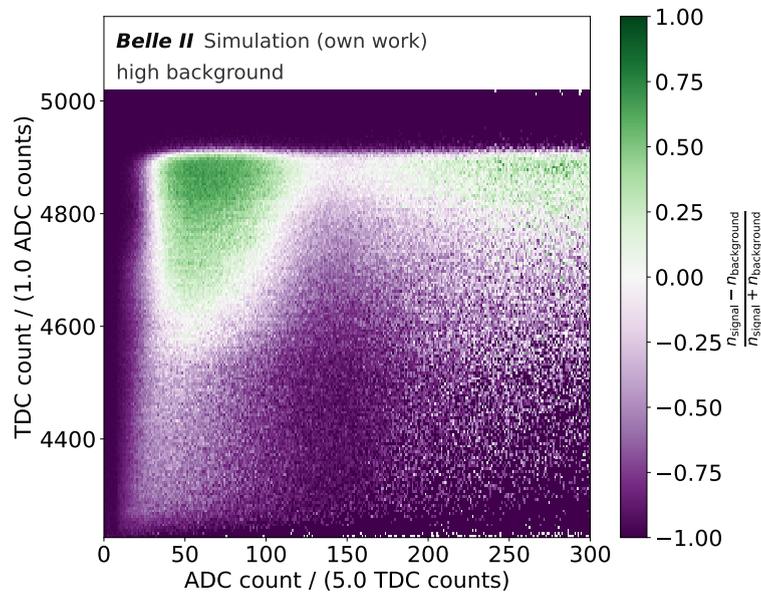
Figure A.28.: Comparison between the signal hits from a proton and high background hits from the simulated data set in a two-dimensional histogram with the TDC count over the ADC count.
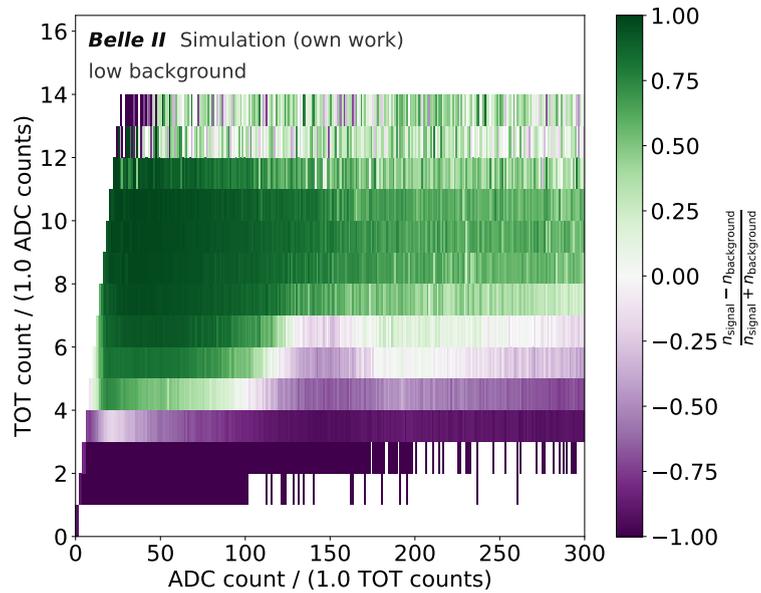
Figure A.29.: Comparison between the signal hits from a proton and low background hits from the simulated data set in a two-dimensional histogram with the TOT count over the ADC count.
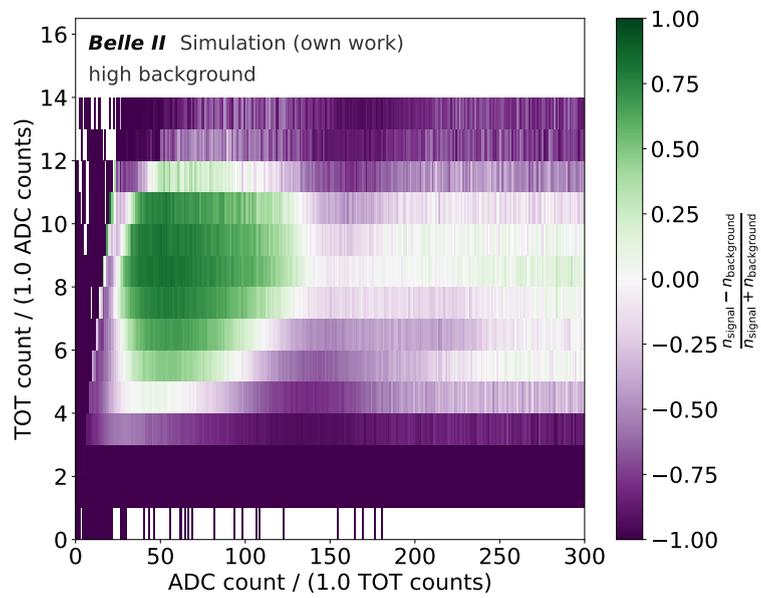
Figure A.30.: Comparison between the signal hits from a proton and high background hits from the simulated data set in a two-dimensional histogram with the TOT count over the ADC count.

A.4. Removed Signal and Background with the Input Feature Criteria

Table A.1.: Ratio of removed signal ϵ_{signal} and background hits $\epsilon_{\text{background}}$ for different feature criteria with lower limits for the `charge_deposit` feature for a muon, electron, and proton signal at the low background.

feature criteria	ϵ_{muon} (%)	$\epsilon_{\text{electron}}$ (%)	ϵ_{proton} (%)	$\epsilon_{\text{background}}$ (%)
$< 5 \times 10^{-8}$ `charge_deposit` counts	0.085	0.091	0.160	3.185
$< 1 \times 10^{-7}$ `charge_deposit` counts	0.101	0.111	0.192	4.006
$< 2 \times 10^{-7}$ `charge_deposit` counts	0.140	0.156	0.287	6.272
$< 3 \times 10^{-7}$ `charge_deposit` counts	0.195	0.212	0.398	8.194
$< 4 \times 10^{-7}$ `charge_deposit` counts	0.411	0.415	0.691	10.387
$< 5 \times 10^{-7}$ `charge_deposit` counts	0.872	0.800	1.187	11.940

Table A.2.: Ratio of removed signal ϵ_{signal} and background hits $\epsilon_{\text{background}}$ for different feature criteria with lower limits for the `charge_deposit` feature for a muon, electron, and proton signal at the high background.

feature criteria	ϵ_{muon} (%)	$\epsilon_{\text{electron}}$ (%)	ϵ_{proton} (%)	$\epsilon_{\text{background}}$ (%)
$< 5 \times 10^{-8}$ `charge_deposit` counts	0.058	0.062	0.121	20.200
$< 1 \times 10^{-7}$ `charge_deposit` counts	0.070	0.078	0.150	22.985
$< 2 \times 10^{-7}$ `charge_deposit` counts	0.116	0.127	0.252	31.387
$< 3 \times 10^{-7}$ `charge_deposit` counts	0.383	0.378	0.514	37.885
$< 4 \times 10^{-7}$ `charge_deposit` counts	1.301	1.073	1.287	43.655
$< 5 \times 10^{-7}$ `charge_deposit` counts	2.190	1.706	2.034	46.662

Table A.3.: Ratio of removed signal ϵ_{signal} and background hits $\epsilon_{\text{background}}$ for different feature criteria with lower limits for the `adc` and `tot` features for a muon, electron, and proton signal at the low background.

feature criteria	ϵ_{muon} (%)	$\epsilon_{\text{electron}}$ (%)	ϵ_{proton} (%)	$\epsilon_{\text{background}}$ (%)
$\geq$ 6 ADC and $\geq$ 1 TOT counts	0.057	0.060	0.100	2.052
$\geq$ 8 ADC and $\geq$ 2 TOT counts	0.101	0.111	0.192	5.627
$\geq$ 8 ADC and $\geq$ 1 TOT counts	0.057	0.060	0.100	2.052
$\geq$ 10 ADC and $\geq$ 2 TOT counts	0.101	0.111	0.192	6.207
$\geq$ 10 ADC and $\geq$ 3 TOT counts	0.125	0.141	0.255	7.321
$\geq$ 10 ADC and $\geq$ 4 TOT counts	0.191	0.208	0.380	8.168
$\geq$ 15 ADC and $\geq$ 2 TOT counts	0.101	0.111	0.192	7.335
$\geq$ 15 ADC and $\geq$ 3 TOT counts	0.125	0.141	0.255	8.958
$\geq$ 15 ADC and $\geq$ 4 TOT counts	0.386	0.410	0.642	10.792
$\geq$ 20 ADC and $\geq$ 3 TOT counts	0.125	0.141	0.255	10.417
$\geq$ 20 ADC and $\geq$ 4 TOT counts	0.807	0.861	1.129	13.224

Table A.4.: Ratio of removed signal ϵ_{signal} and background hits $\epsilon_{\text{background}}$ for different feature criteria with lower limits for the `adc` and `tot` features for a muon, electron, and proton signal at the high background.

feature criteria	ϵ_{muon} (%)	$\epsilon_{\text{electron}}$ (%)	ϵ_{proton} (%)	$\epsilon_{\text{background}}$ (%)
$\geq$ 6 ADC and $\geq$ 1 TOT counts	0.036	0.037	0.073	16.359
$\geq$ 8 ADC and $\geq$ 2 TOT counts	0.070	0.078	0.151	29.762
$\geq$ 8 ADC and $\geq$ 1 TOT counts	0.036	0.037	0.073	16.360
$\geq$ 10 ADC and $\geq$ 2 TOT counts	0.070	0.078	0.151	31.483
$\geq$ 10 ADC and $\geq$ 3 TOT counts	0.096	0.104	0.215	36.894
$\geq$ 10 ADC and $\geq$ 4 TOT counts	0.328	0.330	0.457	37.747
$\geq$ 15 ADC and $\geq$ 2 TOT counts	0.070	0.078	0.151	33.634
$\geq$ 15 ADC and $\geq$ 3 TOT counts	0.096	0.104	0.215	41.880
$\geq$ 15 ADC and $\geq$ 4 TOT counts	1.044	0.953	1.091	44.045
$\geq$ 20 ADC and $\geq$ 3 TOT counts	0.096	0.104	0.215	44.486
$\geq$ 20 ADC and $\geq$ 4 TOT counts	1.776	1.633	1.782	47.568

B. Appendix: Evaluation of the Configurations

B.1. Total ranking of the Feature Configurations

Table B.1.: Total ranking of all CAT finder models at the low background for muons. Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\mathfrak{r}_{\text{wrong ch.}}$
adc-tdc-tot CAT finder	$96.29^{+0.05}_{-0.05}$	$97.62^{+0.04}_{-0.04}$	$1.44^{+0.03}_{-0.03}$	$0.19^{+0.01}_{-0.01}$	$1.36^{+0.03}_{-0.03}$
angle CAT finder	$96.15^{+0.05}_{-0.05}$	$97.46^{+0.04}_{-0.04}$	$1.55^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$1.34^{+0.03}_{-0.03}$
drift-charge CAT finder	$96.08^{+0.09}_{-0.09}$	$97.69^{+0.07}_{-0.07}$	$1.75^{+0.06}_{-0.06}$	$0.22^{+0.02}_{-0.02}$	$1.65^{+0.06}_{-0.06}$
tot clip CAT finder	$96.03^{+0.06}_{-0.05}$	$97.54^{+0.04}_{-0.04}$	$1.61^{+0.03}_{-0.03}$	$0.24^{+0.01}_{-0.01}$	$1.55^{+0.03}_{-0.04}$
adc log CAT finder	$96.01^{+0.06}_{-0.05}$	$97.37^{+0.05}_{-0.04}$	$1.61^{+0.03}_{-0.03}$	$0.16^{+0.01}_{-0.01}$	$1.4^{+0.03}_{-0.03}$
charge CAT finder	$96.01^{+0.06}_{-0.05}$	$97.37^{+0.05}_{-0.04}$	$1.59^{+0.03}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$1.40^{+0.03}_{-0.03}$
adc8 CAT finder	$95.99^{+0.06}_{-0.05}$	$97.27^{+0.05}_{-0.05}$	$1.45^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$1.32^{+0.03}_{-0.03}$
adc15 CAT finder	$95.99^{+0.06}_{-0.06}$	$97.29^{+0.05}_{-0.05}$	$1.57^{+0.04}_{-0.04}$	$0.17^{+0.01}_{-0.01}$	$1.34^{+0.04}_{-0.04}$
tot3 CAT finder	$95.95^{+0.06}_{-0.06}$	$97.41^{+0.04}_{-0.04}$	$1.37^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$1.5^{+0.03}_{-0.03}$
adc6 CAT finder	$95.93^{+0.06}_{-0.06}$	$97.3^{+0.05}_{-0.05}$	$1.53^{+0.03}_{-0.03}$	$0.14^{+0.01}_{-0.01}$	$1.41^{+0.03}_{-0.03}$
tot2 CAT finder	$95.86^{+0.06}_{-0.06}$	$97.37^{+0.05}_{-0.04}$	$1.68^{+0.04}_{-0.04}$	$0.18^{+0.01}_{-0.01}$	$1.56^{+0.04}_{-0.04}$
no clayer CAT finder	$95.35^{+0.06}_{-0.06}$	$96.83^{+0.05}_{-0.05}$	$1.68^{+0.04}_{-0.04}$	$0.16^{+0.01}_{-0.01}$	$1.53^{+0.03}_{-0.04}$
adc10 CAT finder	$95.29^{+0.07}_{-0.07}$	$96.73^{+0.06}_{-0.06}$	$1.49^{+0.04}_{-0.04}$	$0.16^{+0.01}_{-0.01}$	$1.49^{+0.04}_{-0.04}$
tot1 CAT finder	$95.29^{+0.06}_{-0.06}$	$96.4^{+0.05}_{-0.05}$	$1.63^{+0.04}_{-0.04}$	$0.11^{+0.01}_{-0.01}$	$1.16^{+0.03}_{-0.03}$
no tot CAT finder	$95.27^{+0.06}_{-0.06}$	$96.77^{+0.05}_{-0.05}$	$1.72^{+0.04}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.55^{+0.04}_{-0.04}$
no layer CAT finder	$95.24^{+0.06}_{-0.06}$	$96.86^{+0.05}_{-0.05}$	$1.65^{+0.03}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.66^{+0.04}_{-0.04}$
drift CAT finder	$95.05^{+0.06}_{-0.06}$	$96.47^{+0.05}_{-0.05}$	$1.71^{+0.04}_{-0.04}$	$0.12^{+0.01}_{-0.01}$	$1.47^{+0.03}_{-0.03}$
no tdc CAT finder	$94.64^{+0.06}_{-0.06}$	$95.94^{+0.06}_{-0.06}$	$3.63^{+0.05}_{-0.05}$	$0.14^{+0.01}_{-0.01}$	$1.36^{+0.03}_{-0.03}$
middle-of-wire CAT finder	$93.92^{+0.07}_{-0.07}$	$95.19^{+0.06}_{-0.06}$	$2.03^{+0.04}_{-0.04}$	$0.08^{+0.01}_{-0.01}$	$1.33^{+0.03}_{-0.03}$
angle CAT fitter	$93.26^{+0.07}_{-0.07}$	$93.78^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.55^{+0.02}_{-0.02}$
adc-tdc-tot CAT fitter	$93.2^{+0.07}_{-0.07}$	$93.62^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.44^{+0.02}_{-0.02}$
adc15 CAT fitter	$93.16^{+0.08}_{-0.08}$	$93.6^{+0.08}_{-0.08}$	$0.5^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.46^{+0.02}_{-0.02}$
adc log CAT fitter	$93.02^{+0.07}_{-0.07}$	$93.53^{+0.07}_{-0.07}$	$0.58^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.56^{+0.02}_{-0.02}$
tot3 CAT fitter	$92.94^{+0.07}_{-0.07}$	$93.36^{+0.07}_{-0.07}$	$0.5^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.44^{+0.02}_{-0.02}$
adc6 CAT fitter	$92.89^{+0.07}_{-0.07}$	$93.34^{+0.07}_{-0.07}$	$0.49^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.47^{+0.02}_{-0.02}$
charge CAT fitter	$92.87^{+0.07}_{-0.07}$	$93.3^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.46^{+0.02}_{-0.02}$
tot2 CAT fitter	$92.69^{+0.07}_{-0.07}$	$93.13^{+0.07}_{-0.07}$	$0.57^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.47^{+0.02}_{-0.02}$
tot clip CAT fitter	$92.6^{+0.07}_{-0.07}$	$93.11^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.11^{+0.01}_{-0.01}$	$0.54^{+0.02}_{-0.02}$
adc8 CAT fitter	$92.5^{+0.07}_{-0.07}$	$92.88^{+0.07}_{-0.07}$	$0.54^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.42^{+0.02}_{-0.02}$
drift-charge CAT fitter	$92.3^{+0.1}_{-0.1}$	$92.7^{+0.1}_{-0.1}$	$0.49^{+0.03}_{-0.03}$	$0.07^{+0.01}_{-0.01}$	$0.5^{+0.03}_{-0.03}$
adc10 CAT fitter	$92.09^{+0.09}_{-0.09}$	$92.57^{+0.09}_{-0.09}$	$0.57^{+0.02}_{-0.03}$	$0.07^{+0.01}_{-0.01}$	$0.52^{+0.02}_{-0.02}$
no tot CAT fitter	$92.08^{+0.08}_{-0.08}$	$92.65^{+0.07}_{-0.07}$	$0.51^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.61^{+0.02}_{-0.02}$
no clayer CAT fitter	$92.07^{+0.08}_{-0.08}$	$92.44^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.04^{+0.01}_{-0.01}$	$0.41^{+0.02}_{-0.02}$
tot1 CAT fitter	$92.01^{+0.08}_{-0.08}$	$92.42^{+0.07}_{-0.07}$	$0.5^{+0.02}_{-0.02}$	0.02	$0.44^{+0.02}_{-0.02}$
no layer CAT fitter	$91.97^{+0.08}_{-0.08}$	$92.4^{+0.07}_{-0.07}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.47^{+0.02}_{-0.02}$
drift CAT fitter	$91.91^{+0.08}_{-0.08}$	$92.33^{+0.08}_{-0.07}$	$0.56^{+0.02}_{-0.02}$	0.02	$0.46^{+0.02}_{-0.02}$
no tdc CAT fitter	$91.46^{+0.08}_{-0.08}$	$92.06^{+0.08}_{-0.08}$	$1.24^{+0.03}_{-0.03}$	$0.05^{+0.01}_{-0.01}$	$0.65^{+0.02}_{-0.02}$
middle-of-wire CAT fitter	$91.02^{+0.08}_{-0.08}$	$91.47^{+0.08}_{-0.08}$	$0.81^{+0.03}_{-0.03}$	0.02	$0.49^{+0.02}_{-0.02}$

Table B.2.: Total ranking of all CAT finder models at the low background for electrons. Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\mathfrak{r}_{\text{wrong ch.}}$
adc-tdc-tot CAT finder	$85.59^{+0.1}_{-0.1}$	$87.29^{+0.09}_{-0.09}$	$1.42^{+0.03}_{-0.03}$	$0.25^{+0.01}_{-0.01}$	$1.94^{+0.04}_{-0.04}$
adc6 CAT finder	$85.3^{+0.1}_{-0.1}$	$87.03^{+0.1}_{-0.1}$	$1.57^{+0.03}_{-0.03}$	$0.16^{+0.01}_{-0.01}$	$1.97^{+0.04}_{-0.04}$
adc8 CAT finder	$85.4^{+0.1}_{-0.1}$	$87.06^{+0.1}_{-0.1}$	$1.49^{+0.03}_{-0.03}$	$0.18^{+0.01}_{-0.01}$	$1.92^{+0.04}_{-0.04}$
adc10 CAT finder	$84.9^{+0.1}_{-0.1}$	$86.7^{+0.1}_{-0.1}$	$1.7^{+0.04}_{-0.04}$	$0.21^{+0.01}_{-0.01}$	$2.03^{+0.05}_{-0.05}$
adc15 CAT finder	$85.6^{+0.1}_{-0.1}$	$87.2^{+0.1}_{-0.1}$	$1.55^{+0.04}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.84^{+0.05}_{-0.05}$
tot clip CAT finder	$85.4^{+0.1}_{-0.1}$	$87.34^{+0.09}_{-0.09}$	$1.62^{+0.03}_{-0.03}$	$0.24^{+0.01}_{-0.01}$	$2.2^{+0.04}_{-0.04}$
adc log CAT finder	$85.4^{+0.1}_{-0.1}$	$87.16^{+0.1}_{-0.09}$	$1.63^{+0.03}_{-0.04}$	$0.18^{+0.01}_{-0.01}$	$1.98^{+0.04}_{-0.04}$
tot3 CAT finder	$85.3^{+0.1}_{-0.1}$	$87.07^{+0.1}_{-0.1}$	$1.41^{+0.03}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$2.07^{+0.04}_{-0.04}$
charge CAT finder	$85.3^{+0.1}_{-0.1}$	$87.05^{+0.1}_{-0.1}$	$1.59^{+0.03}_{-0.03}$	$0.17^{+0.01}_{-0.01}$	$1.96^{+0.04}_{-0.04}$
tot2 CAT finder	$85.2^{+0.1}_{-0.1}$	$87.13^{+0.1}_{-0.1}$	$1.7^{+0.04}_{-0.04}$	$0.17^{+0.01}_{-0.01}$	$2.16^{+0.04}_{-0.04}$
drift-charge CAT finder	$84.9^{+0.2}_{-0.2}$	$86.9^{+0.2}_{-0.2}$	$1.78^{+0.06}_{-0.06}$	$0.21^{+0.02}_{-0.02}$	$2.29^{+0.07}_{-0.07}$
tot1 CAT finder	$84.8^{+0.1}_{-0.1}$	$86.27^{+0.1}_{-0.1}$	$1.6^{+0.03}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$1.73^{+0.04}_{-0.04}$
no tot CAT finder	$84.8^{+0.1}_{-0.1}$	$86.5^{+0.1}_{-0.1}$	$1.61^{+0.04}_{-0.04}$	$0.19^{+0.01}_{-0.01}$	$1.92^{+0.04}_{-0.04}$
drift CAT finder	$84.5^{+0.1}_{-0.1}$	$86.32^{+0.1}_{-0.1}$	$1.74^{+0.04}_{-0.04}$	$0.13^{+0.01}_{-0.01}$	$2.05^{+0.04}_{-0.04}$
no tdc CAT finder	$84.3^{+0.1}_{-0.1}$	$85.97^{+0.1}_{-0.1}$	$3.47^{+0.05}_{-0.05}$	$0.17^{+0.01}_{-0.01}$	$1.95^{+0.04}_{-0.04}$
middle-of-wire CAT finder	$83.7^{+0.1}_{-0.1}$	$85.3^{+0.1}_{-0.1}$	$2.09^{+0.04}_{-0.04}$	$0.15^{+0.01}_{-0.01}$	$1.88^{+0.04}_{-0.04}$
angle CAT fitter	$82.2^{+0.1}_{-0.1}$	$82.7^{+0.1}_{-0.1}$	$0.51^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.6^{+0.02}_{-0.02}$
adc15 CAT fitter	$82.0^{+0.1}_{-0.1}$	$82.5^{+0.1}_{-0.1}$	$0.52^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.61^{+0.03}_{-0.03}$
adc-tdc-tot CAT fitter	$82.0^{+0.1}_{-0.1}$	$82.4^{+0.1}_{-0.1}$	$0.51^{+0.02}_{-0.02}$	$0.09^{+0.01}_{-0.01}$	$0.51^{+0.02}_{-0.02}$
adc log CAT fitter	$81.9^{+0.1}_{-0.1}$	$82.4^{+0.1}_{-0.1}$	$0.55^{+0.02}_{-0.02}$	$0.08^{+0.01}_{-0.01}$	$0.63^{+0.02}_{-0.03}$
tot3 CAT fitter	$81.7^{+0.1}_{-0.1}$	$82.2^{+0.1}_{-0.1}$	$0.46^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.57^{+0.02}_{-0.02}$
tot2 CAT fitter	$81.7^{+0.1}_{-0.1}$	$82.1^{+0.1}_{-0.1}$	$0.58^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.5^{+0.02}_{-0.02}$
charge CAT fitter	$81.7^{+0.1}_{-0.1}$	$82.1^{+0.1}_{-0.1}$	$0.54^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$0.55^{+0.02}_{-0.02}$
adc6 CAT fitter	$81.6^{+0.1}_{-0.1}$	$82.1^{+0.1}_{-0.1}$	$0.51^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.6^{+0.02}_{-0.02}$
tot clip CAT fitter	$81.5^{+0.1}_{-0.1}$	$82.0^{+0.1}_{-0.1}$	$0.53^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.61^{+0.02}_{-0.02}$
no tot CAT fitter	$81.3^{+0.1}_{-0.1}$	$81.8^{+0.1}_{-0.1}$	$0.53^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.68^{+0.03}_{-0.03}$
adc8 CAT fitter	$81.3^{+0.1}_{-0.1}$	$81.7^{+0.1}_{-0.1}$	$0.54^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.51^{+0.02}_{-0.02}$
no clayer CAT fitter	$81.2^{+0.1}_{-0.1}$	$81.6^{+0.1}_{-0.1}$	$0.54^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.5^{+0.02}_{-0.02}$
adc10 CAT fitter	$81.1^{+0.1}_{-0.1}$	$81.6^{+0.1}_{-0.1}$	$0.59^{+0.02}_{-0.02}$	$0.08^{+0.01}_{-0.01}$	$0.6^{+0.03}_{-0.03}$
tot1 CAT fitter	$81.1^{+0.1}_{-0.1}$	$81.5^{+0.1}_{-0.1}$	$0.53^{+0.02}_{-0.02}$	0.02	$0.53^{+0.02}_{-0.02}$
no layer CAT fitter	$81.0^{+0.1}_{-0.1}$	$81.4^{+0.1}_{-0.1}$	$0.54^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$0.5^{+0.02}_{-0.02}$
drift CAT fitter	$80.9^{+0.1}_{-0.1}$	$81.3^{+0.1}_{-0.1}$	$0.56^{+0.02}_{-0.02}$	$0.03_{-0.01}$	$0.5^{+0.02}_{-0.02}$
drift-charge CAT fitter	$80.7^{+0.2}_{-0.2}$	$81.2^{+0.2}_{-0.2}$	$0.53^{+0.03}_{-0.03}$	$0.07^{+0.01}_{-0.01}$	$0.55^{+0.04}_{-0.04}$
no tdc CAT fitter	$80.7^{+0.1}_{-0.1}$	$81.2^{+0.1}_{-0.1}$	$1.17^{+0.03}_{-0.03}$	$0.06^{+0.01}_{-0.01}$	$0.65^{+0.03}_{-0.03}$
middle-of-wire CAT fitter	$80.4^{+0.1}_{-0.1}$	$80.8^{+0.1}_{-0.1}$	$0.83^{+0.03}_{-0.03}$	$0.04^{+0.01}_{-0.01}$	$0.55^{+0.02}_{-0.02}$

Table B.3.: Total ranking of all CAT finder at the low background for protons. Ranked by the charge efficiency $\varepsilon_{\mathrm{charge}}$, with the track efficiency $\varepsilon_{\mathrm{track}}$, fake rate $\mathfrak{r}_{\mathrm{fake}}$, clone rate $\mathfrak{r}_{\mathrm{clone}}$, and wrong charge rate $\mathfrak{r}_{\mathrm{wrong\ ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\mathrm{charge}}$	$\varepsilon_{\mathrm{track}}$	$\mathfrak{r}_{\mathrm{fake}}$	$\mathfrak{r}_{\mathrm{clone}}$	$\mathfrak{r}_{\mathrm{wrong\ ch.}}$
tot3 CAT finder	$93.53^{+0.08}_{-0.08}$	$94.24^{+0.08}_{-0.08}$	$3.3^{+0.05}_{-0.05}$	$0.47^{+0.02}_{-0.02}$	$0.75^{+0.03}_{-0.03}$
adc15 CAT finder	$93.41^{+0.09}_{-0.09}$	$94.22^{+0.09}_{-0.09}$	$3.3^{+0.05}_{-0.05}$	$0.47^{+0.02}_{-0.02}$	$0.86^{+0.03}_{-0.04}$
adc-tdc-tot CAT finder	$93.32^{+0.08}_{-0.08}$	$94.24^{+0.08}_{-0.08}$	$3.54^{+0.05}_{-0.05}$	$0.54^{+0.02}_{-0.02}$	$0.98^{+0.03}_{-0.03}$
adc log CAT finder	$93.31^{+0.08}_{-0.08}$	$94.28^{+0.08}_{-0.08}$	$3.78^{+0.05}_{-0.05}$	$0.54^{+0.02}_{-0.02}$	$1.02^{+0.03}_{-0.03}$
angle CAT finder	$93.26^{+0.08}_{-0.08}$	$94.25^{+0.08}_{-0.08}$	$3.79^{+0.05}_{-0.05}$	$0.62^{+0.02}_{-0.02}$	$1.05^{+0.03}_{-0.04}$
adc6 CAT finder	$93.25^{+0.08}_{-0.08}$	$94.17^{+0.08}_{-0.08}$	$3.67^{+0.05}_{-0.05}$	$0.5^{+0.02}_{-0.02}$	$0.97^{+0.03}_{-0.03}$
charge CAT finder	$93.14^{+0.08}_{-0.08}$	$94.02^{+0.08}_{-0.08}$	$3.9^{+0.05}_{-0.05}$	$0.46^{+0.02}_{-0.02}$	$0.93^{+0.03}_{-0.03}$
tot clip CAT finder	$93.11^{+0.08}_{-0.08}$	$94.08^{+0.08}_{-0.08}$	$4.0^{+0.05}_{-0.05}$	$0.55^{+0.02}_{-0.02}$	$1.04^{+0.03}_{-0.03}$
drift-charge CAT finder	$93.1^{+0.2}_{-0.2}$	$94.1^{+0.1}_{-0.2}$	$4.7^{+0.2}_{-0.1}$	$0.63^{+0.04}_{-0.04}$	$1.02^{+0.07}_{-0.07}$
adc8 CAT finder	$93.07^{+0.08}_{-0.08}$	$94.1^{+0.08}_{-0.08}$	$3.67^{+0.05}_{-0.05}$	$0.52^{+0.02}_{-0.02}$	$1.09^{+0.04}_{-0.04}$
no tot CAT finder	$93.03^{+0.09}_{-0.09}$	$93.97^{+0.08}_{-0.08}$	$3.85^{+0.05}_{-0.05}$	$0.54^{+0.02}_{-0.02}$	$1.0^{+0.03}_{-0.04}$
adc10 CAT finder	$92.93^{+0.09}_{-0.09}$	$93.87^{+0.09}_{-0.09}$	$3.43^{+0.05}_{-0.05}$	$0.53^{+0.02}_{-0.02}$	$1.0^{+0.04}_{-0.04}$
tot2 CAT finder	$92.93^{+0.09}_{-0.08}$	$93.95^{+0.08}_{-0.08}$	$3.92^{+0.05}_{-0.05}$	$0.46^{+0.02}_{-0.02}$	$1.09^{+0.04}_{-0.04}$
no layer CAT finder	$92.9^{+0.09}_{-0.08}$	$93.79^{+0.08}_{-0.08}$	$3.76^{+0.05}_{-0.05}$	$0.47^{+0.02}_{-0.02}$	$0.95^{+0.03}_{-0.03}$
no clayer CAT finder	$92.78^{+0.09}_{-0.09}$	$93.83^{+0.08}_{-0.08}$	$3.97^{+0.05}_{-0.05}$	$0.54^{+0.02}_{-0.02}$	$1.12^{+0.04}_{-0.04}$
drift CAT finder	$92.49^{+0.09}_{-0.09}$	$93.21^{+0.08}_{-0.08}$	$4.17^{+0.05}_{-0.05}$	$0.43^{+0.02}_{-0.02}$	$0.77^{+0.03}_{-0.03}$
tot1 CAT finder	$92.09^{+0.09}_{-0.09}$	$92.9^{+0.09}_{-0.08}$	$3.91^{+0.05}_{-0.05}$	$0.44^{+0.02}_{-0.02}$	$0.88^{+0.03}_{-0.03}$
no tdc CAT finder	$91.93^{+0.09}_{-0.09}$	$92.81^{+0.09}_{-0.09}$	$5.59^{+0.06}_{-0.06}$	$0.58^{+0.02}_{-0.02}$	$0.95^{+0.03}_{-0.03}$
middle-of-wire CAT finder	$91.35^{+0.09}_{-0.09}$	$92.17^{+0.09}_{-0.09}$	$4.62^{+0.06}_{-0.06}$	$0.56^{+0.02}_{-0.02}$	$0.89^{+0.03}_{-0.03}$
adc15 CAT fitter	$91.3^{+0.1}_{-0.1}$	$92.06^{+0.1}_{-0.1}$	$1.08^{+0.03}_{-0.03}$	$0.25^{+0.01}_{-0.02}$	$0.85^{+0.03}_{-0.04}$
tot3 CAT fitter	$90.65^{+0.1}_{-0.1}$	$91.48^{+0.09}_{-0.09}$	$1.07^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$0.9^{+0.03}_{-0.03}$
angle CAT fitter	$90.5^{+0.1}_{-0.1}$	$91.43^{+0.09}_{-0.09}$	$1.22^{+0.03}_{-0.03}$	$0.3^{+0.01}_{-0.02}$	$1.02^{+0.03}_{-0.04}$
no tot CAT fitter	$90.4^{+0.1}_{-0.1}$	$91.38^{+0.1}_{-0.1}$	$1.23^{+0.03}_{-0.03}$	$0.25^{+0.01}_{-0.01}$	$1.02^{+0.04}_{-0.04}$
adc log CAT fitter	$90.34^{+0.1}_{-0.1}$	$91.27^{+0.09}_{-0.09}$	$1.24^{+0.03}_{-0.03}$	$0.26^{+0.01}_{-0.01}$	$1.02^{+0.03}_{-0.04}$
adc6 CAT fitter	$90.23^{+0.1}_{-0.1}$	$91.09^{+0.09}_{-0.09}$	$1.13^{+0.03}_{-0.03}$	$0.24^{+0.01}_{-0.01}$	$0.95^{+0.03}_{-0.03}$
adc10 CAT fitter	$90.2^{+0.1}_{-0.1}$	$91.2^{+0.1}_{-0.1}$	$1.2^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$1.04^{+0.04}_{-0.04}$
tot2 CAT fitter	$90.2^{+0.1}_{-0.1}$	$91.09^{+0.09}_{-0.09}$	$1.34^{+0.03}_{-0.03}$	$0.22^{+0.01}_{-0.01}$	$0.98^{+0.03}_{-0.03}$
adc8 CAT fitter	$90.13^{+0.1}_{-0.1}$	$91.05^{+0.09}_{-0.09}$	$1.23^{+0.03}_{-0.03}$	$0.23^{+0.01}_{-0.01}$	$1.02^{+0.03}_{-0.04}$
adc-tdc-tot CAT fitter	$90.12^{+0.1}_{-0.1}$	$91.01^{+0.1}_{-0.09}$	$1.2^{+0.03}_{-0.03}$	$0.25^{+0.01}_{-0.01}$	$0.98^{+0.03}_{-0.03}$
charge CAT fitter	$90.08^{+0.1}_{-0.1}$	$90.94^{+0.1}_{-0.09}$	$1.24^{+0.03}_{-0.03}$	$0.2^{+0.01}_{-0.01}$	$0.95^{+0.03}_{-0.03}$
tot clip CAT fitter	$90.01^{+0.1}_{-0.1}$	$90.87^{+0.1}_{-0.1}$	$1.28^{+0.03}_{-0.03}$	$0.23^{+0.01}_{-0.01}$	$0.94^{+0.03}_{-0.03}$
no clayer CAT fitter	$89.98^{+0.1}_{-0.1}$	$90.83^{+0.1}_{-0.1}$	$1.25^{+0.03}_{-0.03}$	$0.25^{+0.01}_{-0.01}$	$0.94^{+0.03}_{-0.03}$
no layer CAT fitter	$89.92^{+0.1}_{-0.1}$	$90.75^{+0.1}_{-0.1}$	$1.21^{+0.03}_{-0.03}$	$0.19^{+0.01}_{-0.01}$	$0.92^{+0.03}_{-0.03}$
tot1 CAT fitter	$89.5^{+0.1}_{-0.1}$	$90.37^{+0.1}_{-0.1}$	$1.19^{+0.03}_{-0.03}$	$0.21^{+0.01}_{-0.01}$	$0.96^{+0.03}_{-0.03}$
drift CAT fitter	$89.4^{+0.1}_{-0.1}$	$90.25^{+0.1}_{-0.1}$	$1.27^{+0.03}_{-0.03}$	$0.19^{+0.01}_{-0.01}$	$0.92^{+0.03}_{-0.03}$
drift-charge CAT fitter	$88.8^{+0.2}_{-0.2}$	$89.6^{+0.2}_{-0.2}$	$1.41^{+0.06}_{-0.07}$	$0.35^{+0.03}_{-0.03}$	$0.93^{+0.07}_{-0.07}$
no tdc CAT fitter	$88.8^{+0.1}_{-0.1}$	$89.87^{+0.1}_{-0.1}$	$1.87^{+0.04}_{-0.04}$	$0.29^{+0.01}_{-0.02}$	$1.14^{+0.04}_{-0.04}$
middle-of-wire CAT fitter	$88.6^{+0.1}_{-0.1}$	$89.4^{+0.1}_{-0.1}$	$1.58^{+0.03}_{-0.04}$	$0.24^{+0.01}_{-0.01}$	$0.9^{+0.03}_{-0.03}$

B.2. Comparison to the standard CAT and the Baseline for all Particles.

Table B.4.: Evaluation with electrons of the best configuration model in comparison to the Baseline and the standard CAT model. Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\mathfrak{r}_{\text{wrong ch.}}$
std. CAT finder	$88.45^{+0.09}_{-0.09}$	$89.21^{+0.09}_{-0.09}$	$1.0^{+0.03}_{-0.03}$	$0.41^{+0.02}_{-0.02}$	$0.85^{+0.03}_{-0.03}$
best config. CAT finder	$88.33^{+0.09}_{-0.09}$	$89.03^{+0.09}_{-0.09}$	$0.81^{+0.02}_{-0.02}$	$0.06^{+0.01}_{-0.01}$	$0.78^{+0.03}_{-0.03}$
Baseline finder	$85.59^{+0.1}_{-0.1}$	$87.91^{+0.09}_{-0.09}$	$0.76^{+0.02}_{-0.02}$	$0.07^{+0.01}_{-0.01}$	$2.65^{+0.05}_{-0.05}$
std. CAT fitter	$86.02^{+0.1}_{-0.1}$	$86.23^{+0.1}_{-0.1}$	$0.41^{+0.02}_{-0.02}$	$0.13^{+0.01}_{-0.01}$	$0.25^{+0.01}_{-0.02}$
best config. CAT fitter	$85.93^{+0.1}_{-0.1}$	$86.08^{+0.1}_{-0.1}$	$0.27^{+0.01}_{-0.01}$	0.02	$0.17^{+0.01}_{-0.01}$
Baseline fitter	$84.7^{+0.1}_{-0.1}$	$86.32^{+0.1}_{-0.1}$	$0.57^{+0.02}_{-0.02}$	$0.05^{+0.01}_{-0.01}$	$1.86^{+0.04}_{-0.04}$

Table B.5.: Evaluation with protons of the best configuration model in comparison to the Baseline and the standard CAT model. Ranked by the charge efficiency $\varepsilon_{\text{charge}}$, with the track efficiency $\varepsilon_{\text{track}}$, fake rate $\mathfrak{r}_{\text{fake}}$, clone rate $\mathfrak{r}_{\text{clone}}$, and wrong charge rate $\mathfrak{r}_{\text{wrong ch.}}$. Uncertainties below 0.01 % are not shown.

model name	$\varepsilon_{\text{charge}}$	$\varepsilon_{\text{track}}$	$\mathfrak{r}_{\text{fake}}$	$\mathfrak{r}_{\text{clone}}$	$\mathfrak{r}_{\text{wrong ch.}}$
best config. CAT finder	$95.01^{+0.07}_{-0.07}$	$95.88^{+0.07}_{-0.07}$	$2.37^{+0.04}_{-0.04}$	$0.32^{+0.01}_{-0.02}$	$0.91^{+0.03}_{-0.03}$
Baseline finder	$92.92^{+0.09}_{-0.08}$	$93.56^{+0.08}_{-0.08}$	$1.62^{+0.03}_{-0.03}$	$0.17^{+0.01}_{-0.01}$	$0.68^{+0.03}_{-0.03}$
std. CAT finder	$90.42^{+0.1}_{-0.1}$	$92.79^{+0.09}_{-0.09}$	$3.27^{+0.05}_{-0.05}$	$4.38^{+0.05}_{-0.05}$	$2.55^{+0.05}_{-0.05}$
best config. CAT fitter	$91.83^{+0.09}_{-0.09}$	$92.62^{+0.09}_{-0.09}$	$0.82^{+0.02}_{-0.03}$	$0.19^{+0.01}_{-0.01}$	$0.85^{+0.03}_{-0.03}$
Baseline fitter	$91.53^{+0.09}_{-0.09}$	$91.94^{+0.09}_{-0.09}$	$1.11^{+0.03}_{-0.03}$	$0.1^{+0.01}_{-0.01}$	$0.45^{+0.02}_{-0.02}$
std. CAT fitter	$87.9^{+0.1}_{-0.1}$	$89.7^{+0.1}_{-0.1}$	$1.09^{+0.03}_{-0.03}$	$2.05^{+0.04}_{-0.04}$	$1.96^{+0.05}_{-0.05}$