Department of Physics and Astronomy
Heidelberg University

Bachelor Thesis in Physics
submitted by

Furkan Selim Cetin

born in Bielefeld (Germany)

2024

Electron track reconstruction studies and improvement for LHCb's real-time analysis trigger

This Bachelor Thesis has been carried out by Furkan Selim Cetin at the
Physikalisches Institut in Heidelberg
under the supervision of
Prof. Dr. Stephanie Hansmann-Menzemer

Abstract

This thesis presents an improved neural-network-based Track-Matching algorithm to recover and reconstruct electron tracks more efficiently in LHCb's real-time analysis trigger. Electrons emit bremsstrahlung, which complicates their track reconstruction. Currently, the track-reconstruction tuning explicitly excludes electrons, since including them significantly lowers the reconstruction efficiency of other particles.

The presented algorithm is intended to be run in a dedicated electron reconstruction in HLT2, thereby circumventing such issues entirely, while allowing all aspects of the algorithm to be optimised to electrons. It is demonstrated that the improved electron Matching algorithm allows for electron tracking efficiencies of above 90%, while simultaneously reducing the fake track fraction to below 15%.

Kurzfassung

Diese Arbeit präsentiert einen verbesserten Track-Matching Algorithmus, unter Verwendung eines neuronalen Netzes, zur effizienteren Rekonstruktion von Elektronen Spuren in dem Echtzeit-Analyse Trigger des LHCb. Elektronen emittieren Bremsstrahlung, was ihre Spurrekonstruktion erschwert. Derzeit werden Elektronen aus dem Tuning der Spurrekonstruktion explizit ausgeschlossen, da deren Einbeziehung die Rekonstruktionseffizienz der anderen Teilchen deutlich verringert.

Der präsentierte Algorithmus soll in einer dedizierten Elektronen-Rekonstruktion im HLT2 ausgeführt werden, um solche Probleme vermeiden und gleichzeitig den Algorithmus vollständig für Elektronen zu optimieren. Es wird gezeigt, dass der verbesserte Elektronen Matching Algorithmus Elektronen Spureffizienzen von über 90% ermöglichen würde, während zeitgleich der Anteil der falschen Spuren auf unter 15% gesenkt wird.

Contents

List of Abbreviations

e.g. example given

i.e. id est (*das heißt*)

s.t. such that

1   Introduction

The field of particle physics deals with the fundamental constituents of matter on the smallest scales. Since the discovery of the electron in 1897 [1], the ability to visualise the trajectories of charged particles traversing some experimental set-up has been indispensable from high-energy physics. To reconstruct a charged particle's trajectory, its position is measured several times along its path, after which the space point measurements are analysed and can then be used to fit a track. Physicists continuously work on new and better ways to measure and reconstruct paths of charged particles. Advancements often affect the underlying computing architecture upon which they are implemented, thus allowing for more efficient and faster data processing.

The Large Hadron Collider at CERN is at today's forefront of experimental particle physics and hosts many high-energy physics experiments, all of which share the need to measure and visualise particles' trajectories. The major experiments are already able to reconstruct tracks of the charged particles that are of interest to the respective physics programmes with efficiencies well above 95%.

In this thesis, the track-finding efficiency of electrons[1] at the LHCb experiment is studied in order to improve their reconstruction, since they underperform significantly compared to other charged particles. Electrons are the lightest stable charged particles. They typically have a lower momentum spectrum because of photon emissions through bremsstrahlung. This leads to more scattering while traversing the LHCb detector, which is especially problematic since large portions of the detector are not instrumented with tracking sub-detectors. Therefore, following electrons' trajectories, *e.g.* through the magnet, is more difficult, and their reconstruction more challenging.

However, considering electrons are relevant for many key measurements, *e.g.* flavour-changing neutral currents [2] or lepton flavour universality [3, 4] in decays such as $B^0 \rightarrow K^* l^+ l^-$, improving their track-finding efficiency is of great interest to the LHCb collaboration.

This thesis presents a neural-network-based approach to recover and reconstruct electron tracks in the real-time analysis trigger at LHCb. Therefore, the reconstruction sequence must be computationally cheap enough to be run in the trigger, while providing offline-quality track reconstruction.

First, the LHCb experiment and the machine-learning methods used in this thesis are described in Section 2. Afterwards, LHCb's real-time analysis trigger is presented in Section 3, followed by a definition of the performance metrics and a first look at current reconstruction efficiencies in Section 4. The physical phenomena needed for the track reconstruction are briefly outlined in Section 5. Then, the differences between properly reconstructed and *lost* electron tracks are presented in Section 6. In Section 7, a currently

---

[1]Charge conjugation is implied throughout this thesis.

used baseline reconstruction algorithm is described, followed by the author's main work: an improved electron reconstruction. Lastly, the performance of both track-reconstruction sequences is presented and compared.

2   The LHCb Experiment

The LHCb experiment is one of the four large particle physics experiments at the Large Hadron Collider (LHC), located at CERN near Geneva, Switzerland. The LHC is a proton-proton (pp) collider. During Run 3 of the LHC, protons with a centre-of-mass energy of $\sqrt{s} = 13.6\,\text{TeV}$ are collided. The Large Hadron Collider beauty (LHCb) experiment operates at an instantaneous luminosity of $\mathcal{L} = 2 \times 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ [5]. In this section, a brief overview of the LHCb detector is given.

2.1   Overview

The LHCb detector (Figure 2.1) is a single-arm forward spectrometer, with an angular acceptance of 10 to 300 (250) mrad relative to the beamline in the bending x-z (non-bending y-z) plane, corresponding to a pseudorapidity[2] range of $2 < \eta < 5$. Bending refers to the effects of the Lorentz force that charged particles encounter in a magnetic field. The entire detector stretches over a length of approximately 20 m. A right-handed coordinate system is used, with its origin in the centre of the Vertex Locator (VELO), a z-axis direction along the beamline pointing downstream of the magnet, x pointing to the centre of the LHC ring, and y pointing up in the vertical direction.

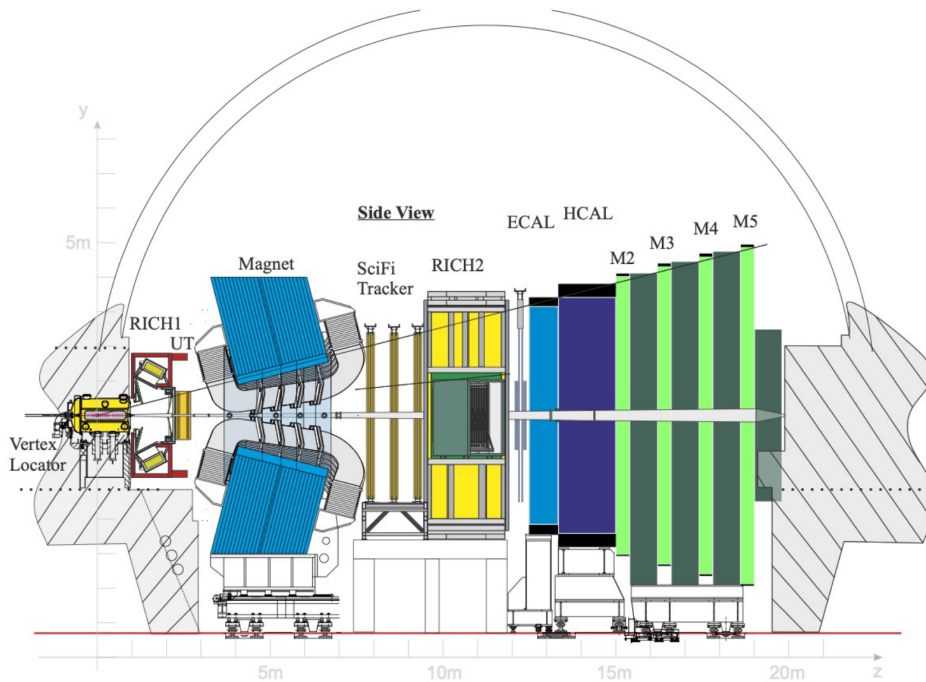**Figure 2.1:** Side view of the LHCb detector in the non-bending y-z-plane [6].

---

[2]Pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$ where θ is the opening angle between a particle's momentum vector and the beam pipe axis.

The interaction point is located in the centre of the VELO, making it the first sub-detector particles encounter. Depending on the lifetime and type of particles, they then travel through the first Ring-Imaging Cherenkov detector (RICH1), the Upstream Tracker (UT), the magnet, the Scintillating Fibre (SciFi) Tracker, the second Ring-Imaging Cherenkov detector (RICH2), the electromagnetic calorimeter (ECAL), and the hadronic calorimeter (HCAL).

Electrons do not typically reach the HCAL, since they get absorbed by the ECAL. Muons also encounter the muon stations (M2-M5) at the end of the detector. Of these sub-detectors, the VELO, UT, and SciFi tracker together with the magnet make up LHCb's tracking system and are described in more detail in the following sections. The RICH detectors, the HCAL, and the muon stations are all part of the particle identification (PID) system, which is only marginally relevant to the work presented in this thesis. While the ECAL is also part of the PID system, its use to identify electrons and photons is explored further in the following sections.

2.2 Magnet

The magnet generates a field that bends charged particles' trajectories, which in turn makes momentum estimates possible. It is located around $z = 5000\,\mathrm{mm}$, with fringe fields extending further into sub-detectors up- and downstream, as shown in Figure 2.2 [8]. A warm dipole magnet with saddle-shaped coils in a window-frame yoke is used. The integrated magnetic field of $4\,\mathrm{Tm}$ for $10\,\mathrm{m}$ tracks in z direction significantly reaches
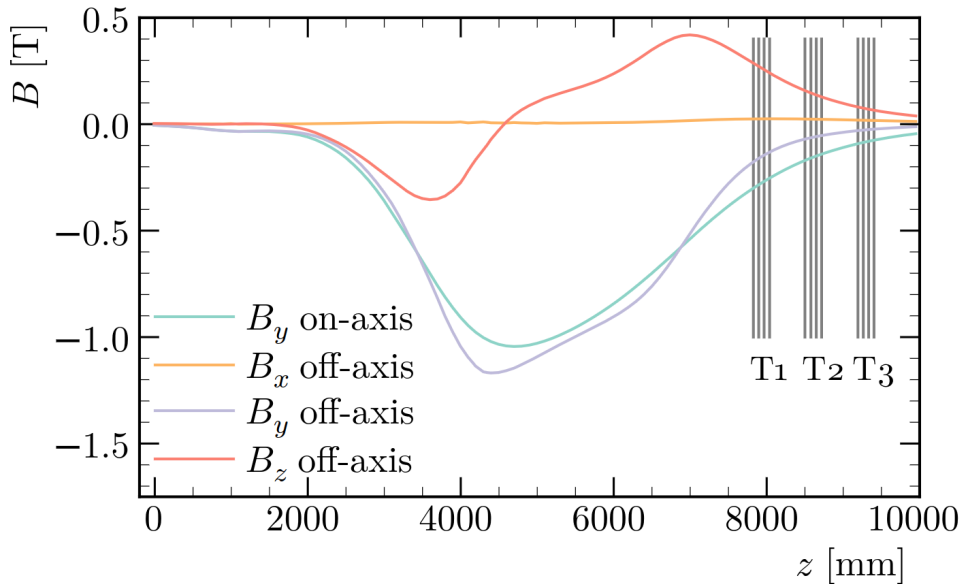

Figure 2.2: Magnetic field strength components as functions of z. The components off-axis are shown for a particle with positive x and y slopes. T1, T2, and T3 illustrate the three stations of the SciFi tracker [7].

into the UT and SciFi tracker, while the field in the VELO is negligible. The magnet polarity (MagUp and MagDown) is reversed regularly during data taking to minimise effects stemming from detector asymmetries [6].

2.3   Tracking Detectors

The tracking system of VELO, UT, SciFi tracker, and magnet is used to find and describe trajectories of charged particles through track state vectors, defined as

$$\boldsymbol{s} = \begin{pmatrix} x & y & t_x & t_y & q/p \end{pmatrix}^{\top}_{z_{\mathrm{ref}}} \tag{2.1}$$

where x and y are the position coordinates, $t_x \equiv dx/dz$ and $t_y \equiv dy/dz$ are the slopes of x and y with regard to z, respectively, and q/p is the signed charge over the particle's momentum. The five track parameters are given at some reference plane z_{ref}, *e.g.* a detector layer.

2.3.1   Vertex Locator

The Vertex Locator (VELO) detects particles coming from the beam collision region. All tracks are first reconstructed as straight lines. Then the location of primary interaction vertices, displaced decay vertices, and the distances between them can be determined [6]. The VELO modules are located $5.1\,\mathrm{mm}$ from the beam to provide a high vertex resolution. The two movable halves consist of 26 L-shaped modules each.

The VELO is a silicon pixel detector reaching a hit resolution of $12.5\,\mathrm{\mu m}$ in the x and y coordinate [9] while adding heavily onto the material budget with a radiation length fraction of $x/X_0 \simeq 21.3\%$, averaged over the nominal pseudorapidity range [10]. The material budget of one module is around $x/X_0 \simeq 0.94\%$ for perpendicular incident tracks.
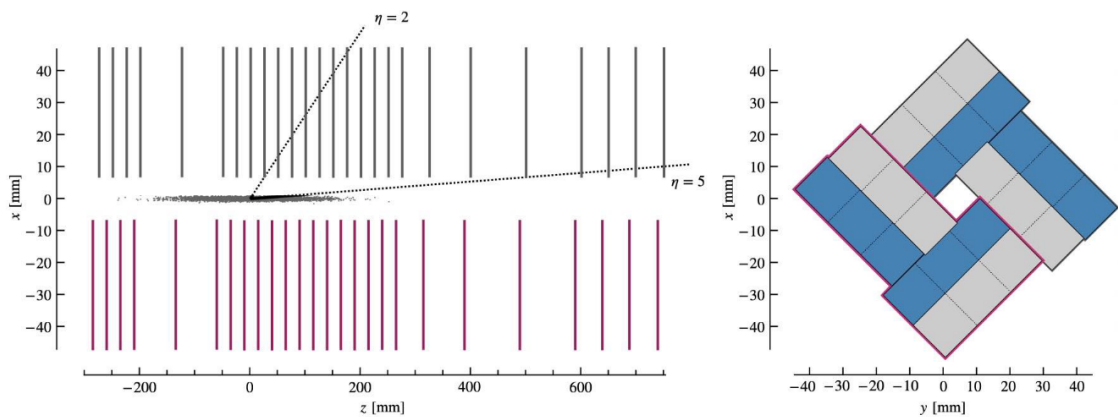

Figure 2.3: Left image: schematic top view of the x-z plane with an illustration of the z extent of the luminous region and the nominal LHCb pseudorapidity acceptance. Right image: nominal layout of the closed VELO configuration [6].

2.3.2 Upstream Tracker

The Upstream Tracker (UT) is the second tracking detector, located around $z = 2485 \, \mathrm{mm}$ upstream of the magnet. The location of the UT puts the sub-detector in a substantial magnetic field, see Figure 2.2. The UT consists of four layers of silicon micro-strip sensors that vary in length and pitch according to the decrease of the particle flux away from the beam pipe. Thus the highest hit resolution is reached by the sensors in the centre around the beam pipe hole with $27 \, \mathrm{\mu m}$. The hit resolution of the outer sensors, coloured in green in Figure 2.4, is $55 \, \mathrm{\mu m}$. Furthermore, the layers are in an x-u-v-x layout, hence measuring only the x coordinate in the first and last layer, and an x and y coordinate with both inner layers using their stereo angle of $\pm 5°$.

Because of the UT's position between the VELO and the SciFi tracker, a low material budget is especially paramount to not hinder the reconstruction of particles. The average radiation length fraction is relatively small with $x/X_0 \simeq 7\%$ [6].
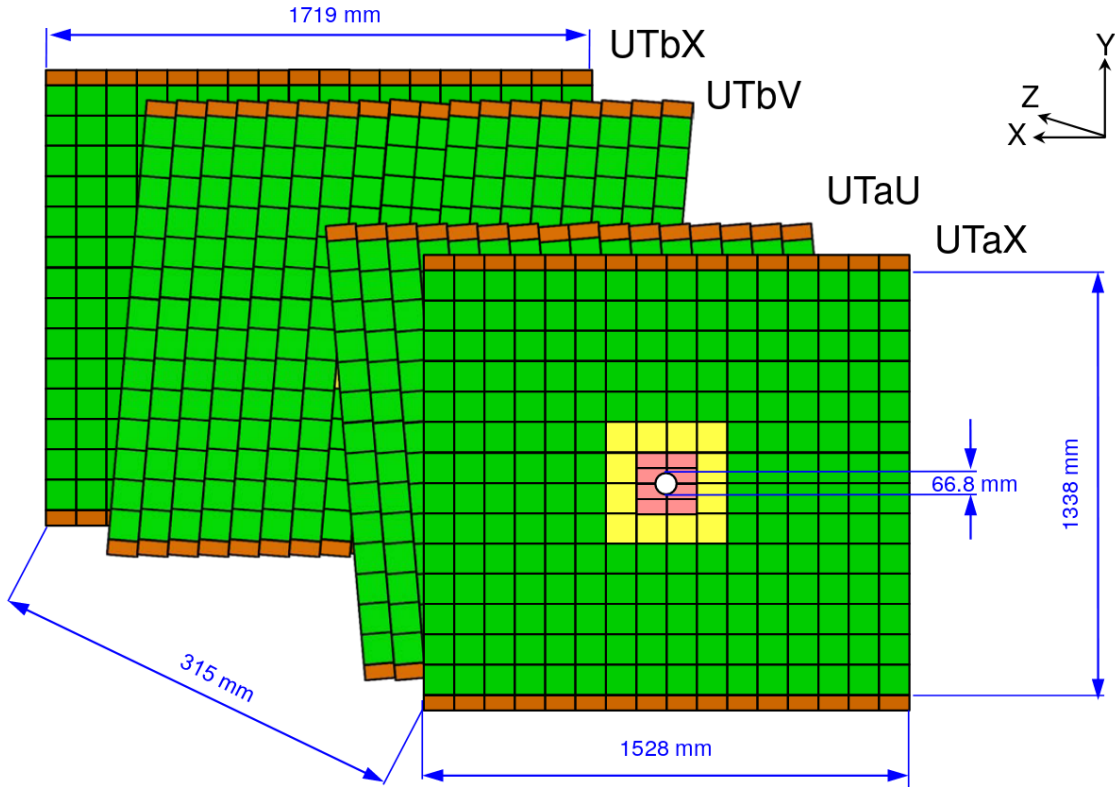

Figure 2.4: UT geometry overview, looking downstream [11]. The pink sensors have a pitch of $93.5 \, \mathrm{\mu m}$ and a length of around $5 \, \mathrm{cm}$, the yellow sensors have the same pitch, but double the length, and the green sensors have a pitch of $187.5 \, \mathrm{\mu m}$ at $10 \, \mathrm{cm}$ of length.
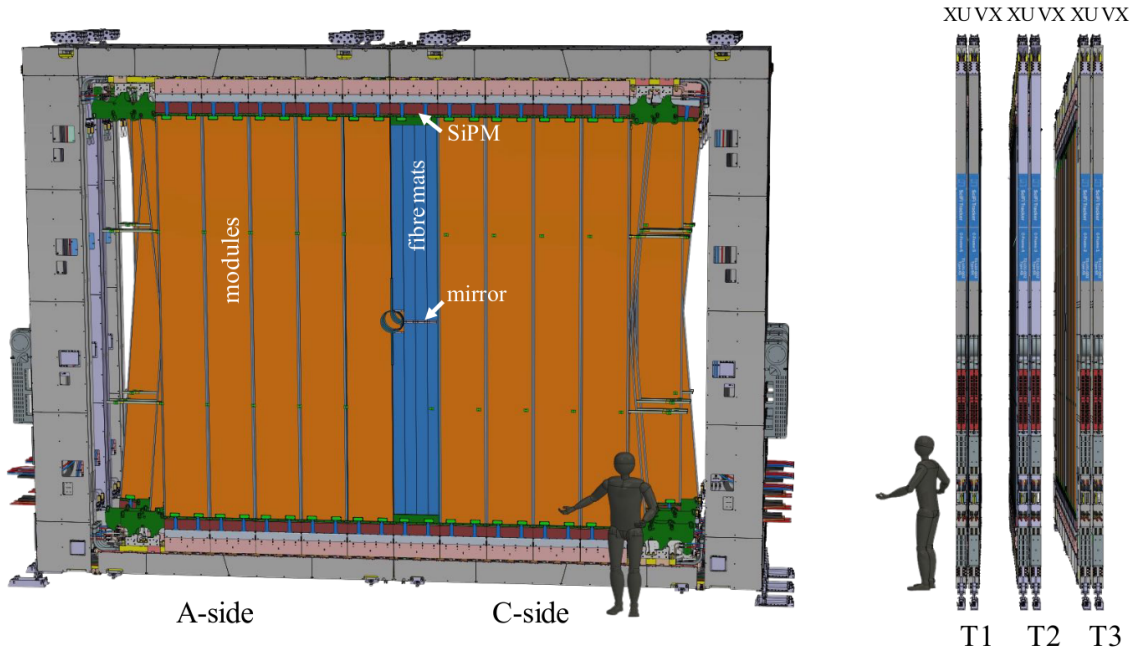
Figure 2.5: Front and side view schematic of the SciFi tracker [6].

2.3.3 Scintillating Fibre Tracker

The Scintillating Fibre (SciFi) Tracker is located around $z = 8520\,\mathrm{mm}$ downstream of the magnet. Measurements from the SciFi tracker combined with hits from the VELO make it possible to reconstruct tracks with the best momentum resolution in LHCb, by exploiting the long lever arm made up of VELO and SciFi tracker, and utilising the bending of the trajectory in the magnetic field. The SciFi tracker's acceptance covers a range of up to $\pm 3186(2425)\,\mathrm{mm}$ horizontally (vertically). Each of the three stations (T1, T2, T3), depicted in Figure 2.5, is composed of four layers that follow the x-u-v-x layout like the UT, with the first and last layer solely used to measure the x coordinate, and the stereo angle of $\pm 5°$ of the inner two layers to also measure a y coordinate. The single hit position resolution, as estimated during a test beam campaign, is about $(64\pm 16)\,\mathrm{\mu m}$ for perpendicular tracks [6]. The SciFi tracker adds to the material budget with a radiation length fraction of $x/X_0 \simeq 12.4\%$.

2.4 Particle Identification

The particle identification (PID) system consists of the Ring-Imaging Cherenkov detectors (RICH1, RICH2), calorimeters (ECAL, HCAL), and the muon stations (M2-M5) [12]. Since only the RICH1 is located upstream of the SciFi tracker, it is the only PID sub-detector that affects the particle trajectories in a manner that could obstruct track reconstruction. The RICH1 is located between the VELO and the UT, so a low material budget is of
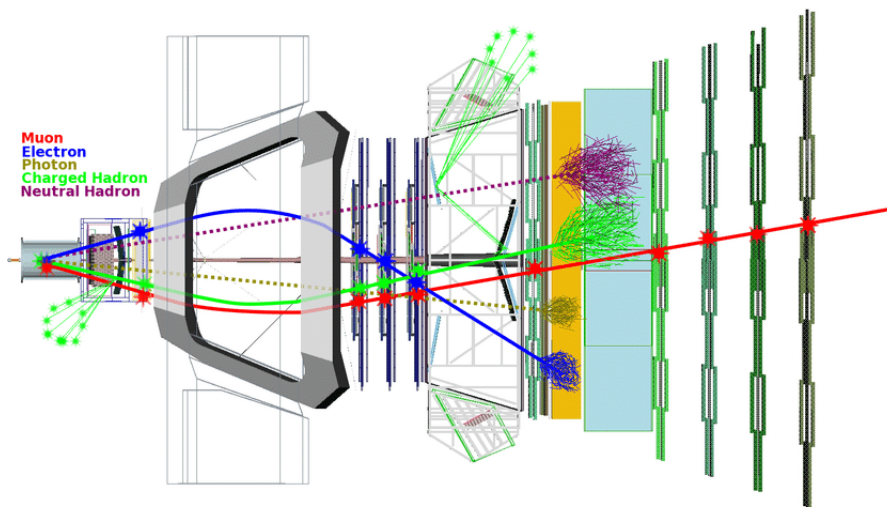
Figure 2.6: Schematic view of different particle signatures in the LHCb detector, including showers in the calorimeters [15]. The ECAL is depicted in yellow. Photons' tracks are gray, and electrons' tracks are coloured blue. The detector layout depicted is not current.

great importance. The total radiation length fraction is $x/X_0 \simeq 4.8\%$, which is not significant for track reconstruction [7].

The ECAL is located behind the SciFi tracker and RICH2 at around $z = 12.5\,\mathrm{m}$ and is part of the calorimeter system [13]. Before particles reach the ECAL (or HCAL), they must traverse through two pre-detectors. The Scintillating Pad Detector indicates the presence of a charged particle, which makes it possible to separate electrons and photons. The PreShower detector marks the start of an electromagnetic (EM) shower, thus making it possible to differentiate between charged hadrons and electrons or photons. The ECAL measures an EM-shower's energy and position and is used to identify electrons and photons. Electrons are distinguished from photons by their shower shape and the fact that their ECAL clusters can be matched to reconstructed tracks, see Figure 2.6. More information on the ECAL can be found in Ref. [14].

2.5   Machine Learning Methods

Machine learning is a subset of artificial intelligence involving algorithms that enable computers to learn from data and perform tasks without explicit programming. The literature separates between supervised, unsupervised, and semi-supervised learning. Unsupervised learning is used to reveal intrinsic structures in the data, *i.e.* the computer is not supplied with any outside information. On the other hand, in supervised learning the computer is provided with labelled data in the form of a training set. The aim is to explore external information that is available during training but is otherwise unknown. For discrete labels, the problem is called classification, for continuous values regression. The last type of learning is semi-supervised learning, where most data are unlabelled and only few are labelled. More information on machine learning can be found in Refs. [17, 16]. This section briefly recapitulates the supervised machine-learning methods used in this thesis.

The Matching algorithm, described in Section 7, uses a machine learning classifier in the form of a neural network. An artificial neural network is a non-linear statistical data-modelling or decision-making tool, comprised of data processing nodes, referred to as artificial neurons. Multi-layer perceptrons (MLP) form the basis for all neural networks. They consist of neurons which are organised into interconnected layers. MLPs contain one or more hidden layers with regard to the input and output nodes. These hidden neurons act as feature detectors. The basic architecture of an MLP with one output variable is shown in Figure 2.7. Each node processes information by performing a computation on the input signal using an activation function and produces an output that influences other neurons in the network. The activation functions allow for non-linearity, making it possible for the network to learn complex patterns in data. The level of influence of a neuron
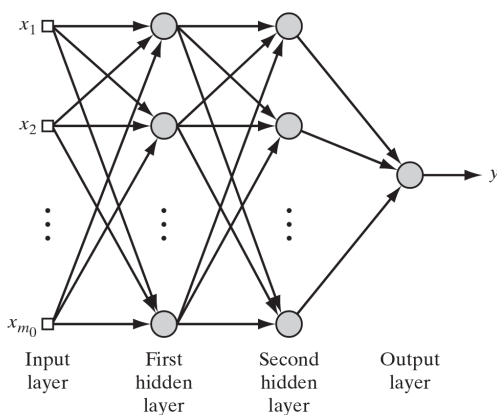

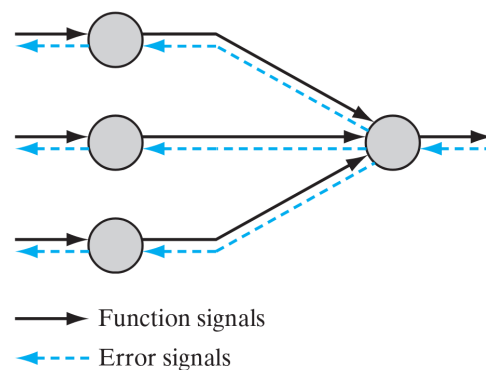**Figure 2.7:** Graph of an MLP with two hidden layers and one output neuron [16].

Figure 2.8: Illustration of the directions of signal flows in an MLP [16].

over another is described by synaptic weights. The weights are adjusted to minimise the loss function, which is used to quantify the level of agreement between the predicted and actual target values.

All MLPs used in this thesis are trained using back-propagation, which means the training stage is split into a *forward* and a *backward phase*. First, the input signal is propagated through the network, layer by layer, with fixed synaptic weights until it reaches the output. Subsequently, the loss function's error signal is propagated through the network in the backward direction, while successive changes to the synaptic weights of the network are made. The forward propagation of function signals and the back propagation of error signals are illustrated in Figure 2.8.

To adequately train the network, a sufficient number of training events is required. The complexity of the data, *i.e.* the number of features and dimensions dictates the volume of the required data to adequately represent the entire phase space. The *curse of dimensionality* refers to the problem that the complexity of the network increases exponentially with the increasing complexity (dimensionality) of the data [18]. However, it is also possible for a network to have too many neurons or training epochs, so that it becomes too complex to model a particular problem. As a result, it learns statistical fluctuations in the training data, *s.t.* it is unable to adapt to and generalise between new events. This is called *overtraining* or *overfitting* [16].

3 Real-Time Analysis in the Trigger

All particle detectors use trigger systems. Triggers use criteria to determine which events are of interest and should be recorded for later analysis, thus reducing the amount of data that has to be stored and processed further. In this section, an overview of LHCb's two-stage software trigger is given. The section follows the structure of the description found in Ref. [7].

3.1 Real-Time Analysis at LHCb

LHCb's trigger for Run 3 is entirely software-based to more efficiently deal with the vast amounts of data recorded by the detector. In the context of LHCb, real-time refers to the time span from a particle-bunch crossing to writing the detector data into permanent storage. For Run 3, LHCb employs a full, real-time aligned and calibrated, software trigger that carries out offline-quality event reconstruction on the incoming data at bunch-crossing frequency (30 MHz). It selects and stores only the reconstructed decay trees that are of interest for further analysis. Event reconstruction is inextricably dependent upon track reconstruction, *s.t.* virtually all variables used in the selection of charged final state particles are derived from the reconstructed tracks.

The high-level trigger (HLT) is split into two parts separated by a storage buffer, as illustrated in Figure 3.1. Both trigger stages perform selections. The High-level Trigger One (HLT1) runs a partial event reconstruction using GPUs while processing events at a rate of about 30 MHz. At this stage, events are mainly selected using one- and two-track-based algorithms. More information on the first level trigger of the LHCb detector can be found in Refs. [20, 21]. Events selected by HLT1 are written to the buffer storage network, where they are accessed by the High-level Trigger Two (HLT2) which performs a full event reconstruction at a rate of around 1 MHz using CPUs [7]. HLT2 is able to select
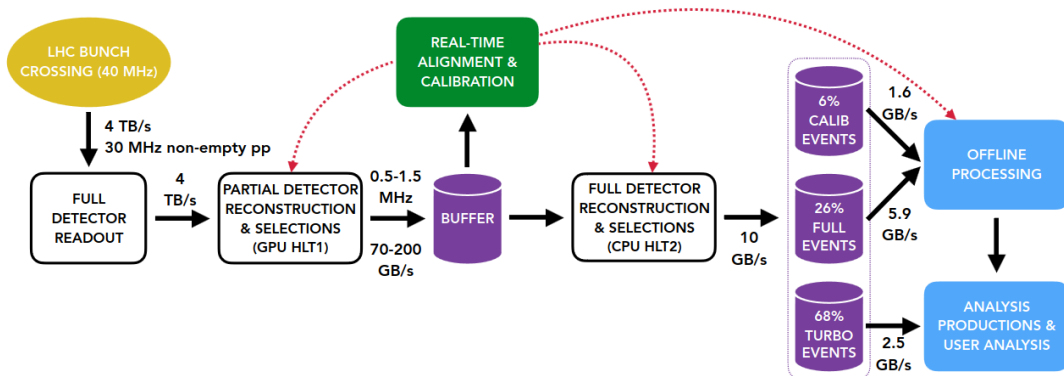

Figure 3.1: LHCb software-trigger dataflow [19]. The bandwidths and percentages on the right-hand side are examples.

specific decay trees that are of interest, rather than entire events. At this point, the data rate is reduced to around $100\,\text{kHz}$ corresponding to $10\,\text{GB/s}$, which can then be stored in permanent storage systems.

3.2   Track Types at LHCb

A track's specific type is determined by what tracking detectors are used to reconstruct a particle's trajectory. The track types relevant to this thesis are:

    VELO tracks, reconstructed from hits in the VELO detector only. VELO tracks are straight (linear). They are used for primary vertex reconstruction and seed other reconstruction algorithms. However, they offer no information on the momentum.

    T tracks, also called *SciFi tracks*, reconstructed from hits in the SciFi detector only. Similar to VELO tracks, they seed other track types, but also offer a rough momentum estimate, with a resolution of 25%-35% [23].

    Long tracks, with hits in the VELO and SciFi detector, and optionally also in the UT stations. These tracks are of most interest because they offer the highest momentum resolution with 0.5%, and are the standard type used in physics analyses [7].

The different track types are visualised in Figure 3.2. Downstream tracks consist of hits in the UT stations and the SciFi tracker, while Upstream tracks are VELO tracks with hits in the UT detector. They are not central to this thesis. Furthermore, there are:

    Fake tracks, or *ghost tracks*, consist of hits of more than one particle or none. They are only known in simulation where the hits' origin is known.

    Clone tracks are tracks that share a significant amount of hits with other tracks. They are associated to a single particle in simulation, and cannot be ghost tracks.

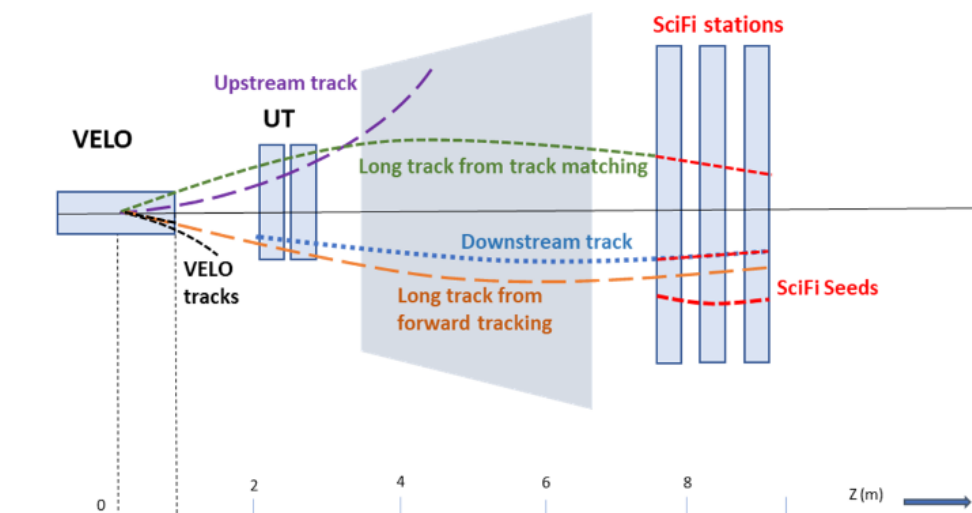These track-like objects are defects of the pattern recognition and can have any track type.

Figure 3.2: Sketch of LHCb track types in the x-z plane [22].

3.3   Full Reconstruction Sequence

HLT2 performs a full event reconstruction, starting with track reconstruction. Reconstructed tracks serve as input for both, the PID system and the selections. The following is a brief overview of how Long tracks are obtained in the baseline track reconstruction sequence, which is illustrated in Figure 3.3. Track reconstruction can be divided into a track-finding and track-fitting stage. In the track-finding stage, the measured hits are combined to track segments using different pattern-recognition methods. In the track-fitting stage, estimates for the track parameters are calculated. Furthermore, fake tracks are rejected, and duplicate tracks removed, *s.t.* only filtered *best* tracks remain [24].

The first tracks to be reconstructed are VELO and T tracks. More information on VELO Tracking can be found in Ref. [25]. T tracks are obtained through the Hybrid Seeding algorithm, which is described in detail in Ref. [26]. Both of these are stand-alone algorithms, meaning they only depend on measurements in the tracking sub-detectors.

LHCb's Long tracks are reconstructed using two algorithms, Forward Tracking and Track Matching. The Forward tracking algorithm uses VELO tracks with the objective of
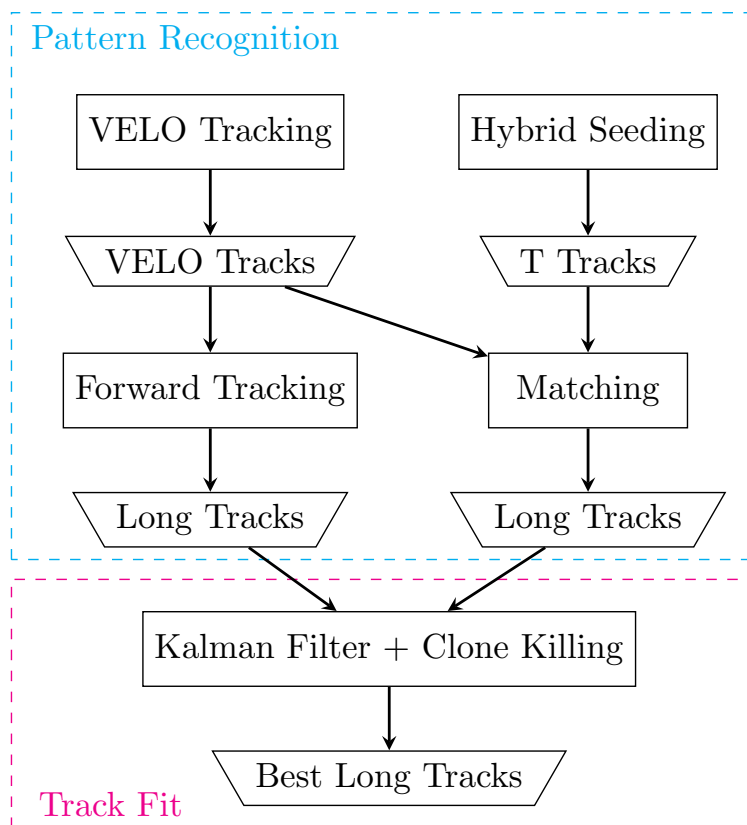
Figure 3.3: Dataflow in the baseline Long-track reconstruction. Algorithms are shown in rectangles, and track containers in trapeziums. Detector hits are not shown.

finding forward extensions of them in the SciFi tracker. A description of Forward tracking in HLT2 as used in Run 3 can be found in Refs. [7, 27]. Track Matching makes use of already reconstructed track segments as input to create Long tracks by combining VELO and T tracks. The Matching algorithm is described in Section 7.2, and in Ref. [7].

Both algorithms supply Long tracks, ending the track-finding stage of the reconstruction. The Long tracks are then fitted using the Kalman filter, and passed through a clone removal algorithm, to select the best track from those that share many hits, thus creating *Best* Long tracks.

4   Tracking

Track reconstruction refers to a system of coordinated hardware and software for detecting and measuring charged particles' trajectories, and providing estimates on the momentum vectors along their path. At LHCb, the tracking sub-detectors consist of multiple layers that are able to measure the position of a particle when passing through. These singular position measurements are called hits. The number of hit-measuring layers cannot be arbitrarily increased for greater precision. More detector material means more interactions between final state particles and the material, which can influence the particles' trajectories or even stop them outright in hadronic interactions [28].

One crucial material effect, which dominates the momentum resolution, is caused by multiple elastic Coulomb scattering (multiple scattering). In thin scatterers, this causes a change in the momentum vector's direction. The scattering angle θ is proportional to $1/p$, where p is the particle's momentum. However, multiple scattering affects all particles, and cannot be addressed properly within the limited scope of this work. More on multiple scattering can be found in Ref. [29]. During the track-finding stage of the reconstruction, special algorithms are used to group a multitude of hits, presumably from one particle, to form a track, which entails the particle's trajectory and momentum estimate. This section gives an overview of the reconstruction performance of electrons and other particles at LHCb.

4.1   Reconstruction Performance

Prior to describing any reconstruction algorithm, the criteria for evaluating the performance of the reconstruction have to be defined. A more detailed description can be found in Ref. [30].

There are two main performance indicators for track reconstruction with pattern recognition: the *track finding efficiency* and the *fake track rate*. The track-finding efficiency is defined as

$$\varepsilon = \frac{N_{\text{reconstructible \& MC-matched}}}{N_{\text{reconstructible}}} \tag{4.1}$$

where N is the number of tracks in the corresponding category. More than 70% of a track's hits must originate from the same simulated particle for a track to be classified as *MC-matched*, or *truth matched*. A particle's reconstructibility depends on the amount of hits it leaves in the tracking detectors. The requirements differ for all track types. The simulated particle of a reconstructible Long track must pass through at least three VELO layers, leaving hits in each, and have a hit in at least one x and one stereo layer in each of the SciFi stations. The fake track fraction is defined as the ratio of tracks that could not be associated

with a simulated particle over the total number of tracks outputted by the algorithm:

$$r_{\text{fake}} = \frac{N_{\text{fake}}}{N_{\text{total}}} \tag{4.2}$$

4.2   Long-Track Efficiency

The Long-track efficiency does not solely depend on the performance of Forward tracking or the Matching algorithm. The efficiencies with which the VELO and T tracks are created directly determine the performance ceiling for the algorithms. VELO tracking performs with the highest efficiency of all track types at LHCb and constitutes the upper boundary in performance observations for Long tracks. The track-finding efficiencies of VELO and T tracks can be found in Section A.3.2.

The integrated Long track efficiencies for electrons and not-electrons are shown in Table 4.1 for several selection categories. The fake track fraction of Long tracks is low at 10%. The track reconstruction of electrons compared to other particles underperforms significantly, *s.t.* a performance discrepancy of around 10% can usually be observed.

The efficiency in dependence on various kinematic variables is plotted in Figure 4.1 for electrons and other particles. The efficiency drops rapidly at low momenta, at least in part because low-momentum tracks experience more multiple scattering which makes them harder to reconstruct.

The efficiency as a function of pseudorapidity and azimuthal angle both illustrate that the respective variable's distributions are fairly similar. The azimuthal angle shows an almost constant efficiency gap of around 10%. In the pseudorapidity, the gap is almost non-existent for small $\eta < 2.5$ but widens with increasing η. Tracks with $\eta = 4.3$ or $\eta > 4.5$ cross beam-pipe material, thus experiencing more scattering or hadronic interactions [7]. This causes a decrease in the distribution of reconstructible particles at high

Table 4.1: Integrated Best Long efficiency of electrons and other particles obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events. The efficiencies corresponding to the *Long from B* category are plotted in Figure 4.1.

Category	$e^{\pm}\ \varepsilon[\%]$	not-$e^{\pm}\ \varepsilon[\%]$
Long	66.72	88.63
Long from B	82.04	91.08
Long $p > 5\,\text{GeV}$ from B	84.15	94.00
Long $p > 3\,\text{GeV}\ p_T > 0.5\,\text{GeV}$ from B	85.58	94.48
Fake track fraction r_{fake}		10.26%

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.
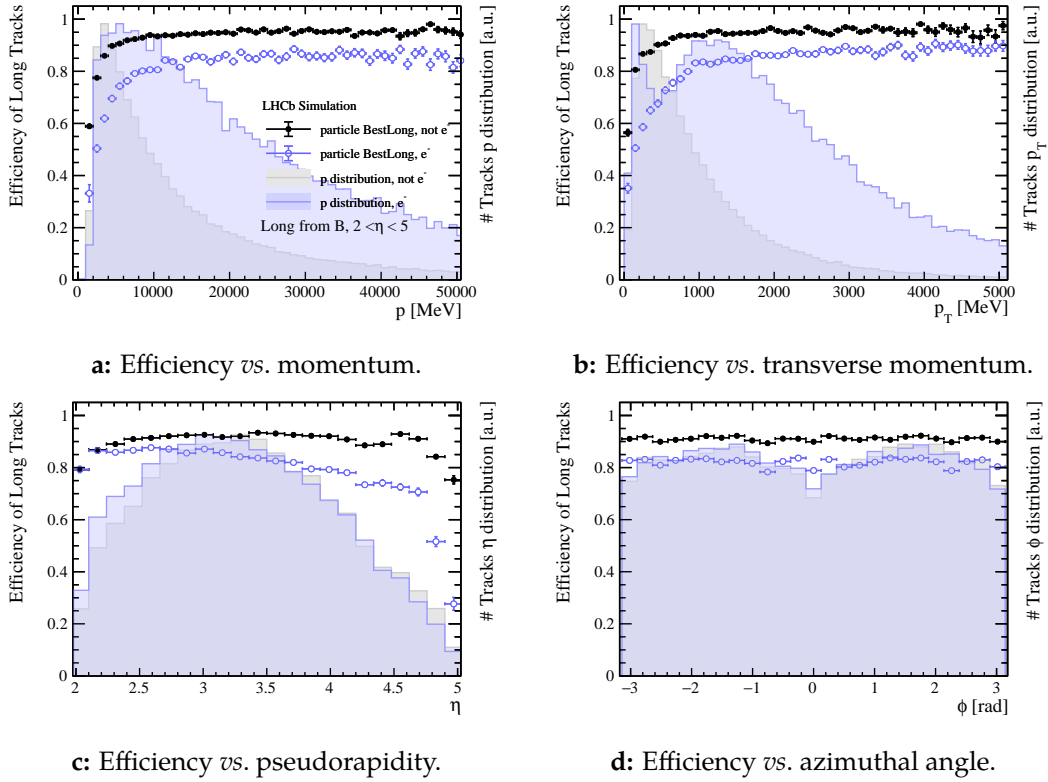
d: Efficiency *vs.* azimuthal angle.

Figure 4.1: Track reconstruction efficiency of Best Long tracks for simulated Long tracks originating from a B meson with regard to various kinematic variables. The variables' distributions and efficiencies are represented in blue for electrons and in grey or black for other particles. The distributions represent the true, simulated variable values at creation. The efficiencies have been obtained using simulated $B^0 \rightarrow K^*(J/\psi \rightarrow e^+e^-)$ events.

pseudorapidity, as well as a rapid drop in efficiency, affecting electrons especially hard. The efficiency drops to around 78% for other particles, and to below 30% for electrons.

5   Tracking Down Electrons

This section deals with the physical phenomena that affect electrons while traversing the LHCb detector. More comprehensive information on the tracking of charged particles at LHCb, and in general, can be found in Refs. [7, 31] respectively.

5.1   Material Effects

In order for the tracking detectors to register a hit, the electrons must interact with them by transferring a small amount of energy to the detector material's electrons. For low energies of around $10\,\mathrm{MeV}$, ionisation dominates the energy loss of electrons. However, ionisation loss rates rise logarithmically, while bremsstrahlung losses increase almost linearly, thus dominating the energy loss above an energy of $\mathcal{O}(10\,\mathrm{MeV})$. The average energy loss can be described by [31]:

$$\langle E(x) \rangle = E_0 \exp\left(-\frac{x}{X_0}\right) \tag{5.1}$$

with the electron's initial energy E_0, and the distance travelled in the material in units of the radiation length x/X_0. The radiation length X_0 is defined as the average distance after which a high-energy electron has lost all but $1/e$ of its energy. This is especially use-
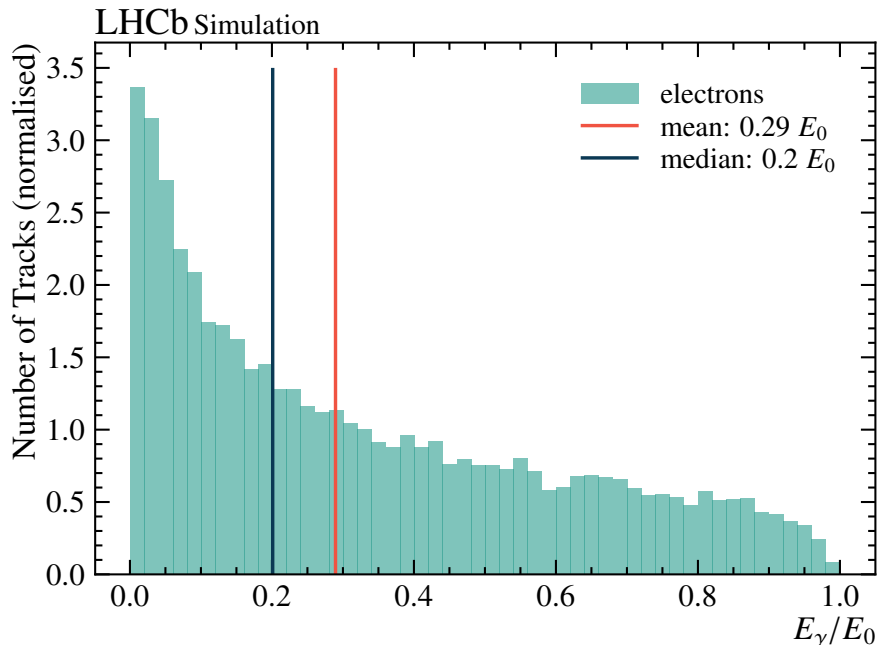

Figure 5.1: Amount of energy emitted by electrons upstream of the magnet in units of the initial energy of the particles. The red line denotes the mean, and the blue line the median energy emission.

ful when discussing the material budget of detectors. The sum of the radiation length fractions of the upstream detectors, provided in Section 2, is $x/X_0 \simeq 33.1\%$. Therefore, an electron[3], on average, loses around a third of its initial energy before reaching the magnet. The mean upstream energy loss, shown in Figure 5.1, concurs with the theoretical predictions.

5.2   Magnetic Field

The dipole magnet generates a field that results in the bending of the trajectories of charged particles as described by the Lorentz force. The amount of bending that takes place can be quantified by the curvature radius

$$r \sim \frac{|\boldsymbol{p}|}{B} \tag{5.2}$$

where $\boldsymbol{p}$ is the particle's relativistic momentum and B the magnetic field's flux density. The magnetic field, depicted in Figure 2.2, is strongest from in front of the UT to in front of the SciFi tracker, with fringe fields reaching into the SciFi stations. The equations of motion show that both magnetic field components B_y and B_z can alter the y-z trajectory of a particle. However, noting $B_y > B_z \gg B_x$, the bending is most significant in the x-z plane.

All charged particles' trajectories are subject to bending and traverse the same magnetic field, *i.e.* the momentum is what mainly affects the trajectories' shapes. Electrons are the only particles whose energy loss is dominated by bremsstrahlung. The emission of a high-energy photon by an electron traversing the magnetic field with initial momentum p_1 causes a change in momentum $p_2 < p_1$. This decrease in momentum also alters the curvature radius $r(p_2) < r(p_1)$ for its remaining passage through the field. This stands in conflict with the parameterisations, defined in Section 7.2, used during reconstruction. New parameterisations for electrons based on the simplified track model are described in Section 7.3.1.

5.3   Calorimetry

Electrons interact with the ECAL converting the particle's energy into a detectable signal in a destructive process, *i.e.* until the particle is absorbed, as can be seen in Figure 2.6. Hadrons shower in the ECAL and HCAL, which means both calorimeters measure a part of their energy. Showers are first matched to reconstructed tracks to dismiss photon showers. Subsequently, the distribution of E/p, where E is the energy measured by the ECAL, and p is the momentum estimate provided by the LHCb tracking system, enables a separation of

---

[3]Unless specifically stated otherwise, the results presented have been obtained using electrons originating from a B meson in the $B^0 \to K^* e^+ e^-$ decay in MC simulated events, with magnet polarity `MagDown`.

a: E/p using Run 1 data.
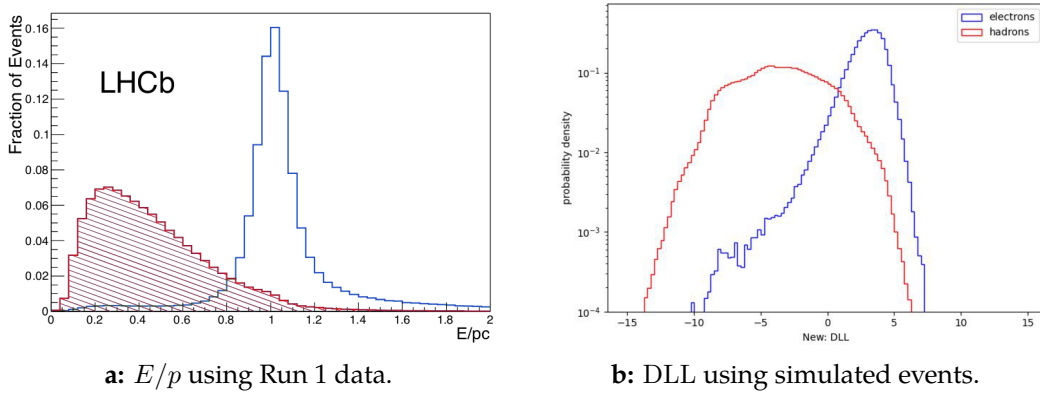
b: DLL using simulated events.

Figure 5.2: Distributions of variables in the ECAL for electrons (blue) and hadrons (red). Figure (a) adopted from Ref. [13], and (b) from (M. V. Veghel, personal communication, January 2024).

hadrons and electrons. Electrons are absorbed by the ECAL in contrast to hadrons, which means the distribution should be centred around $E/p \approx 1$ for electrons, and $E/p < 1$ for hadrons. This is illustrated in Figure 5.2a.

LHCb's software framework contains an ECAL filter designed to separate electron and hadron showers and by extension tracks. The algorithm uses a delta-log likelihood (DLL) of E/p, depicted in Figure 5.2b, as a quality measure to select T tracks (M. V. Veghel, personal communication, January 2024). The tracks are extrapolated to the ECAL's position to find corresponding EM-showers. Selected T tracks must fulfil $DLL(E/p) > -2$. There are currently no publicly available academic works on the functionality or performance of the ECAL filter.

It is also possible to select VELO tracks using information from the ECAL. Because of the VELO's position, however, it is unfeasible to extrapolate VELO tracks to the ECAL directly. Instead, the VELO tracks are matched to photon-induced EM-showers, as opposed to electron-induced EM-showers for T track selection. It is assumed that electrons emit photons while traversing the VELO, causing EM-showers in the ECAL. While such an algorithm exists in the LHCb framework, it is not available out-of-the-box and is thus out-of-scope for this thesis.

6   Understanding Inefficiencies for Electron Reconstruction

The current track reconstruction algorithms are tuned by explicitly excluding electrons. Including them would lower the track-finding efficiency for all other particles. This compromise, however, results in electrons underperforming, as can be seen in Figure 4.1.

This section deals with electrons traversing the LHCb detector while focusing on the properties of reconstructed and not reconstructed electron tracks (in the baseline Matching algorithm). Since the baseline reconstruction already correctly finds many electron tracks, understanding what contributes to them being *found* or *lost* may allow for more efficient track reconstruction.

In the following, *lost* refers to reconstructible but not reconstructed electron tracks, and *found* to tracks that are both reconstructible and MC-matched, according to Section 4.1.

6.1   *The Lost*

The energy loss of electrons, unlike any other particle, is dominated by bremsstrahlung. Almost all electrons emit photons, while traversing the LHCb detector, making this the natural starting point. Found electrons, on average, lose significantly less of their energy due to bremsstrahlung, than lost electrons, as illustrated in Figure 6.1. Especially of interest are energy emissions that occur downstream of the VELO detector, more specifically within the range of the magnet's strong field, see Section 5.2.

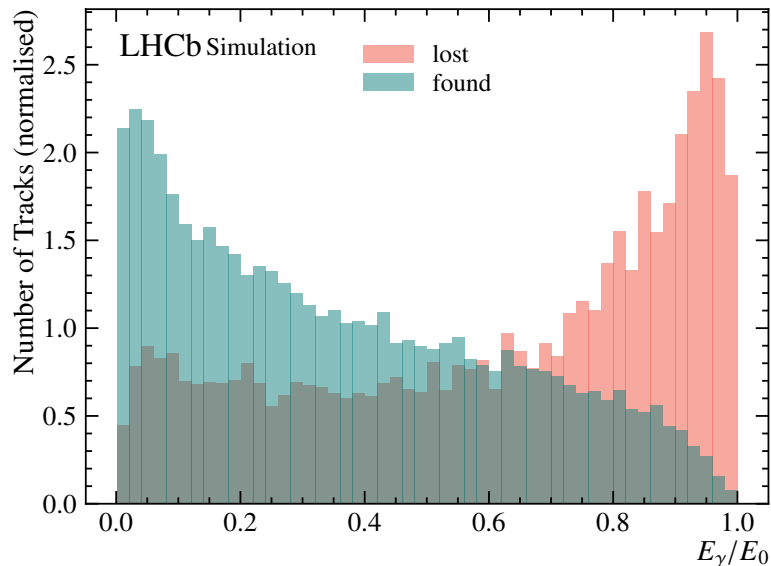Approximately $2/3$ of electrons that emit less than half of their energy in the magnetic

Figure 6.1: Amount of energy emitted by an electron before reaching the RICH2 in units of the initial energy of the particle. Found electrons are represented in green and lost electrons in red.
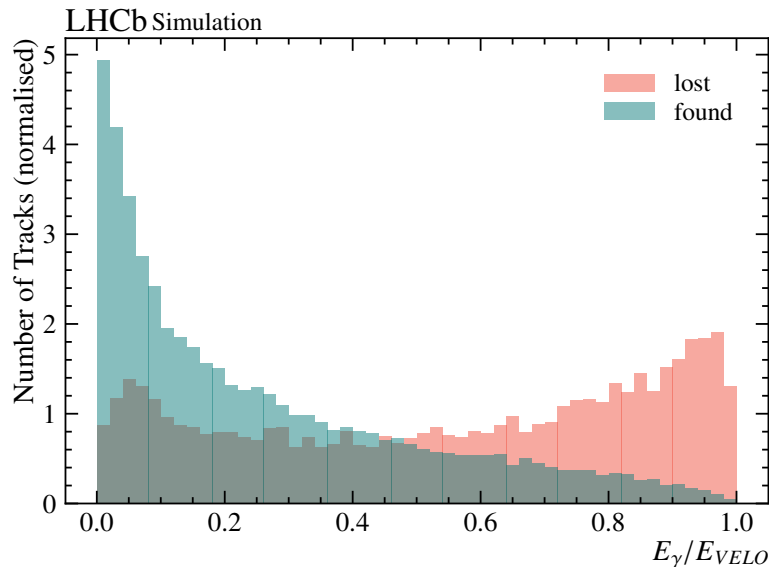
Figure 6.2: Amount of energy emitted by an electron between $770\,\mathrm{mm} < z \leq 9410\,\mathrm{mm}$, *i.e.* from the end of the VELO to the end of the last SciFi station (see Figure 3.2), in units of the energy of the particle at $z = 770\,\mathrm{mm}$. Found electrons are represented in green and lost electrons in red.

field's reach are found. The vast majority of found electrons lose less than 30% of the energy they possess after leaving the VELO. The energy loss distribution, depicted in Figure 6.2, suggests electrons that emit a larger part of their energy after the VELO are, on average, slightly more likely to be lost.

Additionally, the bremsstrahlung vertex distributions plotted in Figure 6.3, show little separation up until the end of the VELO, suggesting that what hinders the track reconstruction occurs there. In both distributions, there are many vertices, where the subdetectors' material is located. However, the distribution on the right is narrower with many vertices immediately around $x = 0\,\mathrm{mm}$, where the beam pipe is located. Both Figure 6.3a and Figure 6.3b depict bremsstrahlung vertices of a similar number of electrons. Yet there are substantially more vertices in the right plot, indicating that lost electrons emit photons more often while traversing the detector.

Summarising, bremsstrahlung after the VELO, especially around the beam pipe can severely obstruct the successful track reconstruction. This is also apparent in the pseudorapidity distributions, plotted in Figure 6.4. Most electrons in the high pseudorapidity region, *i.e.* with a small opening angle, are lost. The ensuing question is, whether it is material in general that leads to the inefficiencies at higher pseudorapidity, or if the beam pipe, specifically, obstructs reconstruction. The radiation length fraction, shown in Figure 6.5, provides a measure of how much material an electron encounters. However, there are no significant differences in the distributions to draw a conclusion from. It appears
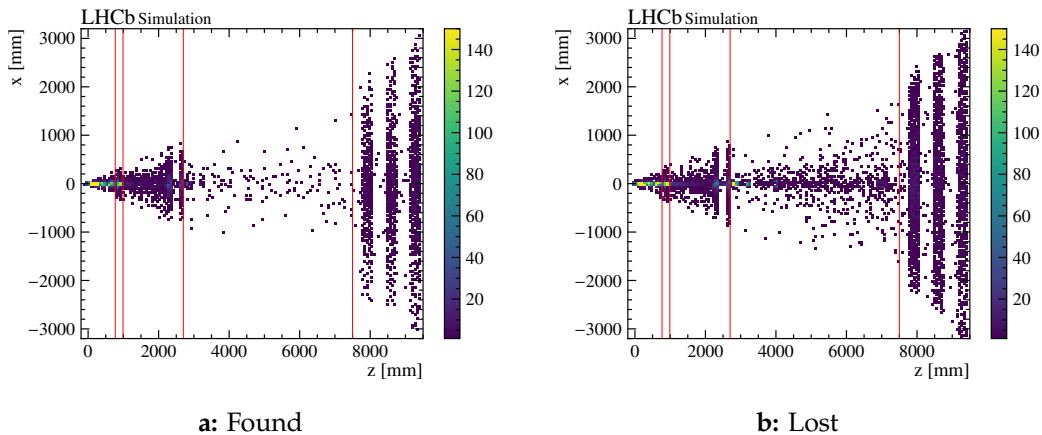
a: Found

b: Lost

Figure 6.3: Distributions of bremsstrahlung vertices where more than 0.1 of the electron's current energy is emitted. The distributions are representative of the same amount of electrons. The red lines in the plots mark the sub-detectors' locations (*l.t.r.* : end of the VELO, beginning of the RICH1, end of the UT, and beginning of the SciFi tracker).
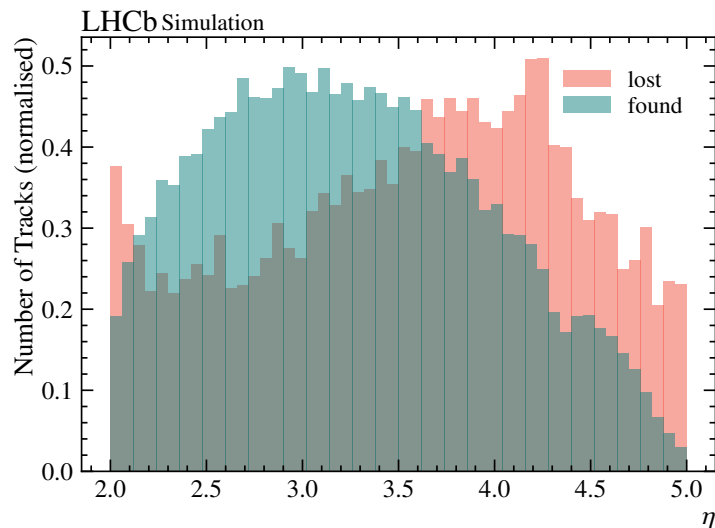

Figure 6.4: Pseudorapidity of an electron at creation. LHCb's acceptance is $2 < \eta < 5$. Found electrons are represented in green and lost electrons in red.

the differences in the amount of material traversed, are too little to impact reconstruction alone, even when considering the total radiation length fraction from the end of the VELO to the end of the last SciFi station.

The radiation length fraction x/X_0 shows correlation to geometrical variables, such as the pseudorapidity η, plotted in Figure 6.6. There is a spike in the radiation length fraction at around $\eta = 4.3$, which is also where a drop in the reconstruction efficiency occurs (see Figure 4.1). As explained in Section 4.2, these tracks cross beam-pipe material. For any such correlation to become feasibly useful in the context of the Matching algorithm, it
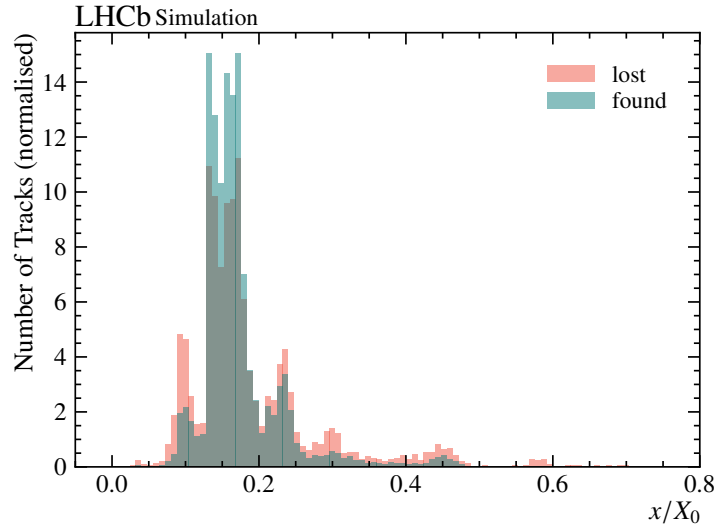
Figure 6.5: Total radiation length fraction of an electron from $z = 770\,\mathrm{mm}$ to $z = 2700\,\mathrm{mm}$. Found electrons are represented in green and lost electrons in red.
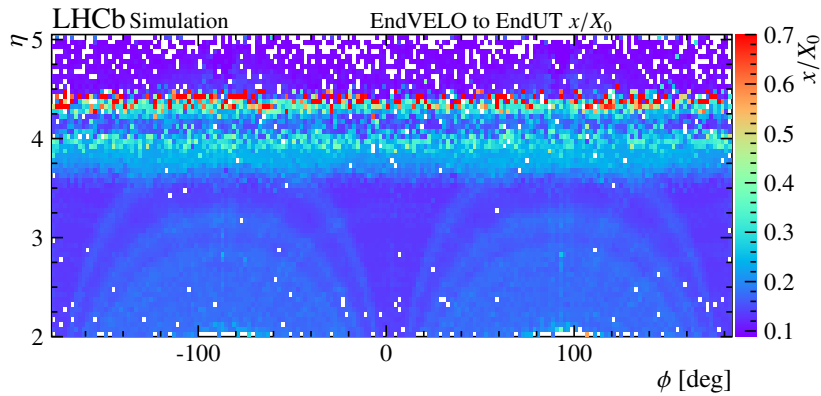
Figure 6.6: Average radiation length fraction x/X_0 seen by particles from $z = 770\,\mathrm{mm}$ to $z = 2700\,\mathrm{mm}$, as a function of the azimuthal angle ϕ and the pseudorapidity η. White spaces are unoccupied (η, ϕ) combinations.

must be parameterisable (see Section 7.2), as well as offer information that prove useful in separating electron tracks from background. Consequently, the use of x/X_0 would not reasonably improve Matching.

6.2   Redefinition of Momentum Dependent Efficiencies

The track-finding efficiencies are illustrated with regard to various kinematic variables, such as the total and transverse momentum. This is useful to understand how particles with certain properties perform in track reconstruction. This section uses the efficiency definition based on simulated events introduced in Section 4.1.

Currently the efficiency plots with regard to both, the total and transverse momentum use the *true*, simulated variable values at generation. Electrons emit bremsstrahlung, as explained in Section 5.1, which means that electrons lose significantly more energy than other particles. Therefore, the momentum of electrons traversing the sub-detectors is often lower than their *true* value at generation suggests. As pointed out in Section 4.2, lower momentum particles' track reconstruction is much less efficient than in the higher-momentum regime. If there is a significant number of high-momentum electrons at generation that lose so much energy in the VELO that they effectively behave like low-momentum particles when entering the RICH1, this would skew the efficiency plots with regard to total and transverse momentum[4].

The momentum of a particle is estimated by its trajectory through the magnetic field, which means bremsstrahlung before reaching the field essentially has the same effect on an electron's track shape that a lower momentum at generation would have.

On average, electrons lose about 16% of their initial energy before leaving the VELO, as shown in Figure 6.7, and 30% of their initial energy before reaching the magnet. However, more than half of the electrons lose less than 20% of their energy upstream (see Figure 5.1). Some electrons lose a significant amount of energy in the VELO and are consequently of lower momentum than at their generation, which could result in a skewed momentum distribution for electrons. To circumvent this, the momenta of simulated electrons used
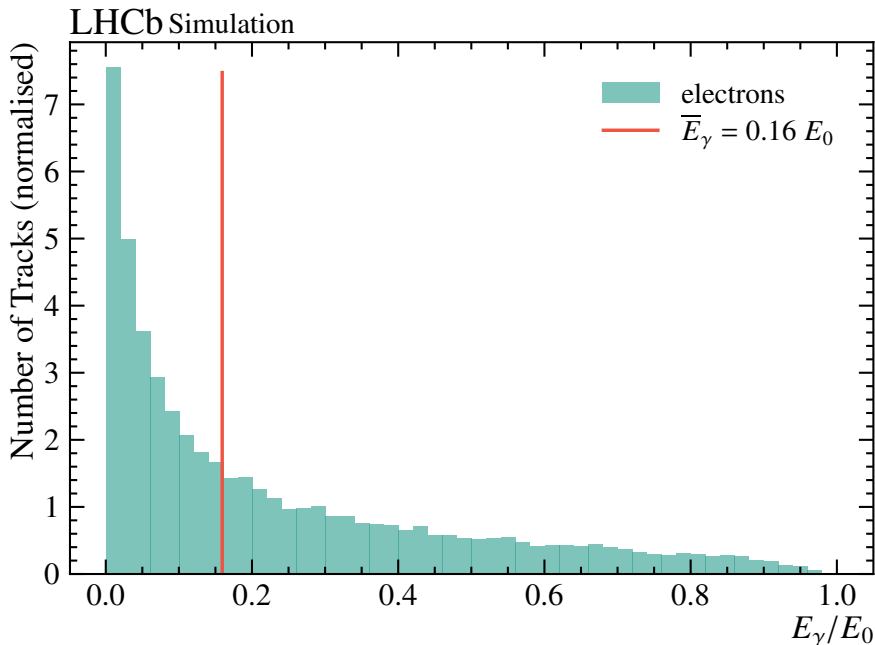

Figure 6.7: Amount of energy emitted by electrons in the VELO in units of the initial energy of the particles. The vertical line denotes the mean energy emission.

---

[4]The end of the VELO has been chosen as reference.

for efficiency plots are redefined to take energy emissions in the VELO into consideration. This allows for the efficiency plots to reflect a more precise measure of performance regarding the tracking algorithms. The efficiencies in dependence of the default and redefined momenta are depicted in Figure 6.8. There is no change in the integrated efficiencies, which is expected since both numerator and denominator in Equation 4.1 remain unchanged. Moreover, the momentum distribution is shifted towards lower momenta, and the depicted efficiencies change according to the particles that now occupy a given momentum region. The biggest change in efficiency is noticeable in the lower-momentum regime, where the efficiencies are higher using the momentum at the end of the VELO. However, the effect of photon emissions in the VELO on the efficiency plots appears to be relatively small compared to the difference of electron to non-electron performance, as shown in Figure 4.1. Although bremsstrahlung in the VELO is not the main reason for the efficiency difference, it accounts for around $1\% - 2\%$. A comparison plot for the Seed track efficiencies can be found in Figure A.11.

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

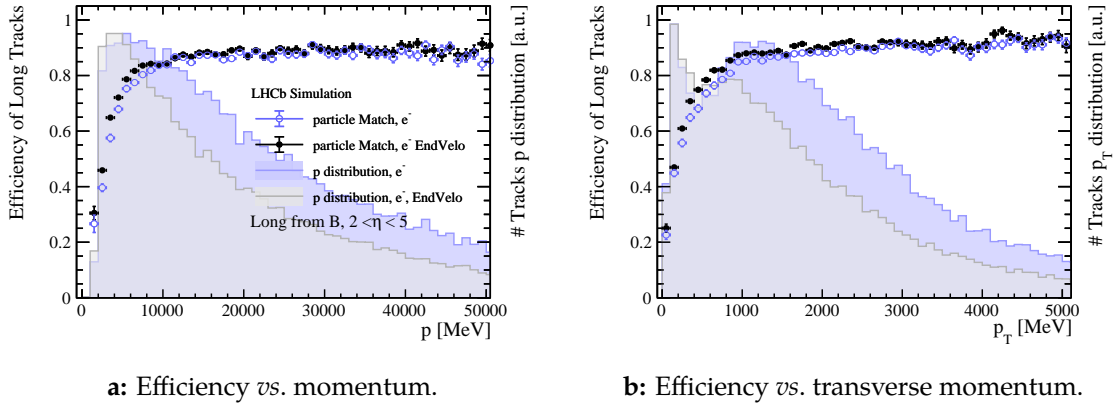**Figure 6.8:** Track finding efficiency of the baseline Matching algorithm with VELO and T input tracks for simulated Long tracks originating from a B meson with regard to the total and transverse momentum. The distribution of the true variable values at creation is shown in blue and the values at the end of the VELO are shown in gray. The corresponding efficiencies are depicted in blue and black respectively.

7   The Matching Algorithm

This section describes the Matching algorithm at LHCb. Since VELO and T tracks are created independently of each other, an algorithm is implemented to take these track segments and try to combine them into Long tracks using an artificial neural network.

7.1   Simplified Track Model

Describing the trajectory of a charged particle in a magnetic field is a solved problem. However, the issue is that there is not enough time in the trigger to solve the differential equation. So instead of solving a complex problem, the problem is simplified as far as possible while maintaining as much information as possible. For that purpose, the hereinafter described model is used. It allows a simpler approximation of the trajectory. The track state vectors of the VELO and T track segments are

$$\boldsymbol{s}(z_V) = \begin{pmatrix} x_V, & y_V, & t_x, & t_y, & \pm q/p \end{pmatrix}^\top_{z_V} \tag{7.1}$$

at the end of the VELO, z_V, and

$$\boldsymbol{s}(z_T) = \begin{pmatrix} x_T, & y_T, & t_x^{\text{EndT}}, & t_y^{\text{EndT}}, & q/p \end{pmatrix}^\top_{z_T} \tag{7.2}$$

at the end of the SciFi, z_T, respectively.

The Matching algorithm uses a simplified track model on the basis of the Optical Model method [32]. The magnetic field is approximated as a thin lens refracting tracks like light rays, such that it is possible to parameterise the state transfer from upstream to downstream of the magnet as two tangents on the bending arc. The exact path of particles through the magnet is not relevant to track finding at LHCb, since there are no tracking sub-detectors in the magnet, making this approach feasible. For a complete definition of the simplified track model, see Ref. [7]. The optical centre of the magnet, illustrated in Figure 7.1, is the intersection between the trajectory tangents upstream and downstream of the magnet. It is also referred to as the *magnet kick position* and is defined by

$$z_{\text{mag}} = \frac{x_V - t_x z_V - x_T + t_x^{\text{EndT}} z_T}{t_x^{\text{EndT}} - t_x} \tag{7.3}$$

The slope differences are defined as $\Delta t_x^{\text{match}} = t_x^{\text{EndT}} - t_x$ and $\Delta t_y^{\text{match}} = t_y^{\text{EndT}} - t_y$ for the x and y components.

However, because of the fringe fields (see Figure 2.2), as well as the material interactions described in Section 5, charged particles do not traverse the detector in a straight line, as illustrated in Figure 7.2. The magnet kick position in the Matching algorithm is thus
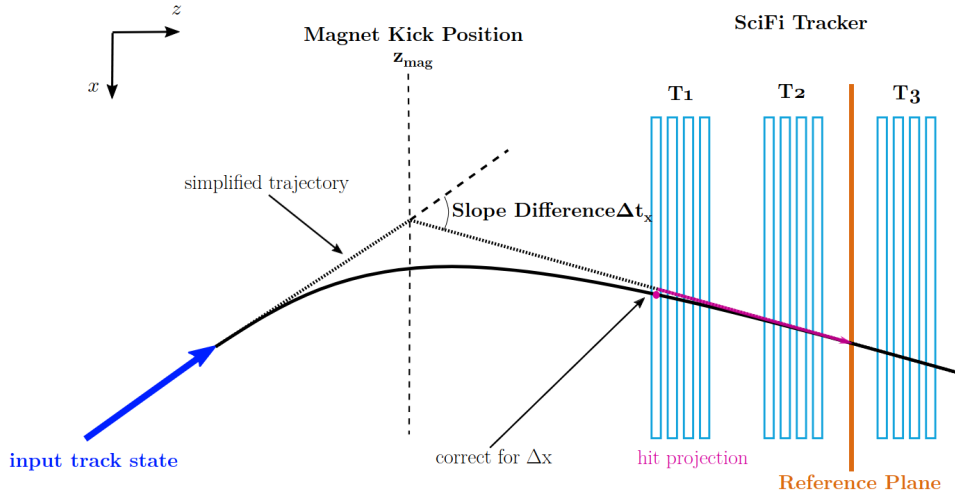
Figure 7.1: Optical Model method used to describe a trajectory through the magnet [7]. The hit projections onto a reference plane, and the Δx correction are merely used in Forward tracking.

estimated by

$$z_{\text{mag}}^{\text{match}} = c_0 + c_2|x_T| + c_3 t_x^2 + |\Delta t_x^{\text{match}}|(c_1 + c_4|\Delta t_x^{\text{match}}|) \tag{7.4}$$

with the coefficients listed in Table A.1.

Before using estimates such as this, the intention behind such parameterisations should be pointed out. A parameterisation is an estimate of some property that is unknown during track reconstruction with real data. Therefore, using simulated particles, the *true* values are estimated using only information given by the track state vectors, or that is otherwise available during reconstruction with data. Often, linear regression is used to find a polynomial for that purpose. Parameterisations are used to provide previously inaccessible information that help separate true from fake tracks without having to solve the differential equation. More complex properties or variables with more non-trivial dependencies become increasingly difficult to parameterise adequately. Sometimes estimates have to be precise but not especially accurate, meaning a small variance is the main goal.

7.2   Baseline Matching

Matching uses several variables defined specifically to quantify the level of agreement between VELO and T tracks. Then, a machine learning classifier can be evaluated on the variables to assess the matching quality in a combined estimate.

The algorithm's structure is simple: a loop over all VELO track state vectors with a nested loop over all T track state vectors, so that the input variables for the neural net-
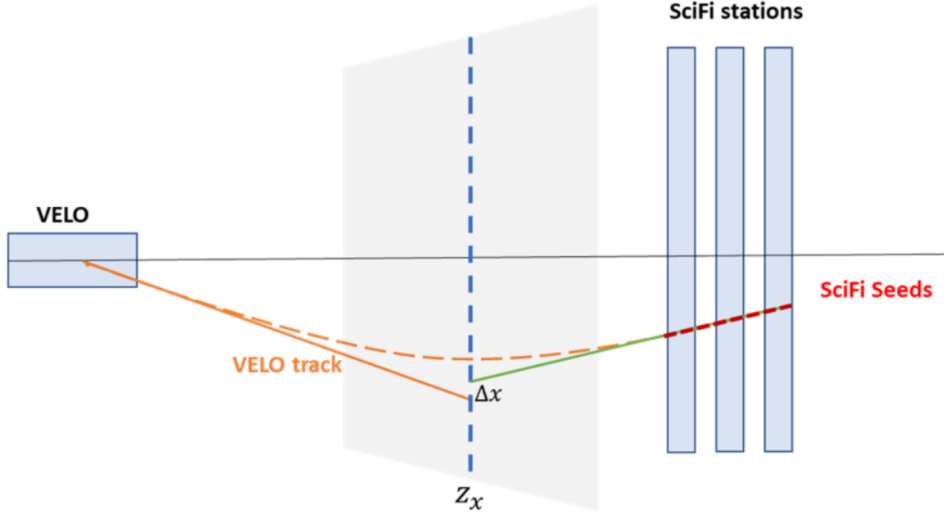
Figure 7.2: Kink approximation in the $z_x = z_{\text{mag}}^{\text{match}}$ plane used in the Matching algorithm [22]. Δx denotes the matching distance D_x^{match}.

work can be calculated for all possible combinations. Alongside the slope differences, the square of the VELO slopes $t_x^2 + t_y^2$ is also used as input for Matching. The tracks are then extrapolated in y onto a reference plane at $z_{\text{match}} = 10\,\text{m}$ to calculate the distance in the y direction, and onto a reference plane at the magnet kick position $z_{\text{mag}}^{\text{match}}$, parameterised in Equation 7.4, in the x direction. The x distance, depicted in Figure 7.2, is defined by

$$D_x^{\text{match}} = \left| x_T - x_V + t_x^{\text{EndT}}(z_{\text{mag}}^{\text{match}} - z_T) - t_x(z_{\text{mag}}^{\text{match}} - z_V) \right| \tag{7.5}$$

and the y distance by

$$D_y^{\text{match}} = \left| y_T - y_V + t_y^{\text{EndT}}(z_{\text{match}} - z_T) - t_y(z_{\text{match}} - z_V) - y_{\text{corr}}^{\text{match}} \right| \tag{7.6}$$

with a y correction

$$y_{\text{corr}}^{\text{match}} = t_y(c_0|\Delta t_x^{\text{match}}|^2 + c_1|\Delta t_y^{\text{match}}|^2) \tag{7.7}$$

for which the coefficients can be found in Table A.2. Both matching distances are subsequently used to define a combined quality measure

$$\chi_{\text{match}}^2 = \frac{D_x^2}{\delta x^2 + t_{\delta x}^2 |\Delta t_x^{\text{match}}|^2} + \frac{D_y^2}{\delta y^2 + t_{\delta y}^2 \left(t_x^2 + t_y^2 \right)} + \frac{|\Delta t_y^{\text{match}}|^2}{\text{var}(t_y^{\text{EndT}})} \tag{7.8}$$

where $\delta x = 8\,\text{mm}$ and $t_{\delta x} = 80\,\text{mm}$ correspond to the x uncertainty, while $\delta y = 6\,\text{mm}$ and $t_{\delta y} = 300\,\text{mm}$ correspond to the y uncertainty. Additionally, the inverse variance of the T track's y slope at the end of the SciFi tracker is $\text{var}(t_y^{\text{EndT}})^{-1} = 625$. This χ_{match}^2, even if named as such, is strictly speaking not a χ^2 [7].

Table 7.1: Matching MLP input variables.

Variable	Preselection
χ^2_{match}	< 15
D_x	$< 250 \, \text{mm}$
D_y	$< 250 \, \text{mm}$
$\left\| \Delta t_x^{\text{match}} \right\|$	< 1.5
$\left\| \Delta t_y^{\text{match}} \right\|$	< 0.15
$t_x^2 + t_y^2$	

The network's architecture remains simple, utilising a multi-layer perceptron, for which the training is done using the Toolkit for Multivariate Data Analysis (TMVA) [33]. The basics of an MLP are described in Section 2.5.

The MLP used for the Matching algorithm is described in the following. The network weights are adjusted with Cross-Entropy[5] as the loss function, using backpropagation, as explained in Section 2.5. Furthermore, the neuron activation function is a rectifier linear unit $\text{ReLU}(x) = \max(0, x)$. The network response is projected onto the interval $[0, 1]$. The MLP has two hidden layers, the first of which consists of $N + 2$ neurons, and the second of N neurons, where N is the number of input variables for the Matching, listed in Table 7.1.

Electrons are currently *excluded* from the training data of the baseline Matching algorithm at LHCb [7]. The distributions of the input variables, excluding electrons, are illustrated in Figure A.1. The current cut on the response value is 0.215, resulting in rejecting 84% of the fake tracks, while keeping 97% of the true matches. The network response is depicted in Figure A.2. If there is more than one pair, the pair with the highest response value is chosen, along with other pairs that deviate less than 0.1 from the best one.

The Matching algorithm was designed to be used in the trigger, which requires it to be computationally cheap. It must provide reasonably accurate results within a short amount of time.

---

[5]$CE(p) = -\log p - (1 - p)\log(1 - p)$ where $p = S/(S + B)$ is the purity of a node, given by the ratio of signal events to all events in that node.

7.3   Improved Matching for Electrons

The baseline Track Matching, described in Section 7.2, serves as the basis for the modified Matching algorithm outlined in this section. Hereinafter, all MLPs have been trained using simulated $B^0 \to K^* e^+ e^-$ events, while the tests and efficiency plots have been obtained using simulated $B^0 \to K^* (J/\psi \to e^+ e^-)$ events.

The separation of training and test samples is a simple but effective way to check for overfitting, while the decay was chosen, in part, because it provides high-energy electrons and is relevant in research regarding the standard model.

7.3.1   Parameterisations

The Matching algorithm utilises parameterisations to estimate certain variables, such as $z_{\mathrm{mag}}^{\mathrm{match}}$ and $y_{\mathrm{corr}}^{\mathrm{match}}$. Currently, electrons are excluded from all parameterisations used during track finding. Consequently, the estimates are optimised for not-electrons, *i.e.* the predicted values show large deviations from the true variable values for electrons, as shown with the example of the magnet kick position in Figure 7.3.

This is addressed by reparameterising the magnet kick position and y correction to fit electrons using linear regression. In the case of $y_{\mathrm{corr}}^{\mathrm{match}}$, adding a *correction* term and finding new coefficients has sufficed to provide estimates with decent accuracy for electrons. The fit has a root mean squared error (RMSE) of $10\,\mathrm{mm}$ and an r-squared (R2) of 0.863. The y correction for electrons is defined by

$$y_{\mathrm{corr}}^{\mathrm{match}} = t_y \left(c_2 |\Delta t_x|^2 + c_3 |\Delta t_y|^2 \right) + \Delta t_y \left(c_0 + c_1 \Delta t_x \right) \tag{7.9}$$

with the coefficients listed in Table A.4. The distribution, as well as the deviation from the
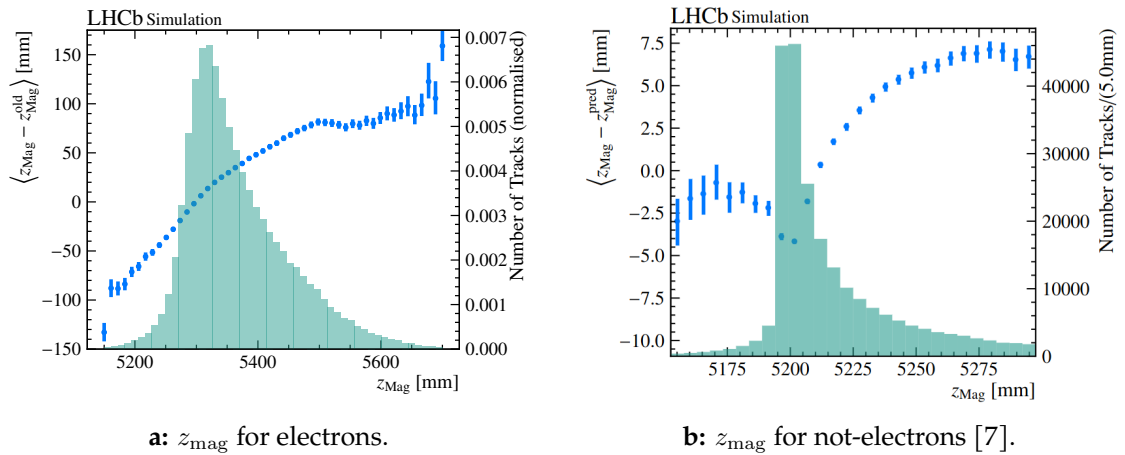

a: z_{mag} for electrons.

b: z_{mag} for not-electrons [7].

Figure 7.3: Regression plot of the baseline z_{mag} parameterisation for (a) electrons, and (b) not-electrons, where $z_{\mathrm{Mag}}^{\mathrm{old}}$ is the predicted value using the current default parameterisation.
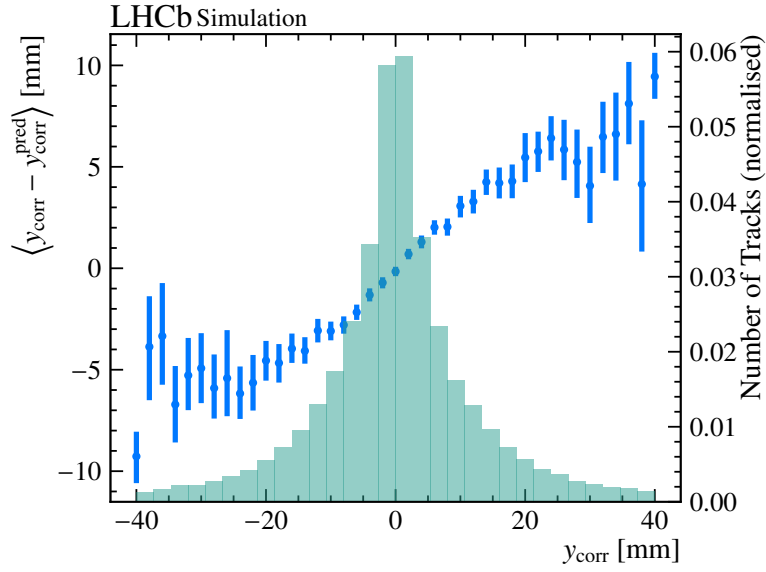
Figure 7.4: Regression plot for new $y_{\text{corr}}^{\text{match}}$ parameterisation for electrons.

true value, is shown in Figure 7.4.

The distribution of the magnet kick position of electrons is broader than that of other final-state particles (Figure 7.5). Since the currently used estimate disregards electrons, it tries to fit the predictions for electrons into a narrower distribution leading to large deviations at the tails. Using the same variables, a new electron-specific parameterisation can be found. The magnet kick position is then defined by

$$
\begin{aligned}
z_{\text{mag}}^{\text{match}} =& c_0 + c_3 t_x^2 + c_7 t_y^2 + |\Delta t_x| \left(c_1 + c_{11} t_y^2 \right) + \Delta t_x^2 \left(c_5 + c_{10}|x_T| \right) \\
&+ \Delta t_x t_x \left(c_4 + c_9|x_T| \right) + |x_T| \left(c_2 + c_8 t_x^2 + c_6|x_T| + c_{12}|x_T|^2 \right)
\end{aligned}
\tag{7.10}
$$

with the coefficients listed in Table A.3. The regression plots depicted in Figure 7.6c and Figure 7.6d show that the magnet kick position is a property of the track slopes at the end of the VELO, but also of the slope difference from before and after the magnet, see Figure 7.6b. The histograms show the correlations between the track properties and z_{mag}. While the new parameterisation provides better estimates than what is currently used, the inaccuracies are still considerably larger than for other particles. The largest deviations can be seen at the left tail of the distribution, but only affecting a limited number of particles. The regression plot for the new parameterisation following Equation 7.10 is shown in Figure 7.6a. Most electrons' z_{mag} estimate deviates less than $50\,\text{mm}$ from the true value, which is tolerable, considering the previously used values. The fit has an RMSE of $37\,\text{mm}$ and an R2 score of 0.815.
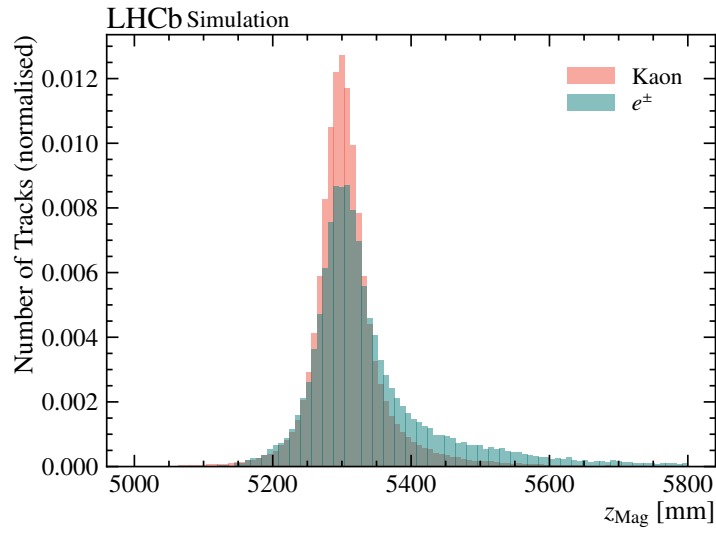
Figure 7.5: Distribution of the magnet kick position z_{mag} (Equation 7.3). Electrons are depicted in teal, kaons in red.

a: z_{mag} model

b: z_{mag} vs. Δt_x

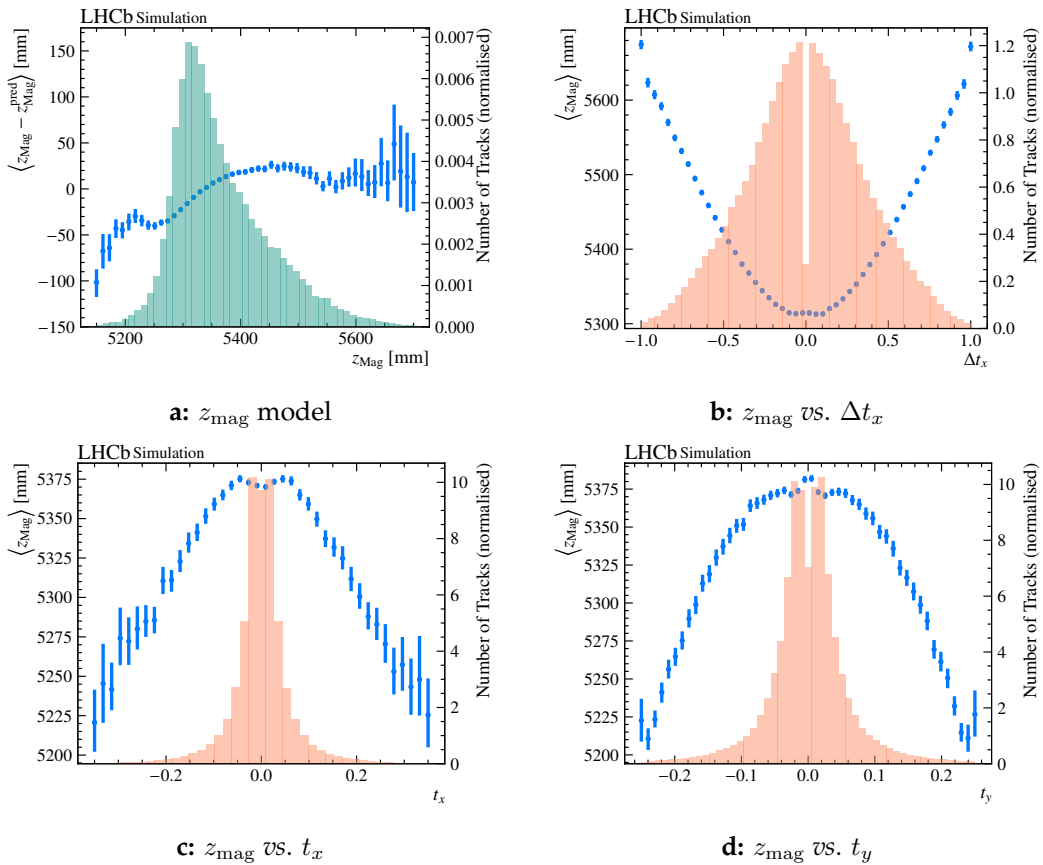**c:** z_{mag} vs. t_x

d: z_{mag} vs. t_y

Figure 7.6: Regression plots for new z_{mag} parameterisation for electrons, and the variables used to determine estimates thereof.

7.3.2   Seed Track Preselection

Matching tries to combine VELO and T tracks[6] to create Long tracks. Reducing the number of input tracks by removing those that likely do not belong to electrons should improve reconstruction performance. This would result in there being fewer track combinations to consider, which contributes to a decrease in the amount of fake tracks as well. T tracks are reconstructed from detector hits in the SciFi tracker, which is located in front of the RICH2 and ECAL, as shown in Figure 2.1. Since electrons and photons primarily interact with the electromagnetic calorimeter, see Figure 2.6, the information provided by the ECAL can be used to determine which T tracks are likely to be of particles that also caused electron showers in the ECAL[7]. The functionality of the ECAL is addressed in Section 2.4, whereas the ECAL filter is briefly described in Section 5.3.

7.3.3   Electron Track Matching

The underlying architecture of the network remains identical to the baseline MLP described in Section 7.2 since the use of different architectures has not led to noticeable performance improvements. Even though other variables have been considered and tried as well, the default input variables have proven to be robust. Moreover, the variables are defined using new parameterisations (Section 7.3.1). This directly affects both matching distances D_x, D_y, and the combined quality measure χ^2_{match}. The preselection cuts on the input variables can be found in Table 7.2. They have been adjusted *s.t.* the network performs optimally. Since there are some outlying values, a cut is also placed on the magnet kick position, confining it to $5100 \, \mathrm{mm} < z^{\mathrm{match}}_{\mathrm{mag}} < 5700 \, \mathrm{mm}$. The MLP is trained with simulated electron Long tracks as signal, and fake tracks of all types of particles as back-

Table 7.2: Matching MLP input variables for electron Matching.

Variable	Preselection
χ^2_{match}	< 15
D_x	$< 300 \ \mathrm{mm}$
D_y	$< 300 \ \mathrm{mm}$
$\left\| \Delta t^{\mathrm{match}}_x \right\|$	< 2.0
$\left\| \Delta t^{\mathrm{match}}_y \right\|$	< 0.15
$t_x^2 + t_y^2$	

---

[6]T tracks are also called Seed tracks since they are obtained through the Hybrid Seeding algorithm. They seed the Matching algorithm.

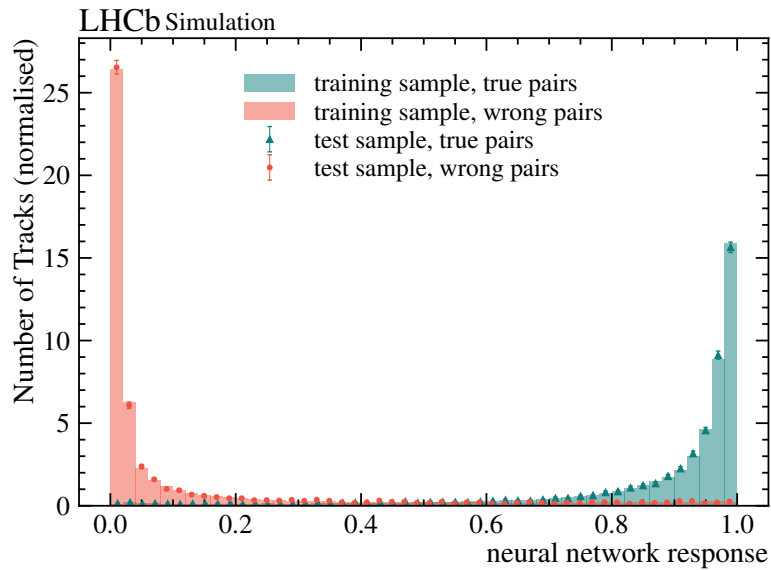[7]The ECAL filter was not implemented by the author of this thesis.

Figure 7.7: Output of the Matching MLP. The bars denote the training sample, and the data points denote the test sample. The distribution of true pairs is teal with triangles, and of wrong pairs red with circles.

ground. It has also been considered to only use fake tracks that possess electron T tracks as background, for the explicit use with the Seed track preselection. However, the Matching MLP performed most reliably using all types of fake tracks. Imposing such a cut on the background sample, while possible, is only feasible in conjunction with a near-perfect T-track filter.

The network response for the training and test sample is depicted in Figure 7.7, which is used to check for overfitting, described in Section 2.5. The network models the data well with a Kolmogorov-Smirnov test score of 0.563 for true pairs and 0.489 for false pairs.

The cut on the response value determines the amount of false matches that are rejected, and the amount of true matches that are kept. The cut is set to 0.25, thereby rejecting 84.1% of fake tracks, while keeping 97% of the true matches. If there is more than one pair, the pair with the highest response value is chosen, along with other pairs that deviate less than 0.1 from the best one. The distributions of the input variables are shown in Figure A.3.

7.4 Performance Comparison

The performance is evaluated on the track-finding efficiency ε and the ghost rate r_{fake}, as explained in Section 4. The Matching algorithm uses VELO and T tracks as input. Referring to Section 4.2, the track-finding efficiency of electron Long tracks must be lower than the T-track reconstruction efficiency, given in Table A.6. By extension, it is also lower than the efficiency of not-electron Long tracks as well (Table 7.3).

The goal is to sustain a ghost rate of below 20%. That may not be possible for electrons, while also maintaining a high efficiency because reducing the ghost rate also lowers the efficiency. It should be pointed out that the ghost rate is a measure overarching all categories. This means that a low ghost rate alone is not an indicator of the amount of ghosts created. Since baseline Matching reconstructs an enormous amount of not-electron Long tracks with a very low ghost rate, the number of fake tracks stemming from mismatching electron track segments is not as prominently featured in the ghost rate because there are many more *reconstructible* not-electron than electron tracks. So it is possible to have an inflated ghost rate while actually reducing the number of fake tracks.

This section presents the electron track-finding efficiencies using baseline Matching and the electron Matching algorithm. All efficiencies have been obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events. The performance of the baseline Matching algorithm is also evaluated in Ref. [7].

7.4.1 Baseline

The integrated efficiencies of the baseline Matching algorithm are listed in Table 7.3 for several selection categories. The fake track fraction is 17%. The track reconstruction of not-electrons performs with an efficiency of around 94% for Long tracks $p > 5\,\text{GeV}$ originating from a B meson. In contrast, electron tracks are reconstructed with around 86%

Table 7.3: Integrated Matching efficiencies obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events using baseline Matching. The efficiency in dependence on various kinematic variables of electron and not-electrons is plotted in Figure A.5.

Category	$e^\pm \ \varepsilon[\%]$	not-$e^\pm \ \varepsilon[\%]$
Long	64.02	87.38
Long from B	83.46	90.35
Long $p > 5\,\text{GeV}$ from B	86.22	94.31
Long $p > 3\,\text{GeV}\ p_T > 0.5\,\text{GeV}$ from B	87.72	94.54
Fake track fraction r_{fake}		17.44%

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

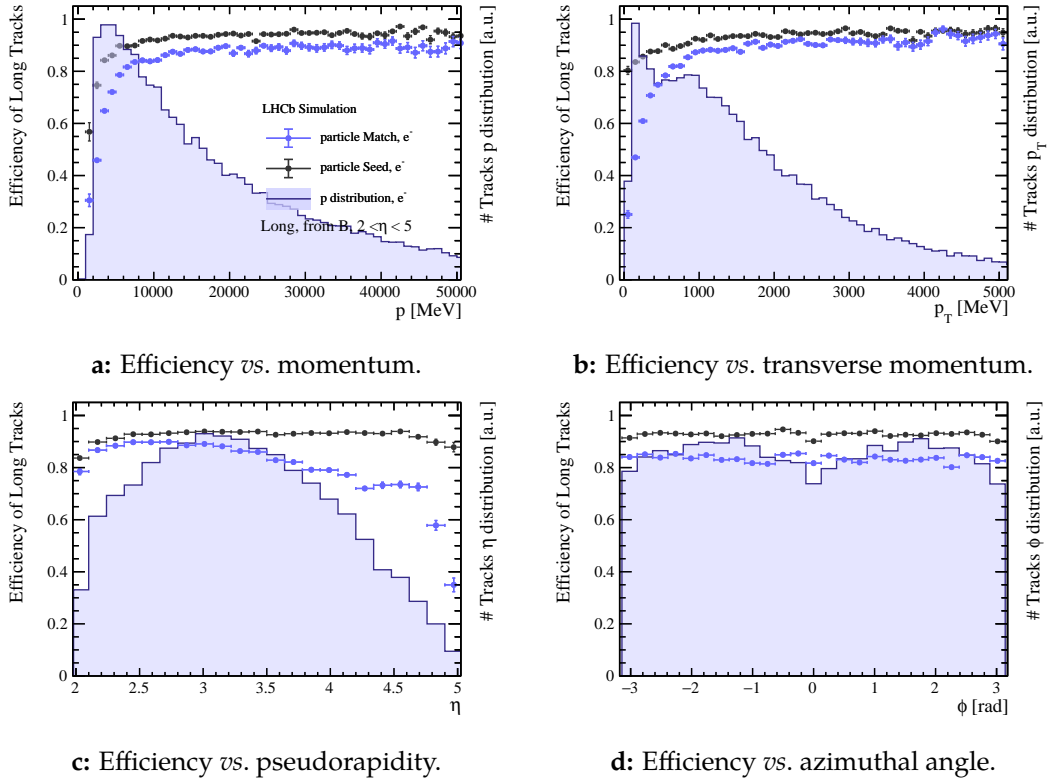**d:** Efficiency *vs.* azimuthal angle.

Figure 7.8: Track finding efficiency of the Seed tracks and the baseline Matching algorithm with VELO and T input tracks for simulated Long tracks originating from a B meson with regard to various kinematic variables. The variables' distributions are represented in blue. Efficiencies of the Seed tracks are depicted in black and the Matching algorithm in blue. The pseudorapidity and azimuthal angle distributions represent the true, simulated variable values at creation. The momenta distributions have been adjusted according to Section 6.2 (for unaltered momentum spectrum see Figure A.8).

efficiency denoting a significant difference in performance. This is also reflected in the efficiency plots in dependence on various kinematic variables in Figure 7.8 when taking into account that the Seed track efficiency for electrons is lower than the Matching efficiency for not-electron tracks. The distributions of the plotted efficiencies resemble the Best-Long efficiencies depicted in Figure 4.1 in Section 4.2, displaying the same detector-related characteristics.

7.4.2 No Preselection

Firstly, the electron Matching algorithm can in theory be used with the same input tracks as baseline Matching, *i.e.* all VELO and T tracks are given as input tracks, and there is no specific track selection prior to the algorithm. Since the algorithm has not been trained on not-electron tracks, a high fake track fraction is expected. The efficiencies are listed in Table 7.4. Electron tracks in the *Long* category experience an efficiency increase by

Table 7.4: Integrated Matching efficiencies obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events, using electron Matching with no preselection.

Category	$e^\pm \ \varepsilon[\%]$
Long	76.17
Long from B	86.60
Long $p > 5\,\mathrm{GeV}$ from B	88.78
Long $p > 3\,\mathrm{GeV}\ p_T > 0.5\,\mathrm{GeV}$ from B	89.68
Fake track fraction r_{fake}	33.74%

more than 10%, whereas the efficiency of tracks originating from a B meson increases by $2\% - 3\%$.

The increase in efficiency is accompanied by a ghost rate of around 34%, which is almost twice as high as in the baseline Matching algorithm. This is in part because the network is trained on electron tracks as signal and fake tracks as background. Therefore, it is unfamiliar with non-electron tracks. The provided VELO and T track segments are matched to pairs that best correspond to electrons' trajectories. However, the majority of input tracks are not of electrons, so the ghost rate is expected to be inflated.

7.4.3 Perfect VELO and T Track Selection

Using simulated events enables the perfect selection of track segments. The algorithm is given electron VELO and T track segments as input tracks to assess the actual efficiency ceiling. The efficiencies and fake track fraction are listed in Table 7.5. The fake track fraction is low at around 14% corresponding to the network-response cut chosen.

Table 7.5: Integrated Matching efficiencies obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events, using electron Matching with perfect VELO and T track selection.

Category	$e^\pm \ \varepsilon[\%]$
Long	78.33
Long from B	87.90
Long $p > 5\,\mathrm{GeV}$ from B	90.05
Long $p > 3\,\mathrm{GeV}\ p_T > 0.5\,\mathrm{GeV}$ from B	90.86
Fake track fraction r_{fake}	13.96%

Table 7.6: Integrated Matching efficiencies obtained from simulated $B^0 \rightarrow K^*(J/\psi \rightarrow e^+e^-)$ events, using baseline Matching with perfect VELO and T track selection.

Category	$e^{\pm} \, \varepsilon[\%]$
Long	65.39
Long from B	84.45
Long $p > 5\,\mathrm{GeV}$ from B	87.19
Long $p > 3\,\mathrm{GeV}$ $p_T > 0.5\,\mathrm{GeV}$ from B	88.66
Fake track fraction r_{fake}	10.95%

The track-finding efficiency shows significant improvements with several categories now reaching efficiencies above 90%. Considering the highest possible efficiency using VELO and Seed tracks as input is about 93% in those categories, the algorithm performs very well. It should also be pointed out that applying a perfect selection filter on the input tracks of the baseline Matching algorithm does not yield comparable efficiency improvements (Table 7.6).

The efficiency in dependence on various kinematic variables is presented in Figure 7.9. While the new algorithm performs better than the baseline at all momenta, the efficiency gain is especially large for low-momentum tracks, as can be seen in Figure 7.9a and Figure 7.9b. In the higher momentum regime, the efficiency gain is larger than the gap between the electron Matching efficiency and the Seed track efficiency. The efficiency as a function of pseudorapidity and azimuthal angle is depicted in Figure 7.9c and Figure 7.9d respectively. The increase in efficiency shows little variance with regard to the azimuthal angle. On the other hand, the efficiency difference between baseline and electron Matching becomes greater with increasing η. The baseline algorithm performs close to the efficiency ceiling at low η, but the distance to the performance ceiling widens as η increases. The new algorithm performs much better in the higher pseudorapidity region, damping the efficiency's downward trend significantly. A decrease in efficiency for $\eta \geq 4$ is expected because of detector material, as explained in Section 4.2. The efficiency as a function of the number of primary vertices, depicted in Figure 7.9e, is relatively flat until the drop in efficiency for tracks with $\mathrm{nPV} \geq 13$. This may indicate that the algorithm does not perform well for busy events. However, the Seed track efficiency also decreases for busy events, which seems to be the actual reason for the drop in performance, and not the Matching algorithm itself.
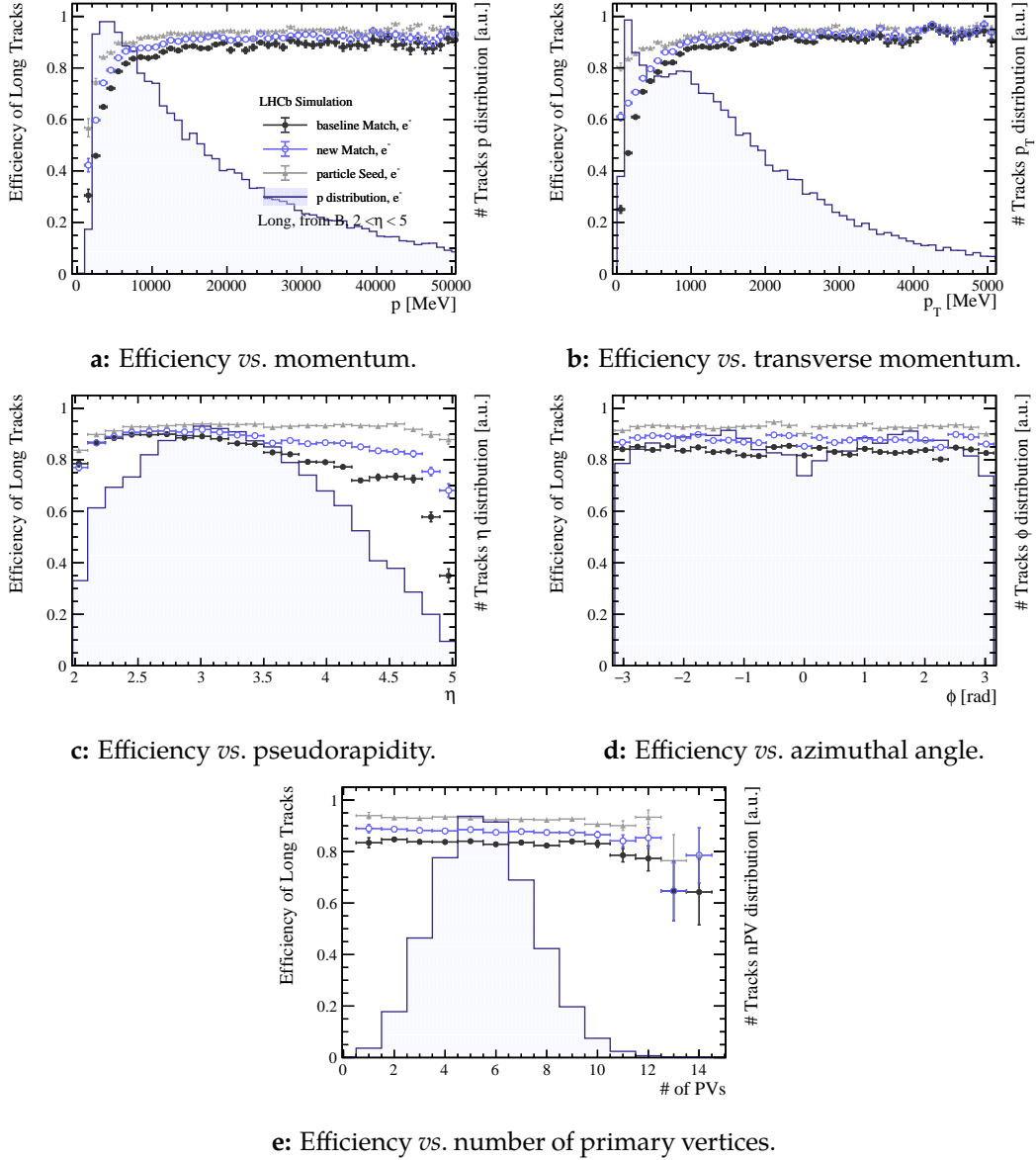
a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.

e: Efficiency *vs.* number of primary vertices.

Figure 7.9: Track finding efficiency of the Seed tracks and the electron Matching algorithm with perfect selection VELO and T input tracks for simulated Long tracks originating from a B meson with regard to kinematic variables and the number of primary vertices. The variables' distributions are represented in blue. Electron Seed track efficiencies are depicted in grey triangles. The black marker with full circles denotes electron tracks of the baseline Matching algorithm, and the blue marker with empty circles denotes electron tracks of the new (electron) Matching algorithm. The pseudorapidity, azimuthal angle, and number of primary vertices distributions represent the true, simulated variable values at creation. The momenta distributions have been adjusted according to Section 6.2 (for unaltered momentum spectrum see Figure A.9).
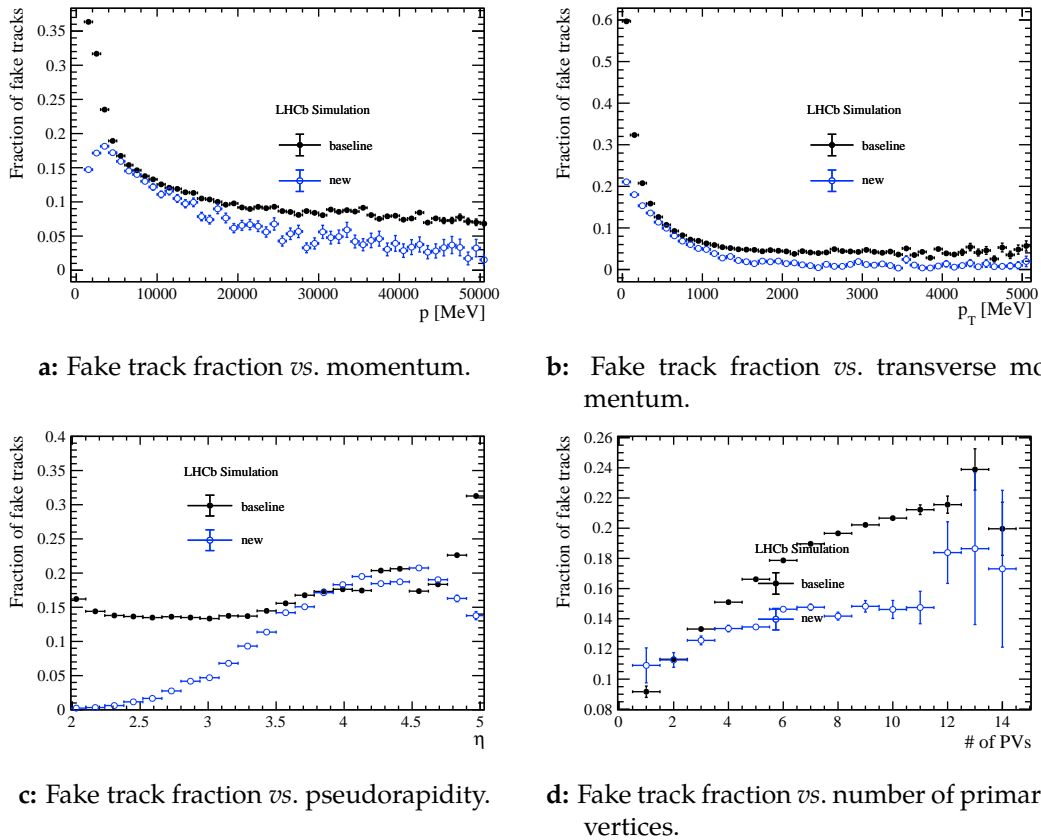
a: Fake track fraction *vs.* momentum.

b: Fake track fraction *vs.* transverse momentum.

c: Fake track fraction *vs.* pseudorapidity.

d: Fake track fraction *vs.* number of primary vertices.

Figure 7.10: Fake track fraction of the electron Matching algorithm with perfect VELO and T track selection of input tracks in dependence on kinematic variables and the number of primary vertices.

The fake track fractions as a function of the total and transverse momentum are presented in Figure 7.10a and Figure 7.10b, respectively. The electron Matching algorithm maintains a lower ghost rate compared to baseline Matching, especially at low momenta, where the improved algorithm produces ghosts at less than half the rate. The ghost rate of the new algorithm in dependence on the pseudorapidity, depicted in Figure 7.10c, is very low for tracks with $\eta \leq 3$ but then rises to around the same rate as the default algorithm, before dropping lower again above $\eta > 4.5$.

7.4.4   ECAL Filter for T Tracks

The ECAL filter, described in Section 5.3, is used to select T tracks that are likely to be of electrons. While this reduces the number of possible track combinations, it also leaves not-electron VELO tracks without their corresponding T tracks. This causes the algorithm to match false track pairs that would otherwise have been dismissed because a better match is available, resulting in a high ghost rate. This could be managed by preselecting VELO tracks similarly to T tracks using the ECAL. This is briefly addressed in Section 5.3. Using perfect VELO track selection and the ECAL filter together yields identical track-finding efficiencies for electrons while reducing the ghost rate by half, from around 53% to 26%.

This configuration provides Long tracks with 76% efficiency, denoting an increase of 12%, and Long tracks originating from a B meson with efficiencies that are around $2\% - 3\%$ higher than in the baseline Matching algorithm. The efficiencies and ghost rates are listed in Table 7.7, and presented in Figure 7.11 and Figure 7.12.

The track-finding efficiency follows the same trends as in Section 7.4.3, with a slightly lower performance gain. The most significant increases are noticeable at lower momenta but even high-momentum tracks perform better in the electron Matching algorithm than in the baseline algorithm. Furthermore, tracks with $\eta > 3$ are also more efficiently reconstructed, whereas a slight drop in the efficiency is visible around $\eta < 2.2$.

Table 7.7: Integrated Matching efficiencies obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events, using electron Matching with implemented ECAL filter for T track selection. A fake track fraction marked with *perfect VELO selection* implies that only VELO tracks associated with electrons are selected, while one marked with *ECAL filter* implies that T tracks are selected using the implemented ECAL filter described in Section 5.3.

Category	$e^{\pm}\ \varepsilon[\%]$
Long	76.41
Long from B	86.40
Long $p > 5\,\text{GeV}$ from B	88.55
Long $p > 3\,\text{GeV}\ p_T > 0.5\,\text{GeV}$ from B	89.39
Fake track fraction	r_{fake}
ECAL filter	52.48%
Perfect VELO selection + ECAL filter	26.37%

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.

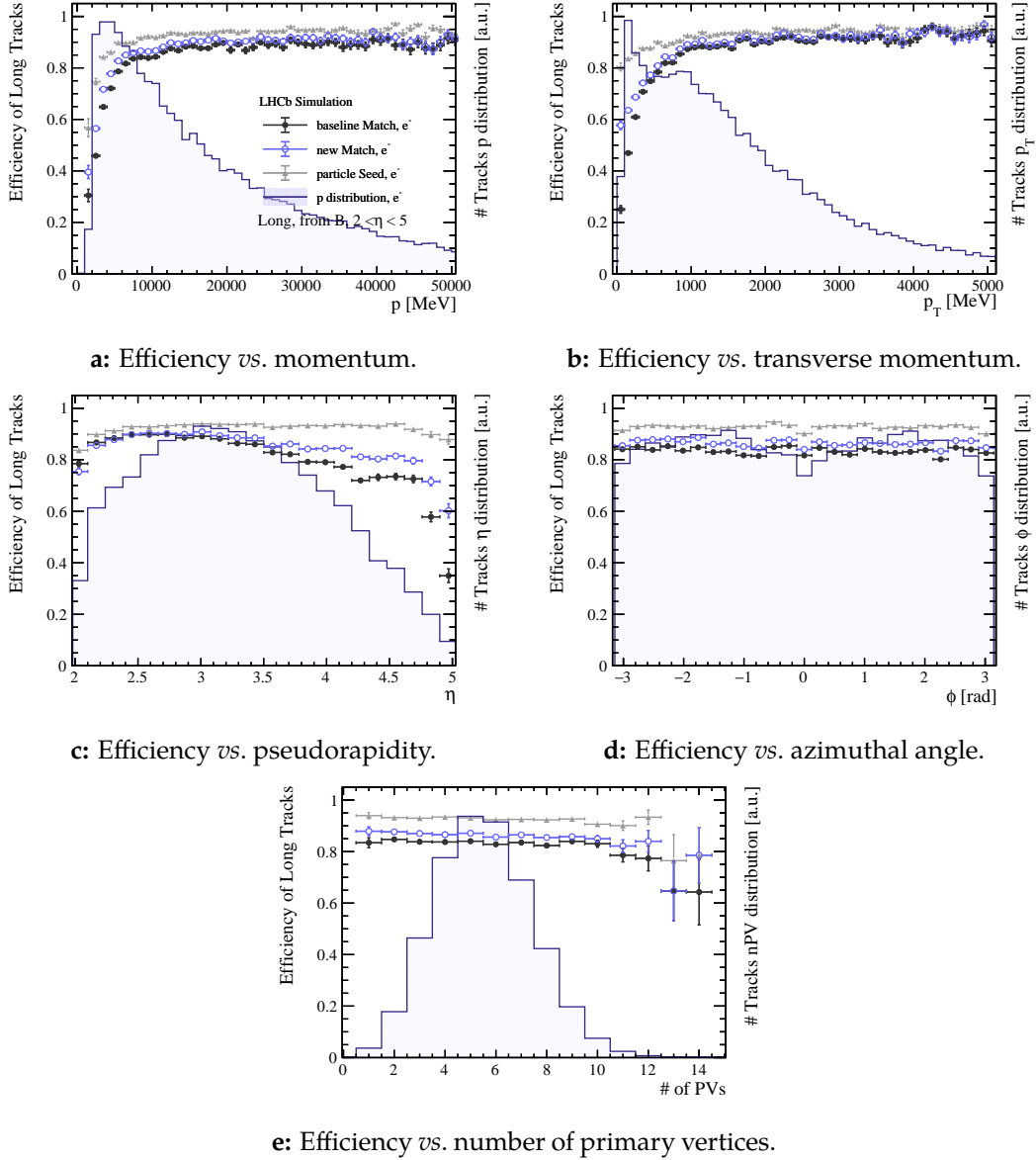**e:** Efficiency *vs.* number of primary vertices.

Figure 7.11: Track finding efficiency of the Seed tracks and the electron Matching algorithm with VELO and ECAL-filtered T input tracks for simulated Long tracks originating from a B meson with regard to kinematic variables and the number of primary vertices. The variables' distributions are represented in blue. Electron Seed track efficiencies are depicted in grey triangles. The black marker with full circles denotes electron tracks of the baseline Matching algorithm, and the blue marker with empty circles denotes electron tracks of the new (electron) Matching algorithm. The pseudorapidity, azimuthal angle, and number of primary vertices distributions represent the true, simulated variable values at creation. The momenta distributions have been adjusted according to Section 6.2 (for unaltered momentum spectrum see Figure A.10).

a: Fake track fraction *vs.* momentum.

b: Fake track fraction *vs.* transverse momentum.

c: Fake track fraction *vs.* pseudorapidity.

d: Fake track fraction *vs.* number of primary vertices.

Figure 7.12: Fake track fraction of the electron Matching algorithm with perfect VELO track selection and ECAL filtered T tracks as input tracks in dependence on kinematic variables and the number of primary vertices.

The fake track fraction is shown in Figure 7.12. The new algorithm performs at a ghost rate that is approximately 10% higher than baseline Matching, as is visible in Figure 7.12d. The fraction of ghosts is high at low momenta but decreases rapidly, before asymptotically approaching a lower limit. This is present in both Figure 7.12a and Figure 7.12b. However, while the ghost rate of the new algorithm as a function of the total momentum is continuously higher compared to baseline Matching, this is not the case for the ghost rate as a function of the transverse momentum. The new algorithm performs with almost 20% fewer ghosts at low $p_T \leq 200\,\text{MeV}$, and a slightly lower ghost rate above $p_T > 2000\,\text{MeV}$. The fake track fraction in dependence on the pseudorapidity, depicted in Figure 7.12c, is very low for tracks with $\eta \leq 3$ but rises with increasing pseudorapidity. The new algorithm produces around 15% more ghosts above $\eta > 4.3$ than default Matching.

7.5   Summary and Integration into LHCb's Real-Time Analysis

The Matching algorithm is tasked with finding Long tracks at LHCb. It is run in the HLT2 event reconstruction sequence using VELO and Seed tracks as input. The baseline Matching algorithm is designed to be most efficient with not-electron tracks. In order to achieve this, electrons are explicitly excluded from all parameterisations (see Section 7.1) and training samples [7]. As a result, electron tracks underperform by a significant margin compared to tracks of other particles. This issue is currently not addressed in LHCb's real-time analysis sequence.

The electron Matching algorithm utilises new parameterisations, optimised for electrons (Section 7.3.1). The newly parameterised variables perform much better for electrons. However, they still show significantly larger deviations from the true values than other particles' predictions using the current parameterisations. This is due to the fallacious assumption that an electron's trajectory in a magnetic field can be adequately approximated using the simplified track model, described in Section 7.1. Electrons lose a significant portion of their energy due to bremsstrahlung (Section 5.1), resulting in more complex trajectories through the magnetic field. To address this issue properly, the simplified track model would have to be abandoned completely. The new neural network is trained on electron tracks exclusively, with adjusted preselection cuts on the input variables, listed in Table 7.2, and the magnet kick position.

There are several ways to integrate this dedicated electron reconstruction into the track-reconstruction sequence of HLT2. One possibility, proposed in Ref. [7], is to run it over *residual* VELO and T tracks, *i.e.* tracks unmatched during baseline Long-track reconstruction. This would, however, prevent the recovery of mismatched electron tracks. Alternatively, it might run parallel to baseline Matching, *s.t.* Forward tracking, baseline Matching, and electron Matching independently find Long tracks that are subsequently passed through the Kalman filter and clone removal process. In order to control the fake track fraction of the new algorithm, a VELO and T track preselection might be necessary, whereby only track segments of electrons are selected (Section 5.3).

8   Conclusion

A dedicated electron reconstruction for the real-time analysis in HLT2 is presented in this thesis. The track reconstruction sequence in LHCb's HLT2 is described, with a focus on the neural-network-based Matching algorithm. Simulated $B^0 \to K^* e^+ e^-$ and $B^0 \to K^*(J/\psi \to e^+ e^-)$ events are used.

The Matching input variables have been defined using electron-specific parameterisations thereby improving their accuracy for electrons. The underlying architecture of the network corresponds to that of the baseline Matching algorithm. The subsequent training of the neural network has been done on electron Long tracks from simulated $B^0 \to K^* e^+ e^-$ events. In order to control the fake track fraction, both VELO and T tracks undergo a preselection to dismiss not-electron track segments. For this, an already implemented ECAL filter is used, which matches electron-induced EM-showers to T tracks. Similarly, VELO tracks can be matched to photon-induced EM-showers, using an algorithm that is currently not available out-of-the-box and thus beyond the scope of this thesis. Therefore, VELO tracks are passed through a *perfect* filter instead.

The new electron reconstruction finds approximately 76% of electron Long tracks, marking an improvement by more than 12%, and around 86.4% of electron Long tracks originating from a B meson, denoting an efficiency gain of more than 2%. The two Long track categories, *Long $p > 5$ GeV from B* and *Long $p > 3$ GeV $p_T > 0.5$ GeV from B*, are often used in physics analyses. Marking similar efficiency improvements, their track-finding efficiencies are determined to be around 88.6% and 89.4% respectively. The fake track fraction is around 26% before the Kalman filter, which is relatively high, but in the context of electron reconstruction acceptable. The ghost rate is expected to be smaller after the track-fitting stage, *i.e.* the Kalman filter and clone removal.

The performance of the new Matching algorithm, which is the centrepiece of the electron reconstruction, is best determined using perfect VELO and T track selection. Both categories containing momentum cuts perform with an efficiency above 90%, which is just short of the upper limit set by the VELO tracking and Hybrid Seeding algorithms while maintaining a ghost rate of below 14% before the Kalman filter.

Track reconstruction consists of both track finding and track fitting. The focus of the work presented in this thesis has been on the former since tracks must first be found before they can be fitted. The electron Matching algorithm in combination with a high-performance VELO and T track preselection would significantly improve electron reconstruction at LHCb in Run 3 and possibly beyond. In addition to implementing and optimising a high-efficiency ECAL filter for both VELO and T tracks, the tracks found by the Matching algorithm would have to be fitted as well.

The implementation of a specialised track fit for electrons would improve the performance of an electron reconstruction and aid in controlling the ghost rate further. The Kal-

man filter is almost optimal for a near-linear track model and approximately Gaussian noise. However, for electron reconstruction, the Gaussian-sum filter might be an alternative estimator to be used during track fit instead of the Kalman filter, because of its superior ability to deal with highly asymmetric and long-tailed noise, such as energy loss by bremsstrahlung. More information on the Gaussian-sum filter can be found in Ref. [31].

A Appendix

A.1 Parameterisations

A.1.1 Baseline Matching Algorithm

Table A.1: Coefficients for z_{mag} (Equation 7.4).

Coefficient	Value
c_0	5286.687877988849
c_1	-3.259689996453795
c_2	0.015615778872337033
c_3	-1377.3175211789967
c_4	282.9821232487341

Table A.2: Coefficients for $y_{\mathrm{corr}}^{\mathrm{match}}$ (Equation 7.7).

Coefficient	Value
c_0	-1974.6355416889746
c_1	-35933.837494833504

A.1.2   Electron Matching Algorithm

Table A.3: Coefficients for z_{mag} (Equation 7.10). Parameterisation only for electrons.

Coefficient	Value
c_0	5308.689740859701
c_1	94.8390840129271
c_2	0.03952176022974936
c_3	70418.55504271125
c_4	62087.4827671762
c_5	13580.830615214854
c_6	-0.000842287765099925
c_7	-2237.8985443626666
c_8	-16.447718852307414
c_9	-13.929546692369508
c_{10}	-3.2033475757333143
c_{11}	3364.6401097763382
c_{12}	$1.936634587279551 \times 10^{-7}$

Table A.4: Coefficients for $y_{\mathrm{corr}}^{\mathrm{match}}$ (Equation 7.9). Parameterisation only for electrons.

Coefficient	Value
c_0	4089.1594362560113
c_1	25.0456971896117
c_2	1049.7443418962382
c_3	77388.96417801932

A.2   Matching Algorithm Input Variables

A.2.1   Baseline Matching
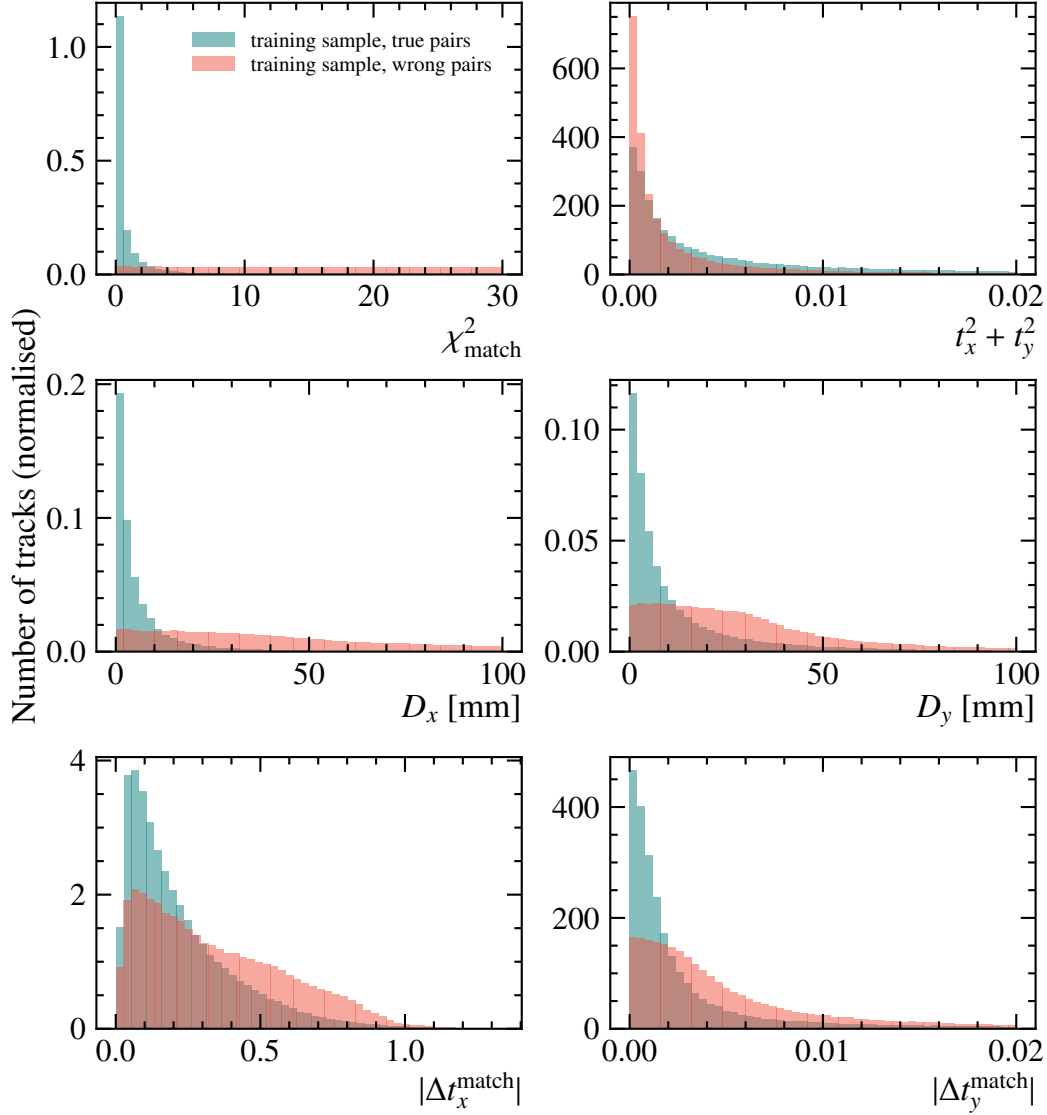
Figure A.1: Distributions of the input variables for the Matching MLP, listed in Table 7.1. The MLP was trained using true VELO, T track-pairs, excluding electrons, as signal, and false pairs, *i.e.* fake tracks, as background.
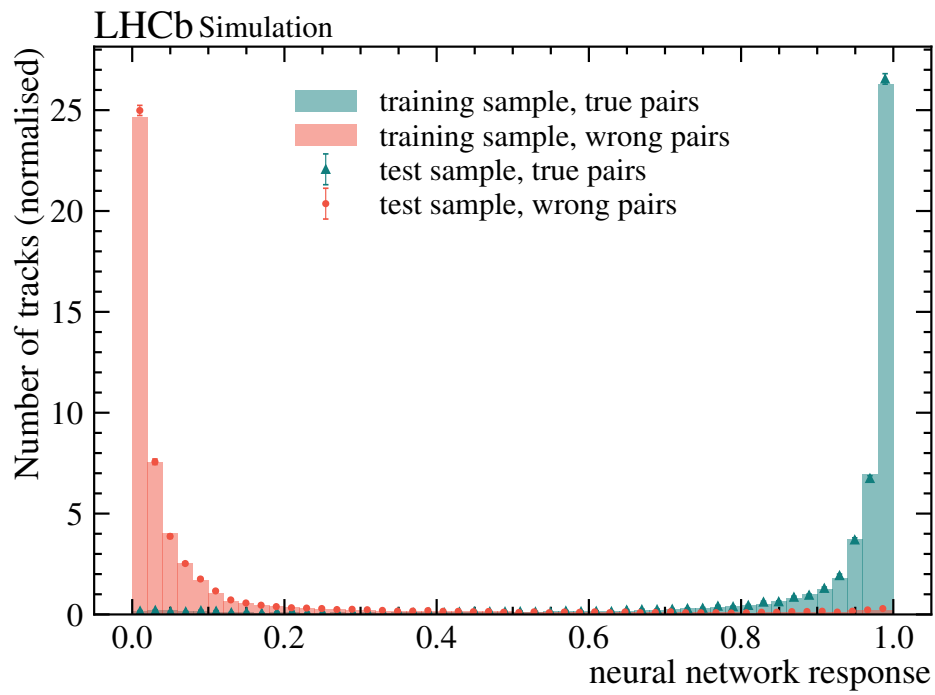
Figure A.2: Output of the Matching MLP. The bars denote the training sample, and the data points denote the test sample. The distribution of true pairs is teal with triangles, and of wrong pairs red with circles.

A.2.2   Electron Matching
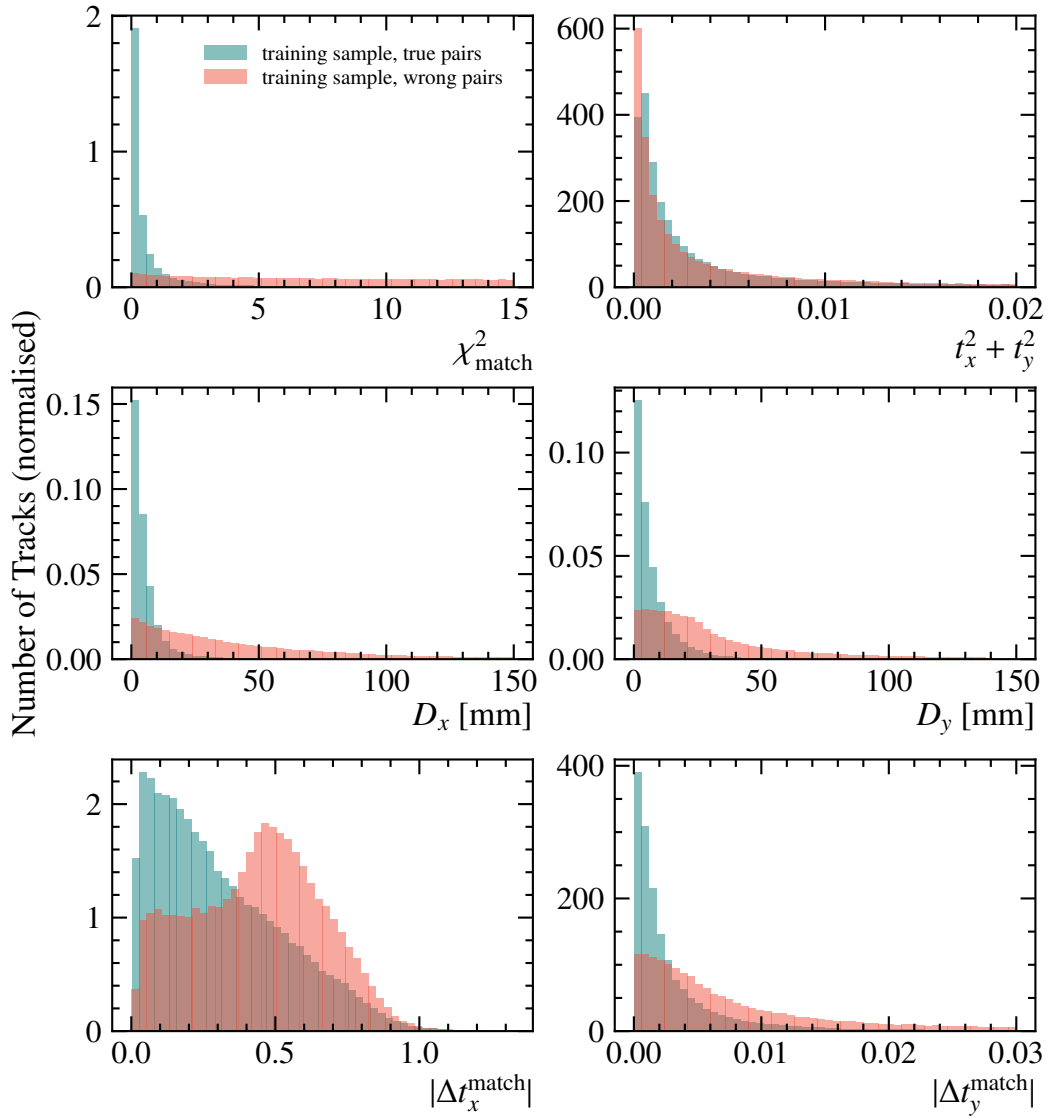
Figure A.3: Distributions of the input variables for the Matching MLP, listed in Table 7.2. The MLP was trained using simulated electron VELO, T track-pairs as signal, and fake tracks of all particles as background.

A.3 Reconstruction Efficiencies

A.3.1 Best Long Tracks

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.

e: Efficiency *vs.* number of primary vertices.
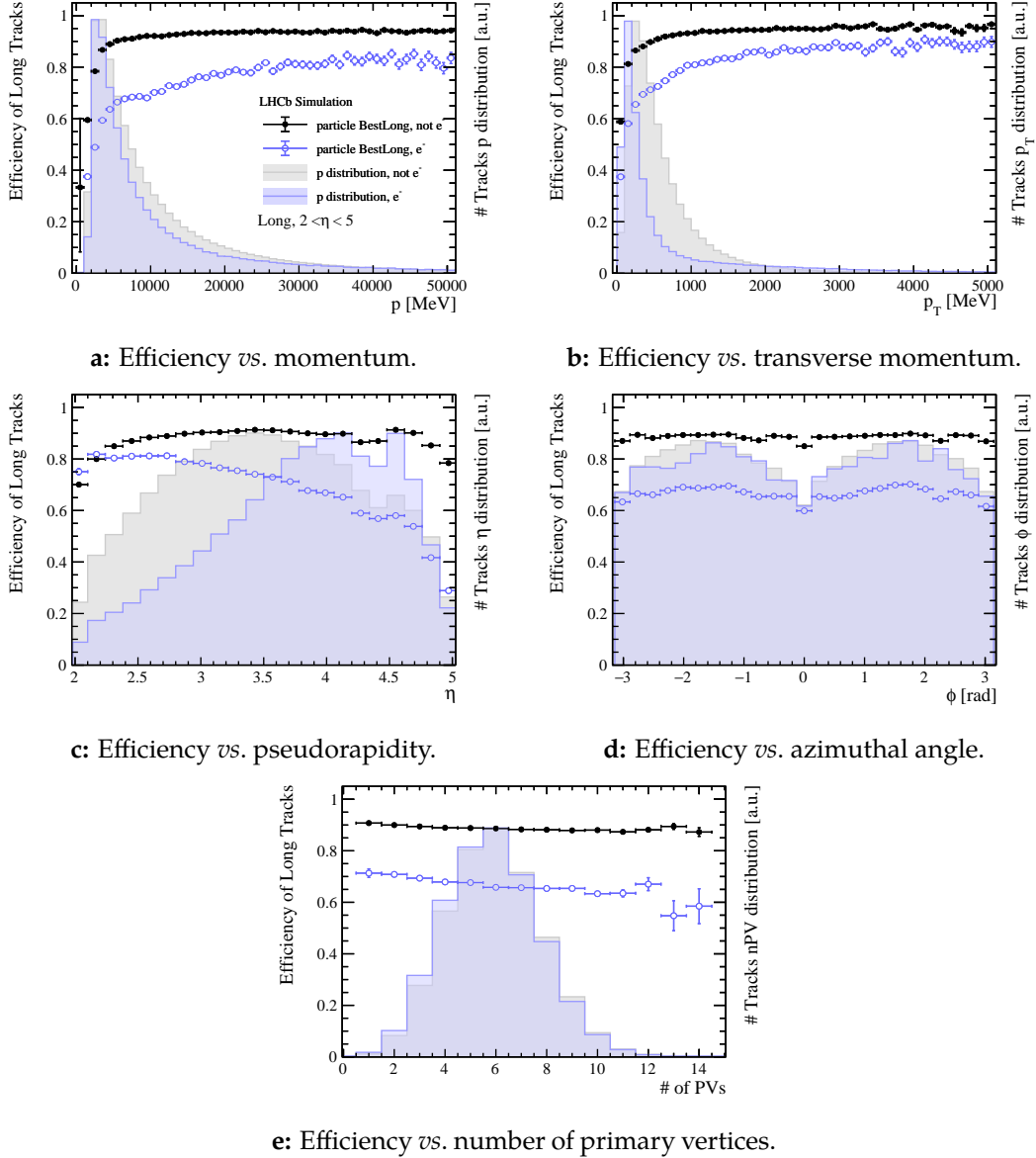
Figure A.4: Track reconstruction efficiency of Best Long tracks for simulated Long tracks with regard to various variables. The variables' distributions and efficiencies are represented in blue for electrons and in grey or black for other particles. The distributions represent the true, simulated variable values at creation.
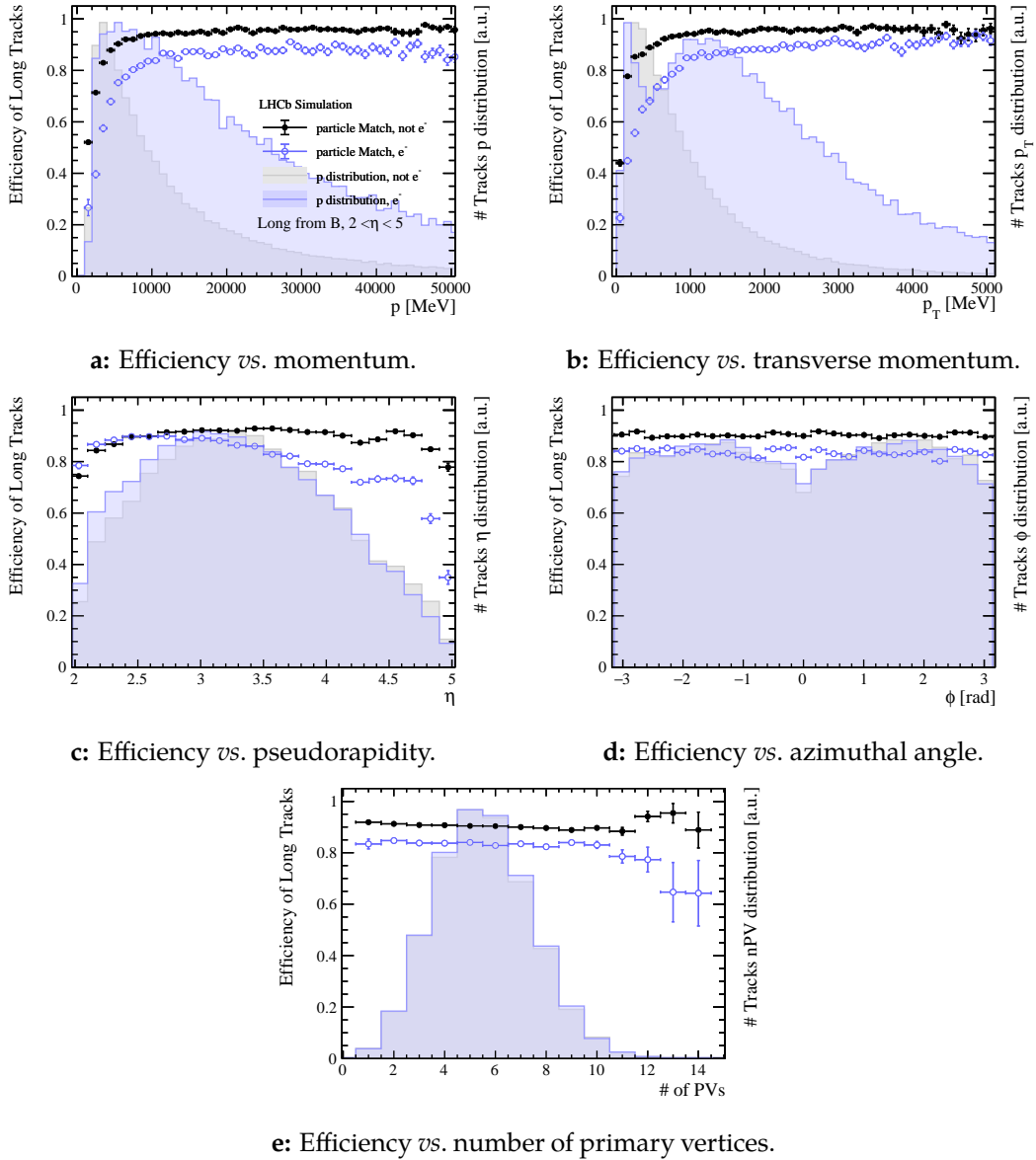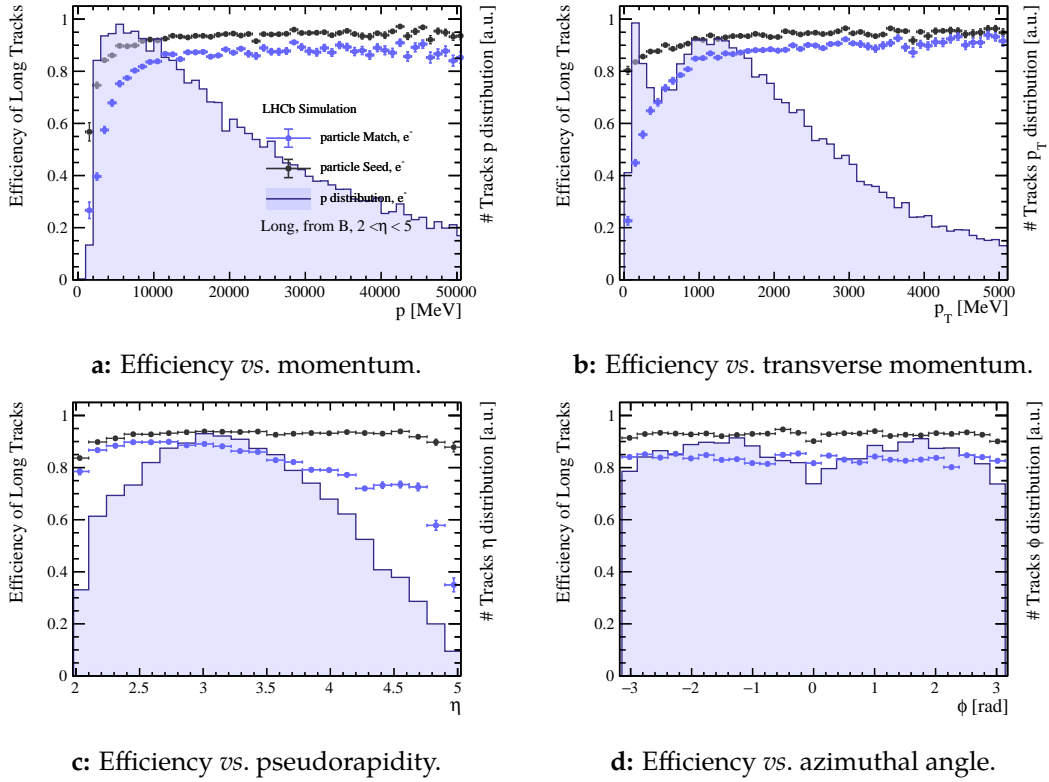
A.3.2 VELO Tracking and HybridSeeding

Track-finding efficiencies for the VELO Tracking and Hybrid Seeding used as input input to the Matching algorithm.

Table A.5: Integrated VELO tracking efficiencies obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events.

Category	$e^\pm \; \varepsilon[\%]$	not-$e^\pm \; \varepsilon[\%]$
Long	96.04	99.20
Long from B	97.41	99.28
Long $p > 5\,\text{GeV}$ from B	98.17	99.57
Fake track fraction r_{fake}	2.47%	

Table A.6: Integrated Hybrid Seeding efficiencies obtained from simulated $B^0 \to K^*(J/\psi \to e^+e^-)$ events.

Category	$e^\pm \; \varepsilon[\%]$	not-$e^\pm \; \varepsilon[\%]$
Long	87.87	94.30
Long from B	92.78	95.58
Long $p > 5\,\text{GeV}$ from B	93.91	97.57
Fake track fraction r_{fake}	2.02%	

A.3.3   Baseline Matching

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.
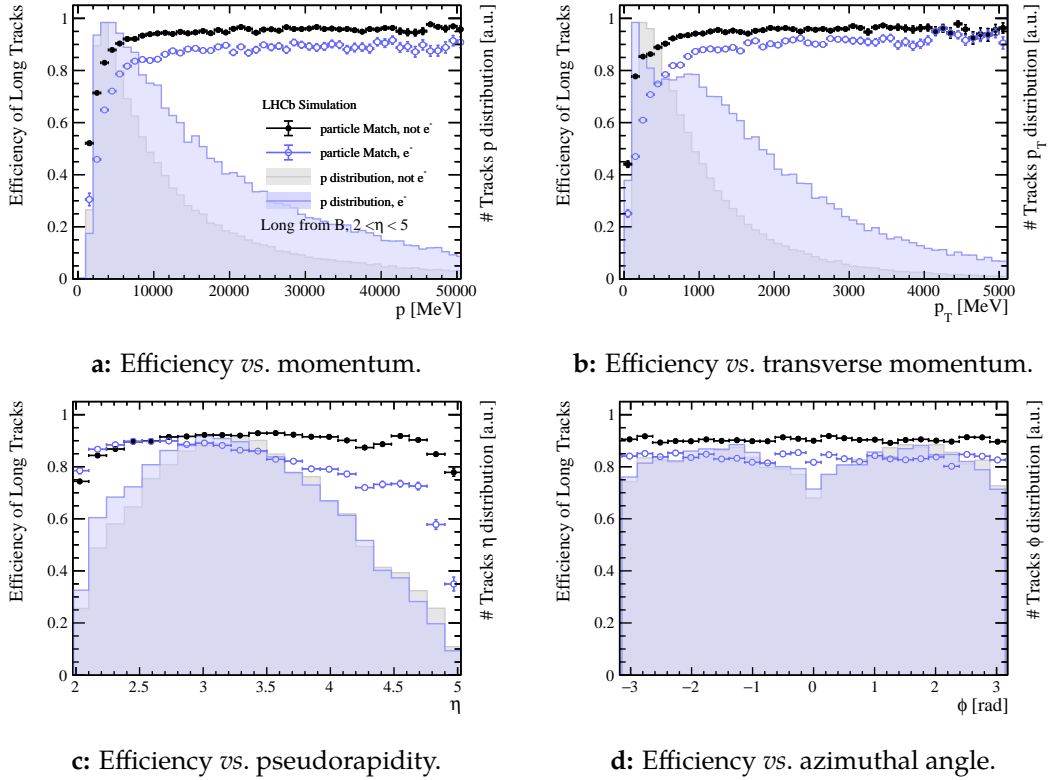
Figure A.5: Track finding efficiency of the baseline Matching algorithm with VELO and T input tracks for simulated Long tracks originating from a B meson with regard to various kinematic variables. The variables' distributions and efficiencies are represented in blue for electrons and in grey or black for other particles. The pseudorapidity and azimuthal angle distributions represent the true, simulated variable values at creation. The momenta distributions have been adjusted according to Section 6.2 (for unaltered momentum spectrum see Figure A.6 and Figure A.7).

a: Efficiency *vs*. momentum.

b: Efficiency *vs*. transverse momentum.

c: Efficiency *vs*. pseudorapidity.

d: Efficiency *vs*. azimuthal angle.

e: Efficiency *vs*. number of primary vertices.
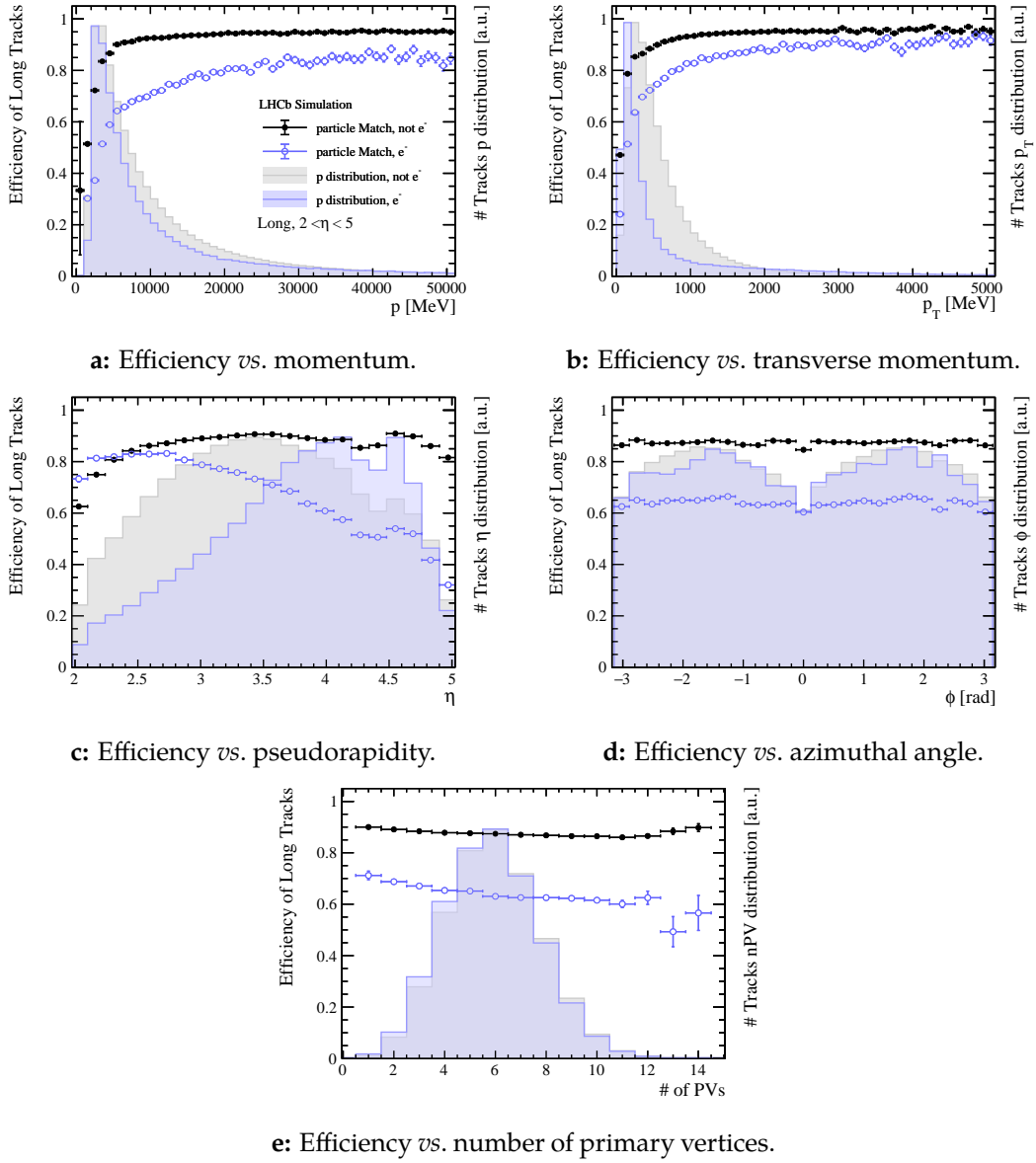
Figure A.6: Track finding efficiency of the baseline Matching algorithm with VELO and T input tracks for simulated Long tracks with regard to various variables. The variables' distributions and efficiencies are represented in blue for electrons and in grey or black for other particles. The distributions represent the true, simulated variable values at creation.

a: Efficiency *vs*. momentum.

b: Efficiency *vs*. transverse momentum.

c: Efficiency *vs*. pseudorapidity.

d: Efficiency *vs*. azimuthal angle.

e: Efficiency *vs*. number of primary vertices.

Figure A.7: Track finding efficiency of the baseline Matching algorithm with VELO and T input tracks for simulated Long tracks originating from a B meson with regard to various variables. The variables' distributions and efficiencies are represented in blue for electrons and in grey or black for other particles. The distributions represent the true, simulated variable values at creation.

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.

Figure A.8: Track finding efficiency of the Seed tracks and the baseline Matching algorithm with VELO and T input tracks for simulated Long tracks originating from a B meson with regard to various kinematic variables. The variables' distributions are represented in blue. Efficiencies of the Seed tracks are depicted in black and the Matching algorithm in blue. The distributions represent the true, simulated variable values at creation.

A.3.4 Electron Matching

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.
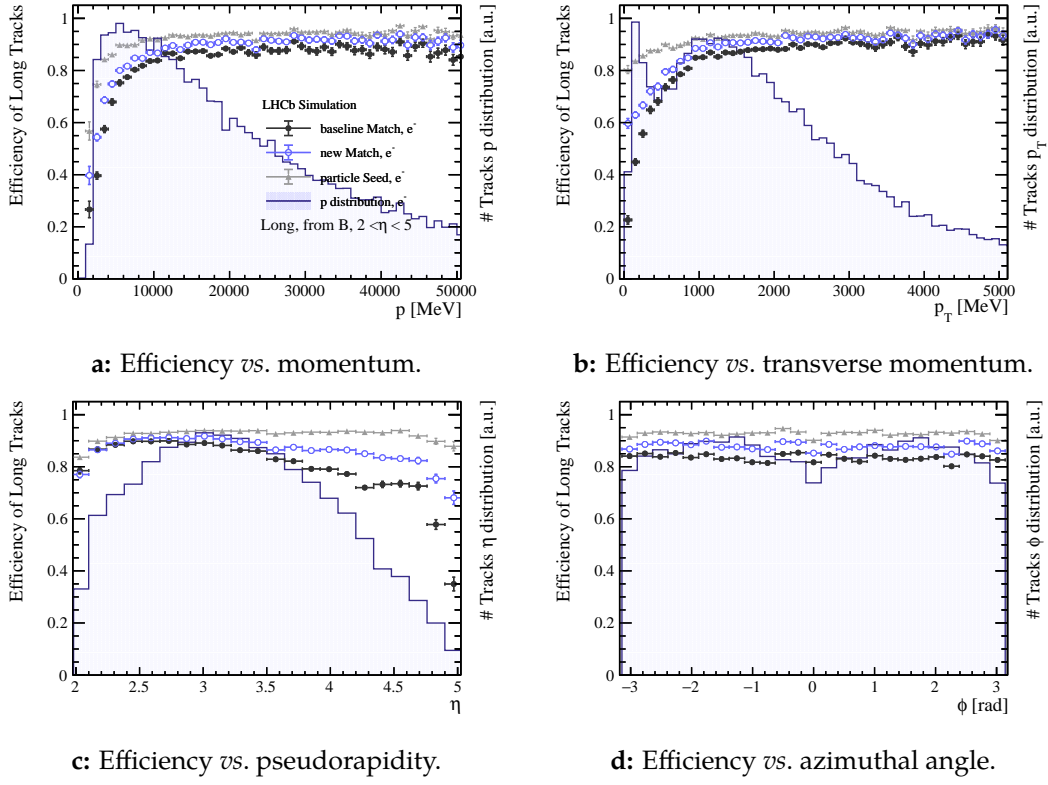
Figure A.9: Track finding efficiency of the Seed tracks and the electron Matching algorithm with perfect selection VELO and T input tracks for simulated Long tracks originating from a B meson with regard to various kinematic variables. The variables' distributions are represented in blue. Electron Seed track efficiencies are depicted in grey triangles. The black marker with full circles denotes electron tracks of the baseline Matching algorithm, and the blue marker with empty circles denotes electron tracks of the new (electron) Matching algorithm. The distributions represent the true, simulated variable values at creation.

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.

c: Efficiency *vs.* pseudorapidity.

d: Efficiency *vs.* azimuthal angle.
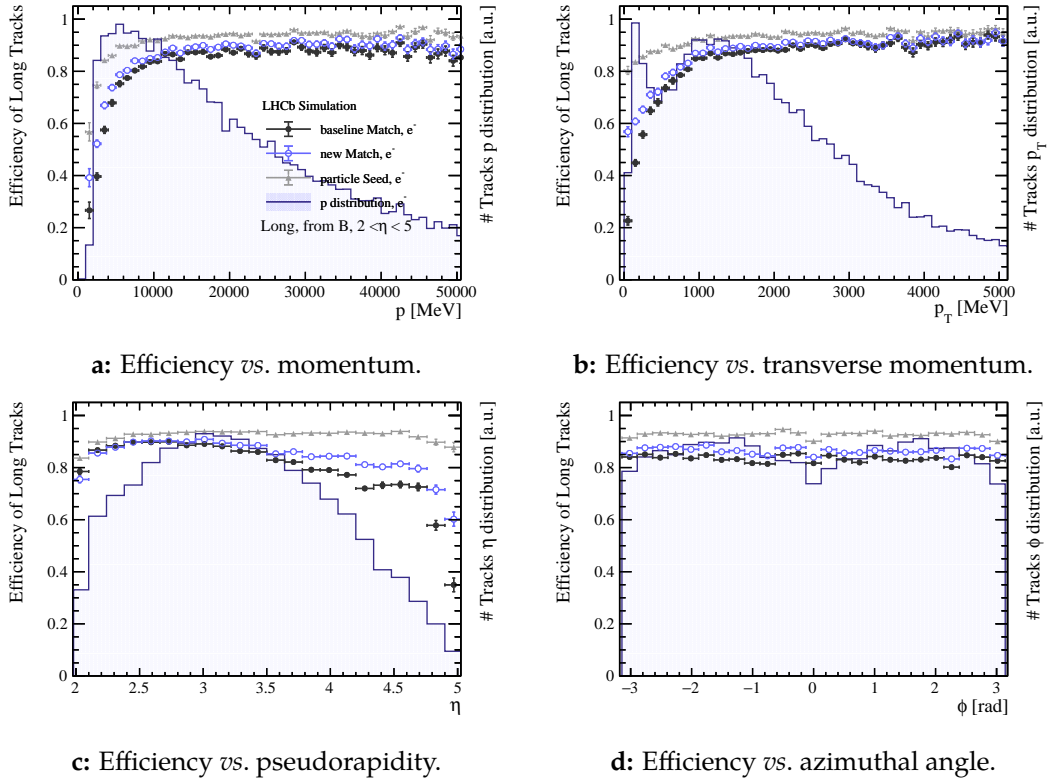
Figure A.10: Track finding efficiency of the Seed tracks and the electron Matching algorithm with VELO and ECAL-filtered T input tracks for simulated Long tracks originating from a B meson with regard to various kinematic variables. The variables' distributions are represented in blue. Electron Seed track efficiencies are depicted in grey triangles. The black marker with full circles denotes electron tracks of the baseline Matching algorithm, and the blue marker with empty circles denotes electron tracks of the new (electron) Matching algorithm. The distributions represent the true, simulated variable values at creation.

A.3.5   Momentum Redefinition

This appendix is related to Section 6.2.

a: Efficiency *vs.* momentum.

b: Efficiency *vs.* transverse momentum.
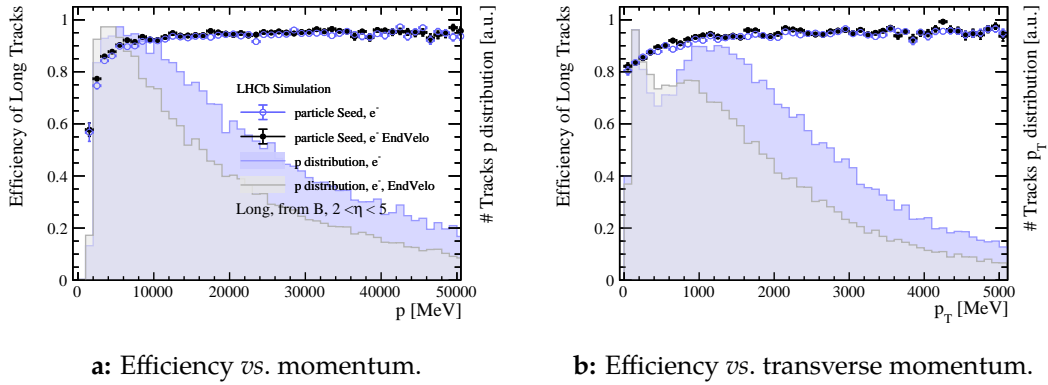
Figure A.11: Track finding efficiency of the baseline HybridSeeding algorithm with SciFi hits as input for simulated T tracks originating from a B meson with regard to the total and transverse momentum. The variables' unaltered distributions and efficiencies are represented in blue and the redefined values in grey and black.

Acknowledgements

All praises to God and His blessing for the completion of this thesis. I thank God for all the opportunities, trials, and strength that have been given to me to finish writing this thesis.

First and foremost, I would like to thank Prof. Stephanie Hansmann-Menzemer for providing me with the opportunity to work for the Heidelberg LHCb group, and Prof. Klaus Reygers for agreeing to be second referee.

Special thanks to André Günther for his guidance and support throughout this entire experience. I am grateful for the interesting topic of my thesis, as well as the motivation he showed towards steering me in the right direction. My experience would have been far less enjoyable without him.

I would also like to thank my office mates, Lau, Marius, and Tobias, for the peaceful and friendly working environment, and the enormous help they have given me. It was a joy to come to work with them in the office.

Thanks to Nils for the many late-night FaceTime conversations that somehow always ended up so far off-topic that we had no idea how we got there.

Danke an meine Familie und Freunde, die mich immer unterstützt haben und mir immer eine große Hilfe sind in allem was ich mache.

Allah annemden, babamdan, kardeşimden, ve bütün ailemden razı olsun. Böyle mübarek bir ayda bu tezi yazabildiğim için Allah'a şükrediyorum.

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 8. April 2024

Furkan Selim Cetin

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die an-

Bibliography

[1] J. J. Thomson, 'XL. *Cathode Rays,*' *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 44, no. 269, pp. 293–316, Oct. 1897.

[2] B. Aubert, "Evidence for the flavor changing neutral current decays b to k l+ l- and b to k* l+ l-," 25th Jul. 2002. arXiv: `hep-ex/0207082`.

[3] Robert Fleischer *et al.*, 'New perspectives for testing electron-muon universality,' *Journal of High Energy Physics*, vol. 2023, no. 6, p. 33, 7th Jun. 2023. arXiv: `2303. 08764[hep-ex,physics:hep-ph]`.

[4] Alex Seuthe, "Lepton flavour universality tests at LHCb," 14th May 2023. arXiv: `2305.08216[hep-ex]`.

[5] F. Follin *et al.*, 'Implementation and experience with luminosity levelling with off-set beam,' *CERN Yellow Report*, CERN-2014-004, pp. 183–187, 2014. arXiv: `1410. 3667[physics]`.

[6] LHCb collaboration *et al.*, "The LHCb upgrade i," 17th May 2023. arXiv: `2305. 10515`.

[7] Paul André Günther, 'Track reconstruction development and commissioning for LHCb's run 3 real-time analysis trigger,' Ph.D. dissertation, Heidelberg, 2023.

[8] CERN. Geneva. LHC Experiments Committee *et al.*, Eds., *LHCb magnet: Technical Design Report*, Technical design report. LHCb 1, Geneva: CERN, 2000, 42 pp.

[9] E. Buchanan *et al.*, 'Spatial resolution and efficiency of prototype sensors for the LHCb VELO upgrade,' *Journal of Instrumentation*, vol. 17, no. 6, P06038, 1st Jun. 2022. arXiv: `2201.12130[physics]`.

[10] LHCb Collaboration, 'LHCb VELO upgrade technical design report,' 2013.

[11] LHCb Collaboration, *LHCb Tracker Upgrade Technical Design Report*. 2014.

[12] LHCb Collaboration, 'LHCb PID upgrade technical design report,' 2013.

[13] C. Abellán Beteta *et al.*, "Calibration and performance of the LHCb calorimeters in run 1 and 2 at the LHC," 26th Aug. 2020. arXiv: `2008.11556[hep-ex,physics: physics]`.

[14] Hermann Kolanoski *et al.*, *Particle detectors: fundamentals and applications*, First edition. Oxford ; New York, NY: Oxford University Press, 2020, 927 pp., OCLC: on1139946145.

[15] Navarro, Albert Puig, 'First measurements of radiative b decays in LHCb,' Ph.D. dissertation, Barcelona, 2012.

[16] Simon S. Haykin *et al.*, *Neural networks and learning machines*, 3rd ed. New York: Prentice Hall, 2009, 906 pp., OCLC: ocn237325326.

[17] Thiago Christiano Silva *et al.*, *Machine Learning in Complex Networks*. Cham: Springer International Publishing, 2016.

[18] Christopher M Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995, 482 pp.

[19] V V Gligorov *et al.*, 'RTA and DPA dataflow diagrams for run 1, run 2, and the upgraded LHCb detector,'

[20] R. Aaij *et al.*, 'Allen: A high-level trigger on GPUs for LHCb,' *Computing and Software for Big Science*, vol. 4, no. 1, p. 7, 30th Apr. 2020.

[21] Christina Agapopoulou, 'Commissioning LHCb's GPU high level trigger,' *Journal of Physics: Conference Series*, vol. 2438, no. 1, p. 012 017, 1st Feb. 2023.

[22] Brij Kishor Jashal, "Standalone track reconstruction and matching algorithms for GPU-based high level trigger at LHCb," 2022.

[23] LHCb collaboration, "Long-lived particle reconstruction downstream of the LHCb magnet," 20th Nov. 2022. arXiv: `2211.10920[hep-ex]`.

[24] Pierre Billoir *et al.*, 'A parametrized kalman filter for fast track fitting at LHCb,' *Computer Physics Communications*, vol. 265, p. 108 026, Aug. 2021. arXiv: `2101.12040[hep-ex,physics:physics]`.

[25] Arthur Hennequin *et al.*, 'A fast and efficient SIMD track reconstruction algorithm for the LHCb upgrade 1 VELO-PIX detector,' *Journal of Instrumentation*, vol. 15, no. 6, P06018–P06018, 15th Jun. 2020. arXiv: `1912.09901[physics]`.

[26] Salvatore Aiola *et al.*, 'HybridSeeding: A standalone track reconstruction algorithm for scintillating fibre tracker at LHCb,' *Computer Physics Communications*, vol. 260, p. 107 713, Mar. 2021. arXiv: `2007.02591[hep-ex,physics:physics]`.

[27] Paul Andre Günther, 'LHCb's forward tracking algorithm for the run 3 CPU-based online track-reconstruction sequence,' 11th Jul. 2023. arXiv: `2207.12965`.

[28] Anton Lechner, 'Particle interactions with matter,' *CERN Yellow Reports: School Proceedings*, vol. 5, pp. 47–47, 20th Dec. 2018.

[29] Particle Data Group *et al.*, 'Review of particle physics,' *Progress of Theoretical and Experimental Physics*, vol. 2022, no. 8, p. 083C01, 8th Aug. 2022.

[30] P Li *et al.*, 'Tracking definitions and conventions for run 3 and beyond,'

[31] Rudolf Fruehwirth *et al.*, *Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors* (Particle Acceleration and Detection). Springer, 2021.

[32] M Benayoun *et al.*, "The forward tracking, an optical model method," Place: Geneva, 2002.

[33]   A. Hoecker *et al.*, "TMVA - toolkit for multivariate data analysis," 7th Jul. 2009. arXiv: physics/0703039.

List of Tables

List of Figures