DISSERTATION

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

M.Sc. Rowina Christiane Caspary

born in Freiberg, Germany

Oral examination: 02.07.2025

Development of a data-driven method to determine reconstruction efficiencies in the LHCb tracking detectors and study of the detector performance in Run 3

Referees:
Prof. Dr. Stephanie Hansmann-Menzemer
Prof. Dr. Norbert Herrmann

Zusammenfassung:

LHCb ist ein Experiment, dass der Messung von Teilchenzerfällen, die b- und c-quarks enthalten, gewidmet ist. Während der dritten Datennahmeperiode des *Large Hadron Collider (LHC)* kollidieren Protonen am Wechselwirkungspunkt von LHCb mit einer instantanen Luminosität von bis zu $2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ und einer Schwerpunktsenergie von $\sqrt{s} = 13.6\,\mathrm{TeV}$. Ein detailliertes Verständnis der Detektor- und Rekonstruktionseffizienzen des Experiments ist für viele geplante Präzisionsmessungen der LHCb-Kollaboration unerlässlich. Um die Spurrekonstruktionseffizienz der LHCb-Subdetektoren zu ermitteln, wurde in dieser Arbeit eine *Tag-und-Probe-Methode* entwickelt, die $J/\psi \rightarrow \mu\mu-$Zerfälle nutzt. Unterschiede zwischen den in simulierten und gemessenen Daten ermittelten Spurrekonstruktionseffizienzen wurden untersucht und daraus Korrekturfaktoren bestimmt, die in späteren Analysen verwendet werden. Die simulierten und aufgezeichneten Spurrekonstruktionseffizienzen stimmen im Sub-Prozent-Bereich über fast den gesamten Phasenraum und in allen relevanten Subdetektoren überein, was das ausgezeichnete Verständnis des LHCb-Detektors und seiner Spurrekonstruktionsdetektoren demonstriert. Um den Effekt von hadronischen Wechselwirkungen im Detektormaterial zu bestimmen, wurde ein neuartiger Ansatz entwickelt. In einer ersten Schätzung wurde daraus eine zusätzliche systematische Unsicherheit in der Größenordnung von wenigen Prozent für die Effizienz der Spurrekonstruktion hergeleitet. Es wird erwartet, dass in weiteren Studien diese systematische Unsicherheit auf ein Prozent oder weniger reduziert werden kann.

Abstract:

The LHCb experiment is a highly complex apparatus dedicated to measuring the decays of particles containing *b*- and *c*-quarks. During Run 3 of the Large Hadron Collider (LHC), protons collide at the interaction point of LHCb with an instantaneous luminosity of up to $2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at a centre-of-mass energy of $\sqrt{s} = 13.6\,\mathrm{TeV}$. A full understanding of the detector and reconstruction efficiencies involved in the data acquisition is essential for many high precision measurements. A tag-and-probe method is developed to estimate the track reconstruction efficiency of the LHCb tracking subdetectors exploiting the decay of $J/\psi \to \mu\mu$. Discrepancies of the measured track reconstruction efficiencies between simulated and recorded data are studied and used to derive correction factors. An agreement between recorded and simulated data at the sub-percent level is achieved over almost the entire phase space and for all tracking subdetectors, illustrating an excellent understanding of the upgraded LHCb detector and its reconstruction sequences. A novel approach to derive the effect of interactions in the detector material is developed and used to provide first estimates, which contribute a systematic uncertainty on the order of a few percent to the track reconstruction efficiency. Further corrections are expected to reduce this systematic uncertainty to the order of one percent or less.

Contents

Chapter 1

Introduction

The principle idea of particle physics, that all matter is made up of particles, was already proposed by the ancient Greeks. Thales of Miletus (624 - 546 B.C.) postulated the idea that all matter is based on one basic element, this being water. Empedocles of Akragas (492 - 440 B.C.) believed that there are four basic elements: earth, water, air and fire. Democritus of Abdera, Thrace (470 - 380 B.C.) finally proposed the idea of small indivisible particles called *átomos*, meaning *indivisible*. However, all these people were philosophers, not scientists and as such based their theories purely on thought rather than empirical evidence. For example, Aristotle (384 - 322 B.C.) believed that heavier objects would always fall faster than lighter objects, a theory that could have been easily checked by him, but wasn't falsified until Galileo Galilei (1564 - 1642) [1].

Nowadays, science, and specifically physics, is based on the evidence-based *scientific method*, which means that experiments are needed to test hypotheses and confirm theories. Often, it is probing the smallest things that requires the largest and most complex experiments, at which point understanding and describing the nature of the experiments itself becomes vital.

The largest of the experiments used for particle physics are located at the *Conseil Européen pour la Recherche Nucléaire (CERN)*, an international centre for particle research situated close to Geneva at the franco-suisse border. There, the *Large Hadron Collider (LHC)*, a large ring of 27 kilometre circumference, is used to accelerate and collide protons and ions at high energies in several interaction points. There, giant detectors are stationed in order to measure the decay products of these collisions and to produce large amounts of data that can be used to test various hypotheses with the final goal of building a complete description of the universe at the particle level.

The LHCb experiment, situated at CERN, is one such highly complex apparatus. At full operation of the LHC, it records proton-proton collisions at an interaction rate of 40 MHz. Each collision potentially results in hundreds of particles flying through the detector, of which the few relevant ones have to be recorded and reconstructed to perform scientific measurements.

Understanding exactly what happened in the detector is vital for all decays that were reconstructed and, crucially, also to get an idea of what was not reconstructible and therefore missed by the algorithms. This is a very complex task and an accurate description of the detector and reconstruction performance in simulated data is difficult.

As a particle propagates through the detector, it leaves traces that can be used to reconstruct the path of the particle, called a *track*. This track can be used to measure the momentum of the particle and, together with other detector information, to assign its energy and type. More details on this are given in Chapter 3. However, not every track is correctly reconstructed. For some measurements, for example cross section measurements, the number of particles from certain decays propagating through the detector is a relevant observable. This means that the probability for a particle to be successfully reconstructed in the detector is essential.

The *track reconstruction efficiency* describes the probability that a particle propagating through the detector is reconstructed by the track reconstruction algorithms. This thesis describes a method of determining the track reconstruction efficiency using recorded and simulated data. Their ratio is calculated to determine correction factors that can be used in any measurement where the track reconstruction efficiency is relevant and taken from simulation.

Chapter 2 provides an introduction into the current theory of elementary particle physics, the Standard Model (SM). The interactions of particles with matter are explained, highlighting the challenges of reconstructing tracks of particles interacting in LHCb. Chapter 3 describes the LHC accelerator complex at CERN and LHCb, the experiment investigated within this thesis, with a focus on the tracking system during LHC Run 3. Chapter 4 describes in more detail how tracks are reconstructed both for recorded and simulated data. Chapter 5 summarises the tools and applications used to measure the track reconstruction efficiency. In Chapter 6, the methods to measure the track reconstruction efficiencies both in recorded and simulated data are introduced. First results of the track reconstruction efficiency derived from data taken in 2023 are discussed in Chapter 7. These are valuable to understand and optimise the procedure and the detector alignment for the description of the track reconstruction efficiencies in 2024 data. Track reconstruction efficiencies measured using 2024 data are discussed and assigned systematic uncertainties in Chapter 8. Chapter 9 investigates the causes for discrepancies between the track reconstruction efficiencies in recorded and simulated data and Chapter 10 verifies the method. Chapter 11 considers the hadronic interactions of particles propagating through the detector and their effect on the track reconstruction efficiency and concludes with an estimate of the additional systematic uncertainty assigned to the track reconstruction efficiency. Chapter 12 summarises this thesis and gives an outlook on the future of this topic.

Chapter 2

Theory motivation

This chapter gives an overview of the Standard Model of particle physics (SM). The search for new physics motivates the precise measurement of detection efficiencies, such as the track reconstruction efficiency. Challenges to track reconstruction and the measurement of the track reconstruction efficiency due to scattering effects and energy loss caused by the interaction of particles with matter are discussed.

2.1 The Standard Model of Particle Physics

The Standard Model (SM) of particle physics is a relativistic, renormalisable quantum field theory developed by particle physicists over the course of the last century to describe the fundamental constituents and forces of the universe [2]. It represents the basis of the current understanding of the universe as measured by experiment.

The fundamental constituents of the universe described in the SM are twelve matter and twelve anti-matter particles, which mirror each other in all properties except their mass, which is the same for each matter-anti-matter pair. The twelve matter particles are six quarks and six leptons, each in three generations.

The quarks are the positively charged up-, charm and top-quark, denoted u, c and t, in ascending mass, each with an electric charge of two thirds, and the negatively charged down-, strange- and bottom-quark, denoted d, s and b, also in ascending mass, each with an electric charge of one third. The charges of quarks are given in thirds as they never appear singularly, but instead immediately form composite states of unitary charge, except for the top-quark, which is too heavy to live long enough to form a bound state.

Those bound states comprise of either triplets of three quarks or three anti-quarks, called baryons and anti-baryons, or doublets of one quark and one anti-quark, called mesons. Evidence has also been found for tetraquarks, built out of two quarks and two anti-quarks, and pentaquarks, built out of three quarks and two anti-quarks or vice versa.

All quark composites are called baryons. They are built in such a way that the composite particle has a multiple of the unitary charge. This is due to colour confinement, as all quarks also have a colour charge. There are three possible colour charges, called red, blue and green, and their anti-charge equivalents of anti-red, anti-blue and anti-green.

The leptons are the electron and its anti-particle, the positron, $e^{\pm}$, the muon and anti-muon, $\mu^{\pm}$, and the tauon and anti-tau, $\tau^{\pm}$, each with a unitary electric charge. There are also three electrically neutral neutrinos, called the electron-neutrino, ν_e, the muon-neutrino, ν_μ, and the tau-neutrino, ν_τ, with their respective anti-neutrino equivalents. Neutrinos, which are electrically and colour neutral, interact very little with known matter and are very light, making them difficult to detect. It is, however, known that they oscillate between the three neutrino states.

Interactions between the matter and anti-matter particles are mediated by five gauge bosons. The electrically and colour neutral, weightless photon, γ, interacts with all electrically charged particles. The electrically neutral, weightless gluon, g, interacts with all coloured particles and has itself eight states of colour-anti-colour combinations. This leads to quark confinement, and prevents the gluon from existing as a free particle. The unitary electrically charged $W^{\pm}$ and the neutral Z boson are heavy gauge bosons that couple to all matter particles, as well as the photon in the case of the $W^{\pm}$, and transmit the so-called weak interaction, which allows particles to decay. Because they are heavy, they themselves have a very short lifetime and decay into other particles.

Finally, the electrically neutral Higgs boson, h, produced as an excitation of the Higgs field from the Higgs mechanism, couples to all particles of non-zero mass, including the $W^{\pm}$ and the Z. Since it is not itself weightless, it can couple to itself, and decays quickly upon creation.

2.2 The search for new physics

While the SM has proven to be a remarkably stable and long-lived theory describing the fundamental laws of nature, it still has several weaknesses, including the ignorance of gravity and the missing description of neutrino masses, as well as no explanation of, for example, the existence of three generations of each matter particle.

The aim of modern particle physics is therefore the test of the SM and the search for new physics. This can be done in two ways: either by probing previously untested phase space, thus going to higher energies, or by doing precision physics and measuring discrepancies between the SM predictions and the natural world at a very fine level.

One of the measurements carried out with this purpose is the very precise measurement of particle cross sections at different energies. This could potentially lead to new physics directly, by measuring a cross section different from the theoretical one, or indirectly, since the cross section of matter particles is an important input parameter in the measurement of channels containing non-SM particles and forbidden decays.

To measure a cross-section, three important quantities must be known: the luminosity of the experiment, the number of measured candidates (both for the channel in question and a reference channel) and the efficiencies going into the measurement. This is necessary to derive the number of actually produced candidates from the number of measured candidates.

Many efficiencies contribute to the overall detection efficiency, including the detector acceptance, the trigger efficiency of the data acquisition system, and the particle identification efficiency. However, one of the most important ones is the *track reconstruction efficiency*, which describes the probability that a track passing through the detector can be successfully reconstructed. There are many challenges to reconstructing the track of a particle in a detector, and therefore to the precise measurement of the efficiency.

2.3 The interaction of particles with matter

2.3.1 Long-lived elementary particles and composites

Due to relativistic time dilation, relativistic particles with lifetimes of $10^{-10}\,$s or longer are able to travel substantial distances of up to several metres within the LHCb detector and can therefore leave reconstructible tracks. As a result they are treated as stable in the context of the experiment. Others with lifetimes up to picoseconds, such as the D^0, which consists of a charm- and an anti-up quark, or its anti-particle, can travel a few millimetres from production to decay at the energies they are produced with at the LHCb experiment, allowing for a measurement of the decay time.

Charged particles that can be reconstructed directly in the detector are:

- electrons and their anti-particles, positrons, $e^{\pm}$, which are stable to the extent of current research,
- protons and anti-protons, which consist of two up- and one down-quark or two anti-up- and one anti-down-quark, respectively, and are likewise stable to the extent of current research,
- muons and anti-muons, $\mu^{\pm}$, with a lifetime of $2.20 \times 10^{-6}\,$s,
- charged pions, $\pi^{\pm}$, which consist of an u- and an anti-d-quark or a d- and an anti-u-quark and have a lifetime of $2.60 \times 10^{-8}\,$s and

- charged kaons, $K^{\pm}$, which have a lifetime of $1.24 \times 10^{-8}\,\mathrm{s}$ [3].

Photons can also traverse through the detector, as can neutrons and anti-neutrons, consisting of two down- and one up-quark, or their anti-particles, which have a lifetime of about 14 minutes [3]. However, photons interact less with the detector material than charged particles because they are themselves electrically neutral. Neutrons do not interact electrically with the detector material, but can interact via the strong force and therefore also leave signals in the detector. Both can only be reliably detected by measuring the energy deposits they leave when decaying or interacting with the detector material.

Finally, neutrinos, for which only a mass-dependent lower lifetime limit of $300\,\mathrm{s}/(\mathrm{eV}/c^2)$ is known [3], can traverse the entire detector, but are not detected as their interaction probability with the detector material is extremely low.

When reconstructing the tracks of these particles, there are two main challenges that prevent the reconstruction from being a simple game of connecting the start and end points of the particle trajectory. These are deflections from the straight - or bent in a magnetic field - path, called *scattering*, and loss of energy and momentum.

2.3.2 Scattering of particles in a material

As they propagate through a material, particles can interact with the particle component of the detector material. **Scattering** is the consequence of elastic particle-particle collisions between all kinds of particles, leading to a change in the energy and momentum of the participants.

Electrically charged particles can scatter off the shell electrons of atoms via electromagnetic interaction, so-called **Coulomb scattering**. Hadronic particles can also interact strongly with the nuclei. Even photons can scatter off electrons or nuclei via **Compton scattering**. Each of these effects, while inelastic, changes the momentum and energy of the contributors.

This can happen both in the active material, such as scintillators, silicon detector layers or gas from a Cherenkov detector, leaving valuable information to the measuring system of the detector, and in inactive material, such as the beam pipe, cables and the structural framework. Scattering interactions happen randomly and are thus only described statistically. Both single and multiple scatterings are possible on a particle's trajectory through the detector. The lower the energy of a particle, the more scattering can deflect it from its previous path, making the reconstruction via a simply bended line harder.

2.3.3 Energy loss in particle-matter interactions

The best way to measure a particle in a detector is to reconstruct the full track it leaves in the detector with as little energy loss as possible, and then have it stopped by depositing all its energy in a later layer, allowing for a precise measurement of the energy it still has at the end of its life. This allows for a very precise measurement, because the bending of the track in the magnetic field can be combined with the final energy deposit of the particle to infer its properties as accurately as possible.

However, while a full deposit of its energy via interaction in a late layer is desirable, energy loss can also occur on the way through the detector and thus complicate the reconstruction of the particle additionally to the aforementioned scattering.

Particles interacting with material can experience different types of interaction that lead to energy loss. The type of interaction that dominates the energy loss is dependent on the mass and the momentum of particles. In experiments such as the one discussed in this thesis, particles are generally assumed to be highly relativistic, meaning that their speed v is very close to the speed of light c:

$$\beta := v^2/c^2 \approx 1 \tag{2.1}$$

For electrons above a critical energy E_c, the dominant source of energy loss is called **bremsstrahlung**, where the electron emits a photon in the electrostatic field of a nucleus. E_c is related to the charge Z of the nucleus and approximately

$$E_c \approx \frac{800}{Z}\,\text{MeV}. \qquad [2] \tag{2.2}$$

Below the critical energy, electrons interact via scattering or are absorbed in the material via **electron capture**.

Low-energetic photons below 1 MeV interact primarily via the **photoelectric effect**, where a photon is absorbed by an electron in an electrostatic field. At around 1 MeV, **Compton scattering** instead becomes the dominant source of energy loss for photons. It describes the elastic scattering and exchange of energy between photons and electrons, $e\gamma \rightarrow e\gamma$. At energies of 10 MeV and above photons lose energy predominantly by e^+e^--**pair production** in the electrostatic field:

$$\gamma \rightarrow e^+e^- \tag{2.3}$$

The combination of these electron-photon interactions leads to electromagnetic showers, in which an electron or positron emits a photon, which decays in pair production into a new electron-positron pair, both of which emit photons again. This continues until the energy of the produced photons is too low to produce a new $e^{\pm}$ pair. Inducing electromagnetic showers is the best way to stop an electron and have it deposit all its energy in a comparatively short segment.

Bremsstrahlung occurs at a rate inversely proportional to the squared mass of the particle and is therefore much less dominant for muons, which are heavier than electrons and positrons by a factor of about 200. The main energy loss of muons comes from **ionisation** of atomic nuclei [2]. The energy loss due to ionisation of a particle traversing a material is described by the *Bethe-Bloch-Formula*:

$$\frac{dE}{dx} \approx -4\pi\hbar^2 c^2 \alpha^2 \frac{nZ}{m_e v^2} \left(ln \left[\frac{2\beta^2\gamma^2 c^2 m_e}{I_e} \right] - \beta^2 \right) \qquad [2] \tag{2.4}$$

Here, $\hbar = 6.6 \times 10^{-34}$ Js is the *Planck constant*, c is the speed of light in vacuum, n_Z the number density of charges, Z the atomic number of the medium, $m_e = 0.51\,\mathrm{MeV}/c^2$ the mass of the electron and γ the relativistic factor:

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} = \frac{1}{\sqrt{1-\beta^2}} \tag{2.5}$$

α is the fine-structure constant:

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137.136}, \tag{2.6}$$

with the permittivity of free space, $\epsilon_0 = 8.85 \times 10^{-12}\,\mathrm{F/m}$, and the electric charge, e. Finally, I_e is the effective ionisation potential of the medium, averaged over all atomic electrons. It has a value of around 10 Z eV.

Fig. 2.1 shows the energy loss due to ionisation for different particles in different media as described by the Bethe-Bloch formula as a function of $\beta\gamma$. The curve has a minimum around a value of 3 $\beta\gamma$. A particle of that momentum is called a *minimum ionising particle (MIP)*. Since the energy loss due to ionisation is the dominant loss of energy for muons and they often have a momentum below $100\,\mathrm{GeV}/c$ in the detector, they are close to a MIP. Therefore muons lose comparatively little energy in the detector and can travel the longest distances. In most cases, muons are not stopped in the detector and their energy and momentum has to be inferred from their track without measuring their full energy at the end of their life.

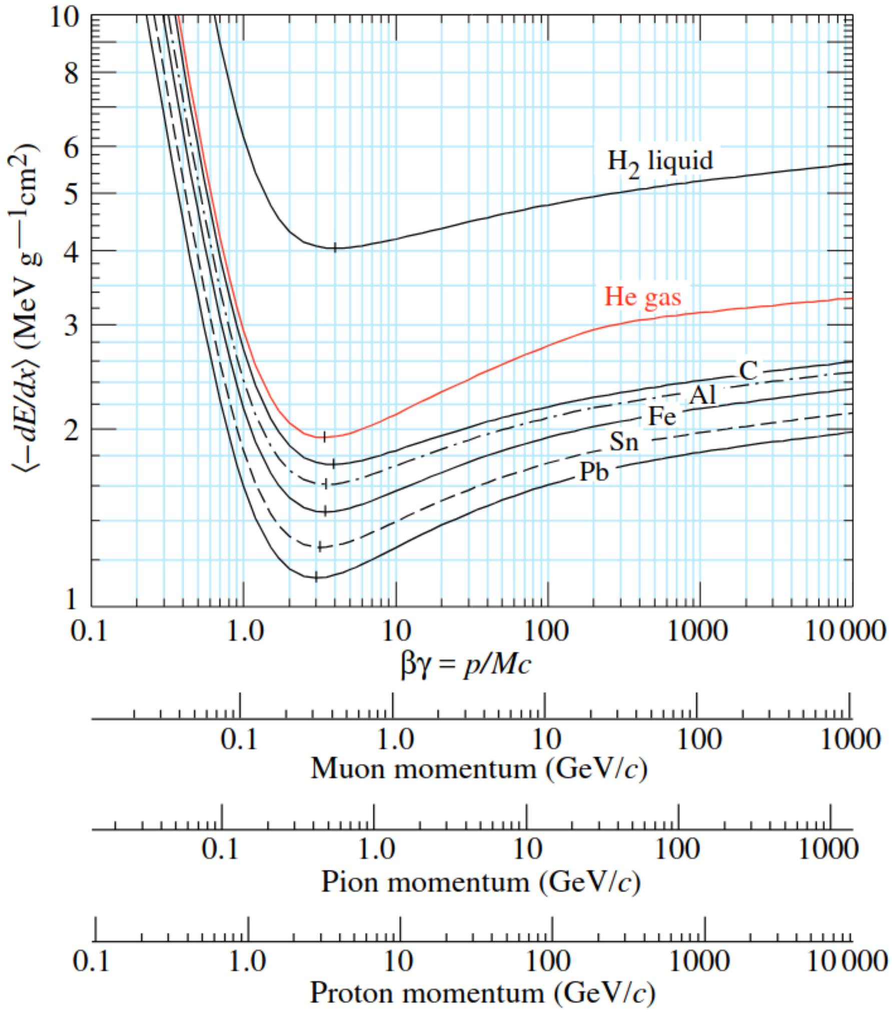


Figure 2.1: Mean energy loss as described by the Bethe-Bloch formula exemplary in liquid hydrogen, gaseous helium, carbon, aluminium, tin and lead, not including radiative effects. [3]

Hadrons, such as kaons and pions, also continuously lose energy through ionisation as they propagate through the detector. However, they can also **interact hadronically** with the atomic nuclei via the strong force, producing other hadronic particles, which themselves travel through the detector, interacting with the material.

A shower of particles produced this way is called a *hadronic* shower. Hadronic showers are much more variable than electromagnetic showers, because they can produce a wide range of particles. A large number of neutral pions, consisting of a superposition of an up-anti-up state and a down-anti-down state, are produced and decay in the most cases into two photons: $\pi^0 \to \gamma\gamma$. Each of these photons then triggers an electromagnetic shower, so that electromagnetic showers are also prevalent within hadronic showers. Additionally, on average 30% of the incident energy of a hadronic shower is effectively lost due to nuclear excitation and the break-up of nuclei.

Not only charged hadrons, but also neutral hadrons such as the neutron can interact with the nuclei via strong interaction and therefore produce hadronic showers.

2.3.4 Material challenges in particle track reconstruction

While many of the particle-matter interactions mentioned above are exploited to detect the track of a particle and measure its final energy, they can also occur over the entire detector length, losing the particle energy and momentum in ways that are hard to account for, and scattering its path. Tracking relies on connecting the point of origin with the point of decay or absorption of a particle via a bent line by the magnetic field. If multiple scattering occurs along the way, the path of the particle might be so far off such a line that it cannot be reconstructed anymore.

Energy loss can have a similar effect, as particles with less energy will bend differently in the magnetic field due to the Lorentz force. The change in radius of the movement makes it harder to predict its path. In addition, a particle may be stopped prematurely in the detector before its path can be reconstructed. This can even be caused by inactive detector material, such as cables or scaffolding, losing any possibility of detecting the interaction.

To correctly represent the track reconstruction efficiency, the probability that all these effects will prevent a particle from being reconstructed must be described. This provides a major challenge to the simulation of track reconstruction. Furthermore, these effects can occur in the interaction with both active and inactive detector material. An exact description of the detector material at each point of the detector is therefore required to correctly simulate the material distribution and hence the likelihood of interactions. However, this can only be known to a limited extent. Therefore, a data-driven measurement of the track reconstruction efficiency has to be employed, as a simulated attempt may fall short of the correct description.

Chapter 3

Experiment

An overview over the LHCb detector at the LHC experiment is given with a special focus on the tracking system. The LHCb software framework is introduced. A brief overview of the method of aligning the LHCb detector for measurements is given.

3.1 LHC

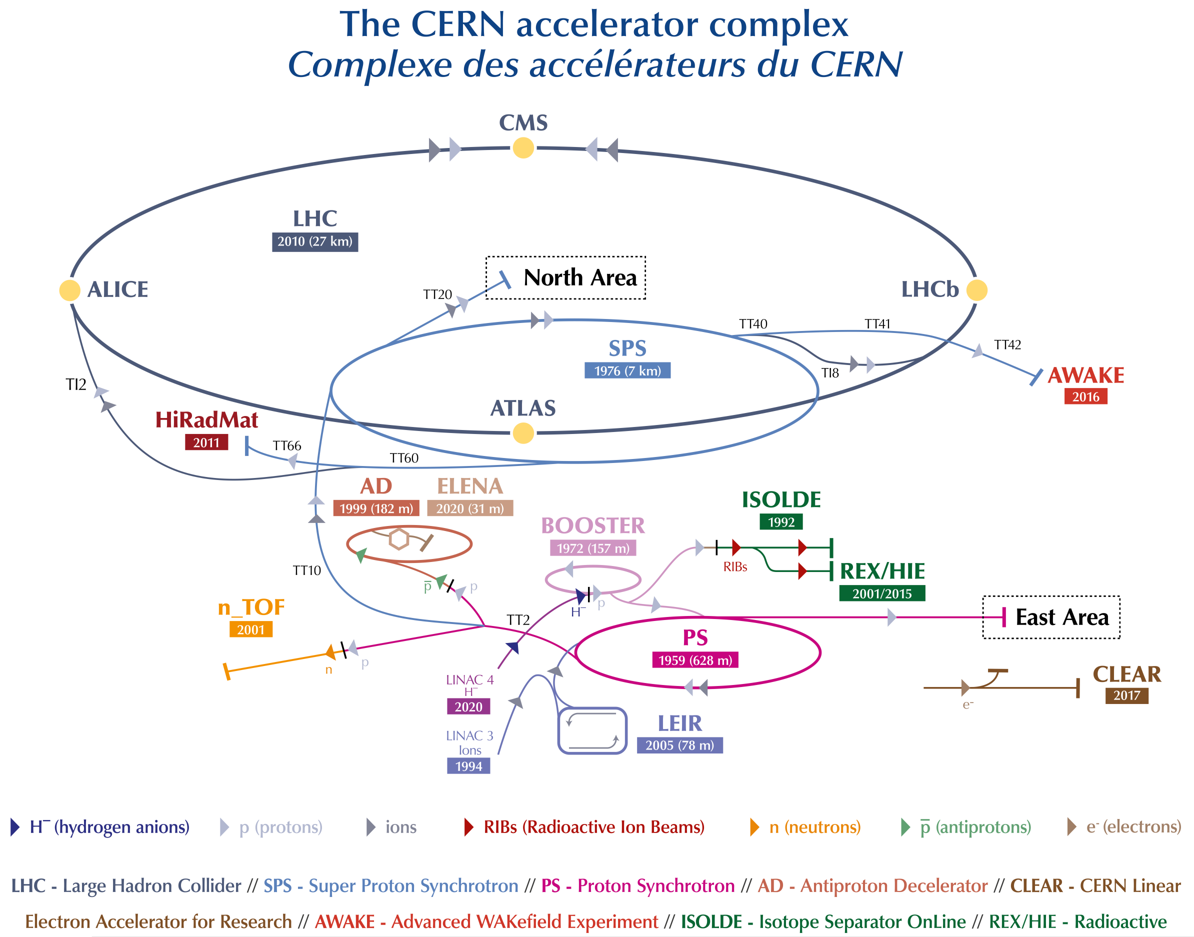
The Large Hadron Collider (LHC) is the world's largest particle accelerator, located at CERN. Its main accelerator ring consists of a 27 km-long storage ring beneath France and Switzerland. Protons and heavy ions are accelerated to energies very close to the speed of light and collided in four interaction points, where the four major experiments of CERN are stationed: ALICE, ATLAS, CMS and LHCb.

The ATLAS and CMS detectors were both built with the primary goal of discovering the Higgs boson. Since its discovery, they concentrate on measuring the Higgs boson and top quarks, and searching for new physics in high energy regimes. They were built with the same purpose in mind in order to confirm each other's measurements. ATLAS is the largest of the four experiments, with a length of 46 metres and a diameter of 25 metres, while CMS is the heaviest with a total weight of 14,000 tonnes. ALICE is dedicated to heavy ion physics and the study of quark-gluon plasmas. LHCb, the LHC beauty experiment, focuses on the measurement of decays from b and c meson decays, with a focus on flavour physics and CP violation. While ATLAS, CMS and LHCb concentrate more on proton collision physics and ALICE on heavy ion collision physics, all four experiments measure data from the same colliding particles at the same time.

This thesis concentrates on collisions involving protons. Protons are produced by stripping hydrogen ions of their electrons in an electric field and accelerating them to an energy of 160 MeV. This is done in the Linear Collider 4 (LINAC 4) and during the injection from

LINAC 4 into the Proton Synchrotron Booster (BOOSTER), which accelerates the protons to 2 GeV. They are then injected into the Proton Synchrotron (PS), which accelerates them to 26 GeV. From the PS they are transferred to the Super Proton Synchrotron (SPS), which accelerates them to 450 GeV and then them injects further into the LHC. The full accelerator complex can be seen in Fig. 3.1.

In the LHC, proton beams circulate in both directions of the ring, being accelerated to an energy of up to 6.5 TeV and then held at that energy for several hours.



[redacted]

Figure 3.1: The CERN accelerator complex layout. [4]

3.2 LHCb detector

The LHCb experiment is the smallest of the four major LHC experiments and differentiates itself from the other experiments at first glance mostly in that it is a single-arm forward spectrometer instead of a $4\pi-$detector. Therefore, it does not cover the collision point symmetrically in both directions of the beam pipe, but is designed to reconstruct the highly energetic forward particles in one direction of the collision point. While this results in a loss of 50% of its acceptance due to only covering one side, it is highly specialised for reconstructing the decay products of b and c meson decays, which are often produced with very high momentum at a small angle in both directions of the collision point. The full detector layout can be seen in Fig. 3.2.

The direction from the collision point into the detector is called *downstream*, while the other direction is called *upstream*. In the commonly used coordinate system, downstream and upstream are defined along the z-axis, which lies parallel to the beam pipe. The x-axis is defined as parallel to the ground and the y-axis is defined as pointing straight up, both perpendicular to each other and to the z-axis.

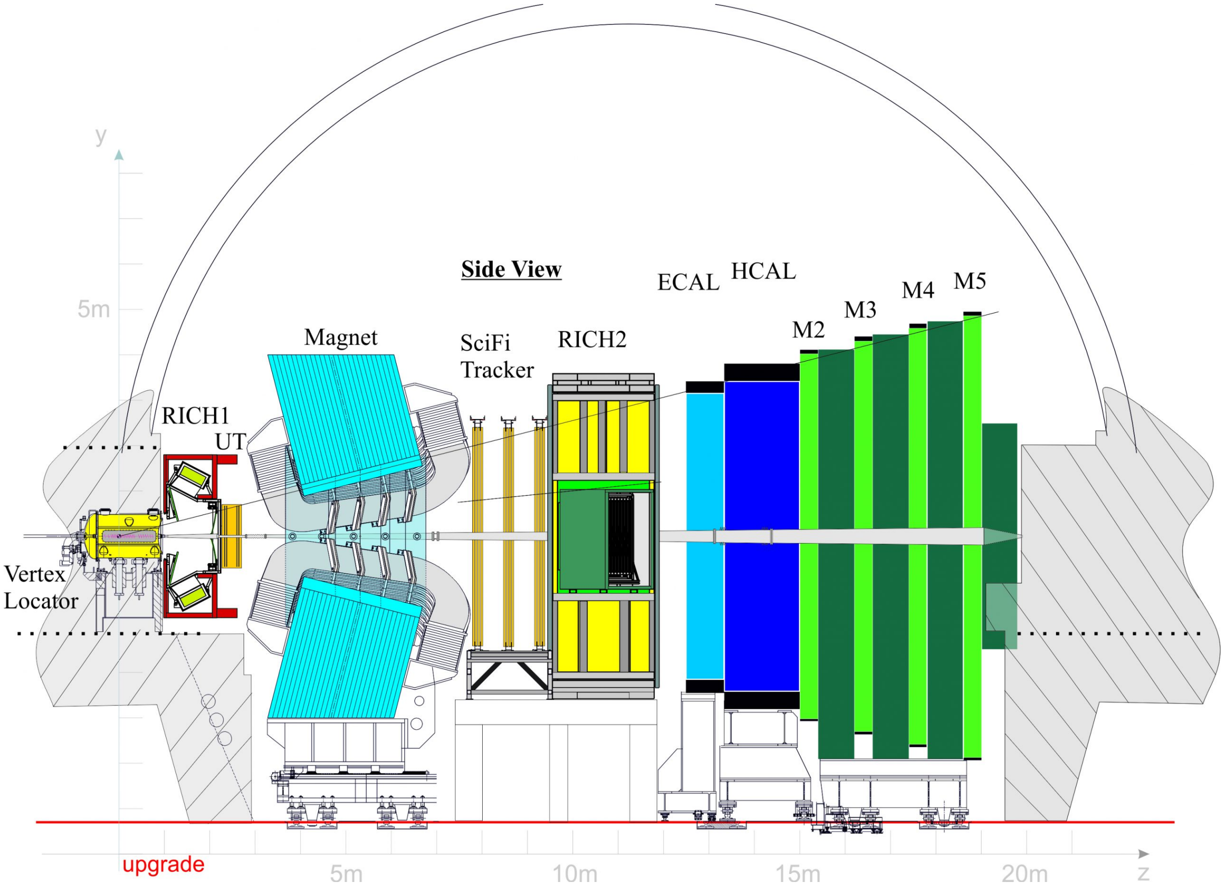


Figure 3.2: Cross-section through the layout of the LHCb detector. [5]

It is often useful to describe the momentum of a particle only in the directions of the x- and y-axes, as the original particles only carry momentum in the z-direction. Momentum in the x-y plane is called transverse momentum:

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{3.1}$$

The azimuthal angle in the x-y-plane is called ϕ. The *pseudorapidity*, η, is introduced to complete the description of the three-dimensional room:

$$\eta = -\ln[tan(\theta/2)] \tag{3.2}$$

An overview of the coordinate system of LHCb is given in Fig. 3.3.
Unlike the other detectors, LHCb does not measure the decays of head-on particle collisions but reduces the incoming particle flux and the number of collisions per interaction by keeping the two beams at a relative distance from each other. This reduces the integrated luminosity in LHCb in comparison to that of the other experiments.
LHCb has measured data in three data-taking periods: Run 1 from 2011 to 2012 with a centre-of-mass energy of 7 and 8 TeV and an integrated luminosity of 15 fb^{-1}, Run 2 from 2016 to 2018 with a centre-of-mass energy of 13 TeV and an integrated luminosity of 2.38 fb^{-1}, and Run 3 since 2022 with a centre-of-mass energy of up to 13.6 TeV, so far with an integrated luminosity of just above 11 fb^{-1}, most of which has been collected in 2024. Run 3 will continue on until 2026 and is the data-collection period on which this thesis focuses. Fig. 3.4 shows the integrated luminosity measured over time for each year.

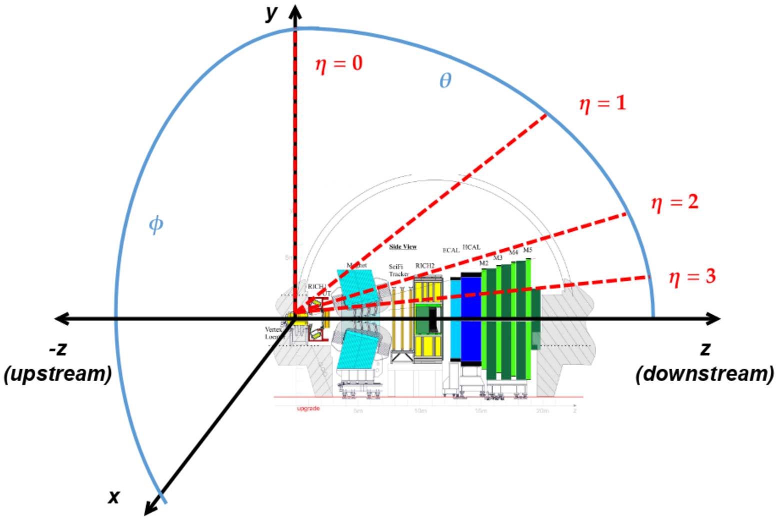


Figure 3.3: Overview of the coordinate system of LHCb, with the definition of the angles η and ϕ.

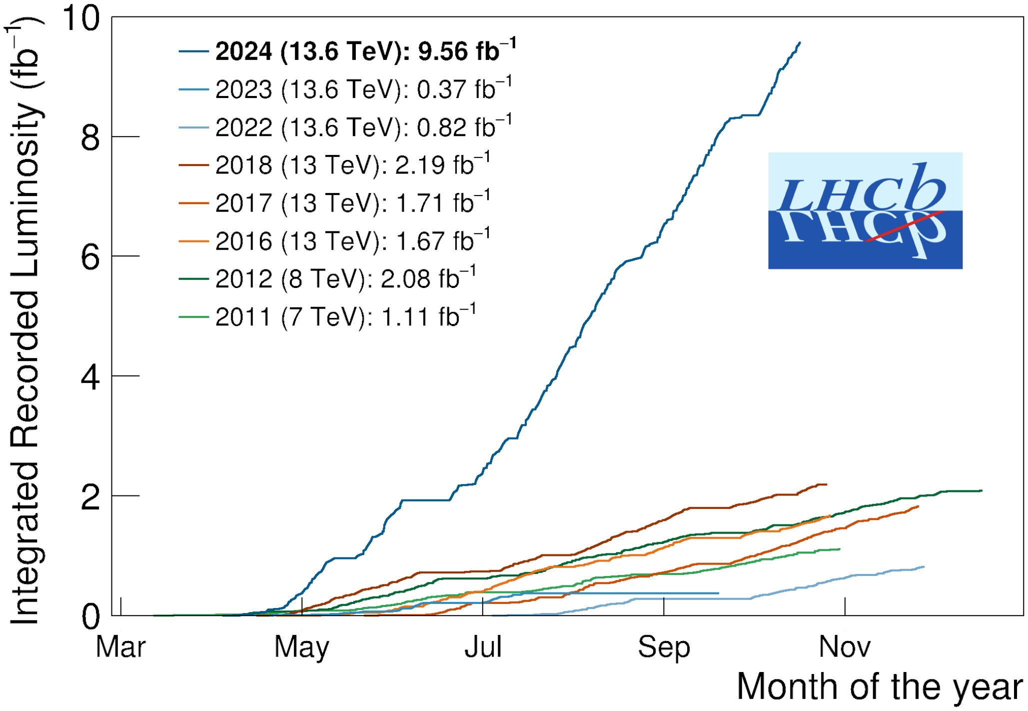


Figure 3.4: Collected integrated luminosity over time for each of the data-taking years at LHCb. [6]

LHCb underwent a major upgrade between the data acquisition periods of Run 2 and Run 3. Most of the tracking subdetectors were replaced, the trigger system was completely overhauled and the software framework for measuring, triggering and analysing the data was updated.

A detailed overview of the detector layout as shown in Fig. 3.2 is given next.

3.3 LHCb magnet

A central feature of LHCb is its magnet [7], located about 5 m downstream of the collision point. It is responsible for bending the path of charged particles in order to measure their mass and charge. The magnetic field has an integrated strength of 4 Tm. The magnet has two configurations called *magnet up* (MU) and *magnet down* (MD) [8], corresponding to the dipole magnetic field being along the positive or negative y-axis. The detector configuration switches between the two polarities to minimise systematic effects of the field on the tracking.

3.4 LHCb Tracking detectors

The LHCb tracking system consists of three tracking subdetectors, called the Vertex Locator, (VELO), the Upstream Tracker, UT, and the Scintillating Fibre tracker, SciFi. Their main task is tracking the path of particles through the detector.

3.4.1 Vertex Locator

The Vertex Locator (VELO) [9], is a silicon vertex detector surrounding the interaction point very closely. Its main task is the precise location of primary vertices in the vicinity of the collision point. It consists of two VELO halves on either side of the beam axis, each consisting out of a number of aluminium foils (RF foils) installed behind each other. A schematic overview of the layout of the RF foils and the two detector halves can be found in Fig. 3.5. A picture of the VELO foils is shown in Fig. 3.6

In order to measure the primary vertices of the collision as precisely as possible, during stable beam measurements the VELO is positioned at a distance of between 4.9 and 5.6 mm around the LHC beams. To avoid damage to the sensitive detector during unstable beam conditions such as during ramp-up or injection, both VELO halves can be moved by about 3 cm on a horizontal axis such that both halves are closed during measurement and open as soon as conditions destabilise. The positions of the VELO halves in closed and open position can be found in Fig. 3.5. While the VELO is closed, there is a small overlap between the two VELO halves for alignment purposes.

The VELO is a very precise detector operating at a very high resolution. As it reconstructs the primary vertices of particle decays even in events with a high pile-up of dozens of decays and is responsible for separating and matching the individual candidates, it plays a crucial

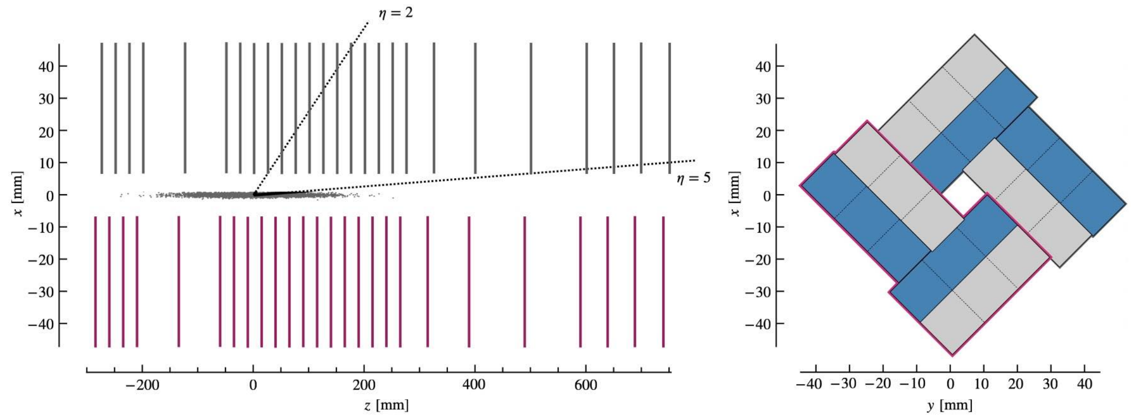


Figure 3.5: Schematic layout over the VELO foils (left) around the interaction point, and the closed VELO (right). [5]

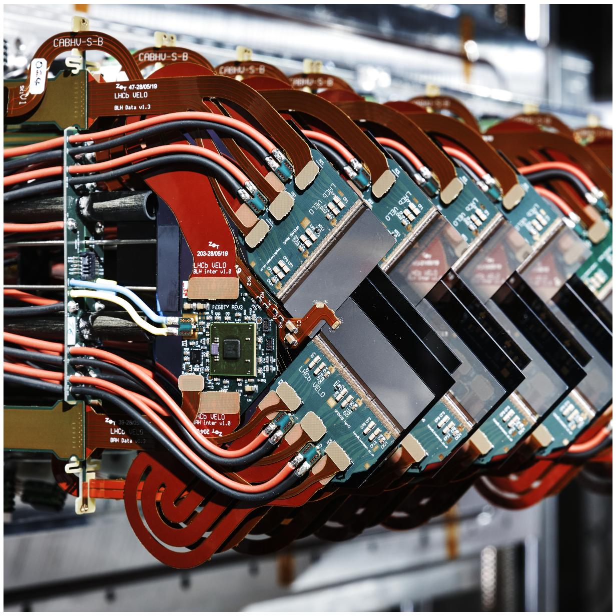


Figure 3.6: Photo of several foils of one half of the VELO. [10]

role in the reconstruction of tracks. In addition, as it is positioned around the interaction point, the VELO presents the starting point for aligning the LHCb subdetectors.

In January 2023, during the end-of-year technical shutdown, an RF foil was damaged during an emergency test [11]. As a result, during the data collection in 2023, the VELO could not be moved for fear of further damaging the detector, and measurements could only be taken with a half-open VELO. This severely limited the acceptance of data taken in that year, as decays with a high pseudorapidity could not be reconstructed.

3.4.2 Upstream Tracker

Downstream of the VELO and directly before the magnet is the UT or *upstream tracker* [12]. It consists of four high-granularity silicon microstrip detector planes. The silicon microstrips run vertically in the y-direction to optimally determine the momentum of particles bent in the x-z plane. The two middle planes are set at an angle of 5° with regard to the outer planes to also provide information about the y-coordinate position of hits in the UT. An overview of the layout can be found in 3.7.

UT information greatly improves the resolution of tracks reconstructed in the VELO and the SciFi. In addition, the UT allows for the measurement of the momentum of so-called upstream tracks which due to their low momentum are bent out of the detector by the magnetic field before reaching the SciFi.

Due to delays in the installation of the LHCb upgrade, the UT installation was only completed during the end-of-year technical shutdown of 2023 and was used for measurements

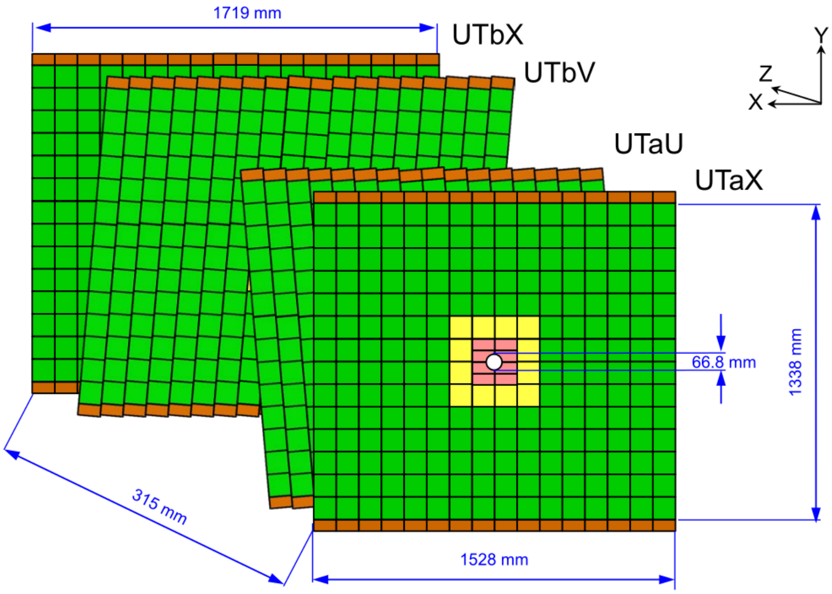


Figure 3.7: Overview of the UT geometry in LHCb [12].

since April 2024. Measurements prior to this were taken without the UT.

3.4.3 Scintillating Fibre tracker

The *Scintillating Fibre tracker*, or SciFi [12], is a detector consisting out of twelve layers of silicon fibre mats arranged behind each other directly downstream of the LHCb magnet. The twelve layers make up three stations, each consisting of four layers, where the two outer layers have no stereo angle with respect to the y-axis, while the two inner layers have an angle of ± 5° with respect to them. The four layers of the last station consist of twelve modules each, while the other eight layers were reduced to only eight modules each for budgetary reasons, giving a total of 112 modules. An overview of the position of the SciFi between the LHCb magnet and RICH can be found in Fig. 3.8.

The SciFi has a very high single hit detection efficiency above 95% and single hit spatial resolution better than 100 µm in the bending plane of the magnet. Its position directly after the magnet allows for the measurement of tracks bent in the magnetic field, and extrapolation towards the VELO allows for a reconstruction of the full track in most cases of particles reaching the SciFi before being absorbed by the detector material or bent out of the detector acceptance.

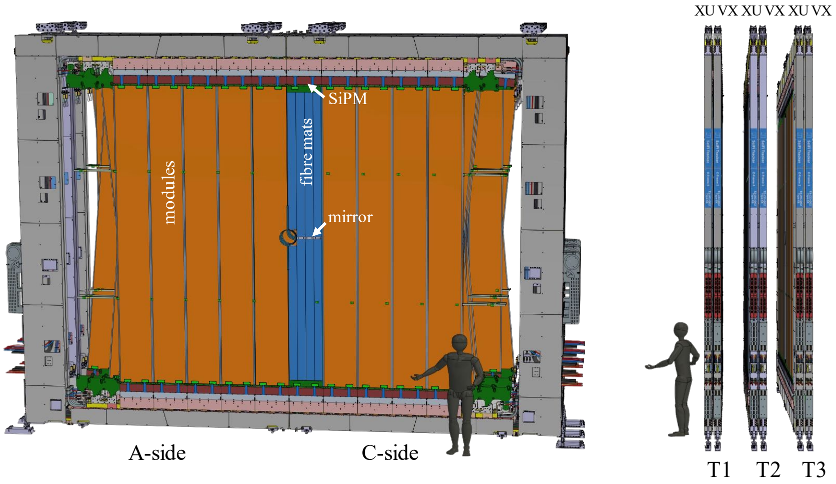


Figure 3.8: Schematic of four SciFi layers making up a station behind each other (left) and the three stations, T1, T2 and T3, behind each other (right), both in comparison to the size of an average human adult. [5]

3.5 Particle identification subdetectors

Correct identification of particle types is essential for flavour experiments. It is mandatory especially for flavour tagging in rare decays involving b and c quarks and for precision measurements of CP violation. At LHCb, two Cherenkov detectors called RICH1 and RICH2 are dedicated to the identification of hadrons, especially the separation of kaons and pions. The electron calorimeter (ECAL) is responsible for electron and photon identification, while the hadronic calorimeter (HCAL) provides additional information to optimise the identification variables. The muon stations are used to identify muons. [13]

The two Ring Imaging Cherenkov detectors, RICH1 and RICH2, are located between the VELO and the UT, and between the SciFi and the calorimeters. They are specialised for the measurement of charged hadrons over a momentum range from 0.5 to $100\,\text{GeV}/c$. They measure the momentum and mass of particles passing through via the emission of Cherenkov radiation as it occurs for a particle passing through a medium faster than the speed of light in that medium. RICH1 uses C_4F_{10} gas as radiator medium, while RICH2 uses CF_4.

The calorimeters, located between RICH2 and the muon stations downstream of the magnet, are an electromagnetic calorimeter, ECAL, and a hadronic calorimeter, HCAL. The ECAL consists of scintillator layers alternating with lead layers as absorption material. Its task is to stop and absorb photons and electrons in the detector, causing them to decay into electromagnetic showers. It also delivers essential information for the identification of electrons and photons.

The HCAL alternates its scintillator layers with thin iron layers. It absorbs hadrons passing through the detector by breaking them into hadronic showers that are stopped in the iron layers. It provides additional information to improve the identification of hadrons, which is done mostly by the RICH detectors.

Both the ECAL and the HCAL collect the light from the scintillating material in wavelength-shifting fibres connected to photomultipliers.

The only particles that pass through the entire detector and the HCAL with relatively little energy loss are, neglecting neutrinos, muons. They are therefore measured in the four muon stations, called M2 - M5, placed at the end of the detector. An additional protective layer is installed between the HCAL and the muon stations to reduce the occupancy. The muon stations rarely absorb the muons completely, but try to measure their momentum as precisely as possible and provide particle identification information for muons. They can also be used for muon tracking as they provide geometrical hit information, although this is almost exclusively used for the measurement of the track reconstruction efficiency.

The muon stations are located behind the HCAL, with another protective layer between them and the HCAL to reduce the occupancy in the muon chambers. They consist of multiwire proportional chambers filled with a mixture of carbon dioxide, argon and tetrafluoromethane.

3.6 LHCb software framework

The LHCb software framework collects electric signals from the various subdetectors, reconstructs tracks and assembles them into event information, identifies decays and presorts the events. All events that are not identified as containing decays of interest are permanently removed to save storage space, highlighting the importance of constantly verifying the quality of event reconstruction and selection. The following paragraphs describe the LHCb trigger system, used to trigger events to be stored on the *LHCb Grid*.

3.7 LHCb trigger

The main focus of the LHCb experiment is on the decay of b- and c-hadron decays, which have a very high cross section of several mb at 13.6 TeV. This is in contrast to the ATLAS and CMS experiments, which target t-quark and Higgs decays occurring with cross sections of few nb. While this small cross section allows for a coarse trigger selection to quickly identify likely t-quark and Higgs candidates, the precision measurements at LHCb need very high statistics to save events at a rate of above 4 MHz. For this reason, while ATLAS and CMS use a coarse hardware trigger on the first stage, LHCb works with a pure software trigger at a rate of up to 40 MHz. The software trigger works on two stages, called the High Level Trigger 1 and 2, HLT1 and HLT2, where HLT1 is programmed on GPUs and HLT2 on CPUs.

Since all events that do not pass one of the trigger stages are removed completely, it is necessary to utilise real-time time alignment and calibration on the detector so that problems during the trigger stage can be detected immediately and corrected for before too much data is lost. This necessitates a fast, precise and efficient reconstruction of the tracks, which is done partially during HLT1 and fully during HLT2. Especially in 2023, when the detector was not yet fully calibrated and commissioned, the track reconstruction efficiency provided valuable feedback for the detector alignment and calibration.

The HLT1 provides an output of approximately 1 MHz, which is written to a buffer of approximately 30 PB, storing the data for up to 80 hours and then delivering it at an output rate of 1 MHz to HLT2. The buffer stores the data to be used immediately for calibration and alignment, providing real-time feedback about the efficiency of HLT1. As shown in the full trigger layout in Fig. 3.9, the data is separated into events where the full event information is stored, marked *Full Events*, events where only the decay of interest is stored to save on storage space, marked *Turbo Events*, and events that are further used for alignment and calibration, marked *Calib Events*. HLT2 reduces the data stream to approximately 10 GB/ s in these three streams. Track reconstruction efficiency measurements are used for the detector calibration and therefore use the *Calib Events* stream, which stores full event information.

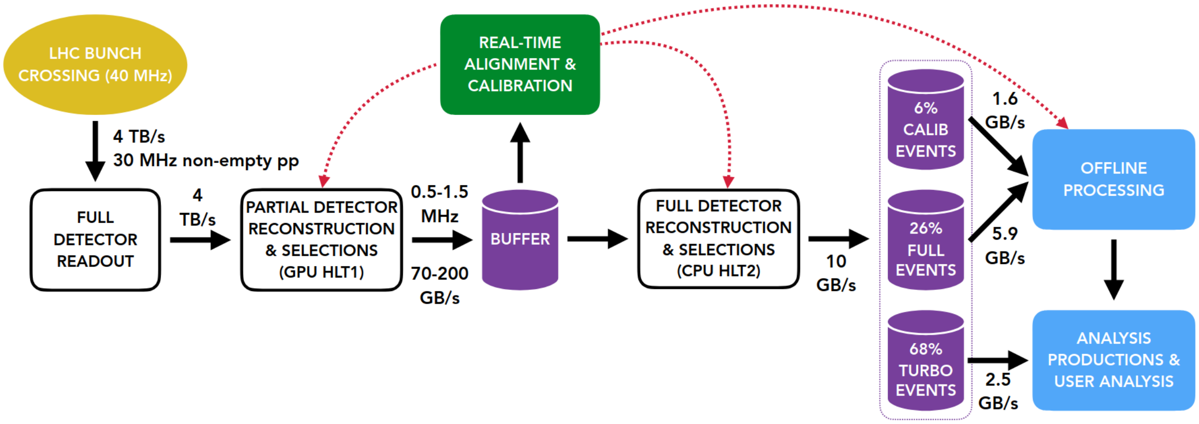


Figure 3.9: Trigger layout of LHCb. [14]

3.8 LHCb detector alignment

For the correct interpretation of the electric signals in the subdetectors and their combination into tracks, it is essential that the time and location of each signal detected in the detector is known with very high precision. The subdetectors must be aligned with regard to each other, and within themselves both spatially and in time.

The *track reconstruction efficiency* is highly susceptible to the alignment of the detector and can suffer significant losses in a poorly aligned detector. It is therefore both limited by the alignment and can provide valuable feedback to it.

The alignment can be separated into *time alignment* and *spatial alignment* of the detector. At LHCb, bunches of protons collide every 25 ns. This means that the next collision occurs before the decay products of the previous collision have fully crossed the detector, been detected and identified. It is therefore essential that each subdetector has precise timing information for each measured signal.

Time alignment is done separately for each subdetector. Instead of about 3,000 proton bunches, only a single proton bunch is sent to collide for time alignment. This increases the time between collisions from 25 ns to several µs. The expected signal is analysed to determine the correct calibration for the subdetector. As the rate is much too low for a normal measurement of physics data, this is done very rarely. In 2024 time alignment was performed at the start of data taking as well as after major changes in the system. Measurements showed that outside of these changes, the time alignment varied by less than one ns.

Unfortunately, in 2023 there were some issues with the time alignment of the muon system, which was not as well time aligned as expected. This lead to loss of data including muons as muonic information was incorrectly assigned.

The *spatial alignment* of the VELO subdetector is a real-time calibration task as opening and closing the VELO can vary the exact position of the VELO by almost a millimetre. For this reason, a new VELO alignment is performed immediately after each closing of the VELO to update the alignment constants for the current runs. The alignment quality is checked constantly by comparing the position of reconstructed primary vertices between the left and the right VELO halves.

The spatial alignment of the detector is performed in sequence: first, the VELO is aligned, then the UT and SciFi are aligned in combination with respect to the VELO. The alignment parameters are the three translational and three rotational degrees of freedom for each alignable detector element. They are determined by looking for well-known track signatures. In 2024, this was done using $Z \to \mu^+\mu^-$ decays. In 2023, the decay $D^0 \to K\pi$

was used.

In the first step, the alignment parameters are kept constant and the positions of the track hits are varied until a track is reconstructed by minimising the trackχ^2. In the second step, the track hit positions are kept constant while the alignment parameters are varied by minimising the trackχ^2. Both steps are repeated in a loop until the optimal alignment is reached. [15]

The RICH mirror alignment, which is done by comparing the Cherenkov angles of the detector photons with their expected values, is performed once per proton fill. It can take up to hours, but can be applied retrospectively, as RICH is not used in HLT1.

The muon detector is aligned very rarely, since for most analysis a very high precision of the muon detector alignment is not required. Unfortunately, the track reconstruction efficiency using muons is one of the rare cases where the muon alignment is relevant, as the exact position of hits in the muon detector is needed to reconstruct partial muon tracks. Therefore, the muon alignment can have an impact on the track reconstruction efficiency.

The ECAL is calibrated about once a month using the π^0 masses from $\pi^0 \rightarrow \gamma\gamma$ decays. The alignment of RICH and ECAL is not relevant for the track reconstruction efficiency.

Chapter 4

Track reconstruction at LHCb

This chapter gives an overview of the reconstruction of tracks at LHCb. Standard tracks reconstructed at LHCb are described. The reconstruction algorithms for standard tracks are explained in detail.

4.1 The unique challenge of track reconstruction at LHCb

The design of the LHCb experiment as a single-arm forward spectrometer aiming predominantly at the study of b-meson decays was heavily inspired by the HERA-B experiment, which took data at the HERA collider at DESY from 2000 to 2003 [16]. HERA-B provided the lesson that in order to achieve the background rejection and track reconstruction efficiency necessary for the analysis of b-decay products, the detector must use as little material as possible to maximise the radiation and interaction lengths. The more detector material exists, the more particles interact with it, either by modifying their trajectory or by being stopped completely. This leads to loss of efficiency, as well as a loss of precision of momentum and energy estimates. LHCb is therefore optimised for minimal detector material, which means that over the length of the detector thin tracking layers and very few subdetectors are interspersed with empty space filled only with air to avoid deflection as well as loss of energy and momentum of the particles. This allows the momentum of low energy particles to be measured much more precisely than at other detectors at CERN.

Track reconstruction refers to the estimate of momentum vectors along the trajectory of a particle through the detector. This is done by combining electric signals caused by the ionisation of detector material by a passing particle. Each of these electric signals is called a *hit* in the detector. The combination of hits from the different subdetectors is used to reconstruct the track.

A high efficiency of the track reconstruction is necessary especially for decay channels where a large number of daughter tracks is reconstructed.

A hit is detected at LHCb when a particle interacts with the detector material via ionisation or atomic excitation. This usually does not significantly change the direction of the momentum. The particle loses energy, as described by the Bethe-Bloch formula, but the effects of ionisation energy losses are negligible.

Since hadrons interact additionally via inelastic scattering with atomic nuclei, the hadronic interaction with the detector material is the most important source of systematic uncertainties for the track reconstruction efficiency of hadrons.

As has been mentioned in Chapter 3, in Run 3 the data produced with LHCb is processed with purely software-based triggers performing real-time analysis. Collisions at the LHC occur at a rate of up to 40 MHz, which is therefore the input rate of the first software based trigger, the High Level Trigger 1 (HLT1). HLT1 has an output rate of approximately 1 MHz to the buffer, which is used for alignment and calibration of the data, and passes the data through a buffer on to the second software based trigger, the High Level Trigger 2 (HLT2), which fully reconstructs the events to the level of detail depending on the stream.

4.2 HLT1 track reconstruction

The first step of LHCb's reconstruction is the HLT1, which runs the Allen application [17], an implementation of partial reconstruction sequences on GPUs. Allen reconstructs tracks based on hits in the VELO, UT and SciFi, as well as primary and secondary vertices. It identifies electrons using ECAL clusters, recovers Bremsstrahlung photons and matches tracks to muon station hits to identify muons. Predefined event criteria in HLT1 *trigger lines* use this information to sort and filter events.

Prior to any reconstruction, a *Global Event Cut* (GEC) is applied, which removes approximately 10% of the incoming events due to their high occupancy, as these events require a disproportionally high amount of computing resources to reconstruct.

Optionally, the VELO information is then decoded and clustered in the VELO's FPGA detector readout boards, called TELL40, before the raw event information is transferred to the Event Builder (EB).

There, a Search By Triplet algorithm [18] is used to find straight trajectories in the VELO from VELO hits. These trajectories are then used to find primary vertices (PVs) in the VELO [19].

In the next step, upstream tracker (UT) hits are decoded if the UT is included in this reconstruction step. This is the case for data collected after August 2024. The UT hits are added to the VELO tracks [20].

After adding the UT hits to the VELO tracks, a first estimate of the charge and momentum of the tracks is obtained by extrapolating the tracks to the SciFi using the magnetic field parametrisation.

For data taken before August 2024, there is no charge and momentum estimate in the HLT1 reconstruction sequence up until this point due to the missing UT. Therefore, the matching algorithm, also called the seeding algorithm, is applied to reconstruct standalone SciFi tracks from SciFi hits, extrapolate them to the VELO and match them to the VELO tracks. This is optimised for tracks with a transverse momentum of at least $1\,\mathrm{GeV}/c$ and a momentum of at least $5\,\mathrm{GeV}/c$ [21].

From August 2024 onwards, the UT reconstruction is included in the Allen application. Complementary to the matching algorithm, the forward algorithm [22] is used, which is optimised for tracks with a transverse momentum greater than $0.5\,\mathrm{GeV}/c$ and a momentum greater than $3\,\mathrm{GeV}/c$ [21]. It uses the charge and momentum estimate from VELO tracks matched with UT hits to extrapolate to the SciFi, find SciFi hits around the extrapolated trajectory and add them to the VELO +UT tracks. The matching algorithm is run after the forward algorithm to pick up any tracks that can be reconstructed without UT hits.

A Kalman Filter with a parametrised noise matrix [23] is then applied to the partial trajectory in the VELO to improve the momentum estimate. Only the partial trajectory is used to save on computational resources.

The track with the improved momentum estimate is then extrapolated to the muon stations to potentially be matched to muon hits [24].

Electrons are identified by matching the tracks to cell clusters in the ECAL. Bremsstrahlung photons are recovered, correcting the momentum estimate for electron tracks.

Finally, secondary vertices are reconstructed from pairs of VELO track segments.

Thus reconstructed tracks are then passed on to the buffer and from there to the HLT2 if they fulfil the conditions of one of the predefined HLT1 trigger lines. The HLT2 performs the reconstruction again from scratch, using no reconstruction information from the HLT1 other than the information which trigger line was triggered in the HLT1. The purpose of reconstruction in the HLT1 is primarily to filter out unnecessary events and to do a preliminary sorting into trigger line categories based on the event content.

4.3 HLT2 track reconstruction

The HLT2 takes events that have been partially reconstructed using the Allen algorithm from the buffer and runs the full online reconstruction on them. Events not selected by any HLT2 line are lost forever, therefore key quantities from the HLT2 output are constantly monitored in the control room to be able to react immediately when something goes wrong.

The HLT2 runs the Gaudi-based Moore application, which has three main components: charged track reconstruction, calorimeter reconstruction and particle identification. The full detector information is used, although the HCAL is mostly useless for the reconstruction and is mainly used for monitoring and to improve the particle identification. Several track types are reconstructed: VELO tracks, seed tracks, upstream tracks, downstream tracks and long tracks. An overview of the track types reconstructed in HLT2 is given in Fig. 4.1. Moore also checks whether these tracks are likely to be fake tracks, also called ghost tracks, or clone tracks.

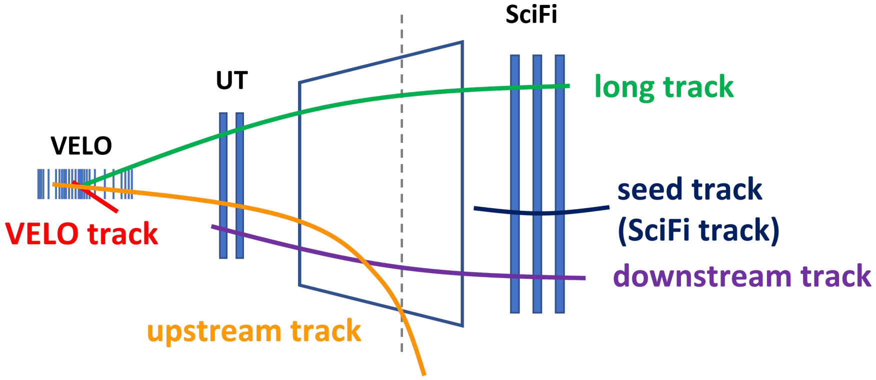


Figure 4.1: A schematic overview of VELO tracks, seed tracks, upstream tracks, downstream tracks and long tracks at LHCb.

4.3.1 Standard charged track reconstruction

Different reconstruction sequences are employed to reconstruct the different track types.

VELO tracks are produced by reconstructing straight lines from three-dimensional hits measured in the VELO. A simplified *Kalman filter* is applied considering multiple scattering and using noise parameters according to the average momentum spectrum.

Seed tracks or SciFi tracks are reconstructed using hybrid seeding. This means that a track is built from hits in two SciFi layers, then a hit is sought from a layer in between. If this is successful, additional hits are added using a third-order polynomial as a fitting model.

Downstream tracks are reconstructed by taking a *seed track* and extrapolating to the UT using a parametric model of the expected track trajectory based on the estimated momentum from the hybrid seeding fit. All UT hits close to the extrapolated position of the track are added. A Kalman filter removes any outlying hits and improves the momentum estimate.

Upstream tracks are reconstructed by exploiting the VeloUT tracking algorithm, using VELO tracks as a baseline and extrapolating linearly to a UT reference surface. All hits within a tolerance around the extrapolated trajectory are collected. A least-square fit is attempted to combine the VELO track with the hits, which are collected into triplets and quadruplets. If the fit quality is good enough, a maximum of one upstream track per VELO track is combined. These tracks are to date scarcely used.

Long tracks can be reconstructed using either the *matching* or the *forward* algorithm. The *matching* algorithm loops over the VELO and SciFi tracks, extrapolates both to a position in the magnet and estimates their distance. Pairwise combinations that are too far apart are rejected, then a machine learning algorithm qualifies the pairwise matching quality. Any number of track combinations can be added if their quality is good enough and the combinations differ enough from each other. UT hits are searched for and potentially added to the long track at this point. After the matching, all used VELO tracks and SciFi hits are removed and the *forward* algorithm is employed to try to construct long tracks from the remaining VELO tracks and SciFi hits.

The *forward* algorithm tries to extend the VELO tracks not used in the matching into the SciFi. If upstream tracks are available for the remaining VELO tracks, their momentum estimate is used, otherwise it is attempted to find SciFi hits without the UT. The VELO track is extrapolated to the SciFi. SciFi hits that can be fitted with a least-square fit are added. The forward algorithm requires a minimum of nine SciFi hits, but in the most cases there are between ten and the maximum of twelve. The momentum is re-estimated after combining the VELO or upstream track with SciFi hits. With the updated momentum information, a straight line is again extrapolated from the VELO to the UT to search again for UT hits. Finally, a Kalman filter is run to remove outlying hits and improve the momentum estimate.

4.3.2 Identification of track-like objects

Fake tracks, also called ghost tracks, are fully reconstructed tracks consisting of hits that cannot be associated with a single simulated particle and represent a possible background source.

Clone tracks are copies of subsets of tracks that share substantial portions of their detector hits with each other, with neither of the clones being a ghost track.

4.3.3 Calorimeter reconstruction

Calorimeter reconstruction, which runs after the reconstruction of standard tracks, forms clusters around energy peaks in the calorimeters using graph-based clustering algorithms [25]. A multivariate network is used to classify events as single-photon or multiple-photon events. Electron clusters are identified by extrapolating reconstructed tracks to the ECAL.

4.3.4 Particle identification

Particle identification (PID) relies on charged track trajectories. Electrons are identified by extrapolating tracks to the ECAL and combining them with ECAL clusters. Muons are extrapolated to the muon stations and combined with hits there. The minimum number of muon hits required for identification as muons depends on the momentum of the tracks. A multivariate classifier is used to reduce misidentifications.

Kaons, pions and protons mostly rely on RICH information for identification.

All PID information is combined to compute a global PID variable using a multivariate classifier that is assigned to each track.

4.4 Track reconstruction sequences

The track reconstruction sequence is optimised to avoid reconstructing the same track twice. VELO tracks are reconstructed first, then upstream tracks using the VeloUT algorithm. Hybrid seeding then reconstructs SciFi tracks. The matching algorithm runs first, trying to reconstruct long tracks. A Kalman filter fits the long tracks reconstructed with matching, then all used VELO tracks and SciFi hits are removed. The Forward algorithm tries to find long tracks from the remaining VELO tracks and SciFi hits. The Kalman Filter is applied to the tracks found by the forward algorithm. Afterwards, all used UT and SciFi hits are removed and the downstream algorithm tries to reconstruct downstream tracks from the unused SciFi hits. The Kalman Filter is applied to the reconstructed downstream tracks.

After the tracks are reconstructed and fitted with the Kalman Filter, each track is extrapolated to the muon stations, where it is attempted to match muon hits to the reconstructed tracks. Finally, calorimeter and particle identification information is added to the tracks. The most important tracks in LHCb analyses are the long tracks reconstructed in HLT2.

This thesis does not discuss the track reconstruction efficiency of HLT1, as it is only used as an intermediate step and the reconstruction is completely re-done in HLT2. Since the UT can be used to optimise long track reconstruction, but its contribution is not mandatory, the *long track reconstruction efficiency* is determined by the combined track reconstruction efficiency of the VELO and the SciFi.

Chapter 5

Tools & Applications

Dedicated tools and applications are required to handle the large data quantities which are analysed in this thesis. This chapter introduces Monte Carlo simulations, which are used to complement measurements on recorded data. The ROOT *framework and its application* ROOFIT*, necessary to extract the number of signal candidates in recorded data, are introduced. Tools for sWeighting, truth-matching and reWeighting to further analyse the data are explained.*

5.1 Monte Carlo Simulation

The most important part of determining efficiencies is simulated data. An efficiency can be easily estimated by simulating the expected signature and the detector's response to it. For this, *Monte Carlo (MC) simulation* is used. The aim of MC simulation is to randomly generate data behaving in the same way as the data recorded in the experiment.

The simulation of data in the LHCb experiment is complex and consists of multiple steps. The MC events are generated by the LHCb software package GAUSS [26]. In the first step, the proton-proton collisions themselves are simulated using an event generator, in this case PYTHIA [27]. The decay of the resulting unstable particles is then simulated using the EVTGEN [28] library. The produced particles are processed through a simulation of the LHCb detector by GEANT4 [29]. The detector response is simulated using the BOOLE package [30]. The simulated data is then subjected to the same reconstruction steps as data recorded in the experiment.

Since the details of the detector are neither known nor simulated with infinite precision, efficiencies derived from simulation do not necessarily reproduce reality exactly. Therefore, the track reconstruction efficiency is evaluated using the same data-driven method on both recorded and simulated data. The differences are used to evaluate correction factors that can be used to correct track reconstruction efficiencies derived on simulated samples for different signal signatures.

5.2 The ROOT data analysis framework and ROOFIT

The analysis of data produced at CERN, both in the experiment and in MC simulation, poses a special challenge. To meet this challenge, the C++-based object-oriented programming language ROOT was developed in the 1990s. It is designed to handle large quantities of data, with individual entries having many associated quantities. It therefore focuses on storing data in so-called *trees* or *tuples* with the substructures *branches* and *leaves*. Data can then be presented in the form of histograms in one or more dimensions, fitted, evaluated, minimised or manipulated in many different ways, providing a simple form of visualisation for the properties of large data sets.

Measured datasets usually contain not only the so-called sought-for *signal candidates*, but also *background candidates* which look similar to the *signal candidates*. Thus they cannot be easily discarded but are not really of interest. An easy way to deal with this statistically is to *fit* the distribution of a *variable of interest* where the variable distribution for signal and background candidates is known. This is done using ROOFIT [31].

The ROOFIT library within ROOT, originally developed for the BarBar collaboration, provides a toolkit for modelling candidate distributions in physics analyses. It can perform both binned and unbinned maximum likelihood fits with a wide range of mathematical functions as probability density functions. For all fits mentioned in this thesis, ROOFIT is used.

A fit to a variable of interest of a given data set not only allows to statistically determine the number of signal and background candidates, but also allows for further manipulation of the data via *sWeighting* and *reWeighting*.

5.3 Background subtraction using sWeighting

Background subtraction in this thesis is performed by so-called *sWeighting* using the *sPlot* tool [32].

The *sPlot* tool is designed to divide a sample into several categories based on the distribution of one or more variables. The distributions of these variables must be known in advance. The categories used in this thesis are the *signal category* and the *background category*. The variables used to determine affiliation with either category are called the *discriminating variables*.

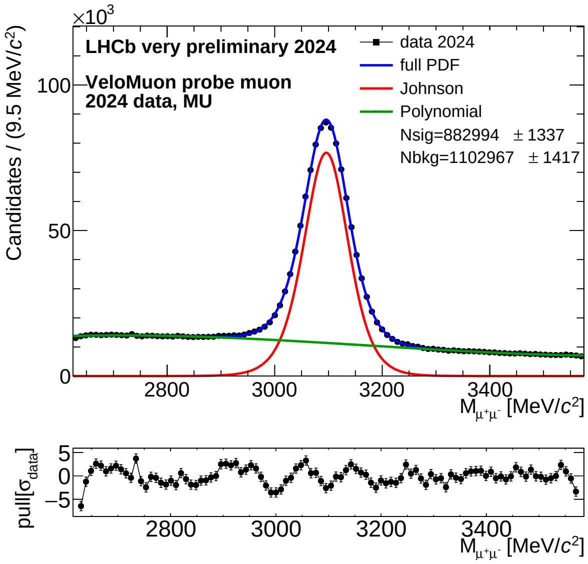


Figure 5.1: Fit to the invariant mass of two muon tracks used to reconstruct a J/ψ to sWeight the sample. The fit to the signal shape is shown in red and the fit to the background shape in green, with the total fit in blue.

In this thesis, the invariant mass of the final decay products is used as *discriminating variable*. Fig. 5.1 shows an example of a fit to a sample of recorded data containing both signal and background candidates. The distribution shown is of the invariant mass of two muons, where the signal candidates contain two muons coming from a J/ψ mother.

The *sPlot* tool can then be used to determine the distributions of variables that are not known for these categories, called the *control variables*. These are for example the momentum of the tracks or the number of tracks in an event. Instead of assigning each entry in the sample to a category, the entries are assigned *sWeights* according to their likelihood of being in each category by fitting the *discriminating variables*. This is used to compute the *sWeight* for each category.

To see the distribution for example for the *signal category* of a control variable, the control variable can then be plotted with each entry multiplied by its *sWeight* according to the category. This results in a distribution that is expected to represent only signal candidates, provided that the *discriminating variable* and the plotted *control variable* are not correlated. As an example, Fig. 5.2 shows the original and *sWeighted* momentum distribution of one of the muons from the sample *sWeighted* in Fig. 5.1. As can be seen, the transverse momentum is shifted to higher values when *sWeighting*, as background candidates tend to have lower transverse momentum.

In this thesis, *sWeighting* always refers to considering the distribution of a control variable for the *signal category*. This way, distributions of simulated data can be compared with the signal distributions in recorded data. To ensure comparability, the distributions of

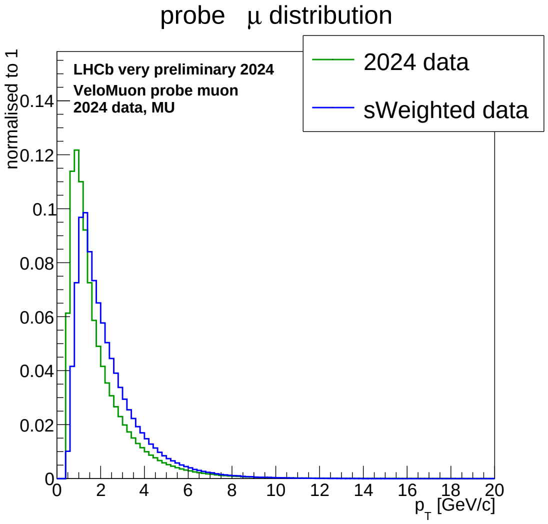


Figure 5.2: Normalised transverse momentum distribution of a muon (green), as well as distribution after sWeighting the data candidates using a fit to the invariant dimuon mass (blue).

simulated data rely on so-called *truth-matching* to match the reconstructed objects to the generated objects they are based on.

5.4 Matching reconstructed to simulated objects

Simulated samples used in this thesis only consist of events containing the desired signal decay or a similar decay with the same final state particles. Nevertheless, the sought-after signal decay is abundant, while the admixture of similar states is small and can be easily filtered out with a requirement on the invariant mass range. However, as the same reconstruction sequences are performed on these simulated samples as on recorded data, it is still possible that ghost tracks are reconstructed. Single tracks belonging to signal candidates may also be lost due to material interactions or due to reconstruction inefficiencies. Therefore not all reconstructed candidates are signal candidates.

To avoid uncontrolled influences when comparing distributions, the simulated candidates are *truth-matched.* When *truth-matching,* only candidates where all tracks are associated with a generated particle belonging to the desired signal decay are allowed to remain. In addition, the invariant mass of the final state particles cannot be reconstructed to be more than 100 MeV above the desired invariant mass.

5.5 Correction of differences between recorded and simulated data samples

Candidates in simulated and recorded data may cover different phase space regions and thus could be distributed differently in the variables relevant to the track reconstruction efficiency. There are two ways to account for this when comparing track reconstruction efficiencies from recorded and simulated data: binning or *reWeighting*.

Binning means the efficiency is given separately for different ranges of values, so-called bins, of the variable. While this may give more reliable results overall, the statistics required for each bin limits the use of this strategy, especially when several different, potentially correlated variables need to be considered.

The other option, reWeighting, means that each candidate is assigned a so-called *weight*. Each candidate contributes with its weight to resemble the reference distribution as close as possible.

The reference distribution is usually an *sWeighted* distribution of recorded data that differs from the target distribution, which consists of simulated data and is truth-matched.

For a one-dimensional reWeighting, the reference and target distribution are considered in one variable. In bins of this variable, the number of target and reference candidates is compared. A weight is computed for each bin. This weight is greater if there are fewer target than reference candidates, and smaller if there are more target than reference candidates. The objective is to reduce the mismatch between the reference and the target distribution.

These weights are stored and can be applied to all distributions of the target sample to see the consequential effects. While distributions of uncorrelated variables are unlikely to be affected in any significant way, distributions of correlated variables may also either improve or deteriorate in approximating the reference distribution. In this case, multi-dimensional reWeighting in several variables is necessary. This approach uses a Gradient Booster Reweighter [33] that trains Binary Decision Trees (BDTs) with a special loss function to determine the ideal weights in several dimensions of variables.

ReWeighted samples are used to check the influence of the disagreement between simulated and recorded data samples of variable distributions on the track reconstruction efficiency.

Chapter 6

Experimental Method

The objective of this thesis is to determine correction factors that describe the discrepancy between track reconstruction efficiencies in simulated and recorded data. This chapter introduces the tag-and-probe method, which is applied to both. Reconstruction sequences used to reconstruct special muon probe tracks are explained. The selection criteria for candidates used to estimate the track reconstruction efficiency are summarised and the fitting procedure to determine the signal yields is explained in detail.

6.1 Tag-and-probe method

The track reconstruction efficiency for a subdetector is determined by taking an event where it is known with a very high confidence that a particle passed through the subdetector, and then testing whether the corresponding track is reconstructed successfully in the subdetector in question.

The decay used for this is the decay of a J/ψ from a B meson decay into two muons, $b \to (J/\psi \to \mu^+\mu^-)X$. The X is in this case an irrelevant side product of the decay which is not considered further; it may be for example a ϕ or kaon. An example for the production of a J/ψ from a B_s^0 with a ϕ as by-product can be found in Fig. 6.1 together with the Feynman diagram of the decay $J/\psi \to \mu^+\mu^-$.

The presence of a track in the muon chambers automatically implies that the track has successfully passed through all the previous subdetectors, provided that its origin lies in the proton-proton collision at the interaction point or a decay product thereof. This can be ensured by combining the muon track with another muon track coming from the same point of origin. The muon chambers can also be used to provide additional information, both on the particle identification and the particle's trajectory.

To ensure that a muon has crossed a subdetector, its track is reconstructed after and

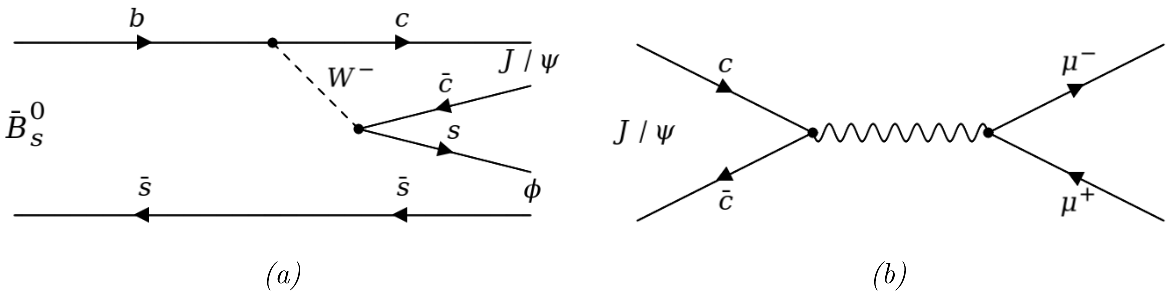


Figure 6.1: Feynman diagram of $\bar{B}_s^0 \to J/\psi\,\phi$ (a) and $J/\psi \to \mu^+\mu^-$ (b).

possibly before the subdetector in question. In addition, the muon is combined with another fully reconstructed muon, called the tag muon, in order to reconstruct a J/ψ. The reconstructed J/ψ mass is used to separate the sample of candidates into *signal* and *background* by fitting the invariant mass distribution.

It is tested whether the partially reconstructed track can be matched to a track fragment in the probed subdetector by checking whether the partially reconstructed muon shares at least 40% of its hits in all subdetectors used for the reconstruction with another muon that is reconstructed as a long track. The fully reconstructed muon is called the long muon. A probe muon that meets this requirement is called successfully *matched*.

The track reconstruction efficiency for the subdetector itself is then calculated by taking the number of events that are successfully matched to a long muon, over all events, both matched and unmatched:

$$\epsilon_{Track\,Reconstruction\,Efficiency} = \frac{N_{signal,matched}}{N_{signal,matched} + N_{signal,unmatched}} \tag{6.1}$$

There are four separate methods used to determine the track reconstruction efficiency:

- The *Downstream method*, shown in Fig. 6.2, uses a downstream muon probe track, consisting of UT, SciFi and muon station hits to create a Downstream probe muon which is matched to a long muon in order to determine the VELO track reconstruction efficiency. An overlap of 40% of the hits found in the SciFi and muon station is required for a successful match. An overlap of 40% of the UT is tested, but not necessarily required, as the UT is not mandatory for the reconstruction of a long track and might skew the VELO track reconstruction efficiency.

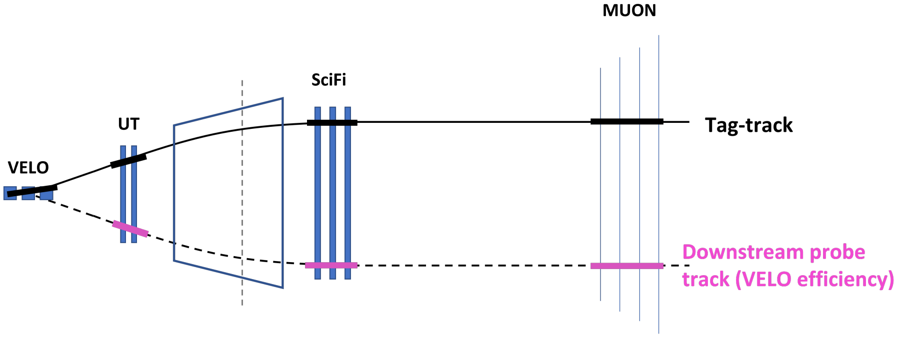


Figure 6.2: The Downstream method, with the Downstream probe track in pink.

- With no UT before 2024, the *SeedMuon method*, shown in Fig. 6.3, uses instead a SeedMuon track, consisting of SciFi and muon station hits, which is matched to a long muon. This method also determines the VELO track reconstruction efficiency. It derives its name from the SciFi tracks, which are also called *Seed* tracks. This method poses a larger challenge than the Downstream method, as only information after the bending field of the magnet is available and the track has to be extrapolated over a long distance to the primary vertex in the VELO to get a sensible momentum estimate. An overlap between the probe track and the long muon track of at least 40% of the hits found in the SciFi and the muon stations is required for a successful match.

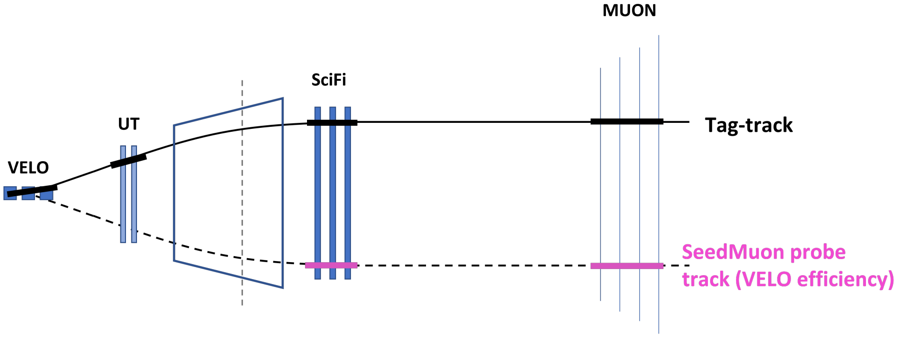


Figure 6.3: The SeedMuon method, with the SeedMuon probe track in pink.

- The *VeloMuon method*, shown in Fig. 6.4, uses a VeloMuon track as a probe by combining hits in the VELO with muon station hits. The resulting VeloMuon track is matched to a long muon to determine the SciFi track reconstruction efficiency. An overlap of 40% of the hits in the VELO and the muon stations is required for a successful match. This method poses a challenge as probe tracks are reconstructed from VELO and muon station hits with a distance of at least 15 metres with no further measurement in between.

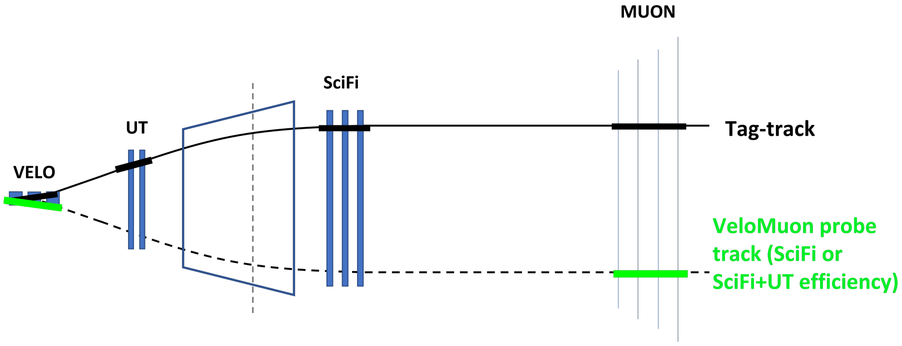


Figure 6.4: The VeloMuon method, with the VeloMuon probe track in green.

- The *MuonUT method*, shown in Fig. 6.5, which can only be exploited when the UT is used in data taking, uses a MuonUT probe track consisting of UT and muon station hits to test the track reconstruction efficiency of the VELO and SciFi combined. An overlap of 40% of the hits in the UT and the muon stations is required for a successful match. As both these efficiencies are already determined by other methods, this method can be used to cross-check and assign corresponding systematic uncertainties.

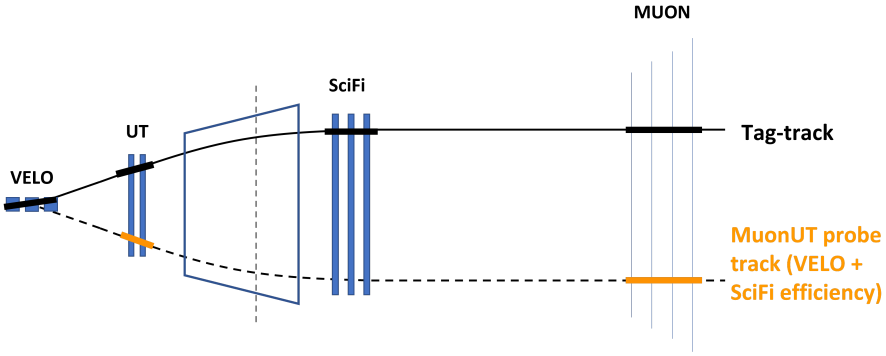


Figure 6.5: The MuonUT method, with the MuonUT probe track in orange.

The long track reconstruction efficiency is determined by multiplying the track reconstruction efficiencies from the Downstream and VeloMuon methods. Alternatively, it is determined using the MuonUT method. CA cross-check is made by comparing the long track reconstruction efficiencies determined by both methods. For data acquisition periods without UT, the long track reconstruction efficiency is determined by multiplying the track reconstruction efficiencies determined by the SeedMuon and VeloMuon methods. Efficiencies can vary over the range of the track momentum and pseudorapidity and are therefore given in bins of these variables.

6.2 Reconstruction of probe muons

The probe muon tracks used to measure the track reconstruction efficiency are reconstructed specifically for this purpose and independently of all other track reconstruction algorithms to avoid any bias or correlation.

Downstream muon tracks, used to measure the VELO track reconstruction efficiency, are built by first reconstructing SciFi tracks, estimating their momentum and then extrapolating to the UT using the first momentum estimate. Between three and eight UT hits are collected around the trajectory. A Kalman filter is applied to remove any outlier hits and to improve the momentum estimate. The track is then extrapolated to the muon station where muon information is added. The minimum number of hits required in the muon chambers is either two or three, depending on the momentum of the track.

SeedMuon tracks are used as an alternative for the measurement of the VELO track reconstruction efficiency when no UT information is available. SciFi and muon tracks are reconstructed separately from hits in the SciFi and the muon stations, respectively. The algorithm then loops over both track containers, calculating the pairwise slope difference and trying to fit the pair with the smallest slope difference for each muon track. The momentum estimate of the SciFi track is inherited for the SeedMuon track.

VeloMuon tracks are primarily used to measure the SciFi track reconstruction efficiency. VELO and muon tracks are reconstructed separately, using hits from the VELO and the muon stations, respectively. The algorithm then loops over both track containers, extrapolating the tracks to a position within the magnet. Track pairs that are too far apart from each other are discarded at this point. For all other pairs, a rough estimate of the momentum is based on the extrapolated trajectory of the combined track. The track is then fitted. If this is successful, the pair with the closest extrapolated distance in the magnet is stored for each muon track.

MuonUT tracks are primarily used to complement the track reconstruction efficiencies measured for the VELO and the SciFi with the other methods in order to cross-check and reduce systematics. They are reconstructed by taking muon tracks reconstructed purely from hits in the muon stations and extrapolating them to the VELO, searching for the best primary vertex. Then, UT hits are sought around the propagated trajectory. If at least two UT hits are found, they are added together. Then an attempt is made to fit the track.

6.3 Candidate selection

Each method excludes subdetectors which would otherwise provide crucial input to the track reconstruction. This results in different levels of background and J/ψ resolution. Therefore, each method presents unique challenges and requires different selection criteria to reduce the background to a manageable level. Selection criteria are applied to the *tag muon*, the *probe muon* and the reconstructed mother, the J/ψ. Some criteria are applied online during the reconstruction, while others are applied offline only for samples which require further background reduction. The online applied selection criteria are summarised in Tab. 6.1, the offline applied selection criteria in Tab. 6.2.
The following variables are relevant in the selection:

- the trackχ^2 per number of degrees of freedom, denoted trackχ^2, a measure of the quality of the track fit. It averages over the distance the hits have to the final fit of the track, normalised by the number of degrees of freedom of the fit;
- the momentum, p, and transverse moment, p_T, which is the momentum in the $x-y-$plane;
- the muon identification variable, ID_μ, which gives the probability that the track if identified as a muon, using information from the muon stations, the RICH detectors and calorimeters;
- the impact parameter, IP, which gives the perpendicular distance between the track and the assigned production vertex of the track;
- the χ^2 of the impact parameter, IPχ^2, a measure of the squared impact parameter significance. It is defined as $(IP/\sigma_{IP})^2$, where σ_{IP} is the uncertainty on IP;
- the ghost probability, GhostProb, which qualifies the probability of a track being a *ghost track*;
- the distance of closest approach, DOCA, which is the minimal distance of the probe and tag muon used to reconstruct the J/ψ;
- the vertexχ^2, a measure of the quality of the fit of the J/ψ production vertex;

methods:	SeedMuon	Downstream	VeloMuon	MuonUT
online				
tag muon				
trackχ^2	< 3 (2)	< 10	< 3	< 5
p [GeV/c]	> 10	> 5	> 7	> 10
p_T [GeV/c]	> 0.5	> 0.7	> 0.5	> 1.3
ID_μ	$>$ -2	$>$ -2	$>$ -1	> 0
IP [mm]	> 0.1	> 0.5	> 0.2	—
IPχ^2	> 38	—	—	—
GhostProb	< 1	< 1	< 1	< 0.3
probe muon				
trackχ^2	< 5 (3)	< 2.8	< 5	< 9999
p [GeV/c]	> 5	> 5	> 5	> 5
p_T [GeV/c]	> 0.5	> 0.5	> 0.5	> 0.5
IPχ^2	> 0	> 0	> 0	< 0
J/ψ				
$DOCA$ [mm]	< 50	< 5	< 0.1	—
IP [mm]	—	< 2	—	> 1
vertexχ^2	—	< 5	< 2	< 2
p_T [GeV/c]	> 0	> 0.5	> 0.5	> 1
m [GeV/c^2]	(2.6,3.6)	(2.8,3.4)	(2.6,3.6)	(2.6,3.6)

Table 6.1: Selection criteria for candidates considered in the four methods. The criteria are applied to the tag muon, the probe muon, or the J/ψ candidate reconstructed from both. A significant change was made to the trackχ^2 selection criterion during 2023 data acquisition, the older value is shown in brackets.

methods:	SeedMuon	Downstream	VeloMuon	MuonUT
offline				
J/ψ				
p_T [GeV/c]	> 2	—	—	—
DOCA [mm]	> 10	—	—	—
τ [ps]	—	> 0.001	—	—
IP [mm]	—	< 0.5	—	< 1
m [GeV/c^2]	(2.7,3.5)	(2.925,3.275)	(2.625,3.575)	(2.625,3.575)

Table 6.2: Optionally applied offline selection criteria for candidates considered in the four methods. The criteria are applied to the J/ψ candidate reconstructed from the probe and tag muon tracks.

- the invariant mass of the reconstructed J/ψ. Signal yields are derived in the fit to the invariant di-muon mass with a constraint to a given mass range. The mass range is given in brackets in Tab. 6.1 and Tab.6.2;
- the decay time, τ, between the production vertex and the decay vertex of the J/ψ.

The tag muon is also required to fulfil a loose muon identification criterion encoded as the boolean variable *ismuon*. Furthermore, it is required to be directly responsible for the triggering of one of two possible trigger lines during the HLT1 trigger stage, called *TrackMVA* and *TrackMuonMVA*.

Both the TrackMVA and the TrackMuonMVA line use a selection tuned by a multivariate analysis tool to select good track candidates and good muon track candidates, respectively. The *TrackMVA* line requires a trackχ^2 value of less than 2.5, and either a transverse momentum of at least 26 GeV/c, the identification of an associated primary vertex and an IPχ^2 of at least 7.4, or a transverse momentum between 2 GeV/c and 26 GeV/c and a tuned IPχ^2 condition depending on the transverse momentum. Additionally, it requires the primary vertex to be inside the VELO and a ghost probability value less than 0.5.

The *TrackMuonMVA* line requires a trackχ^2 value below 100, and either a transverse momentum above 26 GeV/c and an IPχ^2 of at least 7.4, or a transverse momentum between 2 GeV/c and 26 GeV/c and a tuned IPχ^2 selection depending on the transverse momentum. The track must originate in the VELO. A trackχ^2 value of less than 1.8 of a fit to the track segment in the muon chambers is required, thus ensuring that the track is a muon.

However, since these criteria are applied during the HLT1 reconstruction and the reconstruction is run again from scratch during the HLT2 trigger stage, the final parameters of the tag track might lie outside of these requirements.

In addition to the HLT1 trigger requirement on the tag muon, the charge of the probe muon must be opposite to that of the tag muon. The charge of the long muon matched to the probe muon must be the same as that of the probe muon.

No further requirements are placed on the long muon, as the required overlap is sufficient to ensure that a matching track has very similar properties to the probe muon.

6.4 Determination of signal yields

The signal yield of the matched and unmatched candidates, as used in Eq. 6.1, is determined by splitting the sample into two subsamples, called the *matched* and *unmatched* sample, according to whether the probe muon is matched or not. The invariant mass of the J/ψ mother is then fitted simultaneously in both subsamples.

The invariant mass of the J/ψ particle is $3096.900 \pm 0.006\,\mathrm{MeV}/c^2$ and ideally follows a Gaussian distribution with a width of $92.6 \pm 1.7\,\mathrm{keV}/c^2$ [3]. However, while the LHCb experiment has a very good momentum resolution, it is not in the necessary order of magnitude to resolve the natural line width. Thus, the measured shape is not determined by theory. In addition, random combinations of tracks that fulfil the criteria contribute to a *combinatorial background* in the data sample.

As the simulated samples are based purely on generated signal decays, they may differ in shape from data distributions and have a much smaller background.

The background is described by either an exponential function

$$f_{exp}(m; C) = N_{bkg} \times e^{C \times m} \tag{6.2}$$

or a *Chebychev polynomial*, f_{cheby}, with three coefficients, C_A, C_B, C_B:

$$f_{cheby}(m; C_A, C_B, C_C) = N_{bkg} \times (C_A \times m + C_B \times m^2 + C_C \times m^3) \tag{6.3}$$

The background functions are independent between the matched and unmatched subsamples and do not share any parameters. The exponential background has one free floating parameter per subsample, C, in addition to the background yield, N_{bkg}. The parameter C is allowed to vary between -1 and $+1$. The floating parameters of the *Chebychev polynomial* describing the background, C_A, C_B, C_B, are allowed to vary between -1 and $+1$ each except for C_A which has a more restricted lower limit of -0.3. These restrictions are tuned to guarantee stable convergence of the mass fit.

The signal shape is described with either an asymmetrical double-sided Crystal Ball (*DoubleCB*) function

$$f(m;m_0,\sigma_L,\alpha_L,n_L,\sigma_R,\alpha_R,n_R) = N_{sig} \times \begin{cases} A_L \cdot (B_L - \frac{m-m_0}{\sigma_L})^{-n_L}, \; for \; \frac{m-m_0}{\sigma_L} < -\alpha_L \\ \exp\left(-\frac{1}{2} \times \left[\frac{m-m_0}{\sigma_L}\right]^2\right), \; for \; \frac{m-m_0}{\sigma_L} \leq 0 \\ \exp\left(-\frac{1}{2} \times \left[\frac{m-m_0}{\sigma_R}\right]^2\right), \; for \; \frac{m-m_0}{\sigma_R} \leq \alpha_R \\ A_R \cdot (B_R + \frac{m-m_0}{\sigma_R})^{-n_R}, \; otherwise, \end{cases} \tag{6.4}$$

where

$$\begin{aligned} A_i &= \left(\frac{n_i}{|\alpha_i|}\right)^{n_i} \cdot \exp\left(-\frac{|\alpha_i|^2}{2}\right) \\ B_i &= \frac{n_i}{|\alpha_i|} - |\alpha_i| \end{aligned}$$

or a *Johnson* function [34]

$$f(m;m_0,\lambda,\gamma,\delta) = N_{sig} \times \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+(\frac{m-m_0}{\lambda})^2}} \exp\left[-\frac{1}{2}\left(\gamma + \delta \sinh^{-1}\left(\frac{m-m_0}{\lambda}\right)\right)^2\right]. \tag{6.5}$$

Examples for the two signal and the two background fit shapes fitted simultaneously to a matched and an unmatched subsample are shown in Fig. 6.6.

The signal functions of the matched and unmatched subsamples share the mass mean, m_0. When the double-sided Crystal Ball function is used, the five other parameters, σ, α_L, n_L, α_R, n_R, are also shared. The values for n_R, n_L, α_L and α_R are fixed according to the requirements of the specific fit. In case of the Johnson function, the parameters γ and δ are always shared between the matched and unmatched subsample, but the parameter λ is allowed to differ for simulated data. The limits set for the parameters depend on each individual fit.

The signal and background functions are fitted simultaneously to the two subsamples using RooFit. The composite model of the selected background and signal functions for both subsamples is fitted to the distribution of candidates using generally an unbinned maximum likelihood parameter estimation technique by the Root implementation of MINUIT [35]. When fitting a reWeighted distribution, as described in Sec. 5.5, the fit is performed as a binned fit by necessity of the weighting procedure.

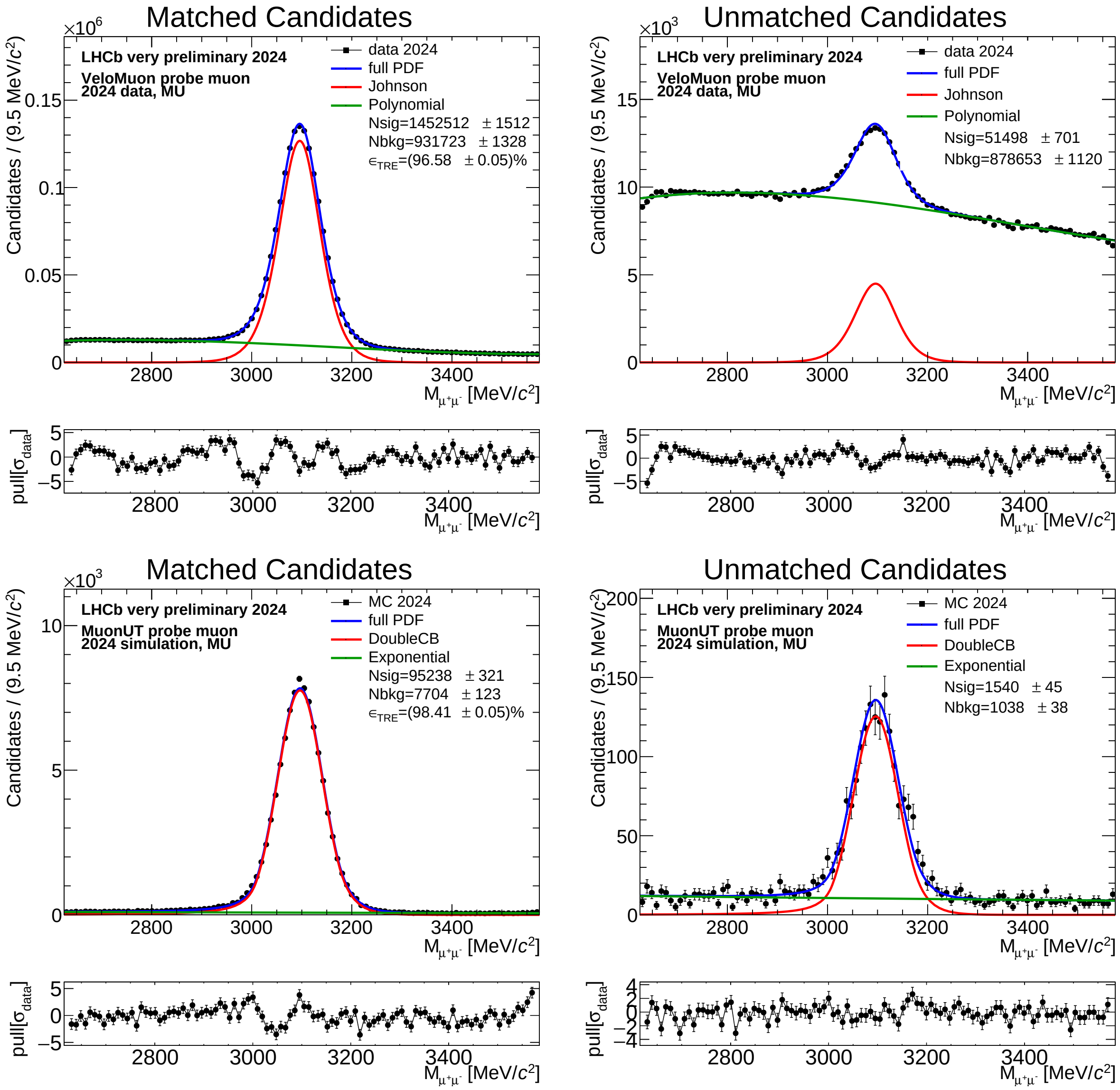


Figure 6.6: Fit to the invariant mass of the two muons reconstructed as a J/ψ of matched (left) and unmatched (right) candidates from recorded VeloMuon method candidates (top) fitted with a Johnson function for the signal (red) and a Chebychev polynomial for the background distribution (green), and of simulated MuonUT method candidates (bottom) fitted with a DoubleCB function for the signal (red) and an Exponential for the background distribution (green). The total fit to the sample is both times shown in blue.

Instead of determining the signal and background yields from the fit and using them to evaluate the track reconstruction efficiency, ϵ, it is incorporated directly as a fit parameter by adapting the signal yields:

$$N_{sig}^{matched} = N_{sig}^{unmatched} \times \frac{\epsilon}{1-\epsilon}$$

This way, the efficiency is immediately given as output of the fitting tool. This includes the propagated statistical uncertainty, σ_{stat}, estimated by the fitting tool using

$$\sigma_{stat}^2(x) = F_a(x) \times Cov(a, a') \times F_{a'}^T(x), \tag{6.6}$$

, where

- $F_a(x) = \frac{f(x,a+da)-f(x,a-da)}{2}$
- $f(x)$ is the corresponding function used to determine the fit parameter in question, in case of the track reconstruction efficiency $\frac{N_{signal,matched}}{N_{signal,matched}+N_{signal,unmatched}}$ as seen in Eq. 6.1
- a are the floating fit parameters used for the calculation of the parameter in question, in case of the track reconstruction efficiency $N_{signal,matched}$ and $N_{signal,not\,matched}$
- da are the 1σ-uncertainties of the relevant fit parameters determined by MINUIT
- $Cov(a, a')$ is the multi-dimensional covariance matrix of the relevant fit parameters determined by MINUIT

The track reconstruction efficiency determined by the fits can then be obtained in dependency of variables of choice, such as momentum, p, and the pseudorapidity, η, by dividing the data samples into subsamples and fitting every subsample separately.

Chapter 7

Challenges and achievements of measuring the track reconstruction efficiency in 2023

This chapter gives an overview of the process of determining and understanding the track reconstruction efficiency in 2023. Discrepancies between recorded and simulated data are investigated, and various potential influences are tested. Lessons for the measurement of track reconstruction efficiencies in 2024 are drawn from the results.

7.1 The detector in 2023

The first data of LHC Run 3 was taken in 2022. However, with an entirely new detector, LHCb was mostly concentrating on aligning the subdetectors and calibrating the detector instead of producing huge quantities of usable data. While hopes were set on the data taking of 2023, the VELO incident [11] during the night from the 9th to 10th of January 2023 left the VELO in a damaged state, meaning it could not be fully closed for the entire data taking period of 2023. The result is a loss of almost half of the usual detector acceptance. Additionally, the UT was installed in the detector cavern, but not used in the data acquisition in 2023, thus providing additional interaction material, but no information usable for track reconstruction.

For the determination of the track reconstruction efficiencies, this affected mostly the VELO. Fig. 7.1 shows the multiplicity of SeedMuon probe and tag muon tracks in η and ϕ in the detector. The SeedMuon probe, which is constructed of SciFi and Muon hits

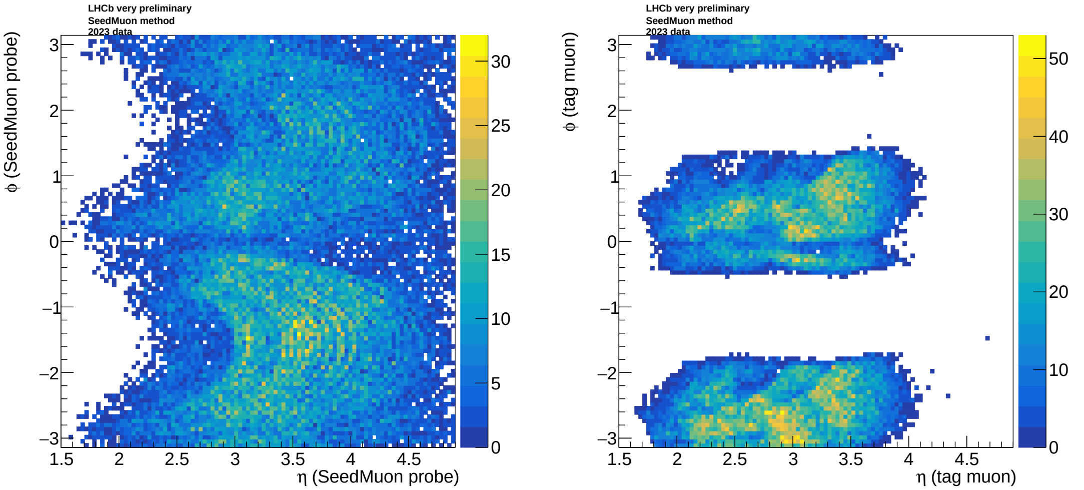


Figure 7.1: Distributions of ϕ and eta of SeedMuon probe (left) and tag (right) tracks in 2023 data.

and does not rely on the VELO at all, covers the whole phase space of the detector. The long tag track however, relying on VELO information, is restricted to the limited VELO acceptance.

The potentially matching long muon tracks suffer from the same issues as the tag muon tracks and have roughly the same distribution. Therefore, the VELO track reconstruction efficiency is significantly below 100%, because in about half the cases where there are track segments in the SciFi and the muon stations, the VELO does not cover the relevant phase space region and therefore no matching long tracks can be found.

This is an effect of the method and does not describe the actual track reconstruction efficiency of the VELO.

The missing UT means that only the SeedMuon and VeloMuon methods can be used to measure the VELO and SciFi track reconstruction efficiency respectively.

The ongoing alignment of the SciFi with respect to the rest of the detector, most prominently the VELO, had a significant impact on the determination of the SciFi track reconstruction efficiency in 2023 data. Furthermore, the process of fully understanding the SciFi, including the single hit efficiencies of the individual SciFi silicon photomultipliers, impacted the track reconstruction efficiency determination. Discrepancies between the efficiency in simulated and recorded data are anticipated. Thus the determination of the track reconstruction efficiencies turned out to be an extremely valuable tool to investigate differences between simulation and data taking conditions. This chapter attempts to explain any remaining discrepancies to the limit of capabilities.

7.2 Track reconstruction efficiency of the vertex locator in 2023 data

The integrated VELO track reconstruction efficiency measured with recorded data in 2023 is:

$$\epsilon_{VELO}^{data} = (47.12 \pm 0.62)\% \tag{7.1}$$

This can be compared to the efficiency using simulated data:

$$\epsilon_{VELO}^{MC} = (52.04 \pm 0.14)\% \tag{7.2}$$

Assigned uncertainties are purely statistical, as no evaluation of systematic uncertainties on 2023 data has been performed. As expected, the efficiency is significantly below 100% due to the open VELO, which only covers parts of the phase space of the SciFi and the muon stations.

The discrepancy between simulated and recorded data is more than four percent. Three possible sources of discrepancies are identified. These are the online applied cut on the trackχ^2, the discrepancy in angular distributions, and the reconstruction algorithm itself.

7.2.1 Limiting the trackχ^2 variable

The first source of discrepancy is the upper limit on the fit quality of the track. This selection criterion was originally set to a maximum trackχ^2 value of 3 for the SeedMuon probe, and to a maximum trackχ^2 of 2 for the tag muon track in the SeedMuon method. Unfortunately, the fit quality of the tracks is much better in simulated than in recorded data, such that the selection only removed insignificant amounts of simulated candidates, but a significant amount of recorded candidates.

This can be seen in Fig. 7.2. While the fit quality is worse in recorded than in simulated data for both tag and probe tracks, it is significantly worse for probe tracks.

This criterion is applied in the selection of the *SeedMuon* candidates for the VELO track reconstruction efficiency determination. The trackχ^2 limit used in analyses is usually looser and thus tracks used in analyses are less affected.

Only data passing the selection criteria defined in the HLT2 reconstruction are stored after the reconstruction. Therefore, candidates that would have survived a looser requirement of the trackχ^2-limit could only be recovered if they had already been selected by similar HLT2 selection lines. The effect of restoring those candidates is shown in Fig. 7.3.

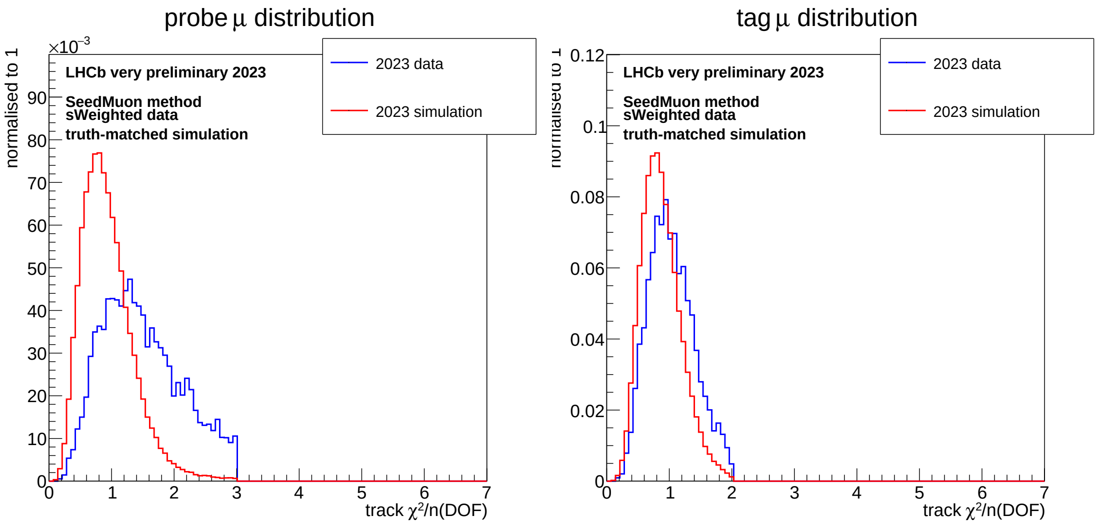


Figure 7.2: trackχ^2 of the SeedMuon probe (left) and tag (right) tracks in recorded sWeighted (blue) and simulated truth-matched (red) data in 2023.

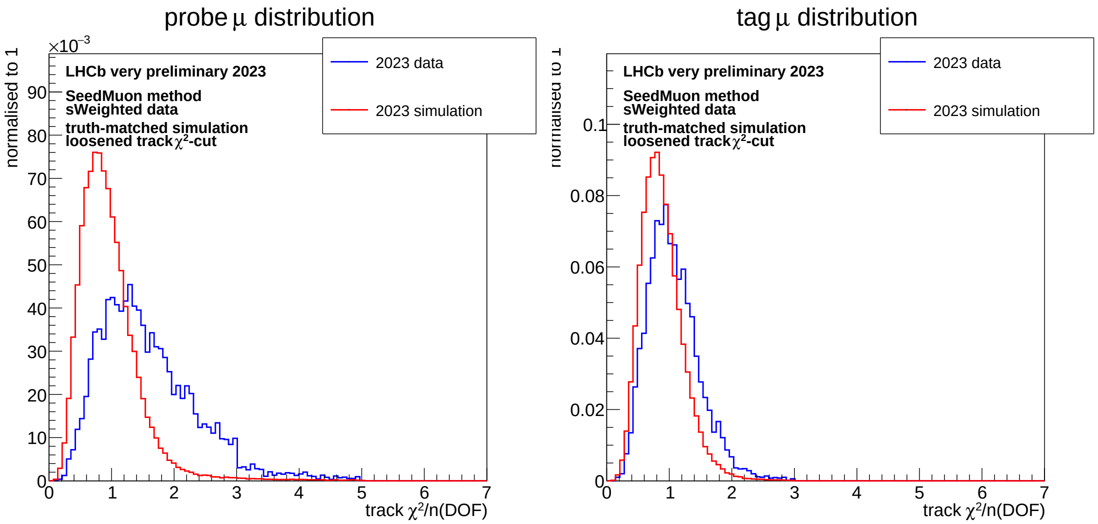


Figure 7.3: Trackχ^2 of the SeedMuon probe (left) and tag (right) tracks in recorded sWeighted (blue) and simulated truth-matched (red) data in 2023 with the trackχ^2-cut loosened.

The recovery of candidates leads to an efficiency in recorded data of

$$\epsilon_{VELO}^{data,looser\,cut} = (50.43 \pm 0.23)\%.$$

The track reconstruction efficiency increases by more than two percent points, reducing the discrepancy. However, this is most likely dominated by the bias introduced by only recovering candidates that have triggered other track reconstruction efficiency lines. This is in most cases the complementary *VeloMuon* line, which relies on the missing VELO segment being reconstructed successfully. The track reconstruction efficiency of the recovered candidates is thus expected to be close to 100%.

By *reWeighting* the simulated sample with looser trackχ^2-cut in bins of probe and tag trackχ^2, the efficiency in simulated data is reduced to

$$\epsilon_{VELO}^{MC,track\chi^2-reweighted} = (48.69 \pm 0.20)\%,$$

which is now significantly below the efficiency in recorded data with a looser cut. This confirms that the improvement due to the recovery of candidates is mostly influenced by the introduced bias and does not reflect the actual effect of loosening the trackχ^2 limit. However, the trackχ^2-limit is still potentially a big factor in the discrepancy between recorded and simulated data.

7.2.2 Angular distributions in data and simulation

The track reconstruction efficiency is often dependent on the position of the track in the detector. This can be caused both by the alignment of the detector, which is better in some regions, and by the different momentum distribution of tracks in different detector regions. The track reconstruction efficiency for high momentum tracks is generally higher. Unfortunately, due to the ongoing understanding of the detector geometry and reconstruction pipeline in 2023, the distribution in η and ϕ of the tracks and the J/ψ candidates is not so well reproduced in simulation. In Fig. 7.4, the distributions of both muons and of the J/ψ mother in η and ϕ in recorded and simulated data are shown. The discrepancy between the distributions is significant. This might have an impact on the track reconstruction efficiency. To test the effect, the simulated sample is *reWeighted*. To avoid a bias in the pseudorapidity of the probe muon, which is a parameter of interest when measuring the track reconstruction efficiency, the *reWeighting* is done in ϕ and η of the tag muon and the J/ψ. The effects of *reWeighting* can also be seen in Fig. 7.4.

By *reWeighting* the simulated data in bins of probe and tag muon trackχ^2, the resulting efficiency is reduced to

$$\epsilon_{VELO}^{MC,\eta,\phi-reweighted} = (47.88 \pm 0.16)\%,$$

which is much closer to the efficiency seen in data in Eq. 7.1. Unfortunately, while the causes for the significant discrepancies between recorded and simulated data are not removed, *reWeighting* can only give a general idea of its impact.

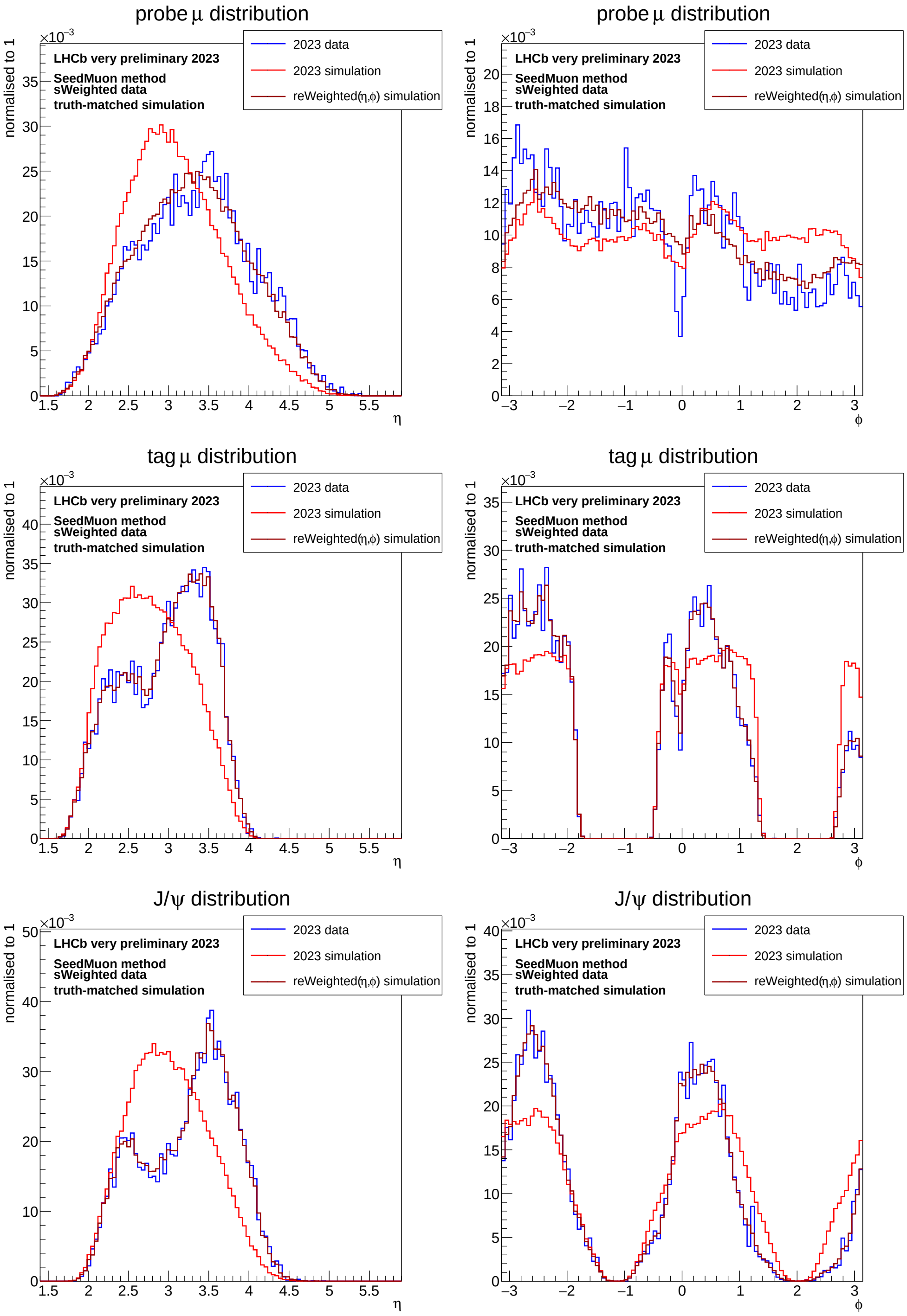


Figure 7.4: Distribution of η (left) and ϕ (right) of the SeedMuon probe (top), the tag muon (middle) as well as the J/ψ mother (bottom) for recorded sWeighted data (blue) and truth-matched simulated data (red), as well as truth-matched simulated data reWeighted to the recorded data in η and ϕ of the tag and J/ψ (dark red).

7.2.3 Improvements in reconstruction strategies

After the data-taking of 2023, several improvements to the reconstruction were made. These include some corrections to the construction of SciFi tracks, improvements to the mechanism to remove clone tracks, debugging tracking losses and an update to the muon alignment. Unfortunately, these improvements could not be tested separately for their effect on the track reconstruction efficiency as they were all included in the same software release used for the reconstruction. However, the effects of the new reconstruction can be clearly seen in the efficiency of smaller test samples produced with the loosened trackχ^2-cut:

$$\epsilon_{VELO}^{data,\, new\, reconstruction} = (52.49 \pm 0.53)\%$$

$$\epsilon_{VELO}^{MC,\, new\, reconstruction} = (52.92 \pm 0.21)\%$$

As can be seen, the track reconstruction efficiencies using the new reconstruction algorithm are compatible within one σ between recorded and simulated data. However, due to again only recovering candidates with an inherent bias, the track reconstruction efficiency in recorded data is biased. Therefore, no clear conclusion can be drawn from the agreement. To get an estimate of the impact of the angular variable distributions again, the simulated sample using the new reconstruction is *reWeighted* in η and ϕ of the tag track and the reconstructed J/ψ mother with regard to the data sample using the new reconstruction. The efficiency of the *reWeighted* sample is:

$$\epsilon_{VELO}^{MC,\, new\, reconstruction,\, \eta,\phi-reweighted} = (50.14 \pm 0.23)\%$$

This is again lower than the not *reWeighted* simulated data sample, and therefore now lower than the corresponding recorded data sample.

7.2.4 Compatibility of the VELO track reconstruction efficiency in data and simulation of 2023

Due to several problems with the alignment and reconstruction, the VELO track reconstruction efficiency of recorded data samples is significantly lower than that of the equivalent simulated data.

Tab. 7.1 summarises the above discussed factors contributing to the discrepancy. As can be seen, the track reconstruction efficiency in recorded data can be significantly improved both by a loosening of the trackχ^2-cut on the reconstructed muon tracks even when only a limited amount of events can be recovered, and an improvement in the reconstruction algorithm. However, this is likely strongly influenced by a bias introduced when recovering candidates from other HLT2 selection lines. Thus, no final track reconstruction efficiency for recorded data can be presented.

The efficiency in simulated data can be lowered to approach the efficiency in data by *reWeighting* in the trackχ^2-variable or in η and ϕ. Several of these scenarios are compared in Tab. 7.1. The actual VELO track reconstruction efficiency for 2023 can never be fully recovered, as data acquisition of 2023 stopped before all improvements could be incorporated in the online processes. It suffices to conclude that under the proper incorporation of all factors both in the reconstruction sequences and data simulation, an accord between both could likely be reached. Further studies are performed on 2024 data.

	data	simulation
original samples	$(47.12 \pm 0.62)\%$	$(52.04 \pm 0.14)\%$
loosened trackχ^2-cut	$(50.43 \pm 0.23)\%$	$(51.92 \pm 0.17)\%$
reWeighted in trackχ^2 w.r.t data	—	$(48.69 \pm 0.20)\%$
reWeighted in η, ϕ w.r.t data	—	$(47.88 \pm 0.16\%$
new reconstruction, loosened trackχ^2-cut	$(52.49 \pm 0.53)\%$	$(52.92 \pm 0.21)\%$
new reconstruction, loosened trackχ^2-cut, *reWeighted* in η, ϕ w.r.t to new reco data	—	$(50.14 \pm 0.23)\%$

Table 7.1: VELO track reconstruction efficiencies for recorded and simulated data under the influence of several factors for 2023.

7.3 Track reconstruction efficiency of the SciFi tracker in 2023 data

The integrated SciFi track reconstruction efficiency in data recorded in 2023 is:

$$\epsilon_{SciFi}^{data} = (90.83 \pm 0.38)\% \tag{7.3}$$

This can be compared to the track reconstruction efficiency of simulated data:

$$\epsilon_{SciFi}^{MC} = (97.15 \pm 0.09)\% \tag{7.4}$$

The discrepancy between the two methods is with almost seven percent points significant, and even larger than the discrepancy in the VELO track reconstruction efficiency. Sources for this are differences in the reconstruction algorithm, discrepancies in angular variable distributions and the detection efficiencies of the single silicon photomultipliers responsible for measuring signals.

7.3.1 Angular distributions in simulated and recorded data

Like the VELO track reconstruction efficiency, the SciFi track reconstruction efficiency is dependent on the angular variables and therefore vulnerable to deviations in those distributions between recorded and simulated data. As can be seen in Fig. 7.5, the discrepancies are significant. To estimate the size of effect of this, the simulated sample is *reWeighted* in η and ϕ of the tag muon and the reconstructed J/ψ mother. The *reWeighted* distributions are also presented in Fig. 7.5

The track reconstruction efficiency in simulated data, *reWeighted* in η and ϕ of tag and reconstructed J/ψ mother, is:

$$\epsilon_{SciFi}^{MC,\eta,\phi-reweighted} = (96.89 \pm 0.09)\%$$

The *reWeighting* lowers the efficiency by almost three standard deviations. This is not as significant as the effect *reWeighting* in angular variables has on the VELO track reconstruction efficiency. This is an effect of the limited coverage of the VELO, as *VeloMuon* tracks are subject to this, while the pseudorapidity of *SeedMuon* tracks is shifted in a wider range.

7.3.2 The hit inefficiency of SciFi modules

An important property of the SciFi is the single hit efficiency of its silicon photomultipliers (SiPMs), describing the efficiency of the specific SiPM to create an electric signal when a light signal caused by a passing charged particle is guided to it. In simulated 2023 data, an average single hit efficiency of 98% is assumed for all active SiPM channels, while in reality several channels had a much lower efficiency, sometimes only up to 65%.

To see the effects of the additional SciFi hit inefficiencies, the actual single hit efficiencies are measured by the SciFi group in recorded data and then applied to a simulated sample. Due to technical difficulties, the corrections could not be applied to the original sample. Thus to avoid mixing up effects, the SciFi track reconstruction efficiency is determined on

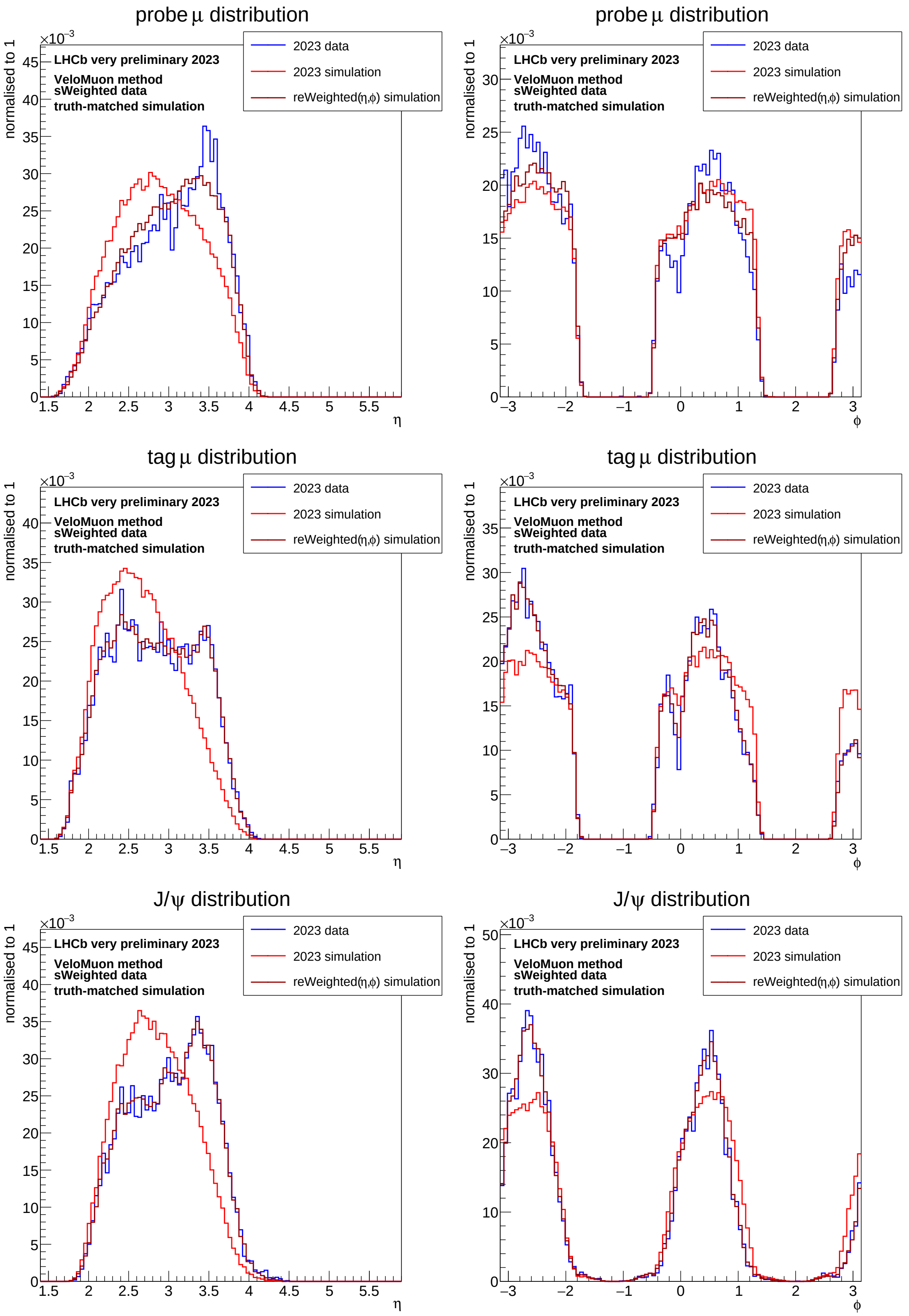


Figure 7.5: η (left) and ϕ (right) distribution of the VeloMuon probe track (top), the tag track (middle) and the J/ψ mother (bottom) for sWeighted data (blue) and truth-matched simulation (red), as well as truth-matched simulation reWeighted to the data sample in η and ϕ of the tag and J/ψ track (dark red).

another sample with and without the correction for the SciFi hit inefficiencies:

$$\epsilon_{SciFi}^{MC,SciFi-hit-corrected} = (91.13 \pm 0.17)\%$$

$$\epsilon_{SciFi}^{MC,SciFi-hit-not-corrected} = (97.34 \pm 0.14)\%$$

The effect of the SciFi hit efficiency corrections is significant, lowering the track reconstruction efficiency in simulated data by more than six percent points and closing the gap to the efficiency measured in recorded data as given in Eq. 7.3.

7.3.3 Improvements in reconstruction

The above measurements of the effect of applying the SciFi hit inefficiencies is performed on a sample reconstructed with the newer reconstruction sequence as mentioned before in Sec. 7.2.3. A recorded data sample, using the same reconstruction sequence, yields a track reconstruction efficiency of:

$$\epsilon_{SciFi}^{data,new\,reconstruction} = (91.80 \pm 0.37)\%,$$

which is now slightly above the efficiency measured in simulated data corrected for the SciFi hit efficiency, but still very close. Again the new reconstruction algorithm improves the track reconstruction efficiency, but as was the case before, candidates recovered using the new reconstruction sequence are inherently biased as they must have triggered other tracking lines, likely the *SeedMuon line*. Thus, no clear conclusion can be drawn from this result.

7.3.4 Misalignment of the simulated sample

Although continuous improvements were made to the alignment of the detector in 2023, the final alignment was not perfect, and there was no definite knowledge of the precision of the alignment parameters.

The best estimate of the uncertainty on the alignment is at around 80 µm in the x-direction and 500 µm each in the y- and the z-direction. To get an idea of the impact of the adjusted resolution on the track reconstruction efficiency, a simulated sample is prepared and the efficiency is measured. Then the same sample is prepared again, this time with a Gaussian smearing of 80 µm in the x-direction and 500 µm in the y- and the z-direction of the position of the SciFi modules. The efficiency is measured again. The resulting efficiencies are:

$$\epsilon_{SciFi}^{MC,new\,reco} = (97.53 \pm 0.24)\%$$

$$\epsilon_{SciFi}^{MC,new\,reco,misaligned} = (97.14 \pm 0.27)\%$$

As both track reconstruction efficiency values are obtained from the same sample, they are correlated. It ca be concluded that the track reconstruction efficiency is slightly reduced due to the misalignment, but the actual size of the effect is within one standard deviation.

7.3.5 Compatibility of the SciFi track reconstruction efficiency of simulated and recoreded data in 2023

Given that the large discrepancy between recorded and simulated data is reduced a little by *reWeighting* in η and ϕ of the tag muon track and the reconstructed J/ψ mother, and reduced a lot by applying the corrections for the SciFi hit inefficiencies, it is a logical conclusion that the *reWeighting* be done to the sample already corrected for the SciFi hit inefficiencies. Since this sample is only available using the newer reconstruction algorithm, all three possible improvements are applied at the same time. The resulting track reconstruction efficiency in simulated data is then:

$$\epsilon_{SciFi}^{MC,new\,reco,SciFi-hit-corrected} = (91.40 \pm 0.18)\%$$

Tab. 7.2 shows the values of the SciFi track reconstruction efficiency for the different corrections. As can be seen, the final recorded data and most improved simulated data values are compatible within one sigma.

	data	simulation
original samples	$(90.83 \pm 0.38)\%$	$(97.15 \pm 0.09)\%$
reWeighted in η, ϕ w.r.t to original data	—	$(96.89 \pm 0.09)\%$
new reconstruction	$(91.80 \pm 0.37)\%$	$(97.34 \pm 0.14)\%$
new reco, corrected for SciFi hit eff	–	$(91.13 \pm 0.17)\%$
new reco, corrected for SciFi hit eff, *reWeighted* in η, ϕ w.r.t. new reco	—	$(91.40 \pm 0.18)\%$

Table 7.2: SciFi track reconstruction efficiency in 2023 for the original data sample and a sample with improved reconstruction algorithm, the corresponding simulated data samples, and both simulated data samples reWeighted in η and ϕ of the tag muon and reconstructed J/ψ with regard to the corresponding recorded data samples.

7.4 Application of the track reconstruction efficiencies in 2023 data

The VELO and SciFi track reconstruction efficiencies measured in 2023, as shown in Tab. 7.1 and Tab. 7.2, are not compatible between simulated and recorded data, and while the influence of several parameters on both efficiencies can be clearly seen, no final conclusion can be drawn on the actual track reconstruction efficiency.

The VELO track reconstruction efficiency is compatible without any further corrections using the newer reconstruction sequence and a looser trackχ^2-cut, but given the clear dependence on the angular variables, a large corresponding systematic uncertainty would need to be assigned. Additionally, this sample uses biased recorded data. The SciFi track reconstruction efficiency is compatible within one sigma when the simulation is corrected for the SciFi hit inefficiencies, but this can only be tested with the newer reconstruction sequence due to technical circumstances. This again introduces a bias to the recorded data. The SciFi track reconstruction efficiency also shows a dependence on the angular variables. A corresponding systematic uncertainty must be assigned, although not as large as for the VELO track reconstruction efficiency.

No further measurements of the VELO and SciFi track reconstruction efficiencies were made in 2023 as no physics analyses requiring these efficiencies were carried out. However, the various measurements and tests with these track reconstruction efficiencies nevertheless delivered valuable information regarding the reconstruction, the detector alignment and the simulation, which was used both to continuously improve the detector and software during the year 2023 and also to ensure a better start for the data taking and simulation for the subsequent measurements in 2024.

Chapter 8

Track Reconstruction Efficiencies in 2024

Based on the insights gained in 2023, track reconstruction efficiencies are measured in 2024 with an improved detector state. The long track reconstruction efficiencies are determined using two methods, which are the combined method of the Downstream and the VeloMuon method, and the MuonUT method. Efficiencies are determined and compared in bins of the pseudorapidity, η, and the momentum, p. Main sources for systematic uncertainties of the track reconstruction efficiencies are discussed.

8.1 The state of the detector in 2024 and used data sets

In the year 2024, data was taken with the LHCb detector from March 5th to November 23rd. Proton-proton collisions were recorded in LHCb between March 21st and November 4th. The damaged VELO, as mentioned in chapter 7, was exchanged during the end-of-year technical shutdown 2023. The VELO was thus fully closed during data taking in 2024. The UT was completely installed, but still had to be commissioned and aligned. Thus the UT is only used in the HLT2 reconstruction sequence since the end of June, and in the HLT1 reconstruction sequence since August 25th.

The detector alignment and the reconstruction algorithms were constantly updated over the course of data taking, with some major improvements until mid July. One VELO module was excluded in July and four SciFi frontend boards had to be excluded on May 17th, July 8th, August 17th and August 27th from data taking. All of these had effects on the track reconstruction. Therefore, the data is separated in several blocks that are considered separately. Data taken between August 6th and August 25th is used in some selected analyses. Most analyses use the data recorded from Augst 28th on.

The track reconstruction efficiency is therefore determined in two data taking periods. *Data set 1* uses data taken between August 6th and August 16th. This data is taken with *Magnet Up* (MU) polarity. The UT is included in HLT2, but not in HLT1. Two SciFi frontend boards are excluded, lowering the SciFi track reconstruction efficiency. *Data set 2* uses data recorded from August 28th to September 11th with MU polarity. This is taken with improved detector alignment in comparison to *data set 1* and the UT is included both in the HLT1 and HLT2 reconstruction sequences. Four SciFi frontend boards are excluded during this period. An overview over these two data sets can be found in Tab. 8.1.

name	polarity	HLT1	HLT2	excl. SciFi FEs	time of data-taking
data set 1	MU	no UT	with UT	2	06.08.−16.08.24
data set 2	MU	with UT	with UT	4	28.08.−11.09.24

Table 8.1: Data sets used for the determination of track reconstruction efficiencies in 2024.

Unfortunately, there were two independent bugs in the HLT2 reconstruction algorithm of *MuonUT probe tracks*.

The *propagation bug* fails to properly propagate the *MuonUT probe track* through the detector to the position of its origin vertex in the VELO. Instead, the initial momentum estimate is left unchanged with an incorrectly assigned charge sign. This leads to the loss of most candidates.

The propagation bug was corrected on August 25th and does not affect *data set 2*. In *data set 1*, the *propagation bug* was corrected in the reprocessing after data taking. However, it is not possible to recover candidates that did not pass the requirements of any HLT2 category at the time of data taking. As only candidates also triggering other lines can be recovered, a bias is introduced in the MuonUT method in *data set 1*. A systematic uncertainty is assigned due to the bias resulting from the *propagation bug*.

Additionally, the *pseudorapidity bug* affects the region of the pseudorapidity from 3.1 to 4.1. *MuonUT probe tracks* coming from this region are propagated incorrectly and often lost during reconstruction. This significantly reduces the number of hits in that region and therefore the number of tracks reconstructed crossing. The *pseudorapidity bug* also cannot be compensated in data and its effect is only estimated using simulated data.

Simulation is produced separately for the different data sets, taking into account the state of the HLT1 reconstruction sequence with or without UT.

The improved simulation excluding all SciFi frontend boards which were not used in data was not ready by the time of writing this thesis. Effects of this will be probed by comparing samples with different numbers of excluded frontend boards.

8.2 Measured track reconstruction efficiency values

8.2.1 Overall efficiencies in comparison between recorded and simulated data

The VELO track reconstruction efficiencies are determined by fitting a sample of candidates where a reconstructed *Downstream muon probe track* is attempted to be matched to a *long muon track*. The SciFi track reconstruction efficiencies are determined by fitting a sample of candidates where a reconstructed *VeloMuon probe track* is attempted to be matched to a *long muon track*. In both cases, the signal yield of both subsamples with and without successful match is determined by fitting a *Johnson distribution*. The background distribution is described with a simple *exponential* in the case of simulated samples and a *Chebychev polynomial* of third order in the case of data samples.

The integrated efficiencies of the VELO track reconstruction are shown in comparison between recorded and simulated data in Tab. 8.2 and the integrated efficiencies of the SciFi track reconstruction in Tab. 8.3. As can be seen, the efficiencies are similar, but do not agree within one standard deviation. A discrepancy between recorded and simulated data is expected, as simulation will never describe the data perfectly. Correction factors for this case need to be introduced. However, at this point, systematic uncertainties are not yet taken into account.

data set	efficiency in recorded data	efficiency in simulated data
data set 1	$(99.44 \pm 0.00)\%$	$(99.59 \pm 0.01)\%$
data set 2	$(99.48 \pm 0.00)\%$	$(99.60 \pm 0.01)\%$

Table 8.2: Integrated VELO efficiencies estimated using the Downstream method in comparison between recorded and simulated data for three different data sets.

The efficiencies are also dependent on the kinematics of the track. Therefore, the agreement in bins of momentum, p, and pseudorapidity, η, is studied and can be found in Fig. 8.1 and Fig. 8.2. As the track reconstruction efficiency agreement in either variable might still be affected by disagreements in the distribution of the other variable, A two-dimensional study is also done and shown in Fig. 8.3 and Fig. 8.4. However, due to statistical constraints, only two pseudorapidity bins are chosen in this case, limiting the

data set	recorded data efficiency	simulated data efficiency
data set 1	$(96.96 \pm 0.01)\%$	$(97.48 \pm 0.02)\%$
data set 2	$(96.57 \pm 0.02)\%$	$(97.49 \pm 0.03)\%$

Table 8.3: Integrated SciFi efficiencies estimated using the VeloMuon method in comparison between recorded and simulated data for three different data sets.

informative value. Other binning schemes are possible and will be used in the future.

The probe muon track distributions in the kinematic variables in comparison between *sWeighted* recorded data and *truth-matched* simulated data can be found in Appendix A in Fig. A.1 and Fig. A.2. As can be seen there, there are some discrepancies between the distribution in recorded and simulated data. The systematic uncertainty due to this is therefore measured by *reWeighting*.

In general, there is good agreement between recorded and simulated data both for the VELO and SciFi efficiencies, though there are remaining discrepancies, which have to be

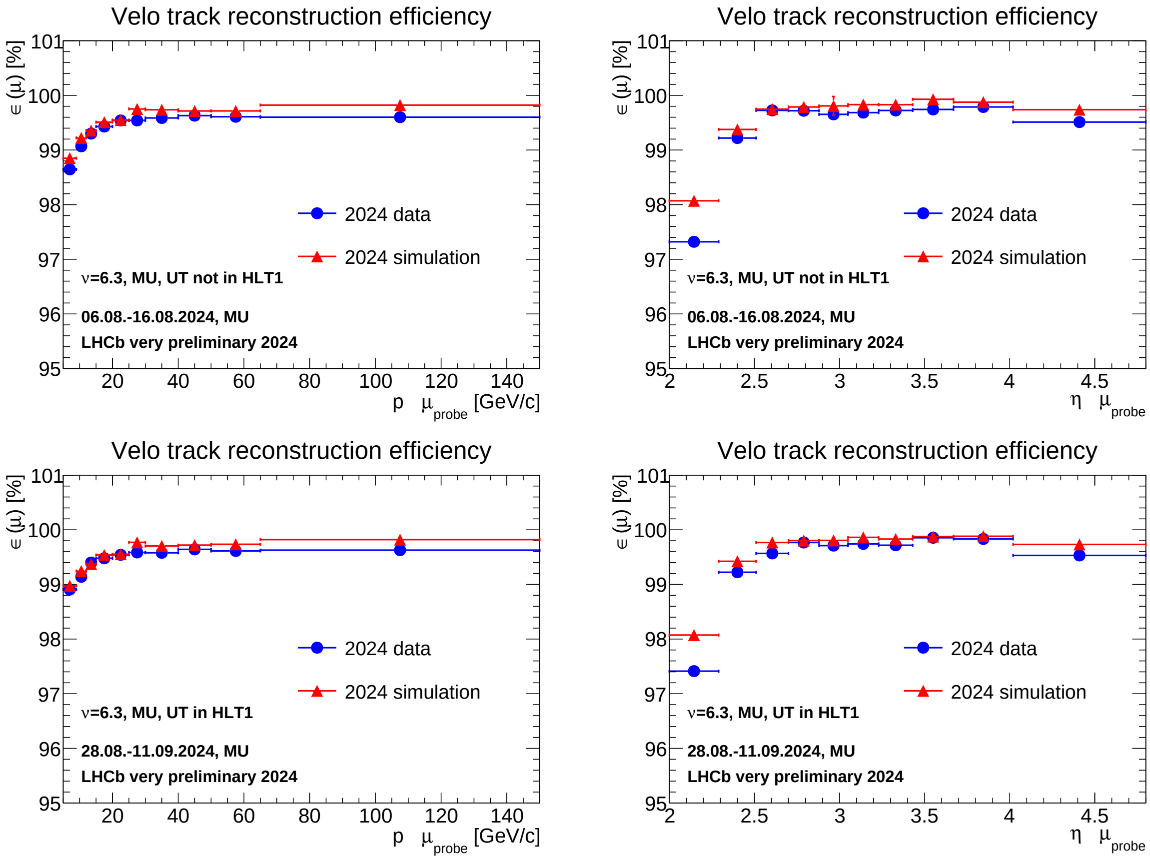


Figure 8.1: VELO track reconstruction efficiencies in bins of momentum, p, (left) and pseudorapidity, η, (right) in comparison between recorded (blue) and simulated (red) data for data set 1 (top) and 2 (bottom).

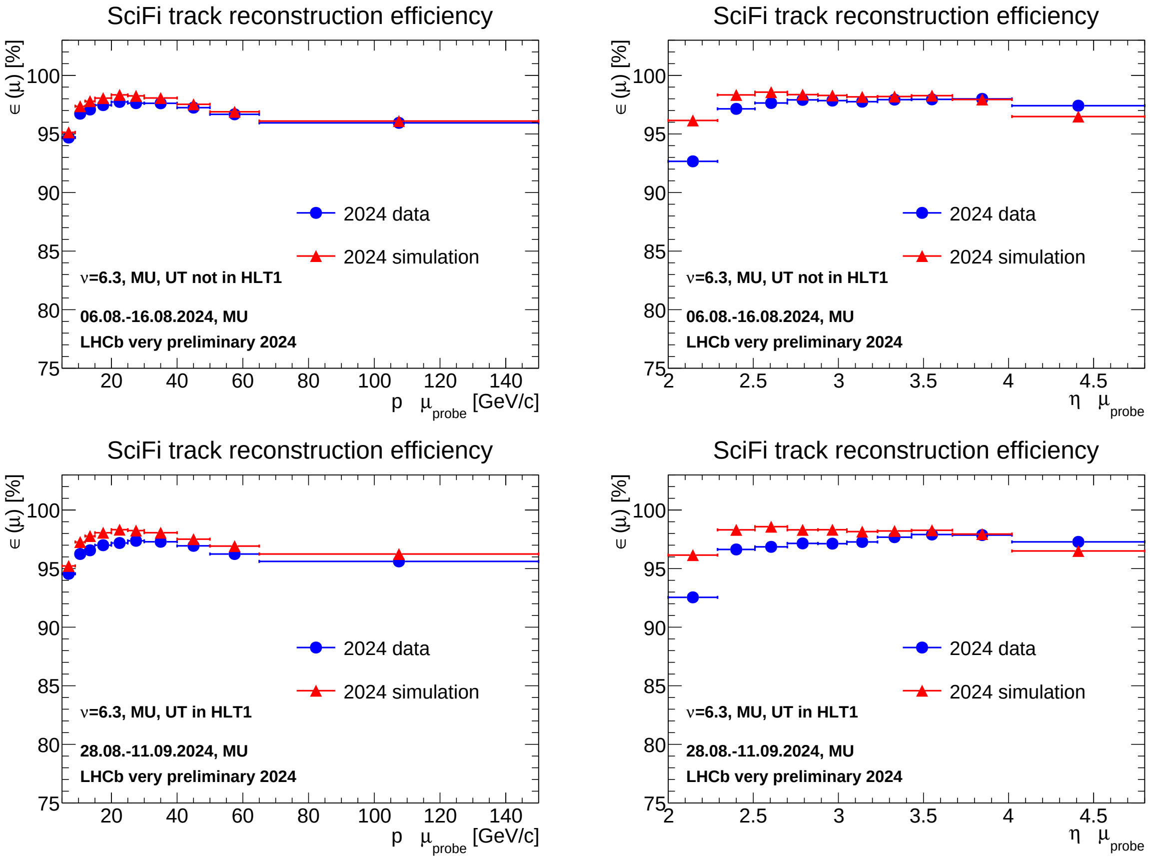
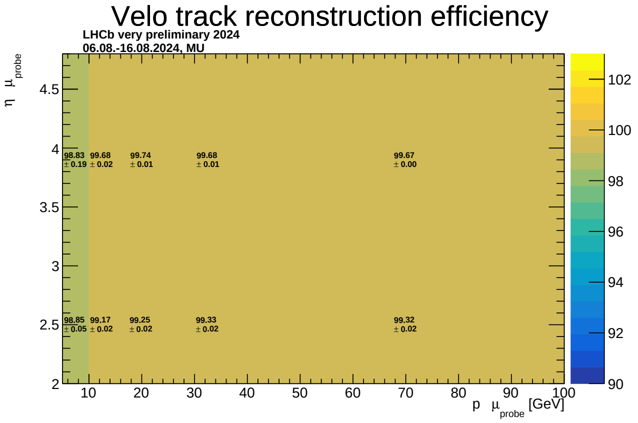


Figure 8.2: SciFi track reconstruction efficiencies in bins of momentum, p, (left) and pseudorapidity, η, (right) in comparison between recorded (blue) and simulated (red) data for data set 1 (top) and 2 (bottom).

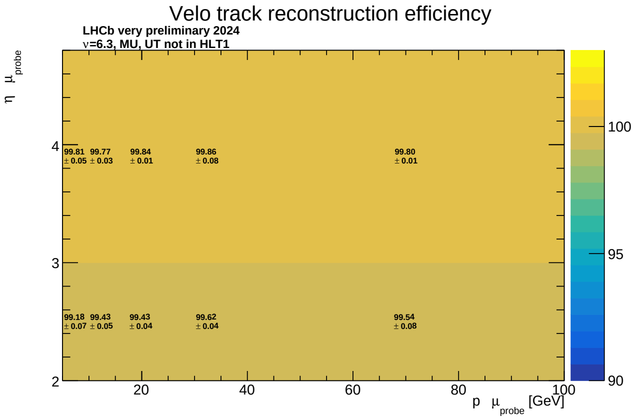
corrected when using simulated track reconstruction efficiencies in analyses. It is clearly visible that the agreement is worst for both efficiencies in the very low η region, which is the outermost region of the detector. This is expected, as this region has the least number of tracks useful for correctly aligning the detector. However, the amount of discrepancy is also a clear hint that the simulation of this region needs improvement.

The uncertainties are again purely statistical. Systematic uncertainties have to still be evaluated.

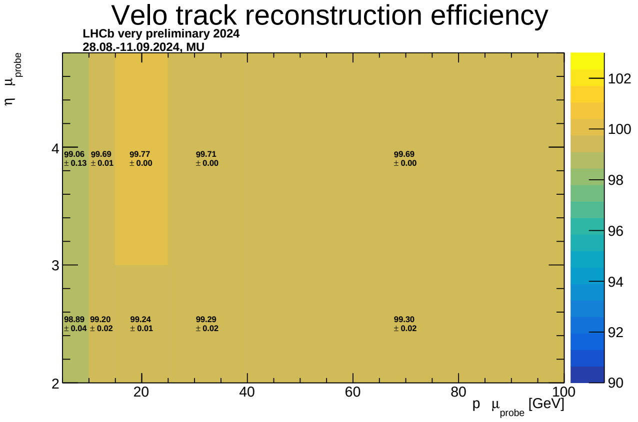
The VELO efficiency shows a slight improvement in the agreement between recorded and simulated data between *data set 1* and *data set 2*. This is most likely due to the ongoing improvements in reconstruction. The SciFi efficiency on the other hand shows worse agreement in *data set 2* than in *data set 1*. This is mostly due to additional SciFi frontend boards that had to be excluded on August 17th and August 27th, which lower the SciFi efficiency in data. This is not yet included in simulated data and the quantity of the effect is evaluated later.



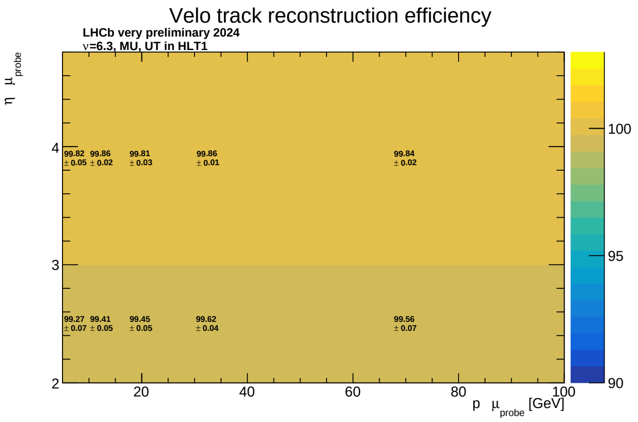
data set 1, recorded



data set 1, simulated



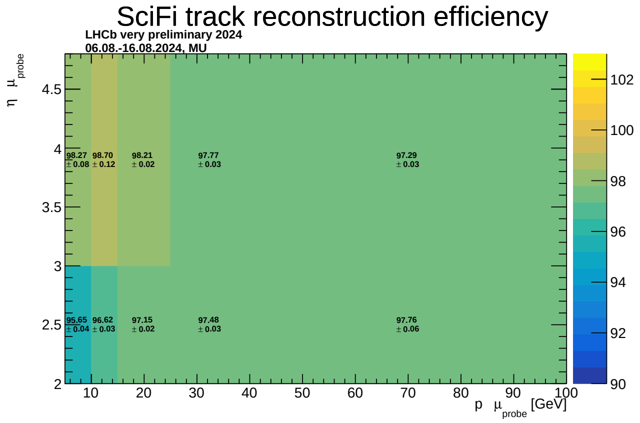
data set 2, recorded



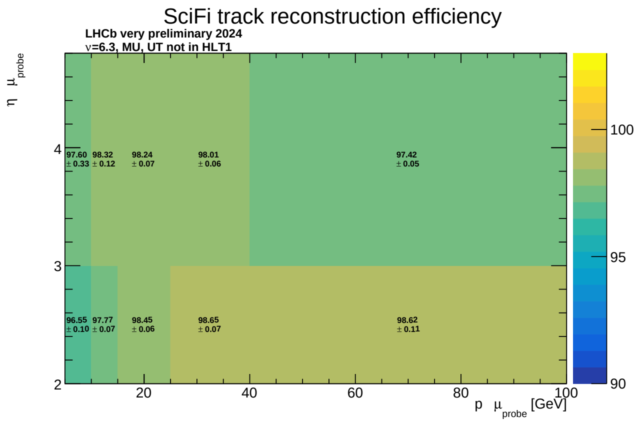
data set 2, simulated

Figure 8.3: VELO track reconstruction efficiencies in two-dimensional bins of momentum, p, and pseudorapditiy, η, in recorded and simulated data for data set 1 and 2.

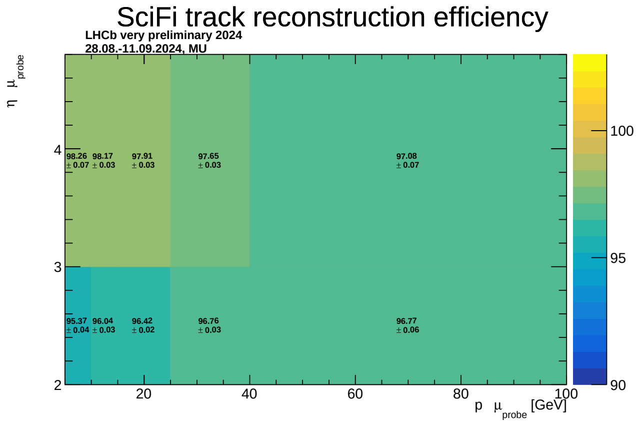
data set 1, recorded



data set 1, simulated



data set 2, recorded



data set 2, simulated

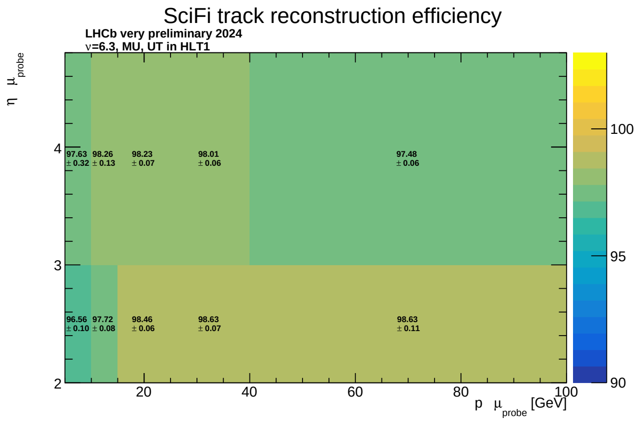


Figure 8.4: SciFi track reconstruction efficiencies in two-dimensional bins of momentum, p, and pseudorapidity, η, in recorded and simulated data for data set 1 and 2.

8.2.2 Comparison with the MuonUT method

The *MuonUT* method, which describes the track reconstruction efficiency of both the SciFi and VELO combined, can be used as a check to the efficiencies produced using the *Downstream* and *VeloMuon* method. Additionally, by averaging over the combined *Downstream* and *VeloMuon* method and the *MuonUT* method to determine the *long track reconstruction efficiency*, common systematic uncertainties cancel and the overall systematic uncertainties can thus be reduced.

As mentioned before, a bias affects the *MuonUT* candidates due to the *propagation bug*. Additionally, *MuonUT probe tracks* are lost in the region $3.1 \leq \eta \leq 4.1$ due to the *pseudorapidity bug*. This is illustrated in Fig. 8.5. The effect of the *pseudorapidity bug* can be estimated by comparing the dip in the distribution of efficiencies over the range of pseudorapidity with the expected distribution gained from extrapolating between the efficiencies below and above that dip.

Fig. 8.6 compares the *track reconstruction efficiency* determined with the *MuonUT* method with the product of the VELO and SciFi *track reconstruction efficiency* determined with the *Downstream* and *VeloMuon* methods in bins of momentum, p, and pseudorapidity, η. Fig. 8.7 compares the same two-dimensionally in p and η for *data set 1* and Fig. 8.8 for *data set 2*. The kinematic distributions of the *MuonUT* probe track in recorded and simulated data can be found in Appendix A in Fig. A.3.

The *long track reconstruction efficiency* determined by the *MuonUT method* tends to

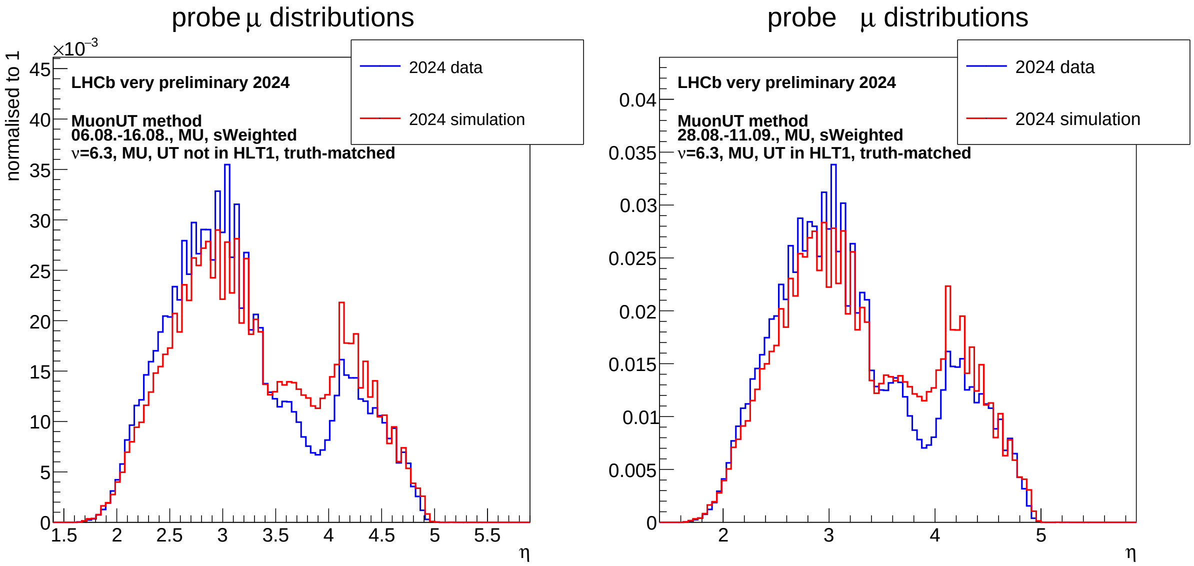
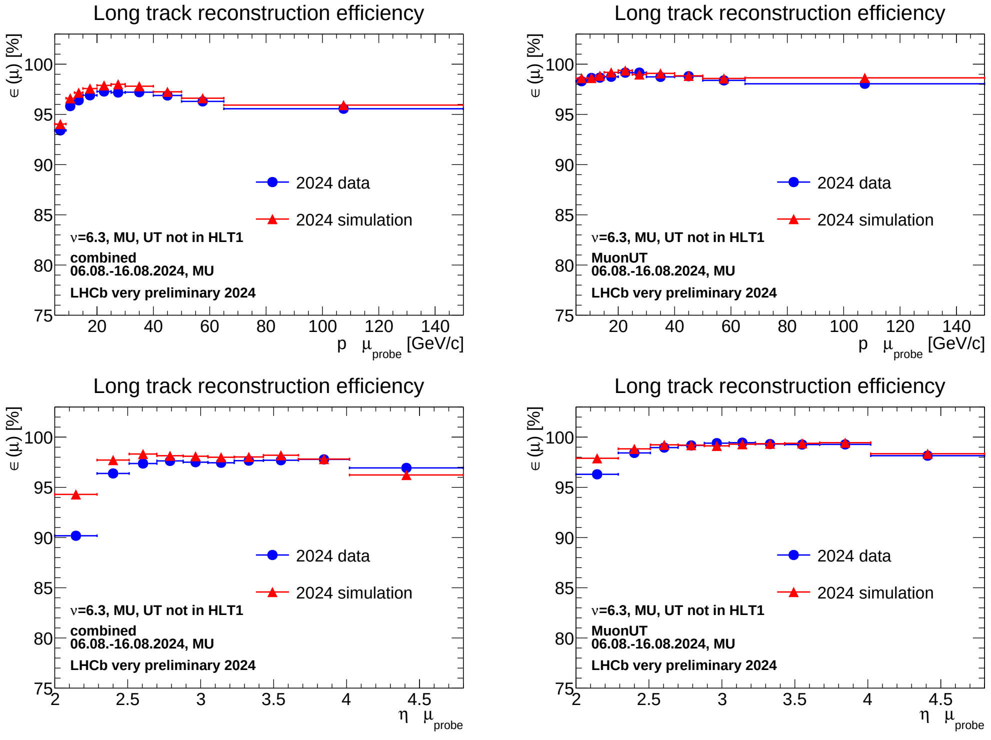


Figure 8.5: Distribution of MuonUT probe pseudorapdity in recorded (blue) and simulated (red) data in data set 1 (left) and data set 2 (right). Clearly visible is the dip between 3.1 and 4.2 where the reconstruction is affected by the pseudorapidity bug.

combined **MuonUT**

data set 1



data set 2

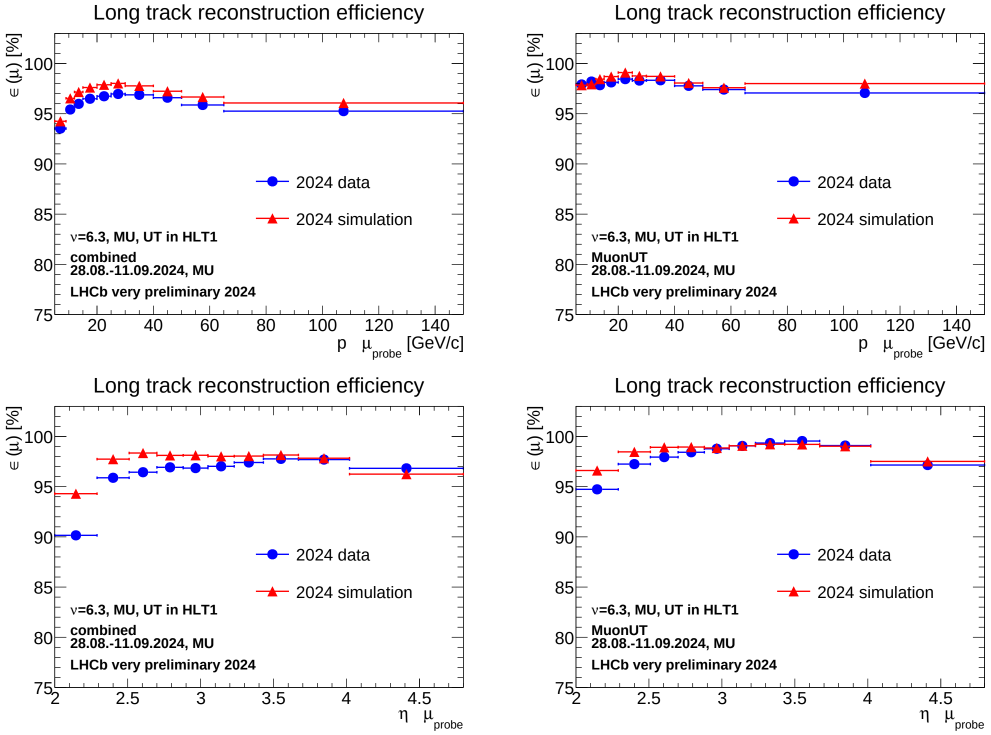
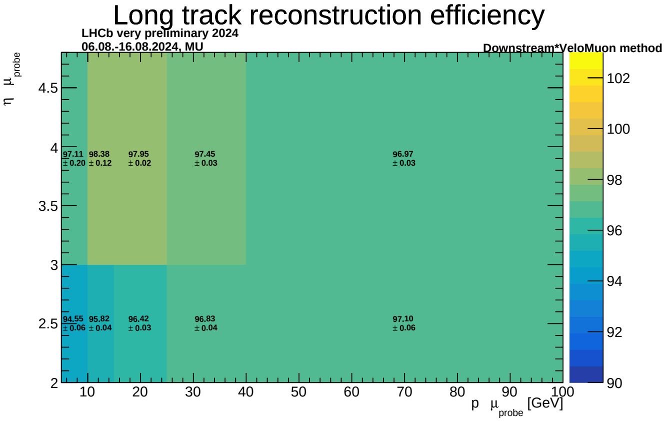
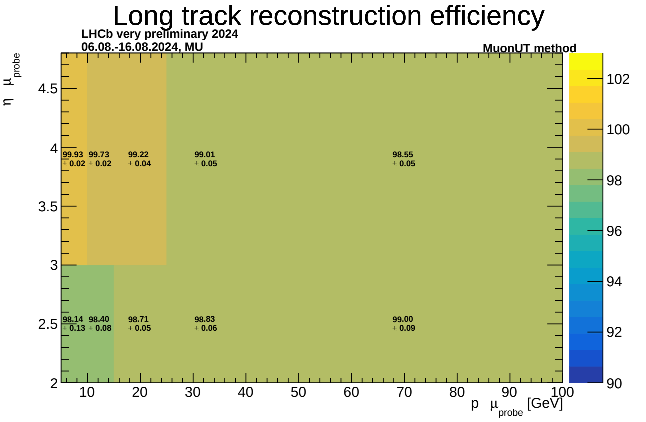


Figure 8.6: Comparison of the combined method, obtained by multiplying the VELO and SciFi efficiencies (left), with the MuonUT method (right), both determining the long track reconstruction efficiency for data set 1 (top four) and data set 2 (bottom four), in bins of momentum (row 1, row 3), and in bins of pseudorapidity (row 2, row 4) in recorded (blue) and simulated (red) data.

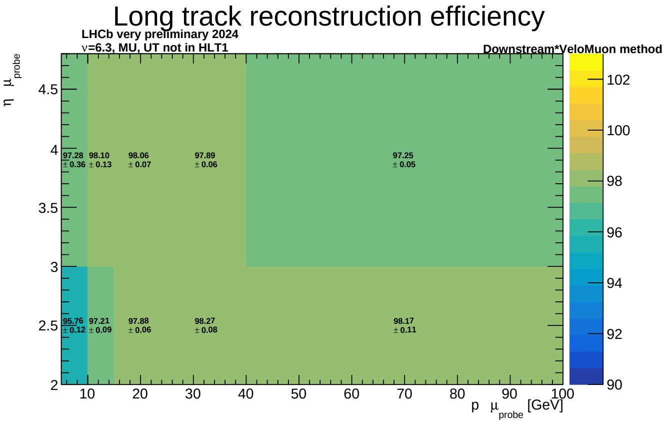
***data set 1,
combined method,
recorded***



***data set 1,
MuonUT method,
recorded***



***data set 1,
combined method,
simulated***



***data set 1,
MuonUT method,
simulated***

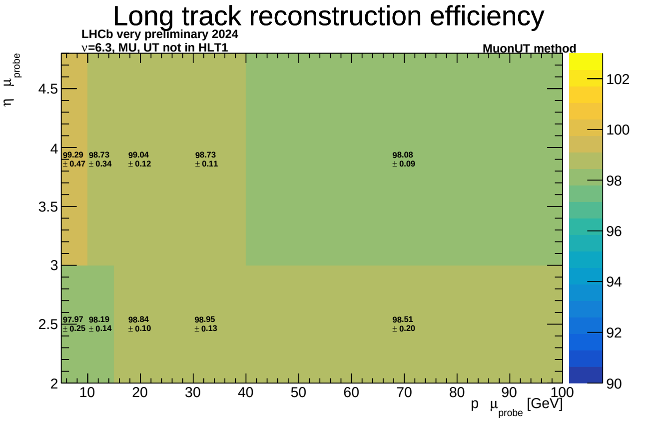
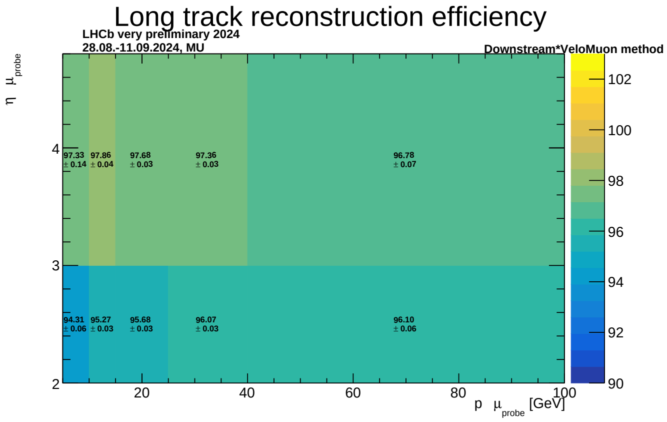
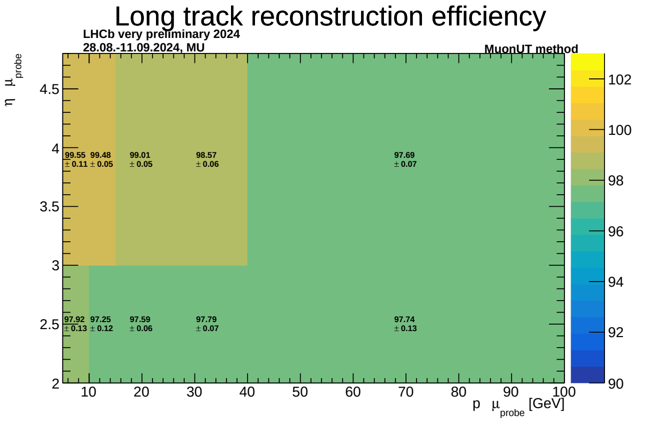


Figure 8.7: Two-dimensional comparison of the long track reconstruction efficiency in p *and* η *from the combined and the MuonUT method in recorded and simulated data of data set 1.*

data set 2,
combined method,
recorded



data set 2,
MuonUT method,
recorded



data set 2,
combined method,
simulated

data set 2,
MuonUT method,
simulated

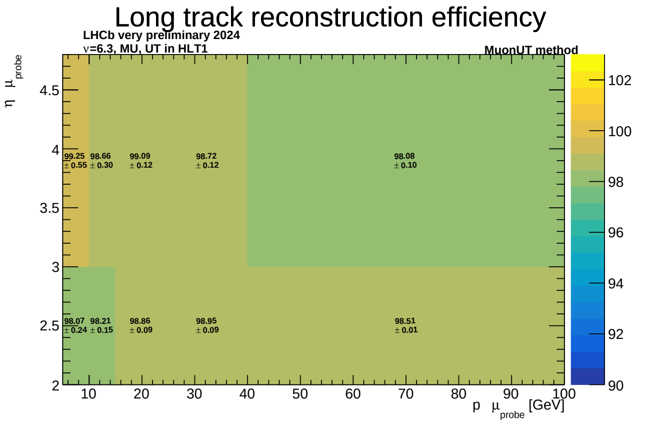


Figure 8.8: Two-dimensional comparison of the long track reconstruction efficiency in p and η from the combined and the MuonUT method in recorded and simulated data of data set 2.

be consistently higher than that determined by the combination of the *Downstream* and *VeloMuon* method. In *data set 1*, a contribution to this is the bias due to the *propagation bug*. However, this is not true for the remaining discrepancies in *data set 2*. Additionally, the track reconstruction efficiencies in both data sets are unreliable in the region $3.1 \leq \eta \leq 4.1$ as well as the region $\eta > 3, p < 100\,\mathrm{GeV}/c$ where the statistics is limited. This is due to the *pseudorapidity bug*. In the region $\eta > 3, p > 100\,\mathrm{GeV}/c$ the background is low enough to allow for a reliable track reconstruction efficiency despite the *pseudorapidity bug*.

For a better comparison between the two methods, the recorded-over-simulated data ratio of the track reconstruction efficiencies determined with the *combined* and the *MuonUT method* are shown in comparison in Fig. 8.9 in bins of momentum, p, and pseudorapidity, η, and two-dimensionally in Fig.8.10.

The *MuonUT* method to complement the other two methods determining the *long track reconstruction efficiency*. However, both the *propagation bug* and the *pseudorapidity bug* affect the *MuonUT* method, making it less reliable. Even estimating the effect of both bugs, the remaining discrepancy between both methods is not fully understood. There-

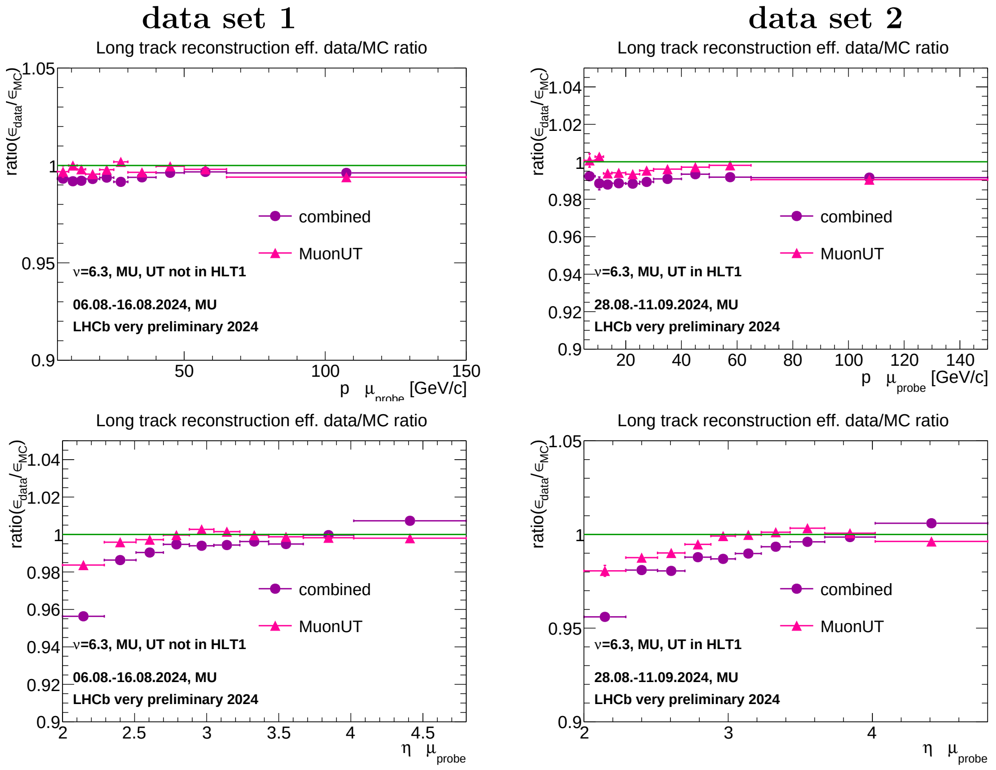
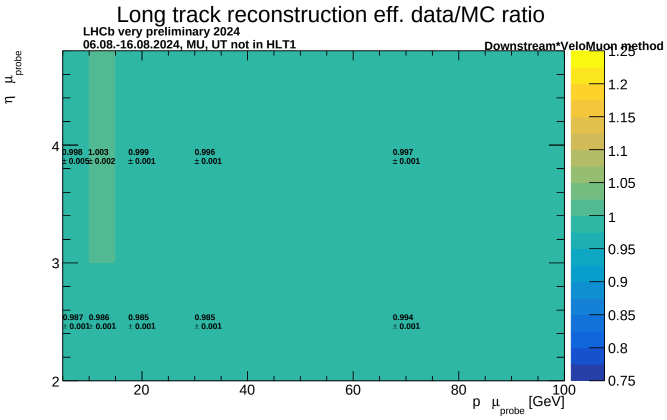
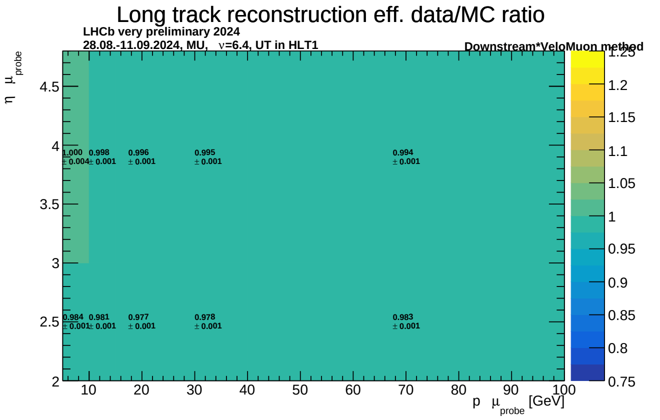


Figure 8.9: Ratio of recorded over simulated track reconstruction efficiencies obtained via the combined method (purple), multiplying the VELO and SciFi efficiencies, and via the MuonUT method (pink) in comparison in bins of momentum, p, (first row) and pseudorapidity, η, (second row) for data set 1 (left) and data set 2 (right).

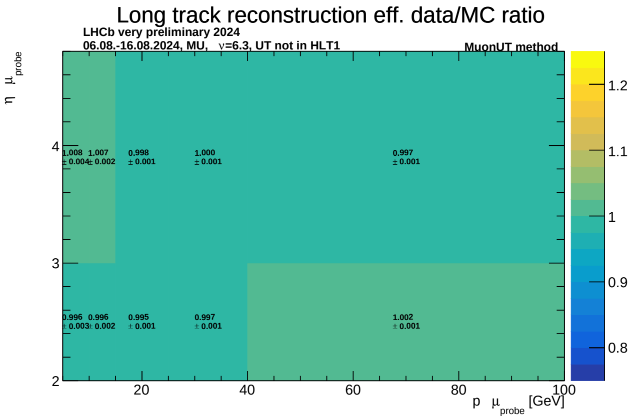
data set 1,
efficiency ratio,
combined method



data set 2,
efficiency ratio,
combined method



data set 1,
efficiency ratio,
MuonUT method



data set 2,
efficiency ratio,
MuonUT method

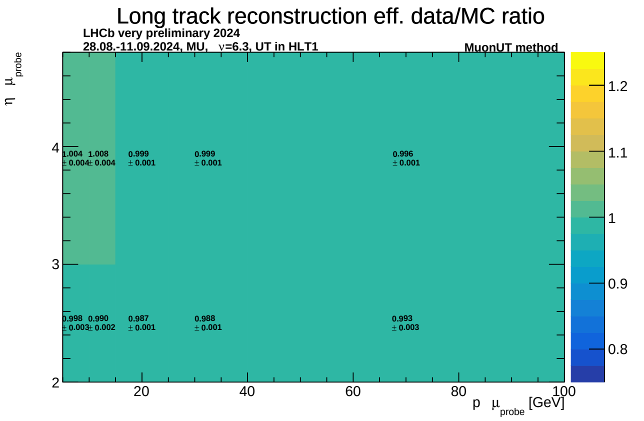


Figure 8.10: Ratio of recorded over simulated track reconstruction efficiencies in p and η in comparison of combined and MuonUT method for data set 1 and 2.

fore, the methods are rather used as comparison, and their discrepancy can be taken as systematic uncertainty. Reasons for the remaining discrepancies are discussed later.

8.3 Systematic uncertainties for the track reconstruction efficiencies

8.3.1 Reconstruction bugs in data set 1

A systematic uncertainty is assigned to the bias introduced into the *MuonUT method* due to the *propagation bug* in *data set 1*. The size of the bias can not be derived from comparing the biased data set 1 and the unbiased data set 2 due to the influence of other effects, most prominently improvements of the alignment, impacting the comparison. Thus the simulated data is reconstructed for all methods with the *propagation bug* present. Afterwards, the reconstruction is run again over all passing candidates, this time with the *propagation bug* corrected. Candidates passing the selection of any other method during the first reconstruction can potentially be recovered for the *MuonUT* method. This is compared with a simulated data sample reconstructed directly using the corrected reconstruction. The remaining difference is an estimate of the effect of the bias in the reprocessed data set 1.

Fig. 8.11 shows as an example the difference between the two samples as a function of pseudorapidity, η. The discrepancy due to the bug varies in dependence of the specific bins, with a maximum deviation of -1.31% in all considered bins. The absolute deviation

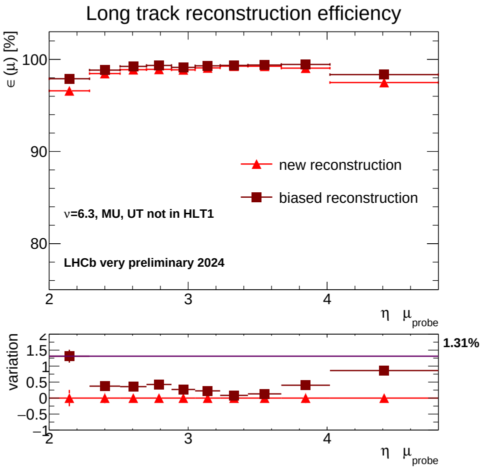


Figure 8.11: Long track reconstruction efficiency using the MuonUT method with the biased reconstruction algorithm (dark red) and the corrected reconstruction algorithm (light red) in comparison. The variation between the two samples in percent points is shown in the lower plot.

of the track reconstruction efficiency estimated with the MuonUT method due to the *propagation bug* for the studied sample is:

$$\sigma^{MuonUT,\,data\,set\,1}_{Track\,Reconstruction}(sys) =^{+0.00\%}_{-0.56\%}$$

This is the systematic uncertainty assigned to any sample with the same kinematic properties as the studied sample. Any other sample needs to take into consideration the bin-dependent systematic uncertainties.

For the recorded/simulated data ratio, the systematic uncertainty is expected to largely cancel out, as both are biased. However, this can not be tested as the systematic uncertainty on data can not be determined. Therefore, the systematic uncertainty determined on simulated data is assigned to the overall ratio, too.

8.3.2 Systematic uncertainties due to the fit model

All track reconstruction efficiency fits for 2024 samples are fitted with a *Johnson* function for the signal and, in the case of simulated data, an *exponential* for the background, in the case of data a *Chebychev polynomial* of the third degree for the background. While the fit automatically provides a statistical uncertainty on the efficiency, the systematic uncertainty due to the fit model is tested by fitting the signal shape alternatively with a *Double Crystal Ball (DoubleCB)* function with several parameters fixed. The track reconstruction efficiency is determined with both fit models in ten bins of momentum, p, ten bins of pseudorapidity, η, and ten bins of both two dimensionally.

The uncertainty due to the fit model is larger for simulated data, where the biggest variation of the *long track reconstruction efficiency* in any bin is $(sys)^{+0.18}_{-0.07}\%$ for data set 1 and $(sys)^{+0.14}_{-0.05}\%$ for data set 2. The absolute deviation for simulated data is for both data sets is:

$$\sigma^{fitmodel,MC}_{Track\,Reconstruction}(sys) =^{+0.15}_{-0.00}\%,$$

dominated by the systematic uncertainty on the SciFi efficiency.

This is the systematic uncertainty expected for a sample equivalent to the considered one. Any sample would need to consider the systematic uncertainty in each bin.

In recorded data, the systematic uncertainty due to the fit model is, thanks to higher statistics:

$$\sigma^{fitmodel,data}_{Track\,Reconstruction}(sys) =^{+0.00}_{-0.02}\%$$

for the *long track reconstruction efficiency.*

However, there are several bins which contribute with a much higher systematic uncertainty. Those are the bins affected by the *pseudorapidity bug* for the MuonUT method. In those bins, the statistics is so limited that the efficiency can vary by up to half a percent point when slightly changing the fit model while still providing a sensible fit. This is visualised in Fig. 8.12.

The same is true for the low momentum bins of the *VELO track reconstruction efficiency* ($p < 10\,\mathrm{GeV}/c$). As these bins are contributing with a high systematic uncertainty, but are in areas with low abundance of signal candidates, they only contribute a negligible amount to a sample with the kinematic distribution of the considered sample.

These bins also experience higher fluctuation in all further tests.

For the ratio of recorded over simulated data, the much smaller contribution to the systematic uncertainty coming from the recorded data sample is neglected. The systematic uncertainty determined on the simulated data sample is taken as systematic uncertainty on the ratio.

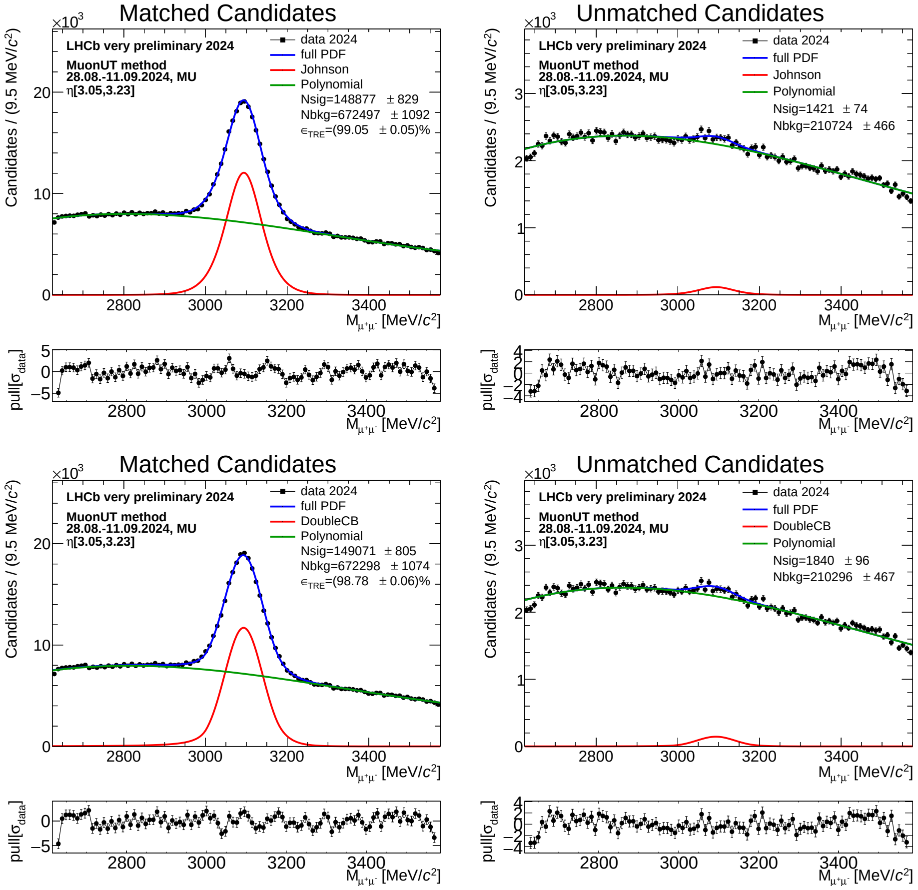


Figure 8.12: Fit to the MuonUT method sample in the $\eta[3.05, 3.23]$ bin of data set 2 with a Johnson function over a Chebychev polynomial (top) and a double-sided Crystal Ball function over a Chebychev polynomial (bottom) to the matched (left) and not matched (right) events, producing efficiencies with an absolute discrepancy of -0.27%.

8.3.3 reWeighting in kinematic and multiplicity parameters

The track reconstruction efficiency is dependent on the kinematics of the probe track. However, the kinematic distributions of probe tracks are not exactly replicated in simulated data, as can be seen in Appendix A. To estimate the systematic uncertainty coming from the different distributions of recorded and simulated data in the kinematic variables, the simulated data is *reWeighted* in the pseudorapidity and angle ϕ of the tag track and the reconstructed J/ψ to *sWeighted data.* The track reconstruction efficiencies in simulated data are compared to those in *reWeighted* simulated data and the maximum variation in all bins is taken as systematic uncertainty.

As an example, Fig. 8.13 presents the effect of *reWeighting* the *SciFi track reconstruction efficiency* in bins of momentum, which shows a consistent trend with the efficiency being corrected to higher values for lower momentum and lower values for higher momentum. The highest systematic uncertainty assigned to any bin due to discrepancies in the kinematic distributions is $(sys)^{+0.50\%}_{-0.34\%}$.

The systematic uncertainty due to *reWeighting kinematically* assigned to the *long track reconstruction efficiency* determined with either method is:

$$\sigma^{kinematic-reWeighting,MC}_{Track\,Reconstruction}(sys) = ^{+0.05\%}_{-0.04\%}$$

This systematic uncertainty does not differ significantly for the two data sets.

Equivalent to *reWeighting in kinematic variables*, the systematic uncertainty corresponding to discrepancy in the *multiplicity* variables is determined. The chosen multiplicity variables are the number of primary vertices, the number of reconstructed tracks, and the number of clusters of hits in the SciFi in the event. A higher multiplicity or occupancy in the detector means a higher chance of hits being misassigned to another track or combined incorrectly, which lowers the track reconstruction efficiency. The multiplicity is often not described correctly in the simulation. Especially the number of SciFi clusters is often underestimated.

Fig. 8.14 shows as an example all three variables *reWeighted* for *data set 2* for the *Velo-Muon method* used to determine the *SciFi track reconstruction efficiency.* As can be seen, the number of reconstructed tracks agrees very well between data and MC, while the number of primary vertices is overestimated slightly in simulated data and the number of SciFi clusters is underestimated significantly in simulated data.

Fig. 8.15 shows that the VELO and SciFi track reconstruction efficiency are both significantly more dependent on the number of primary vertices than the number of SciFi clusters, and even show a significantly increasing discrepancy between recorded and simulated

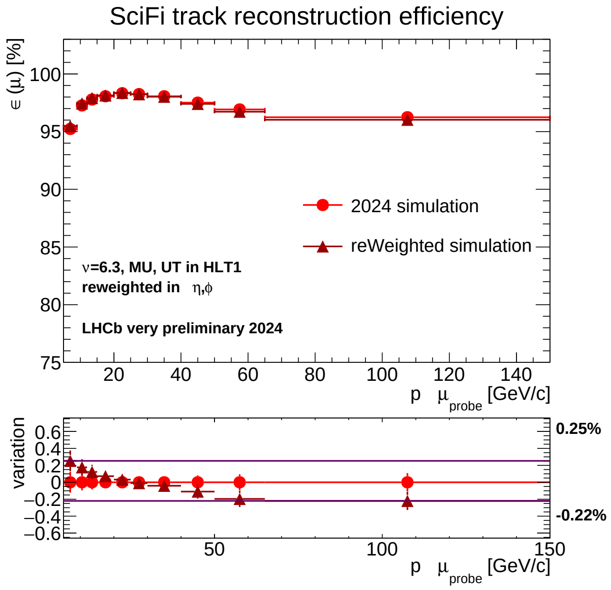


Figure 8.13: Systematic uncertainties due the reWeighting in kinematics of the SciFi reconstruction efficiency in data set 2 comparing the unWeighted simulation (light red) to the reWeighted simulation (dark red).

data for the SciFi track reconstruction efficiency. However, due to the large discrepancy in distributions of simulated and recorded data, the number of SciFi clusters still has the largest impact on the changes of track reconstruction efficiency when *reWeighting*.

The highest systematic uncertainty assigned to any bin due to discrepancies in the multiplicity distributions is $(sys)^{+0.65\%}_{-1.42\%}$ and is assigned to the *MuonUT method*. It is twice as high as the highest systematic uncertainty assigned to the other two methods.
The systematic uncertainty due to discrepancies in the multiplicity distributions assigned to the *long track reconstruction efficiency* determined with either method is:

$$\sigma^{multiplicity-reWeighting,MC,data\,set\,1}_{Track\,Reconstruction}(sys) =^{+0.02\%}_{-0.19\%}$$

for data set 1 and

$$\sigma^{multiplicity-reWeighting,MC,data\,set\,2}_{Track\,Reconstruction}(sys) =^{+0.00\%}_{-0.13\%}$$

for data set 2.
Evidently, *reWeighting* in multiplicity variables has a larger effect and therefore produces a larger systematic uncertainty. However, as both uncertainties are of the same order of magnitude and an order of magnitude higher than uncertainties due to the fitmodel, both need to be taken into account. As can be seen in Fig. 8.16, the *reWeighting* in multiplicity does not affect the kinematic variables significantly. As there is no strong

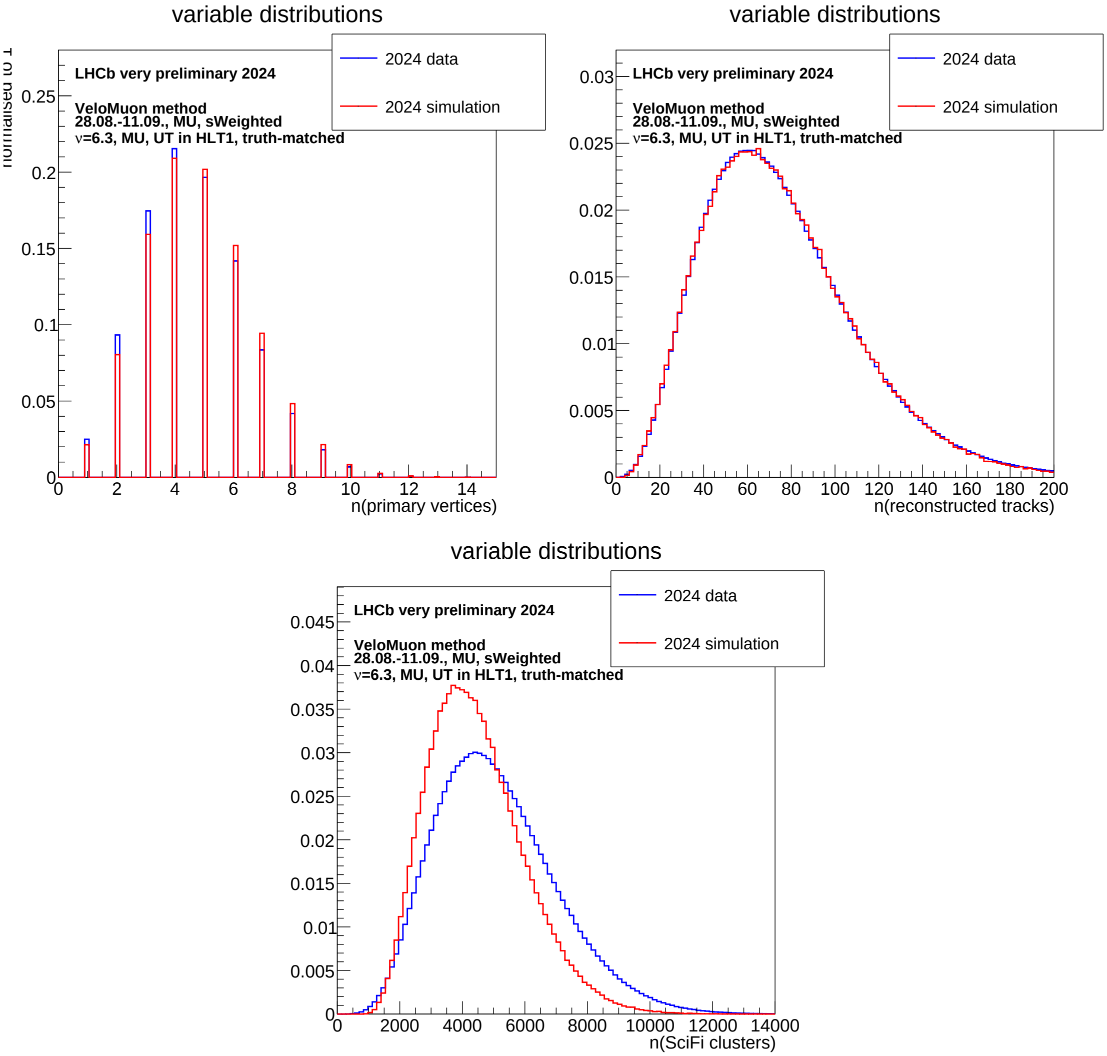


Figure 8.14: Number of primary vertices (top left), number of reconstructed tracks (top right) and number of clusters in the SciFi (bottom) for the VeloMuon method in data set 2 for recorded (blue) and simulated (red) data as an example.

correlation between multiplicity variables and kinematic variables, it is unproblematic to *reWeight* in all variables at the same time: the number of primary vertices, the number of reconstructed trucks, the number of measured SciFi clusters, the pseudorapidity and angle ϕ of the tag muon track, and the pseudorapidity and angle ϕ of the J/ψ reconstructed mother.

After *reWeighting* in all seven of these variables, the efficiencies are again measured and compared with the not *reWeighted* efficiencies. The largest variation in a bin amounts to $(sys)^{+0.56\%}_{-1.57\%}$. The systematic uncertainty assigned to the total sample is:

$$\sigma^{all-reWeighting,MC,data\,set\,1}_{Track\,Reconstruction}(sys) =^{+0.05\%}_{-0.23\%}$$

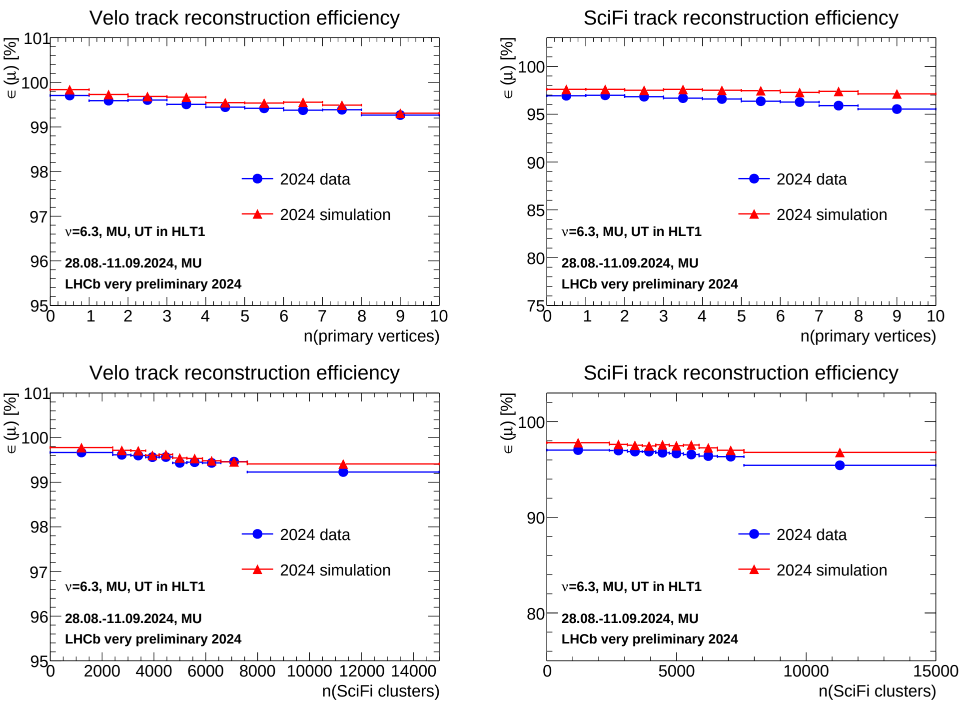


Figure 8.15: VELO (top left) and SciFi (top right) track reconstruction efficiency in dependency of the number of primary vertices (top) and the number of measured SciFi clusters (bottom) in the event in recorded (blue) and simulated (red) data.

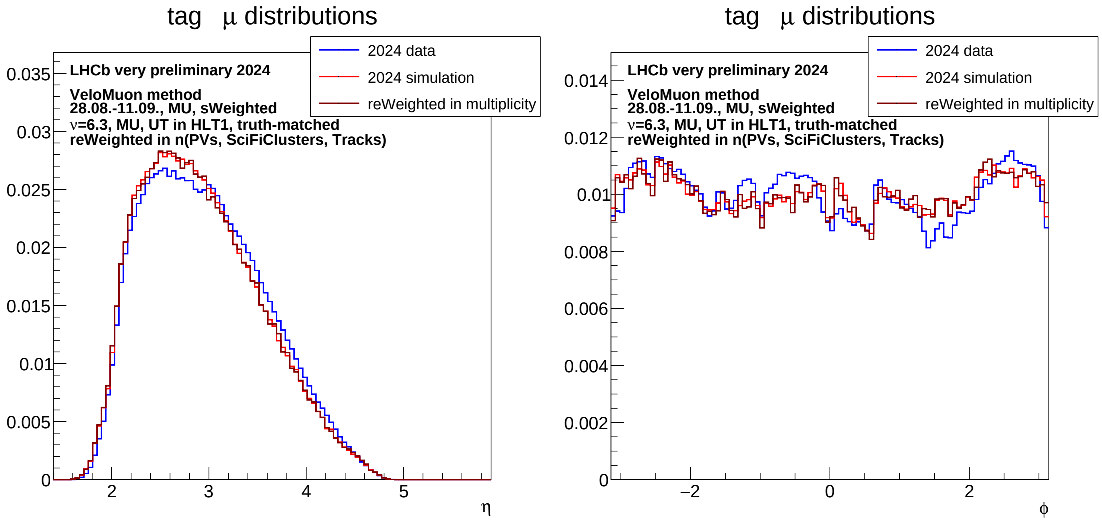


Figure 8.16: Distribution of tag muon η and ϕ with (dark red) and without (red) reWeighting in multiplicity variables for the VeloMuon method in comparison to recorded data (blue) as an example.

for data set 1 and

$$\sigma^{all-reWeighting,MC,data\,set\,2}_{Track\,Reconstruction}(sys) =^{+0.05\%}_{-0.14\%}$$

for data set 2. This is the systematic uncertainty expected for a sample with the same kinematic properties as the considered sample. For a correct assignment of the systematic uncertainty, any other sample has to consider the bin-dependent systematic uncertainties.

8.3.4 Differences between the Combined and the MuonUT method

To determine the *long track reconstruction efficiency*, the VELO efficiency and the SciFi efficiency in every bin are multiplied with each other. The *MuonUT method*, which also describes the *long track reconstruction efficiency*, is constructed both to verify the result and to reduce systematic uncertainty by averaging over both methods.

However, the *MuonUT method* has separate systematic uncertainties due to the *propagation bug* and the *pseudorapidity bug* in its reconstruction algorithm. Additionally, the discrepancies between both methods are still rather large. Thus, the efficiency ratios are compared for both the *combined* and the *MuonUT method.* The largest variation of the recorded/simulated data ratio between the combined and the MuonUT method amount to

$$\sigma^{method,dataset1}_{TrackReconstruction}(sys) =^{+2.74}_{-0.93}\%$$

for *data set 1* and

$$\sigma^{method,dataset1}_{TrackReconstruction}(sys) =^{+2.46}_{-0.97}\%$$

for *data set 2.*

However, the variations are especially large in phase space regions with low abundance, such as the region affected by the *pseudorapidity bug.* As these regions contribute less to the overall efficiency due to less statistics, the systematic uncertainty assigned to the total sample due to discrepancy between the *combined* and the *MuonUT* method amount to

$$\sigma^{method,data\,set\,1}_{TrackReconstruction}(sys) = \pm 0.47\%$$

for *data set 1* and

$$\sigma^{method,data\,set\,2}_{TrackReconstruction}(sys) = \pm 0.69\%$$

for *data set 2.*

Again this is the systematic uncertainty expected for a sample with the same kinematic properties as the considered sample. For a correct assignment of the systematic uncertainty, any other sample has to consider the bin-dependent systematic uncertainties.

8.3.5 Other sources of systematic uncertainties

All other sources of systematic uncertainties are expected to cancel out when taking the ratios between recorded and simulated data. The only exception is the systematic uncertainty due to description of the detector material and interactions with it. However, as muons are mostly minimally interacting particles (MIPs) at LHCb, this is negligible for them. To use the track reconstruction efficiencies for hadrons, an additional systematic uncertainty has to be introduced. This is discussed in Chapter 11.

8.4 Correction factors for the track reconstruction efficiency in 2024

The recorded/simulated data ratios of the *long track reconstruction efficiency* for *data set 1* and *data set 2* to be used in 2024 analyses are presented in Fig. 8.10. Uncertainties presented there are purely statistical.

Statistical uncertainties on the track reconstruction efficiency ratio of simulated over recorded data amount to:

$$(stat) \pm 0.02\% \; (data\, set\, 1) \qquad\qquad (sys) \pm 0.04\% \; (data\, set\, 2)$$

The different sources for systematic uncertainties and their absolute values are summarised in Tab. 8.4.

The systematic uncertainty due to the *propagation bug* is assigned only to the *MuonUT method*. While it is expected to cancel out between recorded and simulated data, this can not be verified. Therefore, it is assigned to the *MuonUT method* efficiency ratios for *data set 1*.

The tiny systematic uncertainty assigned due to the *fit model* is neglected for *recorded data*. The value assigned to simulated data is propagated to the efficiency ratio. As the systematic uncertainty assigned due to *reWeighting* relates only to simulated data, it is also propagated to the efficiency ratio.

The largest source for systematic uncertainty arises from the discrepancy between the *combined method* and the *MuonUT method*. Despite two known bugs in the *MuonUT method*, which are the *propagation bug* and the *pseudorapidity bug*, the source of these discrepancies is not fully understood. It can be assumed that with further research and a better understanding of these discrepancies, this can be reduced or neglected in favour of

source	data set 1	data set 2
propagation bug (MuonUT simulated data)	$(sys)^{+0.00\%}_{-0.56\%}$	—
fit model, simulated data	$(sys)^{+0.15\%}_{-0.00\%}$	$(sys)^{+0.15\%}_{-0.00\%}$
fit model, recorded data	$(sys)^{+0.00\%}_{-0.02\%}$	$(sys)^{+0.00\%}_{-0.02\%}$
reWeighting, simulated data	$(sys)^{+0.05\%}_{-0.23\%}$	$(sys)^{+0.05\%}_{-0.14\%}$
method (recorded/simulated ratio)	$(sys)^{+0.47\%}_{-0.47\%}$	$(sys)^{+0.69\%}_{-0.69\%}$
total (combined method)	±0.52%	±0.70%
total (MuonUT method)	±0.77%	±0.70%

Table 8.4: Systematic uncertainties assigned to the track reconstruction efficiencies in simulated and recorded 2024 data. The combined systematic uncertainties are given in the last two rows separately for the combined and the MuonUT method.

the more reliable method. Thus, the systematic uncertainties due to *reWeighting* and the *fit model* are expected to dominate.

The systematic uncertainty due to the *fit model* can be reduced by using a simulated data sample with more statistics. This is demonstrated by the negligible systematic uncertainty due to the fit model in the recorded data sample, which utilises more statistics.

The simulated samples used in this thesis are not *reWeighted* for the determination of the track reconstruction efficiencies. Instead, the effect of *reWeighting* is taken as systematic uncertainty. By providing efficiency factors using simulated samples already *reWeighted* in at least one variable, this systematic uncertainty can be reduced significantly. Tests show that *reWeighting* the samples in the number of SciFi clusters would reduce the systematic uncertainty due to *reWeighting* to $(sys) \pm 0.05\%$. All samples using track reconstruction efficiency correction values would then be required to also be *reWeighted* in the same variable.

Thus, systematic uncertainties are expected to be reducible with easy measures. Nevertheless, they are expected to stay dominant in comparison to the statistical uncertainties. While the mentioned improvements are not yet implemented, the track reconstruction efficiencies determined by the *combined method* are used and assigned uncertainties of:

$$(99.31 \pm 0.02(stat) \pm 0.52(sys))\% \qquad (data\,set\,1) \tag{8.1}$$

$$(98.94 \pm 0.04(stat) \pm 0.70(sys))\% \qquad (data\,set\,2) \tag{8.2}$$

This is an estimate of the values expected for any sample with similar kinematic distribu-

tions as the studied sample. Any study utilising the values determined here is required to use bin-dependent values and uncertainties. The combined method is assumed to be the more reliable method due to the known bugs in the *MuonUT method*.
Chapter 9 discusses the origins of the discrepancies between recorded and simulated data. Chapter 10 verifies the method and compares it with an alternative method available on simulation.

Chapter 9

Analysis of underlying causes for discrepancies between track reconstruction efficiencies of recorded and simulated data

This chapter analyses the reasons for the observed discrepancy between the track reconstruction efficiency of recorded and simulated data. The effects of deactivating SciFi and VELO modules, SciFi single hit efficiencies and misalignment are discussed. The influence on trackχ^2*-cuts is tested. Possible double-counting effects on the long track reconstruction efficiency are estimated using Ghost Probability cuts. The dependence of the track reconstruction efficiency on the charge of the probe muon is tested.*

9.1 Remaining discrepancies

While the track reconstruction efficiencies measured with recorded data in 2024 generally show good agreement with those predicted by simulated data, there are still some regions where this is not the case. The *VELO track reconstruction efficiency*, measured with the help of *Downstream muon probe tracks*, shows a significant discrepancy in the very low pseudorapidity region, which corresponds to the outer regions of the detector, both in *data set 1* and *data set 2*.

The *SciFi track reconstruction efficiency*, measured using *VeloMuon probe tracks*, shows an increasing discrepancy with decreasing pseudorapidity. It also shows a general discrepancy in all bins in *data set 2* which is not present in *data set 1*.

The *long track reconstruction efficiency*, measured using *MuonUT probe tracks*, generally shows good agreement between recorded and simulated data, but shows a discrepancy in the lowest pseudorapidity region as both of the other methods do, and has a worse discrepancy in *data set 2* than *data set 1*, similar to the *SciFi track reconstruction efficiency*. This is expected. Any effect on the other two methods is expected to impact this method as well.

The largest discrepancy in *track reconstruction efficiencies* is not between recorded and simulated data, but between the combination of the VELO and SciFi track reconstruction efficiency from the *Downstream* and *VeloMuon* methods, and the *long* track reconstruction efficiency from the *MuonUT* method.

There are several reasons for these discrepancies.

9.2 The deactivation of SciFi modules

Four SciFi frontend boards were excluded from data-taking on May 17th, July 8th, August 17th and August 27th. All four frontend boards are excluded in the lower left quadrant of the SciFi, or $x < 0, y < 0$ in the LHCb coordinate system. This means that particles going through this area of the detector are detected on average in one detector layer less. As nine hits in the twelve layers of the SciFi are absolutely required for the reconstruction of a track, and ten are necessary in most cases, this does not mean that all tracks in this region are lost, but it consequently lowers the track reconstruction efficiency in this region of the SciFi. The excluded frontend boards are not yet reflected in the simulation and therefore show up as a discrepancy between recorded and simulated data.

Fig. 9.1 shows the SciFi track reconstruction efficiency in probe muon pseudorapidity, η, separately for the position of the probe muon when crossing the muon stations in four quadrants for *data set 2*. All four SciFi frontend boards are excluded. The position in the muon system is taken as proxy for the position crossing the SciFi as the VeloMuon probe track is measured without information from the SciFi. This approximation is valid as the track crosses most likely through the same quarter in the SciFi as in the muon system.

The discrepancy is significantly higher in the lower left quadrant where the excluded SciFi frontend boards are located. As long as the excluded frontend boards are not properly reflected in simulation, it may be necessary for analyses to additionally separate the track reconstruction efficiency correction in the detector region. The same effect occurs when the track reconstruction efficiency is binned in the probe muon momentum, p and separated into the four quadrants.

Data set 1 is taken with two excluded frontend boards. One more frontend board was

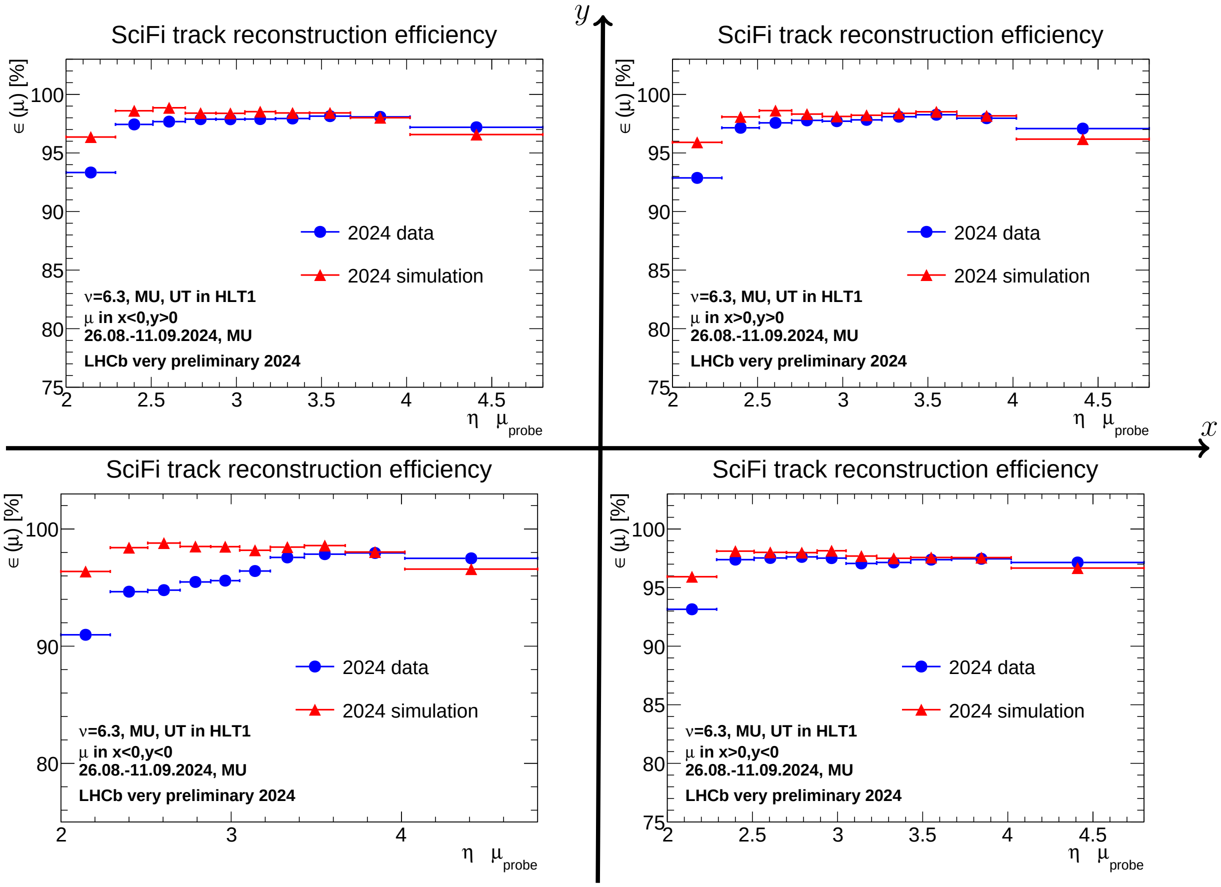


Figure 9.1: SciFi track reconstruction efficiency in bins of pseudorapidity, η, for data set 2 recorded (blue) and simulated (red) data separately in the four quadrants of the detector.

excluded on August 17th, which falls in the data taking period of early analyses, but all data shown so far is taken with only two frontend boards excluded. *Data set 2* is taken with four frontend boards excluded. The fourth was excluded at the time when the UT was included in HLT1 and last improvements where done to the alignment. This makes separating the effects harder, as the efficiency is improved by improved alignment, but worsened by the excluded frontend board. The effects of excluding the third and fourth frontend board on the SciFi track reconstruction efficiency can be seen in Fig. 9.2 in bins of pseudorapidity in the lower left quarter. The effect on the distribution in momentum, p, shows a very similar trend.

Excluding the third frontend board has a significant impact on the efficiency. The discrepancy between recorded and simulated data increases. This is the frontend board closest to the centre of the detector. It is therefore expected to have the greatest influence on the efficiency. For the fourth frontend board, however, the effect is a bit more diverse: while the efficiency worsens slightly in the region between roughly $2.5 < \eta < 3.5$, it improves very slightly in the rest of the range. There are two competing effects here: the sample from *data set 2* has a slightly better alignment than the other two samples. Therefore,

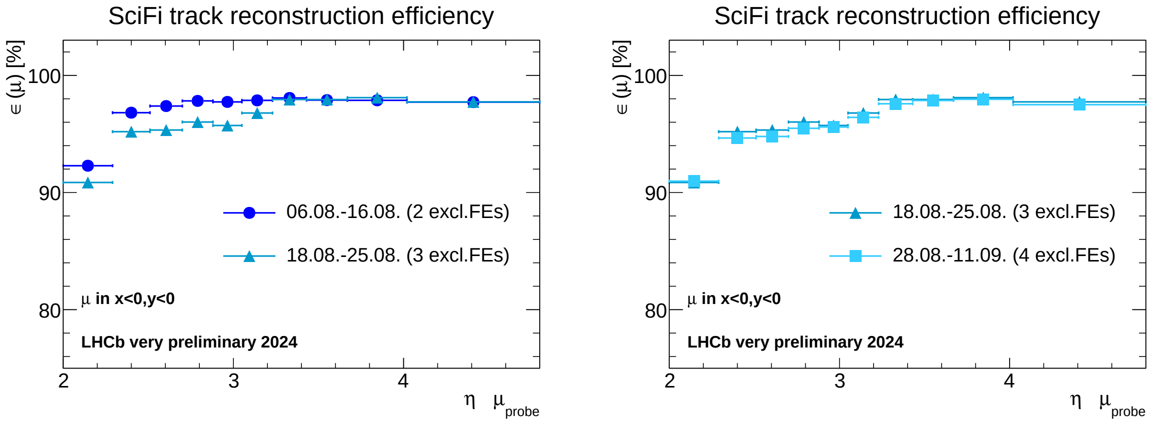


Figure 9.2: Track reconstruction efficiency in dependency of the probe track pseudorapidity, η, for probe tracks passing through the lower left quadrant of the muon system for data taken with two (dark blue) and three (medium blue) excluded frontend (FE) boards (left) and three (medium blue) and four (light blue) excluded frontend (FE) boards (right) in comparison.

a slight overall improvement in the efficiency is offset in the $\eta-$range affected by the excluded SciFi frontend board.

Overall, the excluded frontend boards are most likely the largest cause of discrepancies between recorded and simulated data for the SciFi efficiency, offsetting the slight improvement in efficiency due to the better alignment in *data set 2*.

The effect of excluding the third frontend board on the VELO track reconstruction efficiency and the *long* track reconstruction efficiency measured with the *MuonUT* method in the lower left quarter of the SciFi can be seen in Fig. 9.3. The effect is visible for the *long track reconstruction efficiency*, although not as pronounced as for the SciFi track reconstruction efficiency. As expected, there is no visible effect for the VELO track reconstruction efficiency, although there are some fluctuations between the two samples.

Future simulations are expected to include the missing SciFi frontend boards. This is expected to reduce the seen discrepancies and allow for a better description of the *track reconstruction efficiency*.

The single hit efficiency of the SciFi silicon photomultipliers (SiPMs) can be measured on both recorded and simulated data. This is used to correct the simulated data to correctly simulate the single hit efficiency. This correction also simulates the excluded frontend boards, as the connected SiPMs consequently have a single hit efficiency of 0%. In comparison to other hit inefficiencies, the impact of the excluded SciFi frontend boards dominates the effect seen when correcting the simulation.

This can be seen in Fig. 9.4. The simulated data with corrected single hit efficiencies

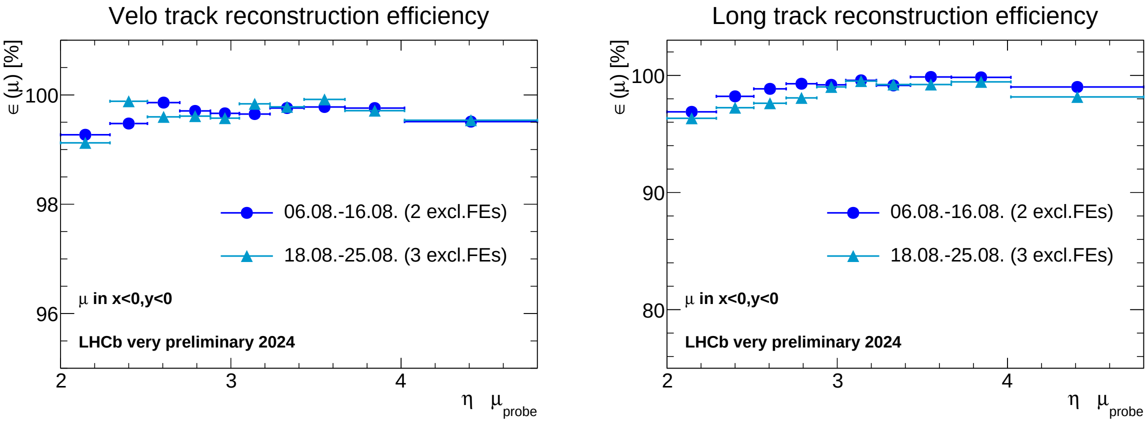


Figure 9.3: Track reconstruction efficiency in dependency of the probe track pseudorapidity, η, for probe tracks passing through the lower left quadrant of the muon system for data taken with two (dark blue) and three (light blue) excluded frontend boards for the Velo track reconstruction efficiency (left) and the long track reconstruction efficiency using the MuonUT method (right).

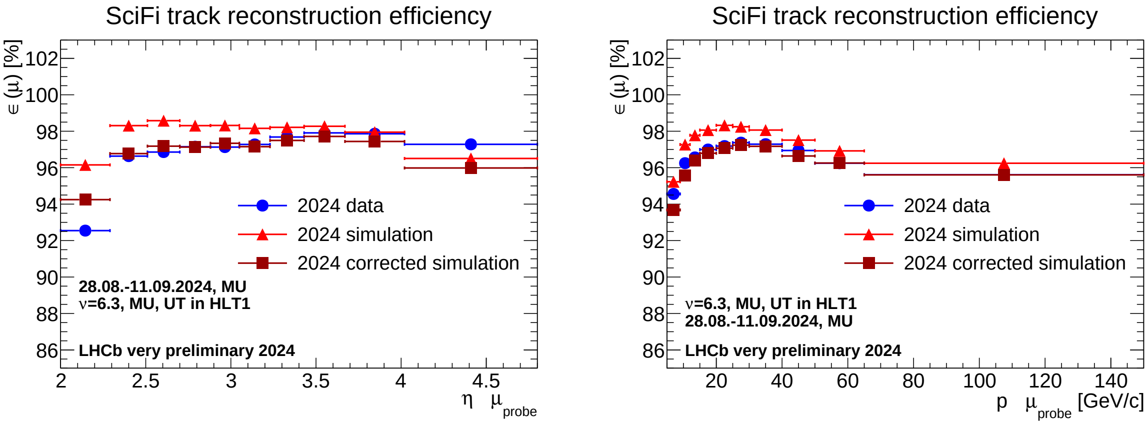


Figure 9.4: SciFi track reconstruction efficiency in dependency of the pseudorapidity (left) and momentum (right) in recorded data (blue) in comparison to simulated data both with (dark red) and without (light red) corrected single hit efficiency.

describes the recorded data more closely than without. The only exception is the highest pseudorapidity bin, where the simulated data continues to underestimate the efficiency, and does so worse with corrected single hit efficiency

9.3 Variations over the detector geometry

Equivalently to the SciFi track reconstruction efficiency in Fig. 9.1, the VELO and long track reconstruction efficiency can be checked in the regions of the detector. As can be seen in Fig. 9.5, the VELO track reconstruction efficiency is significantly worse in the lower pseudorapidity region on the $x > 0$ side of the detector than the $x < 0$ side. The axis range had to be shifted in comparison to other presentations of the VELO track reconstruction efficiency to accommodate the lower values.

This is, however, also predicted by simulation and likely due to a VELO module that is known to be excluded during data taking. Other than the excluded SciFi frontend boards, the excluded VELO module is also described in simulation.

The discrepancy is highest in the lowest pseudorapidity regions. This is visible in all four quadrants. The discrepancy is worse on the $x > 0$ side than the $x < 0$ side. This is likely because the discrepancy is proportional to the inefficiency. Thus, the variation of the VELO track reconstruction efficiency over the detector geometry does not explain any discrepancy observed between recorded and simulated data.

The effects seen in Fig. 9.1 and Fig. 9.5 can also be seen in Fig. 9.6. It shows the variation of the *long track reconstruction efficiency* from the *MuonUT method* over the detector regions. While the wider axis range hides the drop in the low pseudorapidity region in $x > 0$ a bit, it still exists, as does the clearly visible drop in the lower left corner affected by the excluded SciFi frontend boards.

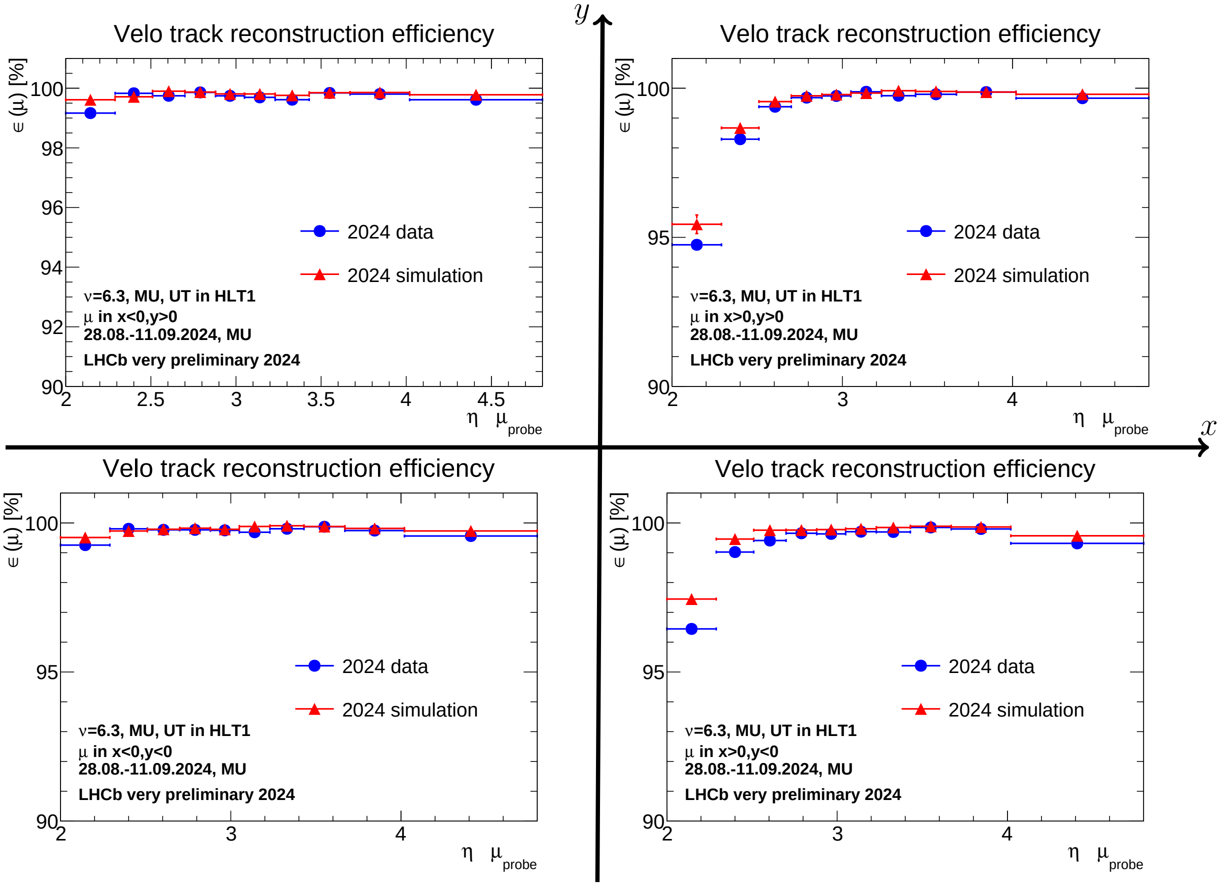


Figure 9.5: VELO track reconstruction efficiency in bins of pseudorapidity, η, for data set 2 recorded (blue) and simulated (red) data separately in the four quadrants of the detector.

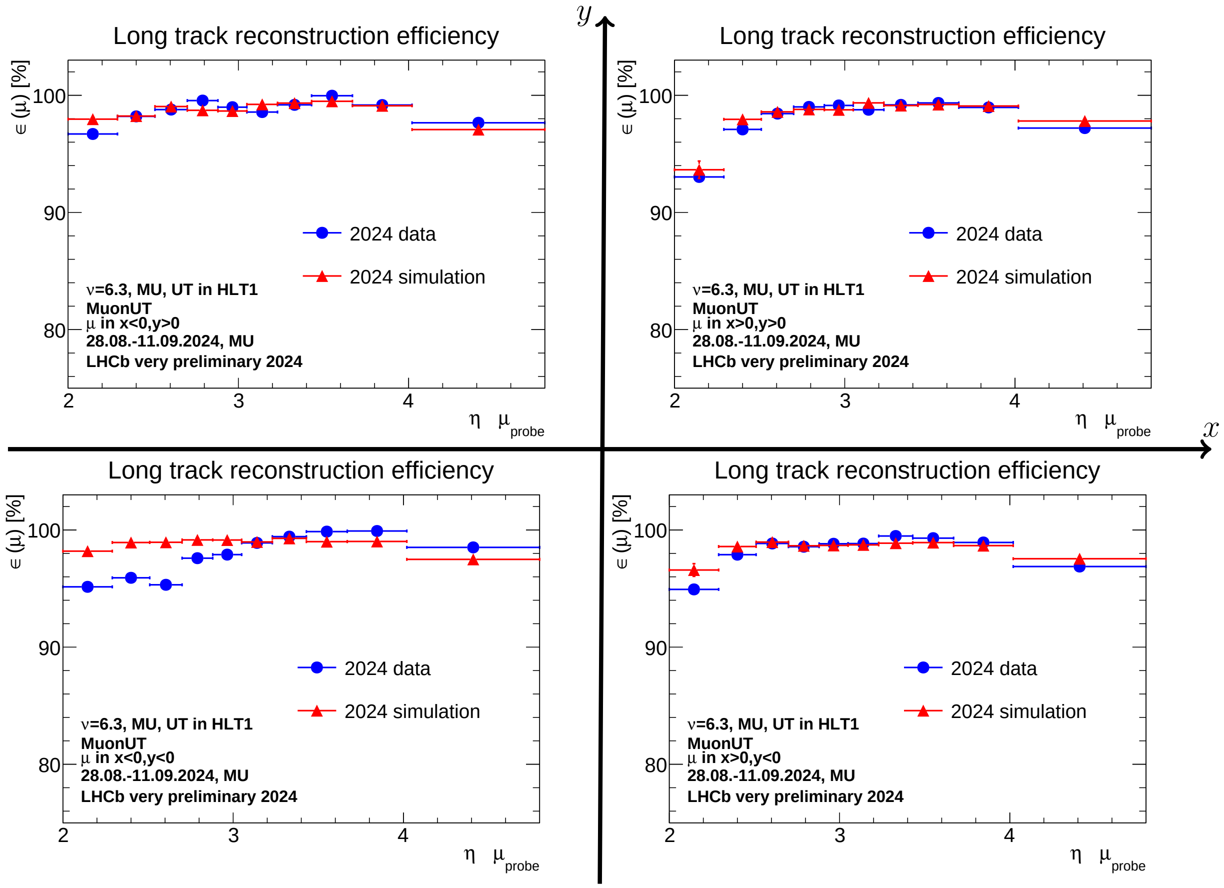


Figure 9.6: Long track reconstruction efficiency using the MuonUT method in bins of pseudorapidity, η, for data set 2 recorded (blue) and simulated (red) data separately in the four quadrants of the detector.

9.4 Trackχ^2 effects

In 2023, a major cause for discrepancies between recorded and simulated data is the distribution of the trackχ^2, which describes the quality of the fit to the track. This is partially caused by too tight cuts on the variable, and partially by a bug that assigned a too large trackχ^2 value to the candidates in recorded data, but not in simulation.

Both these effects are removed for 2024 data, and therefore the trackχ^2 is not anymore a significant source of contribution to the remaining discrepancies between recorded and simulated data. This is tested by *reWeighting* the simulated distribution to *sWeighted* data in bins of the probe and tag trackχ^2.

Fig. 9.7 shows the trackχ^2 and the pseudorapidity distribution *reWeighted* in trackχ^2 as an example for the *VeloMuon method*. Also shown is the effect of *reWeighting* in trackχ^2 on the *long track reconstruction efficiency* determined by combining the *Downstream* and the *VeloMuon* method. As can be seen, the effect of the *reWeighting* is negligible. It can be concluded that the trackχ^2-cuts do not contribute significantly to any discrepancies of the track reconstruction efficiency.

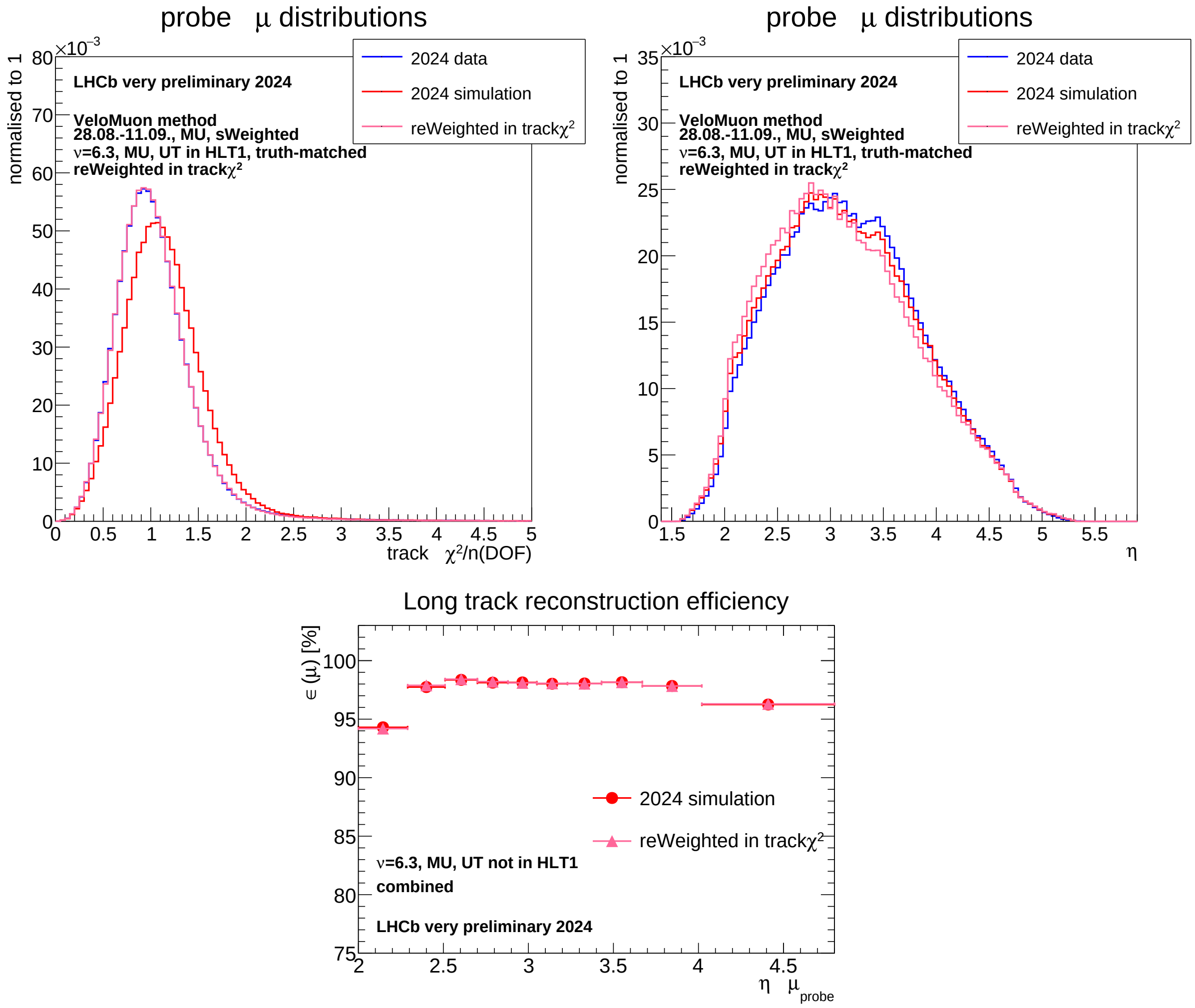


Figure 9.7: Reweighting in the tag and probe trackχ^2 distribution for data set 2: the reWeighted (pink) and not reWeighted (red) probe trackχ^2 distribution in comparison with recorded (blue) data (top left), the effect on the pseudorapidity distribution (top right) and the effect on the track reconstruction efficiency in pseudorapidity bins (below) for the combined VELO and SciFi efficiency as an example.

9.5 Test of double-counting effects using the Ghost Probability cut

While the VELO and the SciFi track reconstruction efficiency are supposed to only describe the track reconstruction efficiency of the respective subdetectors, their can be edge effects affecting both. This can happen for example if an issue in the SciFi makes the reconstruction of a long track passing through a section harder, correcting the momentum estimate in the wrong direction, such that it is still possible to reconstruct a *Downstream probe track*, but not possible anymore to find the correct VELO segment to reconstruct the *long muon track*.

This is an issue for the *combined method*. It might mean that an effect appears twice for the *combined method*, as it influences both the *Downstream* and the *VeloMuon* method. At the same time, the same effect will only influence the *MuonUT method* once. Hints for this can be seen in Fig. 8.6, where in the pseudorapidity the combined efficiency seems to be the square of the *long track reconstruction efficiency* obtained via the *MuonUT method*. To test whether discrepancies between the *combined* and the *MuonUT method* are effects of double-counting, the upper limit on the *GhostProbability* of the *long muon track* as well as the *tag muon track* is tightened. The impact of the limit on the *tag muon track* is expected to contribute little to the overall effect. A tighter limit is expected to lower the *track reconstruction efficiency*, as less *long muon tracks* pass the selection. If the *VeloMuon* and *Downstream* methods are uncorrelated, the *combined method* should be affected by the tightening of the limit the same way as the *MuonUT method*.

The result of this can be seen for the *long track reconstruction efficiency* estimated via the combined *Downstream* and *VeloMuon* methods in comparison to the *MuonUT* method in Fig. 9.8. It is clearly visible that the efficiency in pseudorapidity decreases more in dependence of the *Ghost Probability cut* for the *combined method* than the *MuonUT method*, meaning some effects are probably double-counted. This would make the *MuonUT method* more reliable than the *combined method*. However, it is at this point not yet known what is double-counted. The known bugs in the *MuonUT method* as well as its limited statistics in comparison to the other two methods still lead to a preference of the combined method.

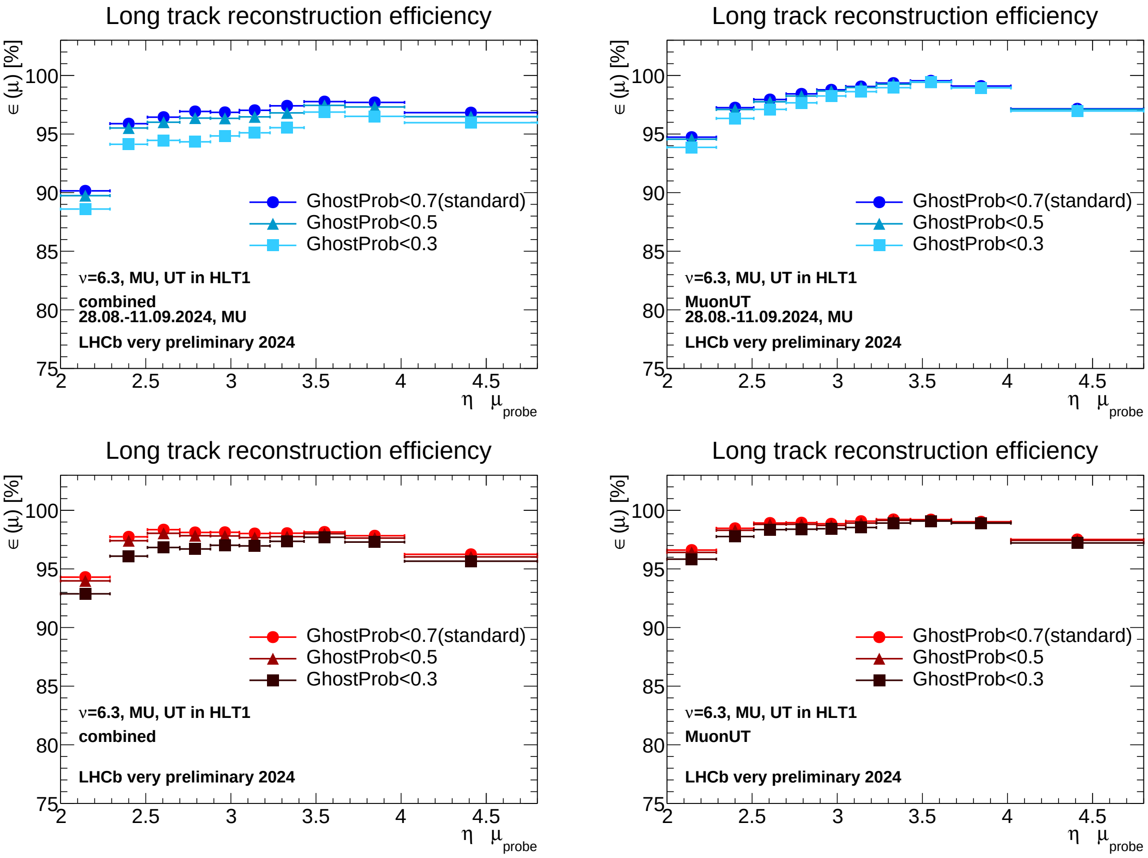


Figure 9.8: Change of the track reconstruction efficiency in data set 2 for recorded (top) and simulated (bottom) data for the combined method (left) and the MuonUT method (right) while tightening the cut on the ghost probability variables.

9.6 Charge of the probe muon

Track reconstruction efficiencies are determined without separating for the charge of the probe muon. To verify that the charge does not affect the track reconstruction efficiency, the samples are separated by the charge of the probe muon. The result for *data set 2* can be found in Fig. 9.9 in bins of momentum, and in Fig. 9.10 in bins of pseudorapidity.

Several things are noteworthy here. As the axis ranges are different for the different methods, the smaller plots below the main plots show the absolute discrepancy between the two charges.

The VELO track reconstruction efficiencies agree very well between the two charges, except for the lowest momentum bin, which has a significant discrepancy between the charges both in recorded and simulated data. As this is the bin that also has a significantly higher discrepancy due to the fitmodel, it can be assumed that this bin is unreliable and should not be included.

The SciFi track reconstruction efficiencies follow the same pattern for both charges but

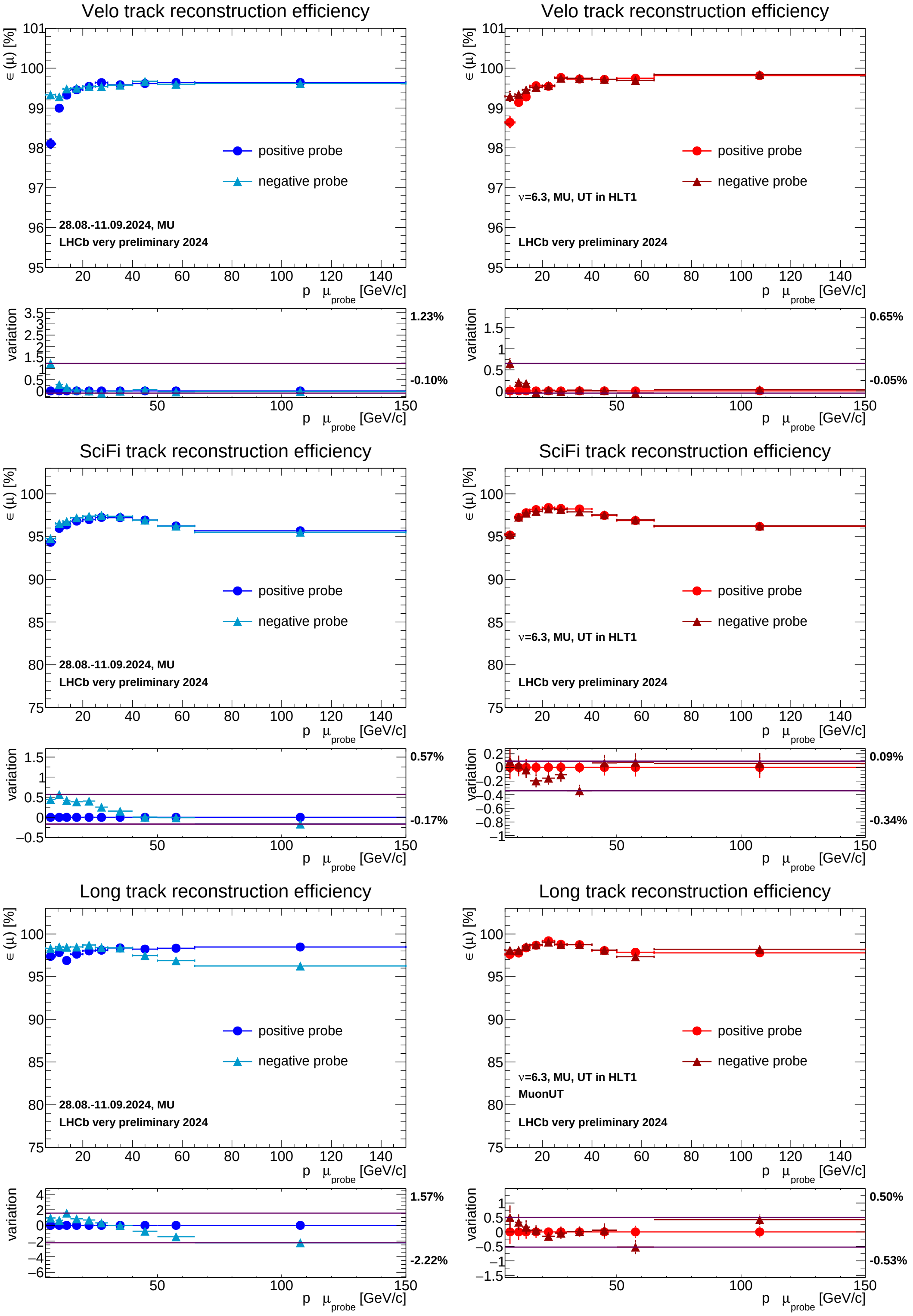


Figure 9.9: Track reconstruction efficiency in bins of momentum, p, separated by positive (dark blue, light red) and negative (light blue, dark red) probe muon charge in recorded (left, blue) and simulated (right, red) data for the Downstream method (top), the VeloMuon method (middle) and the MuonUT method (below).

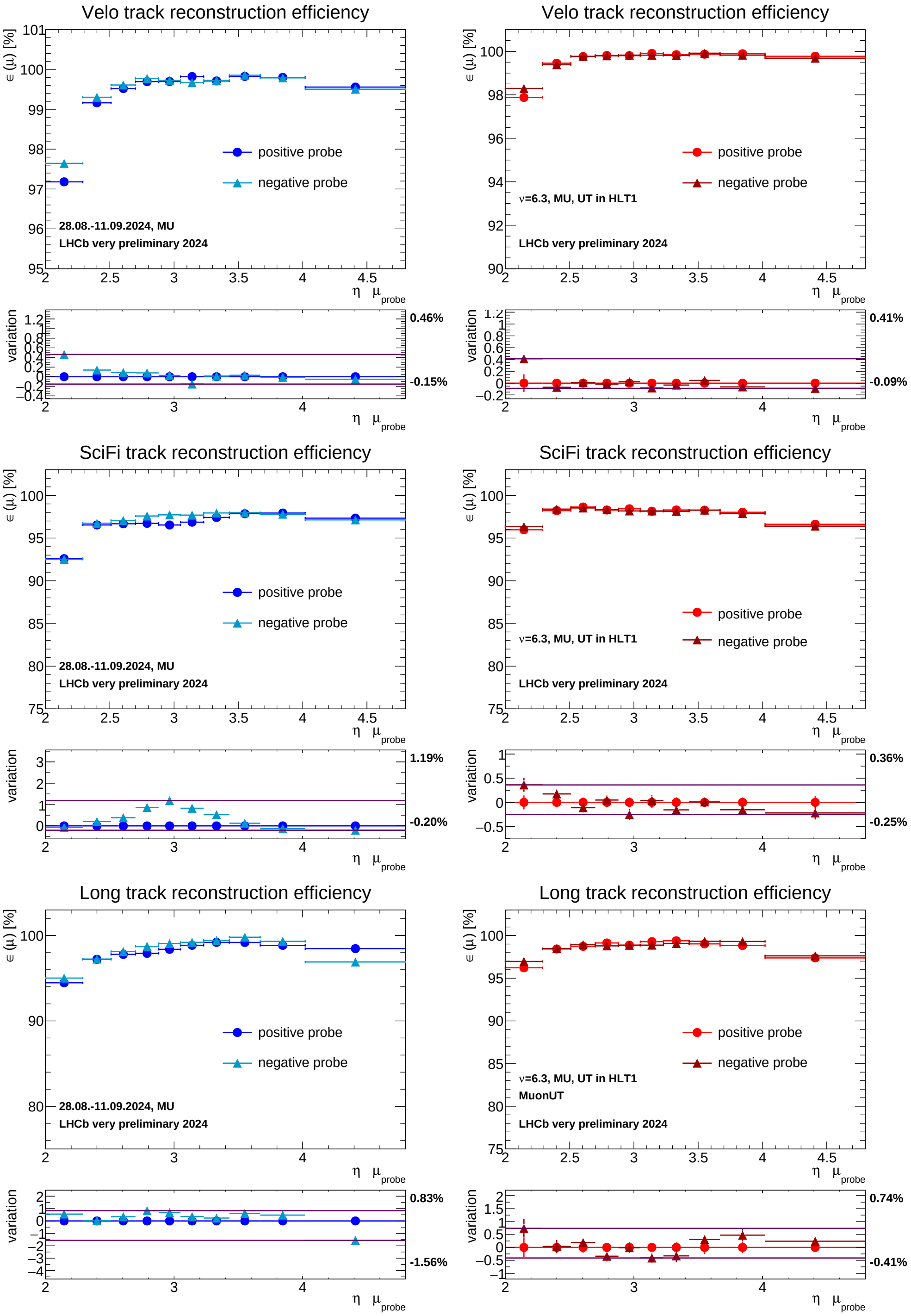


Figure 9.10: Track reconstruction efficiency in bins of pseudorapidity, η, separated by probe muon charge in data (left) and simulation (right) for the Downstream method (top), the VeloMuon method (middle) and the MuonUT method (below).

vary by more than the usual systematic uncertainty. In pseudorapidity, the region affected by the excluded SciFi frontend boards varies markedly between the two charges. This is likely fluctuation, as this region is subject to less statistics. However, it might also be caused by a charge disparity that needs further looking into.

For the *long track reconstruction efficiency* using the *MuonUT method*, there is a significant shift between the two charges which seems to follow a systematic pattern in recorded data.

The differences between the two charges seen for the SciFi and long track reconstruction efficiencies might be further cause for discrepancies as they are larger than the assigned systematic uncertainties. However, the effect might also be random and caused by the lower statistics, as the samples had to be halved for the two probe muon charges. Additionally, effects are generally seen in recorded and simulated data and will therefore likely cancel out to a degree in the ratio.

However, the *MuonUT method* does neither follow the same pattern for both charges, nor show the same discrepancies due to the charge in recorded and simulated data. This needs to be considered further.

The charge of the probe track, which so far has been neglected, could potentially be a source of discrepancy for the *track reconstruction efficiency.*

9.7 Other reasons for discrepancies

While the biggest source of discrepancy between recorded and simulated data in 2024 are the bugs for the *MuonUT method* and otherwise the exclusion of the SciFi frontend boards, discrepancies remain that cannot be explained by those effects. These are, however small. They are most likely dominated by imperfect alignment and imprecisions in the description of the simulation. Simulations better describing the data taken in 2024, including the excluded SciFi frontend boards, will be produced for analyses using 2024 data. In the meantime, the recorded/simulated data ratios of the efficiencies are used as correction factors to correct for these discrepancies.

Chapter 10

Systematic studies of the method

This chapter gives an overview over different checks that have been done to verify the method. A closure test is performed on simulated events to verify that the determined track reconstruction efficiencies agree with two different methods. It is verified that the matching criterion to match the probe muon track and the long muon track determines tracks reconstructed from the same passing muon in the detector with high accuracy. Furthermore, it is verified that the track reconstruction efficiency determined on a data set with magnet up polarity can be used as well for any data set with magnet down polarity by comparing the track reconstruction efficiencies for both polarities.

10.1 Verification of the tag-and-probe method using the hit-and-count method

The *tag-and-probe method* is used to determine the *track reconstruction efficiency* of the LHCb tracking subdetectors. The functionality of the method can be verified by exploiting the knowledge of both the originally generated particles and the final reconstructed tracks in simulated events.

The muon tracks are reconstructed directly from generated simulation of $J/\psi \to \mu^+\mu^-$ events without reconstructing the full event information. Both the probe muon track and the long muon track can be directly linked to the generated particle they were reconstructed from. Thus it is insured that they come from the same particle. It can additionally be insured that this particle is a J/ψ decaying into two muons. The other J/ψ daughter is identified as the tag muon track.

The overlap of hits in the reconstructed long track and the reconstructed probe track is counted for each subdetector. An overlap of at least 40% in the shared subdetectors is defined as a successful match. It is then determined how many of the reconstructed candidates contain successful matches. As all tracks are matched to their respective generated particles, there is not background. This is called the *hit-and-count method*. It can only be applied on simulated data and is used as a verification to the *tag-and-probe method*.

As only the single tracks and not the full J/ψ decay is reconstructed in the case of the *hit-and count method*, not all information is available and thus not all selection criteria that are required for the reconstruction using the *tag-and-probe method* can be applied in the same way in the *hit-and-count method*. However, these selection criteria are designed to remove background, which is non-existent in the *hit-and-count method*. Similarly, the events do not pass either HLT1 nor HLT2 and are not subject to its configuration.

Like the *tag-and-probe method*, the efficiency can be measured with the *hit-and-count method* in bins of momentum and pseudorapidity. The results are shown in Fig. 10.1 for *data set 2 simulation* in comparison to the *tag-and-probe method* used elsewhere in this thesis. As there is no background and neither HLT1 nor HLT2 are run for this reconstruction, the two datasets do not differ for the *hit-and-count method* and are here compared to track reconstruction efficiencies from the *tag-and-probe method* for *data set 2*.

As can be seen, the agreement is very good, especially for the SciFi track reconstruction efficiency. Unfortunately, the produced sample for the *MuonUT* sample is statistically limited such that there are big fluctuations. Nevertheless, the *hit-and-count method* provides a good confirmation for the *tag-and-probe method*

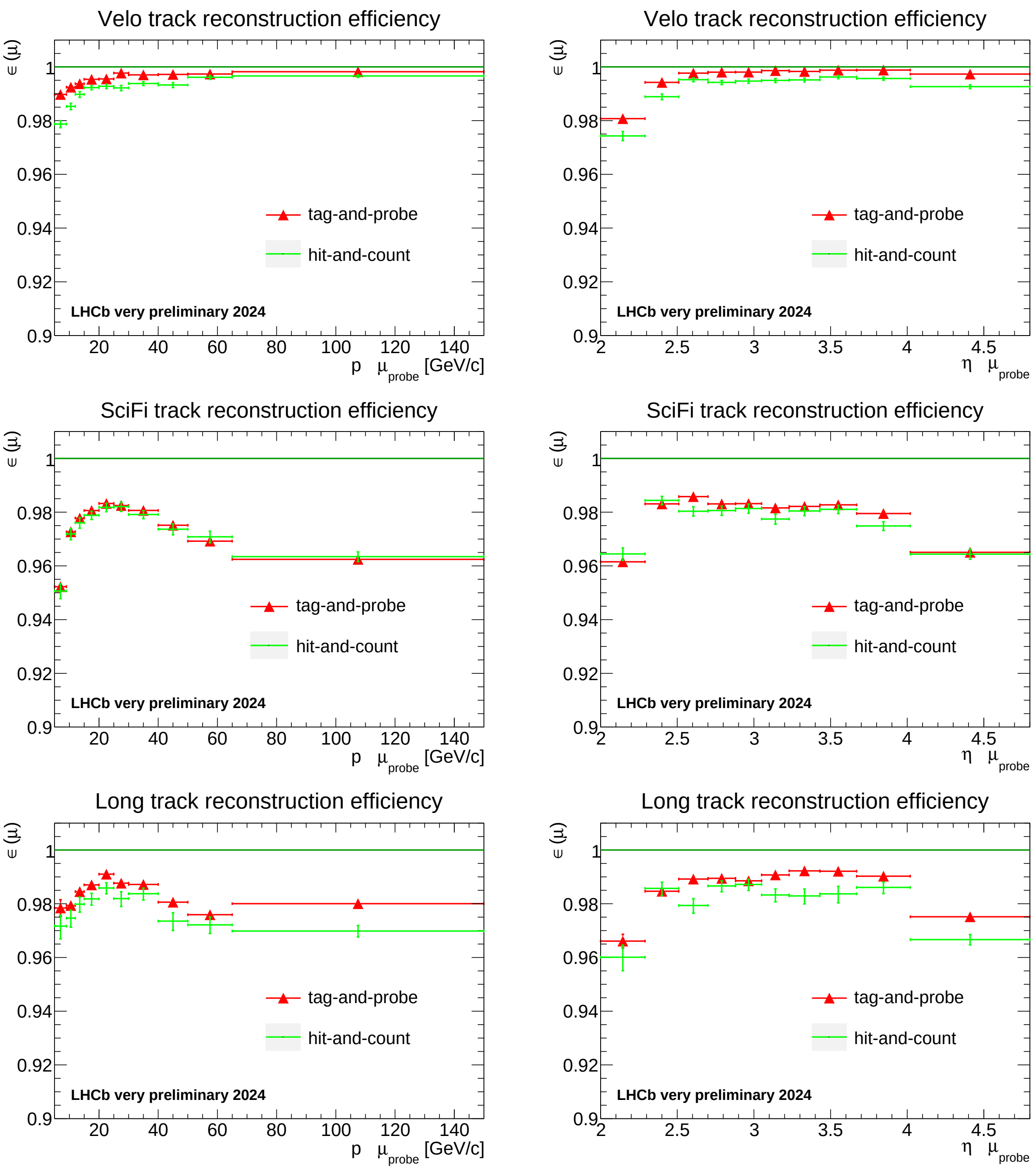


Figure 10.1: Closure test performed on data set 2 Monte Carlo: efficiencies determined with the probe-and-tag method (red) in comparison with the efficiencies determined with closure method (green) for the Downstream method, the VeloMuon method, the MuonUT method and the SeedMuon method in bins of momentum (left) and pseudorapidity (right).

10.2 Investigation of the matching criterion

The track reconstruction efficiency is determined by attempting to match the partially reconstructed probe and the fully reconstructed long muon track. Vital for the procedure is full confidence that the probe track has been reconstructed from the same passing muon as the long muon track and does not belong to a different particle. In order to ensure that, an overlap between the two tracks of at least 40% of their hits in the muon system is required. Additionally, in the case of the *Velo track reconstruction efficiency*, an overlap of 40% of the hits in the SciFi is required for both the *Downstream method* used in 2024 and the *SeedMuon method* used in 2023. In case of the *SciFi track reconstruction efficiency*, an overlap of 40% of the hits in the VELO is required. Additionally an overlap in the UT is checked, but not required, for the *Velo track reconstruction efficiency* using the *Downstream method* as well as the *long track reconstruction efficiency* using the *MuonUT method*. It has to be verified that this is enough to guarantee the identity of the tracks.

However, since the probe and the long track are reconstructed with different algorithms, slightly different hits in the detector might have contributed to each. This is especially a risk in the case of the *VeloMuon*, *MuonUT* and *SeedMuon* probe track, respectively used for the determination of the *SciFi track reconstruction efficiency*, *Long track reconstruction efficiency* and *Velo track reconstruction efficiency without UT information* using the *SeedMuon method*, which base their reconstruction first on a track segment in the muon system and start the reconstruction from there. The long track reconstruction and *Downstream* probe reconstruction on the other hand reconstruct the track fully in the other tracking subdetectors and only add muon hits to the reconstructed track if they surround its trajectory in the muon system.

It is therefore necessary to check that the defined overlap criterion actually ensures the correct matching muon. This is done with two tests.

The momentum of the *long muon track* and the *probe muon track* are not expected to be exactly identical due to the different reconstruction algorithms. Additionally, the *long track information* is not saved directly with the current setup, such that the search for the *long track* afterwards might not yield the track the *probe track* was actually matched too. However, this is assumed to not be the case too often.

The momentum resolution of SciFi tracks is significantly worse than the momentum resolution of fully reconstructed *long tracks*. Nevertheless, they are expected to be similar, as they both ideally are reconstructed from the same muon. In the first test, the momenta of the reconstructed *probe muon* and the reconstructed *long track* matched to it are compared for all candidates where the match was successful.

$$\delta_p = \frac{p_{long} - p_{probe}}{p_{long}} \tag{10.1}$$

This is compared for tracks from *set 2* of 2024 data for samples from all three methods. This comparison is shown and fitted with a simple Gaussian distribution for *sWeighted data set 2 candidates* in Fig. 10.2. As can be seen, the resulting distribution is Gaussian, centred around 0 and has a width of less than 0.025 for the *VeloMuon* and *MuonUT probe* and less than 0.002 for the *Downstream probe.* The *Downstream probe* distributions seems asymmetrical. This needs further investigation.

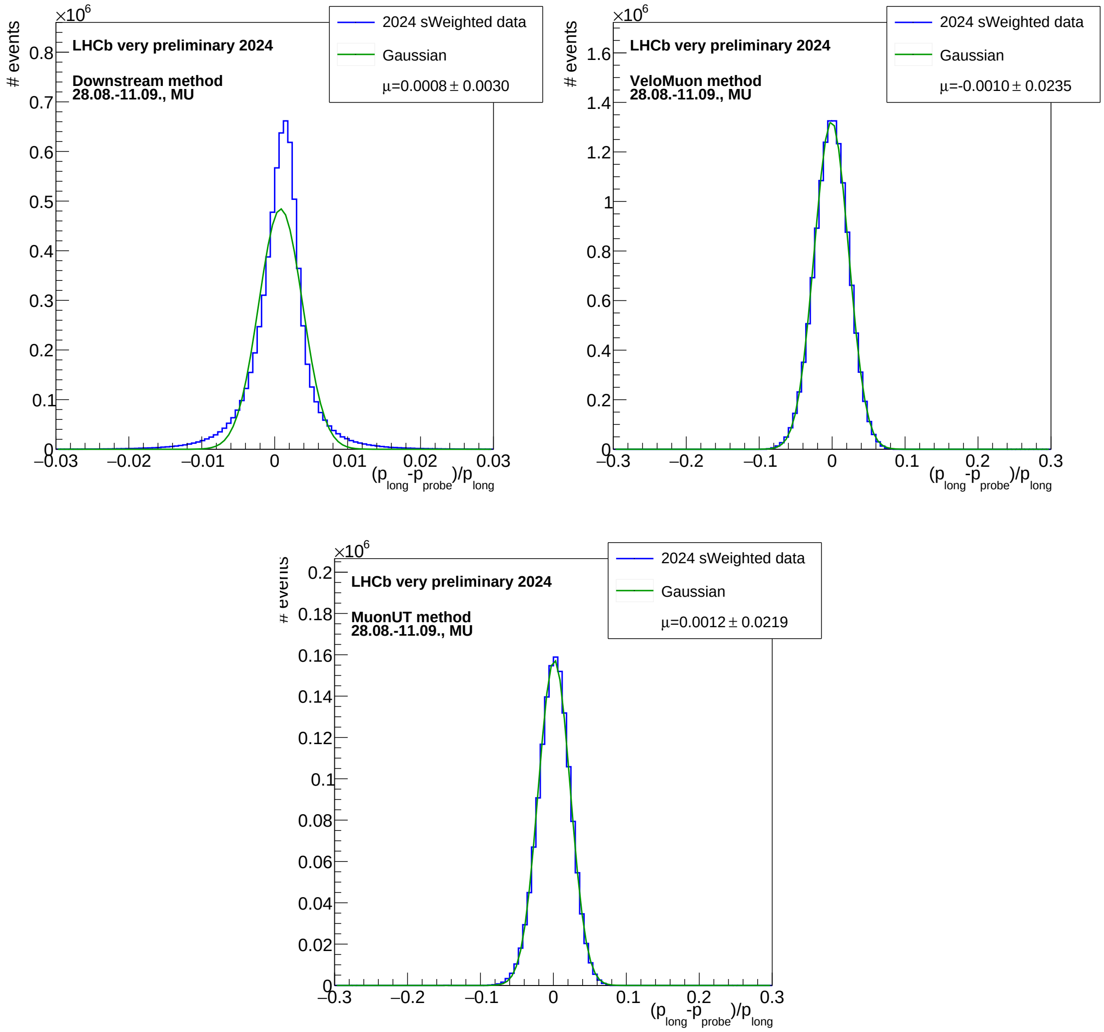


Figure 10.2: sWeighted distribution (blue) of the relative difference of the reconstructed momentum of the long and the probe track in recorded data for the Downstream line (top left), the VeloMuon line (top right) and the MuonUT line (bottom) fitted with a simple Gaussian (green).

To ensure that the correct *long muon track* has been chosen for the comparison, its radial distance ΔR to the *probe muon track* is tested. This is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{10.2}$$

using the difference in pseudorapidity, $\Delta\eta$, and in angle, $\Delta\phi$, between the two tracks. This is plotted against the relative momentum difference in Fig. 10.3. As can be seen, the distribution for Downstream probe tracks is shifted away from the centre and slightly asymmetric. However, all three plots show a regular distribution with no significant outliers, confirming that the correct *long muon track* has been chosen in the majority of the cases.

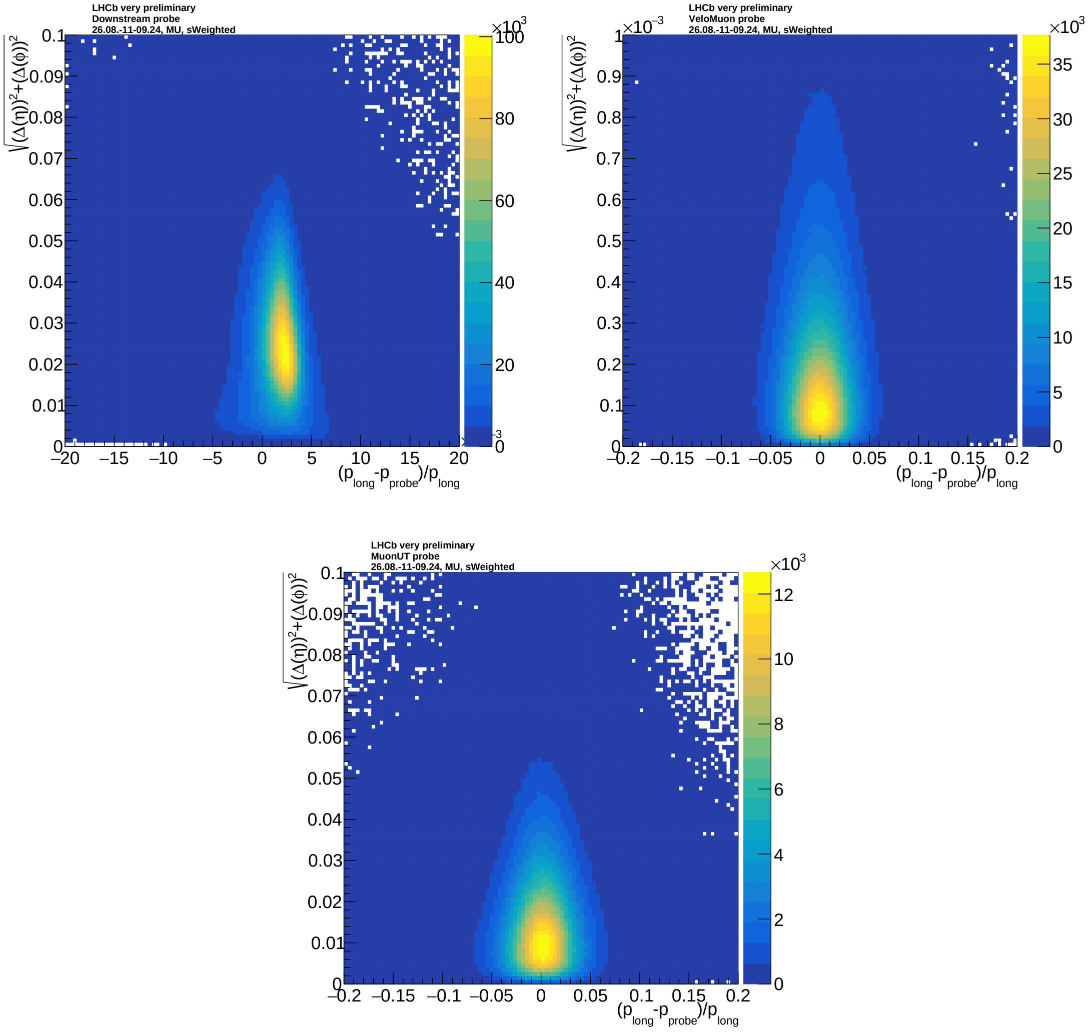


Figure 10.3: sWeighted distribution of the cone-wise distance, $\Delta(R)$, versus the relative momentum difference between the probe and the long track for Downstream probe tracks (top left), VeloMuon probe tracks (top right) and MuonUT probe tracks (bottom).

As a second test, the yield of J/ψ mother particles reconstructed in the matched case from the tag muon and the probe muon is compared with the yield of J/ψ mother particles reconstructed in the same case from the long muon and the tag muon. Should the probe track and the long track not be reconstructed from the same muon, they could not both come from the same J/ψ mother as the tag muon. Therefore, the yield of reconstructed J/ψ would differ significantly. As an example, the fit to the J/ψ invariant mass reconstructed from long and tag versus the fit to the J/ψ invariant mass reconstructed from probe and tag in reconstructed data can be seen in Fig. 10.4. They are compared in Tab. 10.1 together with their relative difference:

$$\delta_N = abs\left(\frac{N_{J/\psi=probe+tag} - N_{J/\psi=long+tag}}{\sqrt{\sigma^2_{J/\psi=probe+tag} + \sigma^2_{J/\psi=long+tag}}}\right) \tag{10.3}$$

As can be seen, while the numbers are always of the same order of magnitude, the deviation can be quite significant. However, this does not necessarily mean that the matching criterion is wrong. As the only information stored in the final data samples is whether a probe track was matched to a long track and not which long track it was matched too, it is possible that the wrong long track is chosen later without invalidating the match itself. Additionally, the reconstructed J/ψ from probe and tag have a much higher background and broader width than those from long and tag, as the probe track has a much worse momentum resolution than the long track. Under these circumstances, the given numbers are a good confirmation that in most cases the probe and the long track both describe the same passing particle. The variation seen between the two J/ψ

	$N_{J/\psi=probe+tag}$	$N_{J/\psi=long+tag}$	δ_N
Downstream data	7085047 ± 5835	6970281 ± 5577	14.22
Downstream MC	309549 ± 573	301125 ± 566	10.46
VeloMuon data	12974553 ± 4529	12554951 ± 4079	68.84
VeloMuon MC	497278 ± 720	487065 ± 709	10.11
MuonUT data	1430107 ± 2480	1434689 ± 1577	1.56
MuonUT MC	93880 ± 322	93500 ± 313	0.85

Table 10.1: Yields of reconstructed J/ψ candidates from the long and tag muon track in comparison with the yields of reconstructed J/ψ andidates from the probe and tag muon track in 2024 data set 2 samples as fitted in Fig. 10.4 as an example.

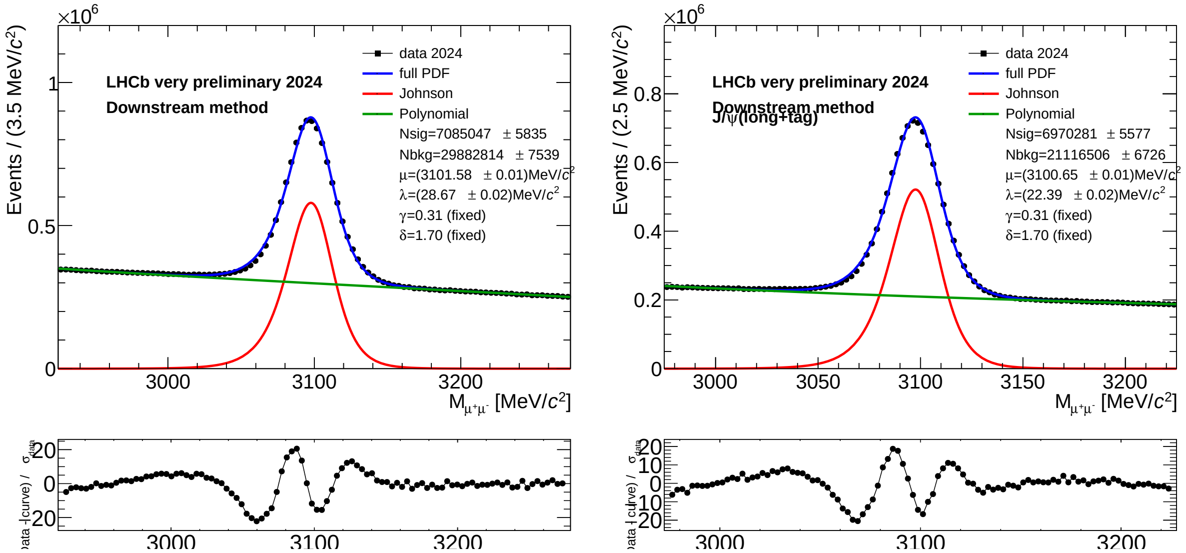


Figure 10.4: Fit to the probe muon track and tag muon track invariant mass (left) and the long muon track and tag muon track invariant mass (right) spectra of the matched sample in comparison to determine the signal yield for the Downstream sample of data set 2.

mass distribution is not large enough to be equivalent to a track reconstruction efficiency deviation of more than one standard deviation.

In conclusion, the matching of probe and long track is reliable.

10.3 Independence from the magnet polarity

All data used in this thesis was taken with *Magnet Up* polarity in the detector. The track reconstruction efficiency is not expected to depend on the sign of the magnet polarity. Nevertheless, this needs to be verified. A data set taken from the 11.09.2024-02.10.2024 with the UT used in the HLT1 reconstruction sequence and *Magnet Down* polarity is compared with *data set 2*. The comparison can be found in Fig. 10.5.

As can be seen, the efficiencies follow the same distribution and are very similar. The exception is again the first momentum bin of the *Velo track reconstruction efficiency*, which is already generally assumed to be problematic as it shows fluctuating behaviour in almost all tests. It can be assumed that the *track reconstruction efficiencies* calculated on data of one polarity can also be used for an equivalent data set of the opposite polarity.

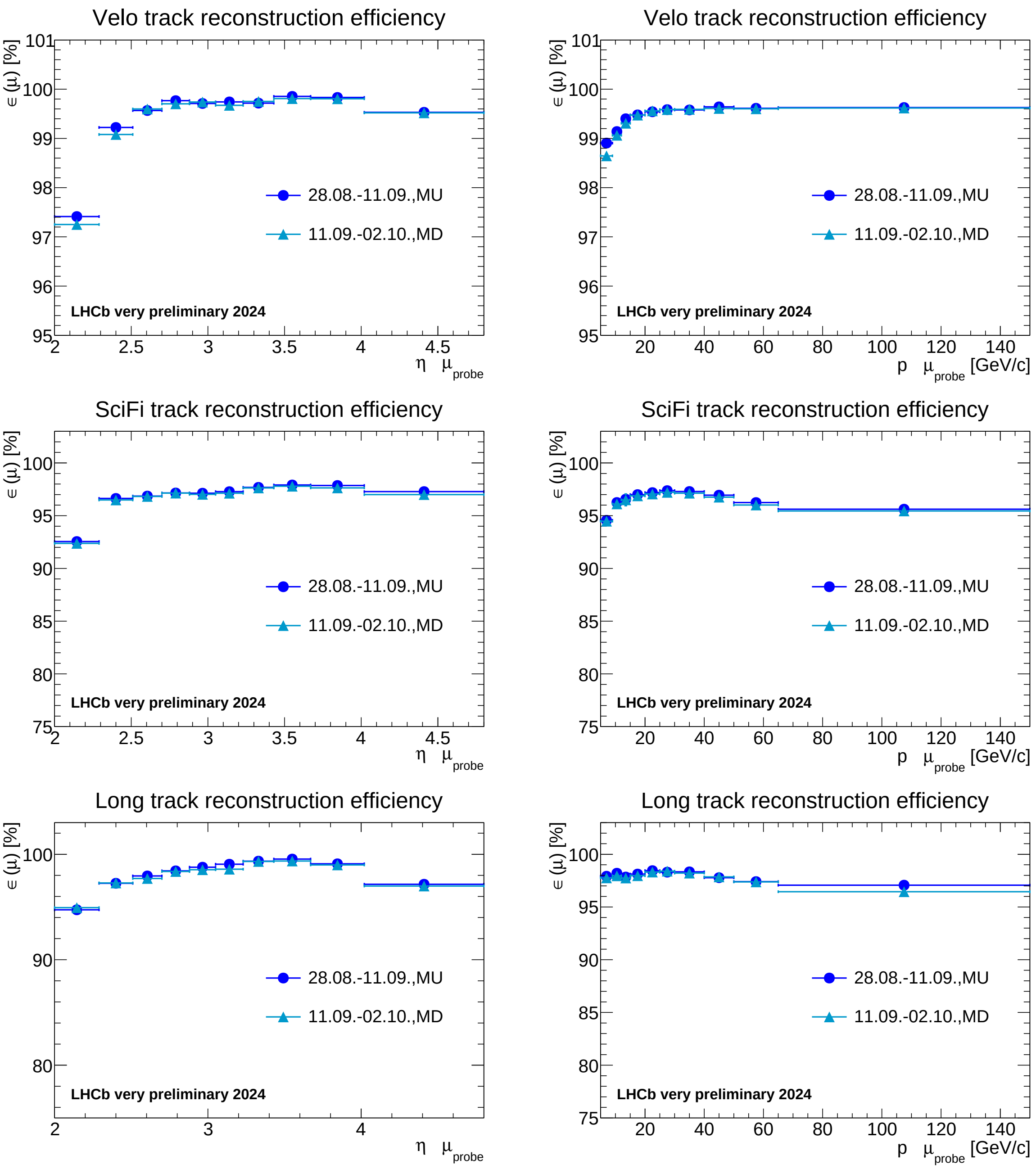


Figure 10.5: Track recconstruction efficiency determined via the Downstream (up), Velo-Muon (middle) and MuonUT (below) methods in bins of pseudorapidity, η, (left) and momentum, p, (right) in comparison between data set 2 (dark blue) and a comparable data set with Magnet Down polarity (light blue).

Chapter 11

Correction for hadronic interactions

Track reconstruction efficiency corrections discussed in the previous chapters are determined using muons and do not consider corrections specific to hadrons. This chapter approaches this issue. Pions from $D^0 \to K^-\pi^+$ *and* $D^0 \to K^-\pi^-\pi^+\pi^+$ *are used to test the agreement of recorded and simulated data concerning the interactions of hadrons with detector material. The differences are studied in bins of momentum and pseudorapidity and the corresponding correction factors are estimated.*

11.1 Method overview

Hadrons such as kaons and pions can be detected in the LHCb detector. However, in addition to energy losses due to ionisation, they can also be stopped in the detector due to hadronic interactions with the detector material. This can occur both with the active detector material and with passive material such as scaffolding, cables and pipes. According to previous simulations an estimated 12% of kaons and 21% of all pions from b and c meson decays cannot be reconstructed due to their hadronic interaction with the detector material [36].

While hadronic interactions are included in the detector simulation, the knowledge of the amount and distribution of the detector material is subject to uncertainty. This chapter aims to determine how accurate the description of the detector material and its interactions with passing hadronic tracks is. Correction factors to the track reconstruction efficiency necessary for hadrons are determined to mitigate the difference between recorded and simulated data.

Since hadrons interact with material more frequently than muons, they have a lower track reconstruction efficiency:

$$\epsilon_{hadron}^{reco} = \epsilon_{muon}^{reco} \times \rho_{hadron}^{hadInt}, \tag{11.1}$$

where ϵ_{reco}^{muon} is the track reconstruction efficiency of muons as calculated previously, and ρ_{hadron}^{hadInt} the correction factor due to hadronic interactions in the detector. It depends on the amount and distribution of the detector material the hadron passes through.

The correction factor ρ_{hadron}^{hadInt} can be determined for pions coming from D^0 decays. To cancel out the D^0 production cross section, the decays $D^0 \to K^-\pi^-\pi^+\pi^+$ and $D^0 \to K^-\pi^+$ are compared. The correction factor for two pions can then be defined equivalently as:

$$\rho_{\pi\pi}^{hadInt} = \frac{\epsilon_{\pi\pi}^{reco}}{\epsilon_{\mu\mu(\pi\pi)}^{reco}}, \tag{11.2}$$

where $\epsilon_{\pi\pi}^{reco}$ is the full track reconstruction efficiency of two pions and $\epsilon_{\mu\mu(\pi\pi)}^{reco}$ the track reconstruction efficiency of two muons with the kinematic properties of the pions, therefore describing the track reconstruction efficiency of the two pions if they do not undergo hadronic interaction.

The track reconstruction efficiency determined using simulated data needs to be corrected by the factor $\rho_{\pi\pi}^{hadInt,data/MC}$, which is derived from comparing the hadronic correction for recorded and simulated data:

$$\rho_{\pi\pi}^{hadInt,data/MC} = \frac{\rho_{\pi\pi}^{hadInt,data}}{\rho_{\pi\pi}^{hadInt,MC}} = \frac{\epsilon_{\pi\pi}^{reco,data}}{\epsilon_{\mu\mu(\pi\pi)}^{reco,data}} \times \frac{\epsilon_{\mu\mu(\pi\pi)}^{reco,MC}}{\epsilon_{\pi\pi}^{reco,MC}} \tag{11.3}$$

This is necessary in addition to the correction factors determined using muons in the previous chapters.

11.2 Estimating the influence of hadronic interaction on the track reconstruction efficiency

A first estimate of the hadronic correction for pions can be reached by comparing the SciFi track reconstruction efficiency for muons and pions. Estimates of the track reconstruction efficiency of pions are determined by other members of the LHCb collaboration, using a similar *tag-and-probe method* as is used in this thesis for muons.

The decay of a K_S^0 into two pions is used, where one pion is fully reconstructed, while the other is only reconstructed as a VELO track. The VELO track is then matched to a fully reconstructed pion to test the SciFi track reconstruction efficiency. This method has not been developed for the track reconstruction efficiency of the VELO.

A comparison of the track reconstruction efficiency of muons and pions in the SciFi in bins of momentum and pseudorapidity gives a rough estimate of how large the factor ρ_{hadron}^{hadInt} is for all hadronic interactions occuring after the VELO. The SciFi track reconstruction efficiency for muons as shown already in Fig. 8.4 is presented in comparison to the SciFi track reconstruction efficiencies for pions in Fig 11.1. The numbers for the track reconstruction efficiency of pions have been provided by other members of the LHCb collaboration.

The track reconstruction efficiency for pions draws from a much shorter time period of *data set 1* as it uses decays with a much higher abundance in LHCb. The hadronic correction factor ρ_{pion}^{hadInt} is shown in the same plot. It has values in a range from 74% to 88%, meaning that between 12% and 26% of pions are lost after the VELO before reconstruction in the SciFi. The 21% mentioned above, lost over the entire reconstruction to the VELO +SciFi track reconstruction efficiency in simulated data, are within this range.

The pion/muon track reconstruction efficiency ratios shown in Fig. 11.1 are determined using recorded data samples. The same procedure has to be repeated on simulated data to compare them to hadronic correction factors in simulation and determine the correction factor $\rho_{pions}^{hadInt,data/MC}$. This correction factor gives an estimate of how well simulated data described the hadronic interaction, and therefore the material of the detector. It can be used to correct track reconstruction efficiencies for hadrons additionally to the ratios from the previous chapters.

However, this is a very rough estimate, as it does not take into account the VELO track reconstruction efficiency, as well as differences between the muon and pion samples compared. A more sophisticated method needs to be developed.

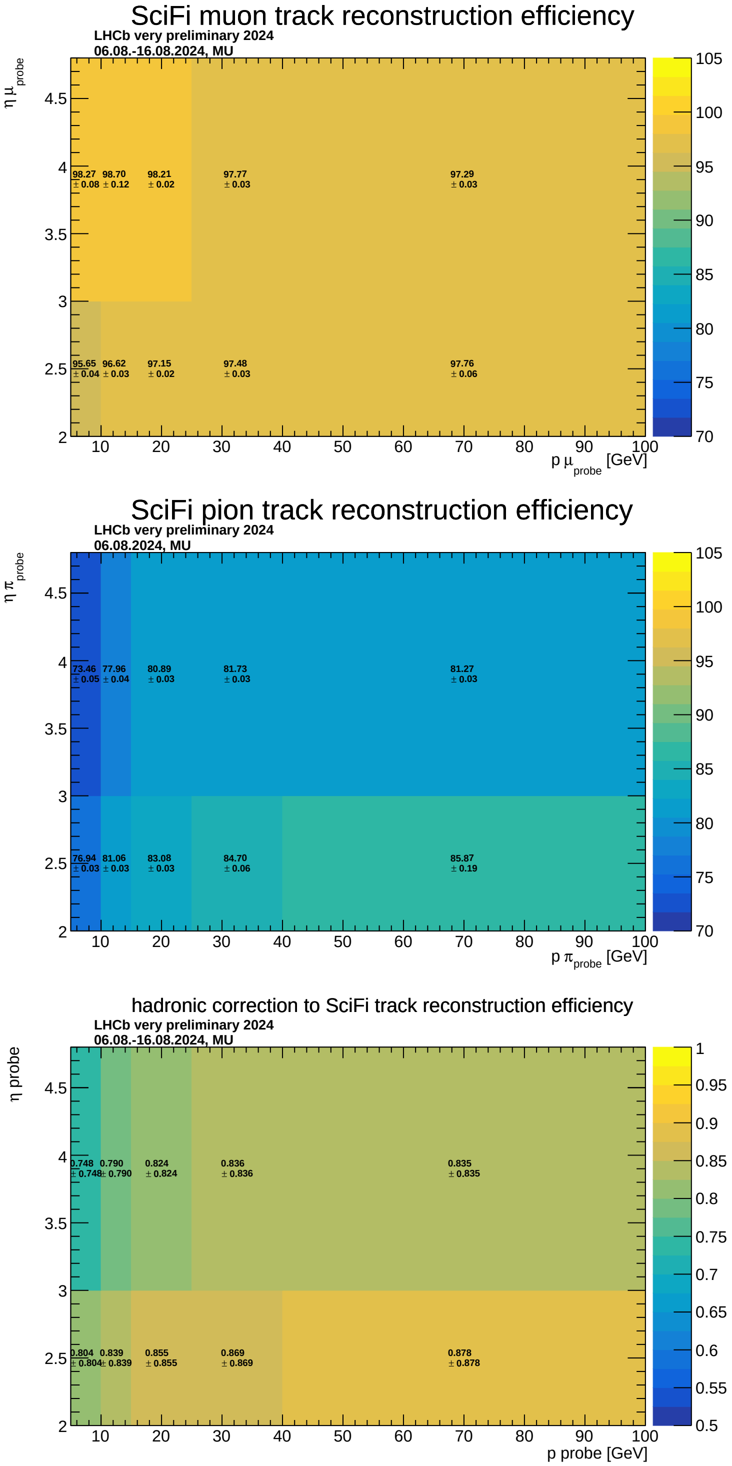


Figure 11.1: Muon (top) and pion (middle) SciFi track reconstruction efficiencies in percent in comparison, as well as pion/muon SciFi track reconstruction efficiency ratios not given in percent determined by dividing pion by muon track reconstruction efficiencies (bottom). Numbers for the pion track reconstruction efficiency have been provided by other members of the LHCb collaboration.

11.3 Using the decays $D^0 \rightarrow K^-\pi^+$ and $D^0 \rightarrow K^-\pi^-\pi^+\pi^+$ to estimate the hadronic interaction correction for two pions

The track reconstruction efficiency of two pions, $\rho_{\pi\pi}^{reco}$, can be extracted from the decay process $D^0 \rightarrow K^-\pi^-\pi^+\pi^+$. As a first step, a sample of $D^0 \rightarrow K^-\pi^-\pi^+\pi^+$ decays is reconstructed, selected and fitted in both data and simulation to gain the signal yield. The signal yield of a decay channel in simulated data, $N_{K\pi\pi\pi}^{yield,MC}$, is equal to the number of originally generated events, $N_{K\pi\pi\pi}^{gen,MC}$ multiplied with all efficiencies of selection and reconstruction, $\epsilon_{K\pi\pi\pi}^{MC}$:

$$N_{K\pi\pi\pi}^{yield,MC} = N_{K\pi\pi\pi}^{gen,MC} \times \epsilon_{K\pi\pi\pi}^{MC} \tag{11.4}$$

In case of recorded data, the number of original events can not be counted, but calculated from the production cross section of the mother particle, σ_{D^0}, the integrated luminosity corresponding to the collected sample, $\mathcal{L}_{int}$, and the branching fraction $\mathcal{B}(D^0 \rightarrow K^-\pi^-\pi^+\pi^+)$:

$$N_{K\pi\pi\pi}^{yield} = \mathcal{L}_{int} \times \sigma_{D^0} \times \mathcal{B}(D^0 \rightarrow K^-\pi^-\pi^+\pi^+) \times \epsilon_{K\pi\pi\pi}^{data} \tag{11.5}$$

The efficiency $\epsilon_{K\pi\pi\pi}$ describes the overall reconstruction and selection efficiency of the process for data and simulation:

$$\epsilon_{K\pi\pi\pi} = \epsilon_{K\pi\pi\pi}^{acc} \times \epsilon_{K\pi\pi\pi}^{HLT1} \times \epsilon_{K\pi\pi\pi}^{HLT2} \times \epsilon_{K\pi\pi\pi}^{reco} \times \epsilon^{offline} \tag{11.6}$$

The other efficiencies contributing to the overall selection and reconstruction efficiency are:

- the acceptance efficiency, $\epsilon_{K\pi\pi\pi}^{acc}$, describing the efficiency of the LHCb detector acceptance
- the HLT1 selection efficiency, $\epsilon_{K\pi\pi\pi}^{HLT1}$, describing the efficiency of the HLT1 trigger line selection applied to the sample
- the HLT2 selection efficiency, $\epsilon_{K\pi\pi\pi}^{HLT2}$, describing the efficiency of the HLT2 trigger line selection applied to the sample
- the offline selection efficiency, $\epsilon_{K\pi\pi\pi}^{offline}$, describing the efficiency of all selection criteria applied offline

The formula can be simplified by summarising the acceptance, HLT1, HLT2 and offline selection efficiency as the nuisance efficiencies:

$$\begin{aligned}
&\epsilon^{nuis}_{K\pi\pi\pi} := \epsilon^{acc}_{K\pi\pi\pi} \times \epsilon^{HLT1}_{K\pi\pi\pi} \times \epsilon^{HLT2}_{K\pi\pi\pi} \times \epsilon^{offline}_{K\pi\pi\pi} \\
&\Rightarrow \epsilon^{reco}_{K\pi\pi\pi} = \frac{\epsilon_{K\pi\pi\pi}}{\epsilon^{nuis}_{K\pi\pi\pi}}
\end{aligned}$$

The track reconstruction efficiency can then be separated into the muonic track reconstruction efficiency, $\epsilon^{reco}_{\mu\mu\mu\mu(K\pi\pi\pi)}$, and hadronic correction factors, ρ^{hadInt}, for all four tracks, separating the two targeted pions from the other two hadrons, and using Eq. 11.2:

$$\epsilon^{reco}_{K\pi\pi\pi} \approx \epsilon^{reco}_{\mu\mu\mu\mu(K\pi\pi\pi)} \times \rho^{hadInt}_{K\pi} \times \rho^{hadInt}_{\pi\pi}, \tag{11.7}$$

where $\epsilon^{reco}_{\mu\mu\mu\mu(K\pi\pi\pi)}$ is the track reconstruction efficiency of four muons with the kinematic properties of the kaon and the three pions in the decay. The hadronic correction factor for the two pions can then be calculated as:

$$\rho^{hadInt}_{\pi\pi} \approx \frac{\epsilon^{reco}_{K\pi\pi\pi}}{\epsilon^{reco}_{\mu\mu\mu\mu(K\pi\pi\pi)} \times \rho^{hadInt}_{K\pi}} = \frac{\epsilon_{K\pi\pi\pi}}{\epsilon^{nuis}_{K\pi\pi\pi} \times \epsilon^{reco}_{\mu\mu\mu\mu(K\pi\pi\pi)} \times \rho^{hadInt}_{K\pi}} \tag{11.8}$$

It describes the probability that a pion pair cannot be reconstructed because of hadronic interactions with the detector material.
Using Eq. 11.4 and Eq. 11.5:

$$\begin{aligned}
\rho^{hadInt,MC}_{\pi\pi} &\approx \frac{N^{yield,MC}_{K\pi\pi\pi}}{N^{gen,MC}_{K\pi\pi\pi} \times \epsilon^{nuis,MC}_{K\pi\pi\pi} \times \epsilon^{reco,MC}_{\mu\mu\mu\mu(K\pi\pi\pi)} \times \rho^{hadInt,MC}_{K\pi}} \\
\rho^{hadInt,data}_{\pi\pi} &\approx \frac{N^{yield,data}_{K\pi\pi\pi}}{\mathcal{L}_{int} \times \sigma_{D^0} \times \mathcal{B}(D^0 \to K^-\pi^-\pi^+\pi^+) \times \epsilon^{nuis,data}_{K\pi\pi\pi} \times \epsilon^{reco,data}_{\mu\mu\mu\mu(K\pi\pi\pi)} \times \rho^{hadInt,data}_{K\pi}}
\end{aligned}$$

In order to cancel out the integrated luminosity, $\mathcal{L}_{int}$, the cross section, σ_{D^0}, and the hadronic correction factors for the kaon and pion appearing in both decays, the similar decay $D^0 \to K^-\pi^+$ is used:

$$\begin{aligned}
N^{yield,MC}_{K\pi} &= N^{gen,MC}_{K\pi} \times \epsilon^{nuis,MC}_{K\pi} \times \epsilon^{reco,MC}_{\mu\mu(K\pi)} \times \rho^{hadInt,MC}_{K\pi} \\
N^{yield,data}_{K\pi} &= \mathcal{L}_{int} \times \sigma_{D^0} \times \mathcal{B}(D^0 \to K^-\pi^+) \times \epsilon^{nuis,data}_{K\pi} \times \epsilon^{reco,MC}_{\mu\mu(K\pi)} \times \rho^{hadInt,MC}_{K\pi}
\end{aligned}$$

The hadronic corrections factor in recorded data, $\rho_{\pi\pi}^{hadInt,data}$, can then be cancelled out as following:

$$\rho_{\pi\pi}^{hadInt,data} \approx \frac{N_{K\pi\pi\pi}^{yield,data}}{\mathcal{L}_{int} \times \sigma_{D^0} \times \mathcal{B}(D^0 \to K^-\pi^-\pi^+\pi^+) \times \epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,data} \times \rho_{K\pi}^{hadInt,data}} \times \tag{11.9}$$

$$\frac{\mathcal{L}_{int} \times \sigma_{D^0} \times \mathcal{B}(D^0 \to K^-\pi^+) \times \epsilon_{K\pi}^{nuis,data} \times \epsilon_{\mu\mu(K\pi)}^{reco,MC} \times \rho_{K\pi}^{hadInt,MC}}{N_{K\pi}^{yield,data}} \tag{11.10}$$

$$= \frac{N_{K\pi\pi\pi}^{yield,data}}{N_{K\pi}^{yield,data}} / \left(\frac{\mathcal{B}(D^0 \to K^-\pi^-\pi^+\pi^+)}{\mathcal{B}(D^0 \to K^-\pi^+)} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,data}}{\epsilon_{K\pi}^{nuis,data} \times \epsilon_{\pi\pi(K\pi)}^{reco,data}} \right), \tag{11.11}$$

as long as the same data sample is used for both decays and the kinematics of the kaon and pion of the two decays is similar enough that the hadronic interaction corrections for them cancel out. Analogue, the hadronic interaction correction factor in simulated data, $\rho_{\pi\pi}^{hadInt,MC}$, reads:

$$\rho_{\pi\pi}^{hadInt,MC} \approx \frac{N_{K\pi\pi\pi}^{yield,MC}}{N_{K\pi\pi\pi}^{gen,MC} \times \epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC} \times \rho_{K\pi}^{hadInt,MC}} \times \tag{11.12}$$

$$\frac{N_{K\pi}^{gen,MC} \times \epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\mu\mu(K\pi)}^{reco,MC} \times \rho_{K\pi}^{hadInt,MC}}{N_{K\pi}^{yield,MC}} \tag{11.13}$$

$$= \frac{N_{K\pi\pi\pi}^{yield,MC}}{N_{K\pi}^{yield,MC}} / \left(\frac{N_{K\pi\pi\pi}^{gen,MC}}{N_{K\pi}^{gen,MC}} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC}}{\epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\pi\pi(K\pi)}^{reco,MC}} \right) \tag{11.14}$$

The hadronic correction factors for recorded and simulated data can now be determined by measuring the yield of signal candidates after full reconstruction and passing of all selection criteria, and dividing by respectively the number of generated candidates or the branching fraction, as well as all known efficiencies, including the track reconstruction efficiencies estimated for muons.

11.4 Sample selection

For efficiencies to cancel as much as possible, the selection of the two samples, $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^-\pi^+\pi^+$, has to be as similar as possible. Additionally, particle identification (PID) criteria are avoided, as they rely heavily on kinematics.

The D^0 mother particle is required to come from the decay $D^{*+} \to D^0\pi^+$, where the pion stemming from the D^{*+} is called the slow pion, ${\pi^+}_{slow}$. This allows to further constrain the selection for both decay processes. The optimised selection for both decay processes is shown in Tab. 11.1. Selection criteria marked as *offline* are only applied on the fully reconstructed and selected samples. All other selection criteria are part of the HLT2 selection algorithm.

The following variables are relevant for the selection:

- the kaon identification variable, PID_K, determined using information from the RICH subdetectors and the hadronic calorimeter as well as kinematic information of the track to give a value of likelihood for the track to belong to a kaon
- the invariant masses of the D^0, m_{D^0}, and the D^{*+}, $m_{D^{*+}}$, as well as their difference $\Delta(M) = m_{D^{*+}} - m_{D^0}$
- the sum of all transverse momenta of the daughter tracks, Σp_T
- the direction angle θ_{DIRA}, which is the cosine of the angle between the propagation vector of the particle and the vector connecting its primary and decay vertex
- the sum of the track quality of the impact parameters, IPχ^2, of all daughter tracks, $\Sigma IP\chi^2$
- IPχ^2, p, p_T, $DOCA$, trackχ^2 and vertexχ^2, which have already been defined in Sec. 6.3

Additionally, only events where the K^- track triggered the HLT1 trigger line *TrackMVADecision* are accepted. Details on this trigger line are given in Sec. 6.3. This is to make sure that no bias is introduced by events triggered on one of the two pion tracks whose hadronic correction is to be determined. At least one track having a high enough quality to trigger the line is now ensured to be shared between both decay channels.

	$D^{*+} \to D^0(\to K\pi)\pi$	$D^{*+} \to D^0(\to K\pi\pi\pi)\pi$
π, K	$p > 2\,\mathrm{GeV}/c$	$p > 2\,\mathrm{GeV}/c$
	$p_T > 250\,\mathrm{MeV}/c$	$p_T > 250\,\mathrm{MeV}/c$
	$IP^2_\chi > 9$	$IP^2_\chi > 9$
K	$PID_K > 0$	$PID_K > 0$
D^0	$m_{D^0} \pm 90\,\mathrm{MeV}/c^2$	$m_{D^0} \pm 90\,\mathrm{MeV}/c^2$
	$m_{D^0} \pm 80\,\mathrm{MeV}/c^2$	$m_{D^0} \pm 80\,\mathrm{MeV}/c^2$
offline:	$1800\,\mathrm{MeV}/c^2 < m_{D^0} < 1925\,\mathrm{MeV}/c^2$	$1800\,\mathrm{MeV}/c^2 < m_{D^0} < 1925\,\mathrm{MeV}/c^2$
	$\Sigma p_T > 2000\,\mathrm{MeV}/c$	$\Sigma p_T > 2000\,\mathrm{MeV}/c$
	$DOCA < 0.15\,\mathrm{mm}$	$DOCA < 0.15\,\mathrm{mm}$
	$track\chi^2 > 49$	$track\chi^2 > 49$
	$vertex\chi^2 < 10$	$vertex\chi^2 < 10$
	$\theta_{DIRA} < 17\,\mathrm{mrad}$	$\theta_{DIRA} < 17\,\mathrm{mrad}$
	$\Sigma IP\chi^2 > 50$	$\Sigma IP\chi^2 > 50$
		$m_{2\,children} < m_{D^0} + 90\,\mathrm{MeV}/c^2 - 2m_\pi$
		$m_{3\,children} < m_{D^0} + 90\,\mathrm{MeV}/c^2 - m_\pi$
π_{slow}	$p_T > 100\,\mathrm{MeV}/c$	$p_T > 100\,\mathrm{MeV}/c$
	$p > 1.5\,\mathrm{GeV}/c$	$p > 1.5\,\mathrm{GeV}/c$
D^{*+}	$\Delta(M) < 165\,\mathrm{MeV}/c^2$ (combination)	$\Delta(M) < 165\,\mathrm{MeV}/c^2$ (combination)
	$\Delta(M) < 160\,\mathrm{MeV}/c^2$ (composite)	$\Delta(M) < 160\,\mathrm{MeV}/c^2$ (composite)
	$vertex\chi^2 < 25$	$vertex\chi^2 < 25$
offline:	$142.43\,\mathrm{MeV}/c^2 < \Delta(M) < 148.43\,\mathrm{MeV}/c^2$	$142.43\,\mathrm{MeV}/c^2 < \Delta(M) < 148.43\,\mathrm{MeV}/c^2$

Table 11.1: Selection criteria for the selection of the two samples used to determine the hadronic corrections. Selection criteria marked as ***offline*** *are applied after the HLT2 trigger selection. Everything else is part of the HLT2 trigger selection.*

11.5 Signal yield extraction

To obtain the signal yields, $N_{K\pi\pi\pi}^{yield,data}$, $N_{K\pi\pi\pi}^{yield,MC}$, $N_{K\pi}^{yield,data}$ and $N_{K\pi}^{yield,MC}$, the invariant mass of the D^0 mother particle is plotted and fitted with a symmetric Johnson function:

$$PDF[Johnson\ S_U] = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1+(\frac{x-\mu}{\lambda})^2}} exp[-\frac{1}{2}(\gamma + \delta sinh^{-1}(\frac{x-\mu}{\lambda}))^2], \tag{11.15}$$

where $\gamma := 0$ to account for the symmetry of the distributions. The background is described with a simple exponential function.
The invariant mass distributions, fitted with the *Johnson* function, are shown in Fig. 11.2 for simulated data, as well as data recorded on the 6th of August 2024, which is during the time period of *data set 1*. Due to a larger abundance of the decays presented here in comparison to $J/\psi \rightarrow \mu^+\mu^-$, the smaller sample is sufficient. The signal yields determined with these fits are summarised in Tab. 11.2.
The branching fraction values as noted by the particle data group[3] are, with charge conjugation implied:

$$\mathcal{B}_{D^0\rightarrow K^-\pi^+} = (3.947 \pm 0.030) \times 10^{-2} \tag{11.16}$$
$$\mathcal{B}_{D^0\rightarrow K^+\pi^-} = (1.50 \pm 0.07) \times 10^{-4} \tag{11.17}$$
$$\mathcal{B}_{D^0\rightarrow K^-\pi^-\pi^+\pi^+} = (8.22 \pm 0.14) \times 10^{-2} \tag{11.18}$$
$$\mathcal{B}_{D^0\rightarrow K^+\pi^-\pi^-\pi^+} = (2.65 \pm 0.06) \times 10^{-4} \tag{11.19}$$

The contributions from the Cabibbo-suppressed decays $D^0 \rightarrow K^+\pi^-$ and $D^0 \rightarrow K^+\pi^-\pi^-\pi^+$ are negligible as they are smaller than the uncertainty on the not-Cabibbo suppressed branching ratios. The number of generated Monte Carlo events used for the presented samples are:

$$N_{K\pi}^{gen,MC} = 17,120,322 \tag{11.20}$$
$$N_{K\pi\pi\pi}^{gen,MC} = 17,149,135 \tag{11.21}$$

Using the branching fraction values from Eq. 11.16 and Eq. 11.18, the number of generated events from Eq. 11.20 and Eq. 11.21, and the signal yields from Tab. 11.2 as input, the

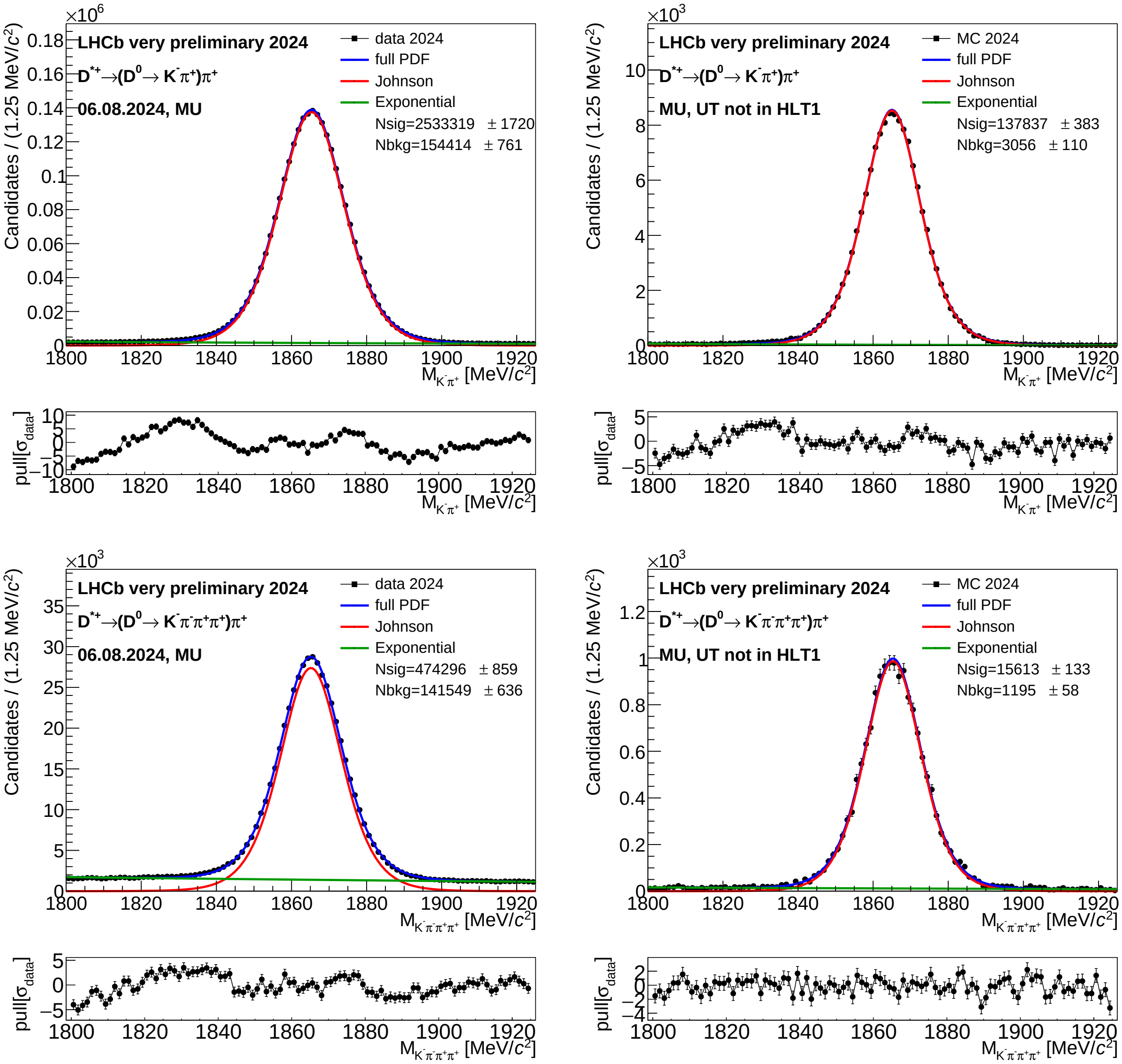


Figure 11.2: Fits of the invariant mass of the final tracks of the $D^0 \to K^-\pi^+$ decay (up) and the $D^0 \to K^-\pi^-\pi^+\pi^+$ decay (down) for 2024 data samples recorded on the 6th of August(left) and simulated with equivalent conditions (right).

	$D^0 \to K^-\pi^+$	$D^0 \to K^-\pi^-\pi^+\pi^+$
$N^{yield,MC}$	137837 ± 383	15613 ± 133
$N^{yield,data}$	2533319 ± 1720	474296 ± 859

Table 11.2: Signal yields as determined by fitting the invariant D^0 mass distributions in 11.2.

first part of $\rho_{\pi\pi}^{hadInt,data/MC}$ can be calculated using Eq. 11.12 and Eq. 11.9:

$$\begin{aligned}
\rho_{\pi\pi}^{hadInt,data} &\approx \frac{N_{K\pi\pi\pi}^{yield,data}}{N_{K\pi}^{yield,data}} / \left(\frac{\mathcal{B}(D^0 \to K^-\pi^-\pi^+\pi^+)}{\mathcal{B}(D^0 \to K^-\pi^+)} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,data}}{\epsilon_{K\pi}^{nuis,data} \times \epsilon_{\pi\pi(K\pi)}^{reco,data}} \right) \\
&= (8.99 \pm 0.17)\% / \left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,data}}{\epsilon_{K\pi}^{nuis,data} \times \epsilon_{\pi\pi(K\pi)}^{reco,data}} \right) \\
\rho_{\pi\pi}^{hadInt,MC} &\approx \frac{N_{K\pi\pi\pi}^{yield,MC}}{N_{K\pi}^{yield,MC}} / \left(\frac{N_{K\pi\pi\pi}^{gen,MC}}{N_{K\pi}^{gen,MC}} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC}}{\epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\pi\pi(K\pi)}^{reco,MC}} \right) \\
&= (11.31 \pm 0.10)\% / \left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC}}{\epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\pi\pi(K\pi)}^{reco,MC}} \right)
\end{aligned}$$

The values of 8.99% and 11.31% do not have any physical meaning while the differences between the track reconstruction and nuisance efficiencies in $D^0 \to K^-\pi^-\pi^+\pi^+$ and $D^0 \to K^-\pi^+$ are unknown. It can however be assumed in first approximation that those differences cancel out between simulated and recorded data:

$$\left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi\pi)}^{reco,data}}{\epsilon_{K\pi}^{nuis,data} \times \epsilon_{\mu\mu(K\pi)}^{recp.data}} / \frac{\epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC}}{\epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\mu\mu(K\pi)}^{reco,MC}} \right) \approx 1$$

In this approximation, the ratio between 8.99% and 11.31% is the recorded/simulated hadronic interaction correction factor of:

$$\rho_{\pi\pi}^{hadInt,data/MC} \approx (79.54 \pm 1.65)\% \tag{11.22}$$

This means that statistically, 79.54% of all pion pairs reconstructed in simulation are also reconstructed in recorded data. This translates to roughly 89% for a single pion.

11.6 Track reconstruction efficiency correction

Fig. 8.6 shows that the track reconstruction efficiency depends on the momentum and pseudorapidity of the tracks. Therefore, the track reconstruction efficiency of the kaon and pion final state particles varies depending on their momentum and pseudorapidity. This might not cancel out in the hadronic correction factor, $\rho_{\pi\pi}^{hadInt,data/MC}$. It needs to be corrected for the track reconstruction efficiency, ϵ^{reco}, as shown in Eq. 11.9 and 11.12.

The track reconstruction efficiency of each candidate can be determined by checking the momentum and pseudorapidity of all D^0 daughter tracks and multiplying the corresponding track reconstruction efficiencies taken from Fig. 8.3 and Fig. 8.4. The corrected signal yields are then determined by *reWeighting* the samples with one over the track reconstruction efficiency:

$$weight_{TrackReco}^{D^0 \to K^-\pi^+} = \frac{1}{\epsilon_{\mu(K)}\epsilon_{\mu(\pi)}} \tag{11.23}$$

$$weight_{TrackReco}^{D^0 \to K^-\pi^-\pi^+\pi^+} = \frac{1}{\epsilon_{\mu(K)}\epsilon_{\mu(\pi)}\epsilon_{\mu(\pi)}\epsilon_{\mu(\pi)}} \tag{11.24}$$

The fitted distributions of *reWeighted* candidates are presented in Fig. 11.3, and their signal yields are summarised in Tab. 11.3. As each candidate is corrected by the absolute track reconstruction efficiency of its final state particles, the yields are equivalent to the signal yields expected at a track reconstruction efficiency of 100%, not correcting for hadronic interactions.

	$D^0 \to K^-\pi^+$	$D^0 \to K^-\pi^-\pi^+\pi^+$
$N^{yield,MC}$	146815 ± 396	17854 ± 142
$N^{yield,data}$	2696535 ± 1774	540996 ± 915

Table 11.3: Signal yields as determined by fitting the invariant D^0 mass distributions in 11.3 reWeighted in track reconstruction efficiency of the daughter tracks.

With this *reWeighting* for the track reconstruction efficiency, the hadronic interaction correction factors are:

$$\begin{aligned}
\rho_{\pi\pi}^{hadInt,data} &\approx \frac{N_{K\pi\pi\pi}^{yield,data}}{N_{K\pi}^{yield,data}} / \left(\frac{\mathcal{B}(D^0 \to K^-\pi^-\pi^+\pi^+)}{\mathcal{B}(D^0 \to K^-\pi^+)} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,data}}{\epsilon_{K\pi}^{nuis,data} \times \epsilon_{\pi\pi(K\pi)}^{reco,data}} \right) \\
&= (9.64 \pm 0.18)\% / \left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,data}}{\epsilon_{K\pi}^{nuis,data}} \right) \\
\rho_{\pi\pi}^{hadInt,MC} &\approx \frac{N_{K\pi\pi\pi}^{yield,MC}}{N_{K\pi}^{yield,MC}} / \left(\frac{N_{K\pi\pi\pi}^{gen,MC}}{N_{K\pi}^{gen,MC}} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC}}{\epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\pi\pi(K\pi)}^{reco,MC}} \right) \\
&= (12.14 \pm 0.10)\% / \left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,MC}}{\epsilon_{K\pi}^{nuis,MC}} \right)
\end{aligned}$$

Under the assumption that the differences between the nuisance efficiencies of $D^0 \to K^-\pi^-\pi^+\pi^+$ and $D^0 \to K^-\pi^+$ cancel out between simulated and recorded data, this leads to a correction factor of:

$$\rho_{corr.TrackReco}^{hadInt,data/MC} \approx (79.40 \pm 1.63)\% \tag{11.25}$$

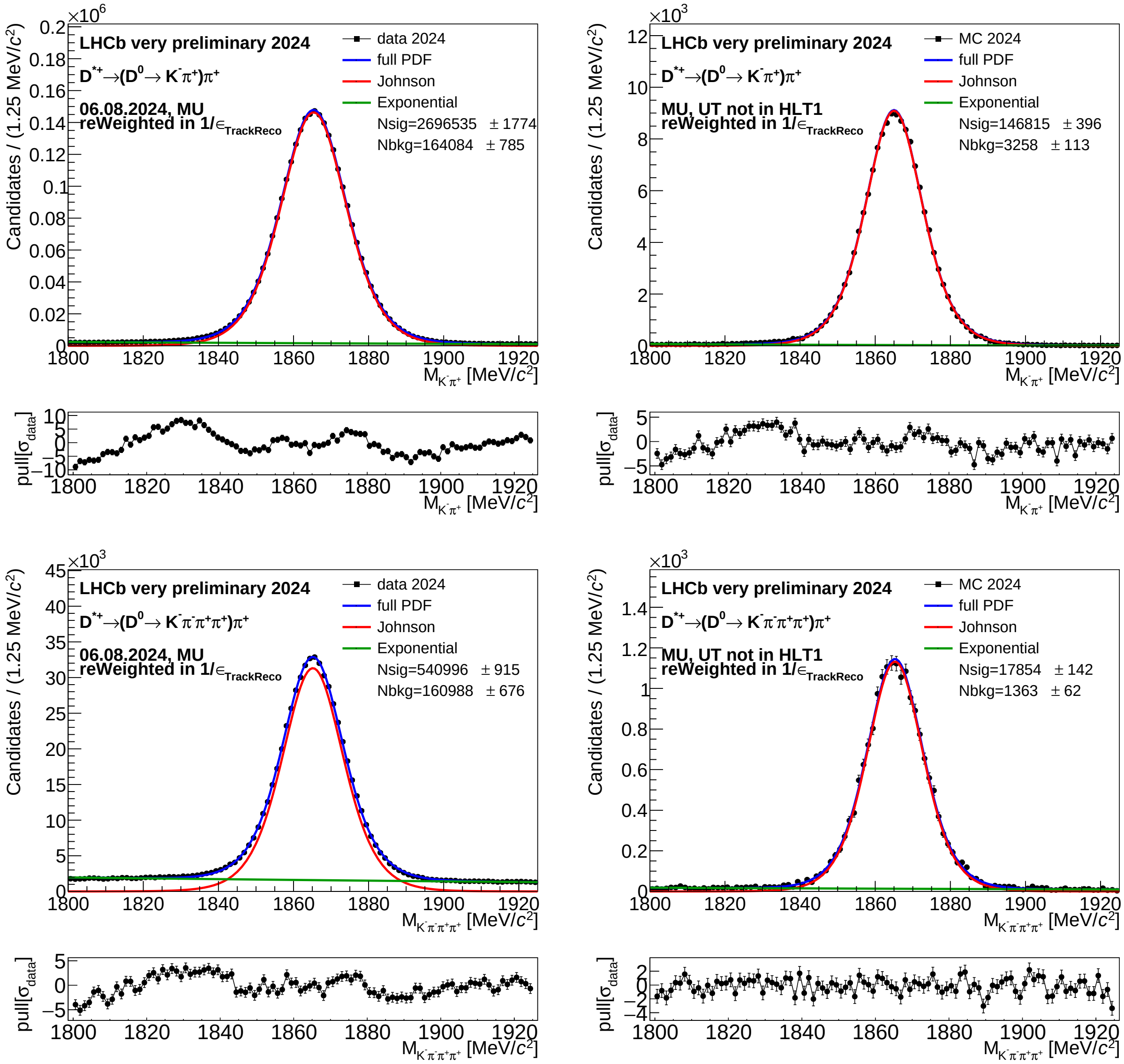


Figure 11.3: Fits of the invariant mass of the final tracks of the $D^0 \to K^-\pi^+$ decay (up) and the $D^0 \to K^-\pi^-\pi^+\pi^+$ decay (down) for 2024 data samples recorded on the 6th of August(left) and simulated with equivalent conditions (right), reWeighted in the reverse of the multiplied track reconstruction efficiency of all daughter tracks.

This is within one standard deviation of the value in Eq. 11.22. The track reconstruction efficiencies are well described in simulated data, especially in the well-populated regions, as was shown in Fig. 8.6 - Fig. 8.8. Therefore they do not have a significant impact on the overall hadronic interaction correction factor.

11.7 Kinematic dependencies

For the nuisance efficiencies to be effectively the same, it is necessary that the kinematics of the kaon and one pion track are very similar. To take into account the distribution of the hadronic correction over the kinematic range, it is necessary to consider the correction factors in bins.

The decay $D^0 \to K^-\pi^-\pi^+\pi^+$ produces one pion of the same charge as the kaon and two pions of opposite charge. Of the two pions that have the same charge as each other, one is expected to be coupled to the kaon and needs to have similar kinematics as the pion track from $D^0 \to K^-\pi^+$, while the other is expected to be coupled to the third pion. To identify which pion is matched to which track, the *sWeighted* distributions of momentum and pseudorapidity are compared between the pion from $D^0 \to K^-\pi^+$ and the two pions of the same charge of $D^0 \to K^-\pi^-\pi^+\pi^+$. They are separated into the pion with the higher transverse momentum, $\pi^+_{high\,p_T}$, and the pion with lower transverse momentum, $\pi^+_{low\,p_T}$.

Fig. 11.4 shows the distributions of the hadrons between both decay chains in comparison. It is clearly visible that the kinematics of the kaons are very similar between $D^0 \to K^-\pi^-\pi^+\pi^+$ and $D^0 \to K^-\pi^+$, confirming that influences on the ratio stemming from the kinematics of the kaon should be minimal. It is also clearly visible that the momentum and pseudorapidity of the pion with higher transverse momentum is closer to the momentum and pseudorapidity of the pion from $D^0 \to K^-\pi^+$. Therefore, the two pions taken into account for the kinematic dependency of the hadronic correction are the pion charged same as the kaon and the low transverse momentum oppositely charged pion.

The track reconstruction efficiency is given in ten different bins of momentum and pseudorapidity. However, by binning in the momentum of both pions, already three bins per pion amount to a total of nine bins, and the statistics of the simulated data sample is much more limited than that of the sample used for the track reconstruction efficiency. A large number of bins is not possible. Nine irregular bins are chosen such that the statistics of simulated data is sufficient for every bin. The two-dimensional distributions and chosen bins are shown in Fig. 11.5.

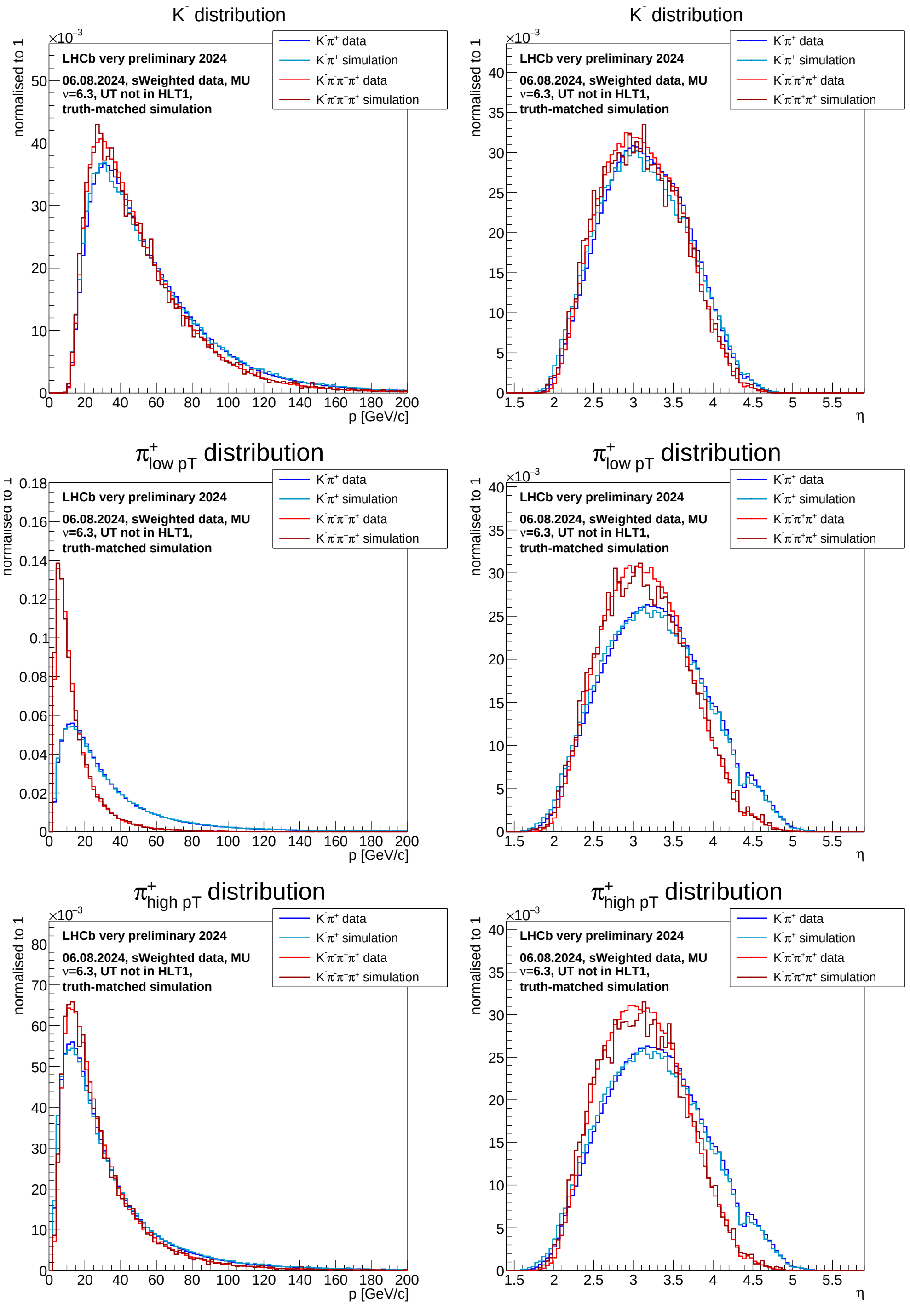


Figure 11.4: *Comparison of momentum (left) and pseudorapidity (right) distributions of the kaon from* $D^0 \to K^-\pi^+$ *(blue) with the kaon from* $D^0 \to K^-\pi^-\pi^+\pi^+$ *(red)(top), of the pion from* $D^0 \to K^-\pi^+$ *(blue) with the low transverse momentum pion (red)(middle) and the high transverse momentum pion from* $D^0 \to K^-\pi^-\pi^+\pi^+$ *(red)(bottom) in recorded sWeighted and simulated truth-matched data.*

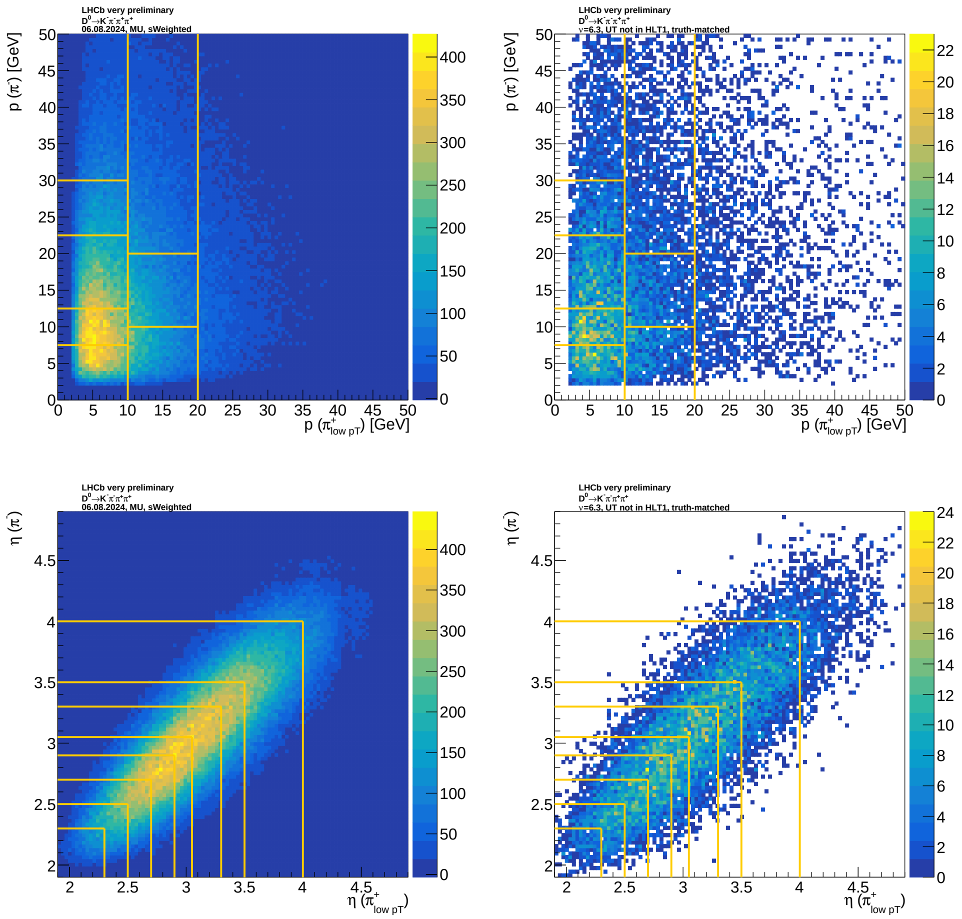


Figure 11.5: Twodimensional distribution of momentum (top) and pseudorapidity (bottom) of the two additional pions in $D^0 \to K^-\pi^-\pi^+\pi^+$, the π^- and the π^+ of lower p_T, in sWeighted recorded data (left) and truth-matched simulated data (right) with chosen bins marked in yellow.

The ratio is calculated in each bin using the yield from the full $D^0 \to K^-\pi^+$ samples as well as the yield in the given bin of the $D^0 \to K^-\pi^-\pi^+\pi^+$ samples. Every candidate is weighted again by $weight_{TrackReco}$. The resulting values in the momentum bins marked in Fig. 11.5 are shown in Tab. 11.4, and the ones in the pseudorapidity bins in Tab. 11.5.

It is clearly visible that the hadronic correction ratio increases with the momentum of either track. It is also clearly visible that it increases with pseudorapidity, such that it is especially low in the outmost detector region.

Values are generally below 100%. That means, that the simulation consistently underestimates material in the detector. However, further corrections need to be considered before drawing conclusions from this dependency.

bin	π^- range p[GeV]	${\pi^+}_{low\,pT}$ range p[GeV]	$\rho^{hadInt,data/MC}_{corr.\epsilon(TrackReco)}$
1	0-7.5	0-10	$72.45 \pm 2.35\%$
2	7.5-12.5	0-10	$72.26 \pm 2.26\%$
3	12.5-22.5	0-10	$75.83 \pm 2.18\%$
4	22.5-30	0-10	$74.11 \pm 2.92\%$
5	30-150	0-10	$82.55 \pm 2.94\%$
6	0-10	10-20	$81.68 \pm 2.86\%$
7	10-20	10-20	$82.79 \pm 2.67\%$
8	20-150	10-20	$86.62 \pm 2.49\%$
9	0-150	20-150	$81.72 \pm 2.09\%$

Table 11.4: Track reconstruction efficiency corrected hadronic interaction recorded/simulated data correction factors in nine bins of momentum as presented in Fig. 11.5.

bin	$(\pi^-\mathbf{or}\,{\pi^+}_{low\,pT})\eta >$	$(\pi^-\mathbf{and}\,{\pi^+}_{low\,pT})\eta <$	$\rho^{hadInt,data/MC}_{corr.\epsilon(TrackReco)}$
1	1.9	2.3	$60.40 \pm 3.22\%$
2	2.3	2.5	$63.72 \pm 2.34\%$
3	2.5	2.7	$71.02 \pm 2.27\%$
4	2.7	2.9	$79.89 \pm 2.41\%$
5	2.9	3.05	$78.14 \pm 2.49\%$
6	3.05	3.3	$80.77 \pm 2.23\%$
7	3.3	3.5	$83.65 \pm 2.48\%$
8	3.5	4.0	$84.62 \pm 2.16\%$
9	4.0	4.9	$82.71 \pm 2.81\%$

Table 11.5: Track reconstruction efficiency corrected hadronic interaction recorded/simulated data correction factors in nine bins of pseudorapidity as presented in Fig. 11.5.

11.8 Corrections through reWeighting

To cancel nuisance efficiencies between $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^-\pi^+\pi^+$, it is essential that the kaon-pion couple shared between both samples comes from the same phase space region. The momentum and pseudorapidity distribution of the pion of higher p_T is slightly different from that of the $D^0 \to K^-\pi^+$ pion. Likewise, the distributions of either kaon from $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^-\pi^+\pi^+$ differ.

To correct for effects of these differences, the $D^0 \to K^-\pi^+$ is *reWeighted* both in recorded and simulated data to recorded and simulated data of $D^0 \to K^-\pi^-\pi^+\pi^+$, respectively. The variables used for *reWeighting* are the momentum and pseudorapidity of the kaon and the pion. The recorded data samples are both *sWeighted.*

Additionally, the simulated samples can both be *reWeighted* to the samples of *sWeighted* recorded data. This is to ensure the simulated data samples cover the same phase space regions as the recorded data samples. To avoid interference with the first step of *reWeighting*, the $D^0 \to K^-\pi^+$ simulated sample is *reWeighted* to the $D^0 \to K^-\pi^+$ recorded data sample that has already been *reWeighted* in kaon and pion kinematics to the $D^0 \to K^-\pi^-\pi^+\pi^+$ recorded data sample. All recorded data samples are *sWeighted.*

The simulated data of $D^0 \to K^-\pi^+$ are *reWeighted* in the momentum and pseudorapidity of the kaon and the pion to the recorded and *sWeighted* $D^0 \to K^-\pi^+$ data, which is already *reWeighted* in the kaon and pion momentum and pseudorapidity to the kaon and high transverse momentum oppositely charged pion from $D^0 \to K^-\pi^-\pi^+\pi^+$. The resulting momentum and pseudorapidity distributions in all steps of *reWeighting* are shown in Fig. 11.6.

The kaon and all three pions from the $D^0 \to K^-\pi^-\pi^+\pi^+$ simulated data are *reWeighted* in their momentum and pseudorapidity to recorded *sWeighted* $D^0 \to K^-\pi^-\pi^+\pi^+$ data. The resulting momentum and pseudorapidity distributions of the kaon and the high transverse momentum oppositely charged pion are shown in Fig. 11.7. The distributions of the other two pions, which are the ones relevant for determining the hadronic corrections, are shown in Fig. 11.8.

With simulated data *reWeighted* to recorded data, $D^0 \to K^-\pi^+$ *reWeighted* in recorded and simulated data to $D^0 \to K^-\pi^-\pi^+\pi^+$, and everything *reWeighted* to correct for track

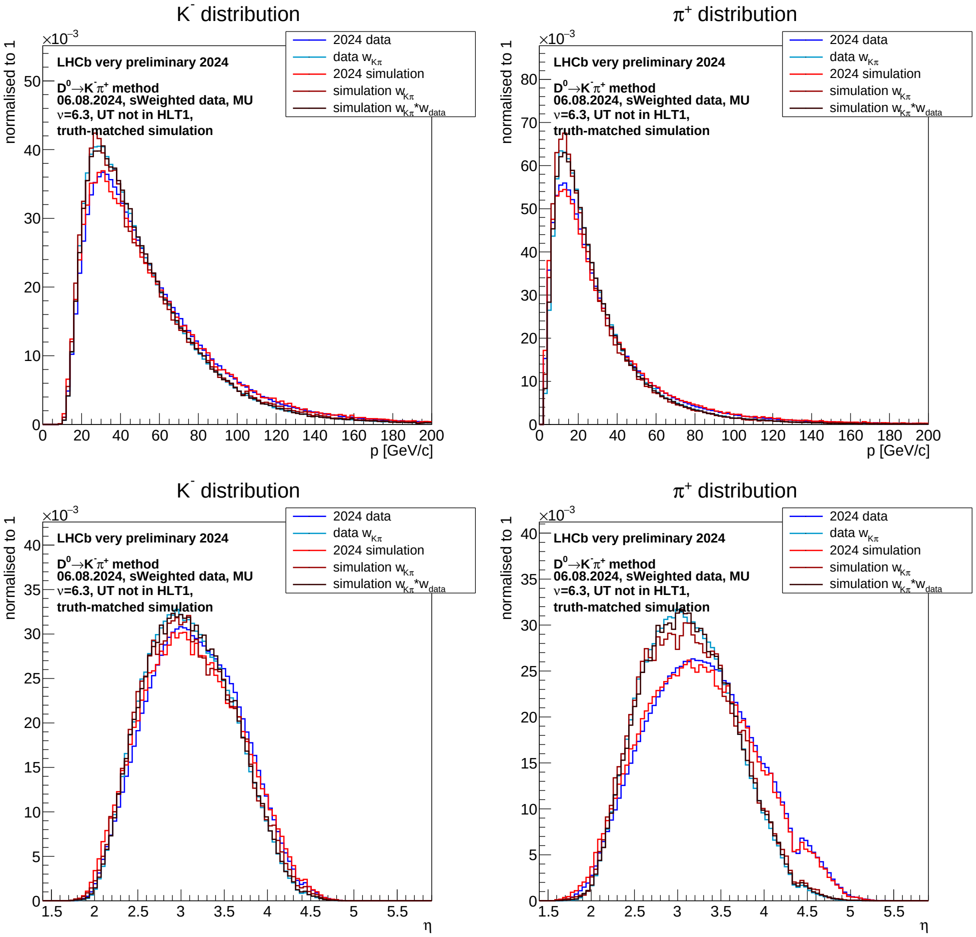


Figure 11.6: Kaon (left) and pion (right) from $D^0 \to K^-\pi^+$ distributions of momentum (top) and pseudorapidity) bottom in recorded data (blue), recorded data reWeighted to $D^0 \to K^-\pi^-\pi^+\pi^+$ (light blue), truth-matched simulated data (red), truth-matched simulated reWeighted to $D^0 \to K^-\pi^-\pi^+\pi^+$ (darker red) and truth-matched simulated data reWeighted to both $D^0 \to K^-\pi^-\pi^+\pi^+$ and to reWeighted recorded $D^0 \to K^-\pi^+$ data (darkest red).

reconstruction efficiencies, the hadronic corrections are:

$$\rho_{\pi\pi}^{hadInt,data} \approx \frac{N_{K\pi\pi\pi}^{yield,data}}{N_{K\pi}^{yield,data}} \Big/ \left(\frac{\mathcal{B}(D^0 \to K^-\pi^-\pi^+\pi^+)}{\mathcal{B}(D^0 \to K^-\pi^+)} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,data} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,data}}{\epsilon_{K\pi}^{nuis,data} \times \epsilon_{\pi\pi(K\pi)}^{reco,data}} \right)$$

$$= (14.63 \pm 00.27)\% \Big/ \left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,data}}{\epsilon_{K\pi}^{nuis,data}} \right)$$

$$\rho_{\pi\pi}^{hadInt,MC} \approx \frac{N_{K\pi\pi\pi}^{yield,MC}}{N_{K\pi}^{yield,MC}} \Big/ \left(\frac{N_{K\pi\pi\pi}^{gen,MC}}{N_{K\pi}^{gen,MC}} \times \frac{\epsilon_{K\pi\pi\pi}^{nuis,MC} \times \epsilon_{\mu\mu\mu\mu(K\pi\pi\pi)}^{reco,MC}}{\epsilon_{K\pi}^{nuis,MC} \times \epsilon_{\pi\pi(K\pi)}^{reco,MC}} \right)$$

$$= (15.58 \pm 0.13)\% \Big/ \left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,MC}}{\epsilon_{K\pi}^{nuis,MC}} \right)$$

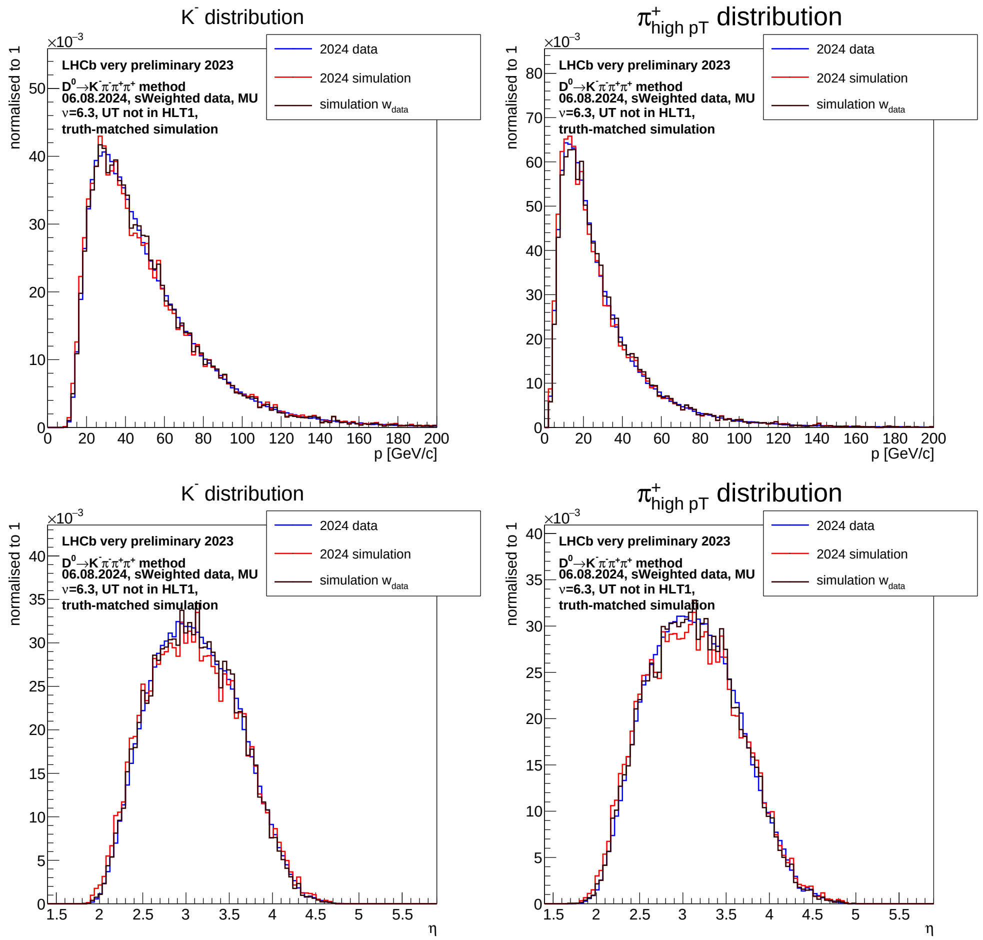


Figure 11.7: Kaon (left) and high-p_T oppositely charged pion (right) from $D^0 \to K^-\pi^-\pi^+\pi^+$ distributions of momentum (top) and pseudorapidity) bottom in recorded data (blue), truth-matched simulated data (red), and truth-matched simulated data reWeighted to both $D^0 \to K^-\pi^-\pi^+\pi^+$ and to reWeighted recorded $D^0 \to K^-\pi^+$ data (darkest red).

Assuming again that the nuisance efficiency discrepancies between $D^0 \to K^-\pi^-\pi^+\pi^+$ and $D^0 \to K^-\pi^+$ cancel out between recorded as in simulated data, this leads to a correction factor of:

$$\rho^{hadInt,data/MC}_{corr.TrackReco,reWeight.K\pi} \approx (93.90 \pm 1.93)\% \tag{11.26}$$

This is a significant change from the value in Eq. 11.25. If 94% of the pion pairs reconstructed in simulated data are also reconstructed in recorded data, about 97% of single pions are expected to be reconstructed, leading to a discrepancy of just 3%. That means, 79% reconstructed simulated pions implies a reconstruction of 76.6% of recorded data pions and therefore an uncertainty on the track reconstruction efficiency of 2.4%.

The main influence on the change in $\rho^{hadInt,data/MC}$ due to *reWeighting* comes from *reWeighting* the kinematics of the simulated kaon and pion to the recorded kaon and pion tracks. The *reWeighting* of the $D^0 \to K^-\pi^+$ kaon and pion to the $D^0 \to K^-\pi^-\pi^+\pi^+$

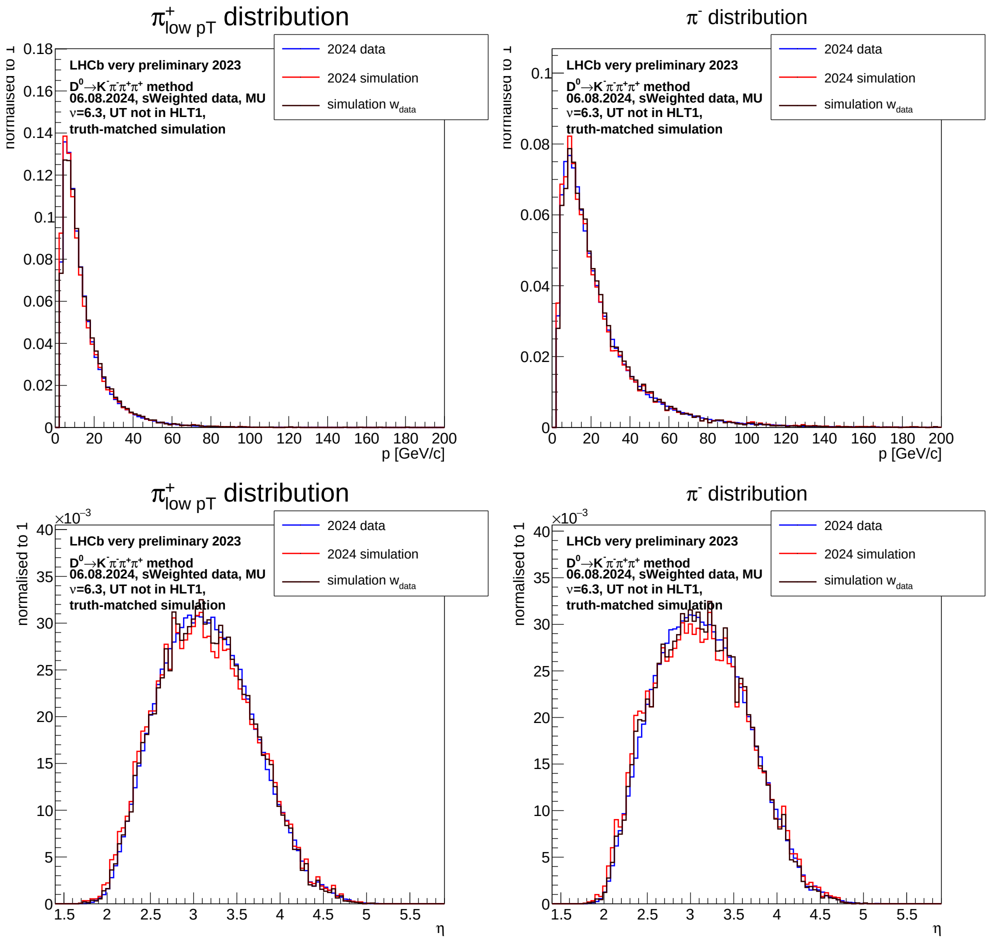


Figure 11.8: low–p_T positive pion (left) and negative pion (right) from $D^0 \to K^-\pi^-\pi^+\pi^+$ distributions of momentum (top) and pseudorapidity) bottom in recorded data (blue), truth-matched simulated data (red), and truth-matched simulated data reWeighted to both $D^0 \to K^-\pi^-\pi^+\pi^+$ and to reWeighted recorded $D^0 \to K^-\pi^+$ data (darkest red).

kaon and pion does not show an effect above one standard deviation. Neither does the *reWeighting* of the two pions between simulated and recorded data show a significant effect.

Further investigation out of the bounds of this thesis is needed on why *reWeighting* simulated to recorded data leads to such a significant improvement in the agreement between data and simulation, given that this step does not seem to influence the distributions of momentum and pseudorapidity significantly.

11.9 Estimation of nuisance efficiencies

All hadronic correction factors presented thus far assume the cancellation of nuisance efficiency discrepancies of $D^0 \to K^-\pi^-\pi^+\pi^+$ and $D^0 \to K^-\pi^+$ between simulated and recorded data:

$$\left(\frac{\epsilon_{K\pi\pi\pi}^{nuis,data}}{\epsilon_{K\pi}^{nuis,data}} / \frac{\epsilon_{K\pi\pi\pi}^{nuis,MC}}{\epsilon_{K\pi}^{nuis,MC}} \right) \approx 1$$

However, differences in the efficiency of $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^-\pi^+\pi^+$ cancel out in the overall recorded/simulated correction factor, $\rho^{adInt,data/MC}$, only if these differences are the same in recorded and simulated data. This might not be the case. It is for example possible that simulated data describes the efficiencies of $D^0 \to K^-\pi^+$ better than those of $D^0 \to K^-\pi^-\pi^+\pi^+$, which due to the higher multiplicity of its decay products is a more complicated decay.

The efficiencies contributing to the nuisance efficiency are:

- the acceptance efficiency, ϵ^{acc}
- the HLT1 selection efficiency, ϵ^{HLT1}
- the HLT2 selection efficiency, ϵ^{HLT2}
- the offline selection efficiency, $\epsilon^{offline}$

These efficiencies are not easily measurable on recorded and simulated data. The acceptance efficiency, for example, can be determined on simulated data using information from the generation, but not on recorded data, where the exact luminosity, cross section and branching fraction of all contributing production channels and the decay channel need to be known to estimate the number of original candidates. However, the acceptance efficiency is expected to depend mostly on the kinematics of the decay, which agrees very well between recorded and simulated data, and therefore is not expected to contribute significantly.

11.9.1 The HLT1 selection efficiency

The HLT1 requirement on the samples used to determine the hadronic correction is the triggering of the *TrackMVADecision* line by the K^+ track. As this criterion is already applied online to save disk space when storing the selected candidates, the efficiency of this line can not be measured on the used samples directly. However, a similar sample can be used that does not have the requirement of this specific trigger line online.

Due to the structure of the LHCb trigger system, only candidates that pass any HLT1 trigger line with very few exceptions are handed over for reconstruction with HLT2. This means that even if the sample did not trigger this specific line, a bias is expected when comparing the candidates triggered by this line with all triggered candidates. To avoid this bias, the TISTOS method as described in [37] can be attempted.

This will have to be tested out of the bounds of this thesis.

11.9.2 The offline selection and D^{*+} selection efficiency

The only criterion applied offline to the sample is the requirement on the mass difference of the D^{*+} and the D^0 meant to reduce background and produce a signal shape easier to fit:

$$142.43\,\mathrm{MeV}/c^2 < \Delta(M) := m_{D^{*+}} - m_{D^0} < 148.43\,\mathrm{MeV}/c^2$$

The ratio for this selection by which the hadronic interaction recorded/simulated data ratio needs to be divided, $R^{\epsilon,offline}$, is determined by fitting the signal yields with and without the $\Delta(M)-$selection:

$$R^{\epsilon,offline} = \frac{\epsilon^{offline}_{K\pi\pi\pi,data} \times \epsilon^{offline}_{K\pi,MC}}{\epsilon^{offline}_{K\pi\pi\pi,MC} \times \epsilon^{offline}_{K\pi,data}} \tag{11.27}$$

This selection can have a slightly different effect on data and on simulation. To verify that $R^{\epsilon,offline}$ is reliable, an independent sample analysing the same decays but with a different selection is tested. This sample is called the *charm* sample as it is used in analyses to measure *charm* cross sections. Additionally, $R^{\epsilon,offline}$ is determined both with and without the correction for the track reconstruction efficiency.

$$\begin{aligned} R^{\epsilon,offline} &= 1.066 \pm 0.002 \\ R^{\epsilon,offline}_{corr.TrackReco} &= 1.061 \pm 0.002 \\ R^{\epsilon,offline}_{charm} &= 1.065 \pm 0.006 \\ R^{\epsilon,offline}_{charm,corr.TrackReco} &= 1.065 \pm 0.006 \end{aligned}$$

The values agree in all four cases within one standard deviation.

The *charm* sample includes, additionally to the samples for $D^{*+} \to (D^0 \to K^-\pi^+)\pi$ and $D^{*+} \to (D^0 \to K^-\pi^-\pi^+\pi^+)\pi$, also samples which do not require the D^0 to come from a D^{*+} decay. Therefore, it can also be used to test the efficiency of selecting these specific decays. This is usually a part of the HLT2 selection.

The combined nuisance efficiency ratio of the offline selection and the $D^{*+}-$origin selection is determined by taking the signal yield of candidates coming from a $D^{*+} \to D^0\pi$ decay and passing the $\Delta(M)$ selection over the signal yield of $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^-\pi^+\pi^+$ candidates without $D^{*+}-$origin requirement or $\Delta(M)-$requirement.:

$$\epsilon^{offline,D*+} = \frac{N(passing\,offline\&from\,D^{*+})}{N(no\,offline\&no\,origin\,requirement)}$$

The corresponding nuisnace efficiency correction ratio is:

$$R^{\epsilon,offline,D*+} = 0.9896 \pm 0.0111$$

When correcting for the track reconstruction efficiency of the daughter tracks by *reWeighting* with $weight_{TrackReco}$ from Eq. 11.23 and Eq. 11.24, the nuisance efficiency correction ratio is instead:

$$R^{\epsilon,offline,D*+}_{corr.TrackReco} = 0.9895 \pm 0.0104 \tag{11.28}$$

The hadronic correction ratio from Eq. 11.26, still under the assumption that the other nuisance efficiencies are the same in recorded as in simulated data, amounts then to:

$$\rho^{hadInt,data/MC}_{corr.TrackReco,reWeight,K\pi,offline} \approx (94.90 \pm 2.19)\% \tag{11.29}$$

Eq. 11.26 is already corrected for the track reconstruction efficiency and *reWeighted* kinematically. This is with an absolute difference of 1.00% within one standard deviation of the value given in Eq. 11.26. While it needs to be taken into account in any finalised numbers, this does not seem to be a major effect on the correction factor. At this point, it is taken into account as a systematic uncertainty.

11.9.3 Other HLT2 selection efficiencies

The selection criteria summarised in Tab. 11.1 are, where possible, the same between $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^-\pi^+\pi^+$, and always the same between data and simulation. However, due to discrepancies in the distributions of variables in recorded and simulated data, it is possible that some selection criteria have different efficiencies that do not cancel out in the overall ratio.

Distributions of variables affected by these selection criteria may be compared to find potential variables of interest which might be affected differently between the decay channels and between data and simulation. These are necessary further checks out of the bounds of this thesis to finalise the hadronic interaction ratios.

11.10 Systematic uncertainty on the track reconstruction efficiency due to hadronic interactions

The track reconstruction efficiency corrections determined as ratios of recorded over simulated data efficiencies on muons do not include necessary corrections due to hadronic interactions, as muons are minimally ionising particles. The agreement of hadronic interactions between recorded and simulated data of two pions is, with correction due to track reconstruction efficiency and *reWeighting*, given in Eq. 11.26:

$$\rho^{hadInt,data/MC}_{corr.TrackReco,reWeight} \approx (93.90 \pm 1.93(stat) \pm 1.00\%(sys))\%$$

The systematic uncertainty is currently only coming from correcting for the offline $\Delta(M)$ as well as the D^{*+}−origin selection. Further systematic studies are essential.

In a very rough estimate of drawing the square root, that leads to an agreement for one pion of:

$$96.90 \pm 1.00(stat) \pm 0.52(sys)\%$$

The agreement between recorded and simulated data of two pions in dependency of momentum and pseudorapidity, track reconstruction efficiency corrected and kinematically *reWeighted*, are given in Tab. 11.6 and Tab. 11.7. As can be seen, they vary by up to 20%, which is a significant variation. The clear momentum dependence of Tab. 11.4 is not present anymore in Tab. 11.6.

The clear dependence on the pseudorapidity, on the other hand, is still visible. While the first pseudorapidity bin has a very large systematic uncertainty, all the following ones show a clear tendency of a higher correction factor for higher pseudorapidity. Tracks with a higher pseudorapidity, which fly at a higher angle from the beam pipe, have a bigger discrepancy between the hadronic correction of recorded and simulated data. This is expected, as particles flying at a higher angle are expected to be slower and see more material. Any discrepancy between recorded and simulated data therefore shows a larger absolute effect.

These are very rough estimates of the discrepancy between recorded and simulated data concerning the hadronic interactions of particles with material. Systematic uncertainties and the influence of nuisance efficiencies still need to be studied in further detail. The reasons for the large effects of *reWeighting* the samples need to be understood.

The final objective of estimating the correction due to the hadronic uncertainty is to determine a systematic uncertainty on the recorded/simulated track reconstruction efficiency ratio when it is used for hadrons. However, for that, the hadronic correction needs to be determined not only with higher precision, but for a single pion track. After the above mentioned missing checks are done and potential further adjustments have been made, the next step will be determining the dependence of the correction factor on each of the pion kinematics.

This can also be translated into an absolute uncertainty on the knowledge of the material budget of the LHCb detector. Reference values from detector experts are not yet available. Further tests will also have to be made on the pion track reconstruction efficiency in comparison to the muon track reconstruction efficiency in simulated data. This will complete the comparison between the track reconstruction efficiency of pions and muons made in Fig. 11.2.

In LHC Run 2, between 2015 and 2018, a 10% uncertainty on the material budget was assumed. The current hadronic correction values amount to a larger uncertainty than this still, but it can be assumed confidently that the listed modifications and enhancements to the method will allow for a value of 10% uncertainty or even better to be reached.

bin	π^- range p[GeV]	${\pi^+}_{low\,pT}$ range p[GeV]	$\rho^{hadInt,data/MC}_{corr.TrackReco,reWeight}$
1	0-7.5	0-10	$101.33 \pm 3.49\%$
2	7.5-12.5	0-10	$93.21 \pm 2.99\%$
3	12.5-22.5	0-10	$93.12 \pm 2.99\%$
4	22.5-30	0-10	$87.34 \pm 3.45\%$
5	30-150	0-10	$100.99 \pm 3.45\%$
6	0-10	10-20	$95.82 \pm 3.37\%$
7	10-20	10-20	$88.89 \pm 2.80\%$
8	20-150	10-20	$94.90 \pm 2.68\%$
9	0-150	20-150	$91.57 \pm 2.32\%$

Table 11.6: Hadronic correction factor as data-over-simulation ratio in nine bins of momentum as presented in Fig. 11.5. The samples are corrected for the track reconstruction efficiency determined on muons for all daughters, and reWeighted in kinematic variables to align the kaon and pion from $D^0 \to K^-\pi^+$ to $D^0 \to K^-\pi^-\pi^+\pi^+$, as well as simulated kaon and pion tracks to recorded kaon and pion tracks.

bin	$(\pi^{-}\mathbf{or}\,\pi^{+}{}_{low\,pT})\eta >$	$(\pi^{-}\mathbf{and}\,\pi^{+}{}_{low\,pT})\eta <$	$\rho^{hadInt,data/MC}_{corr.TrackReco,reWeight}$
1	1.9	2.3	$89.37 \pm 5.28\%$
2	2.3	2.5	$82.19 \pm 3.13\%$
3	2.5	2.7	$87.37 \pm 2.84\%$
4	2.7	2.9	$93.56 \pm 2.83\%$
5	2.9	3.05	$90.11 \pm 2.86\%$
6	3.05	3.3	$91.99 \pm 2.52\%$
7	3.3	3.5	$94.72 \pm 2.78\%$
8	3.5	4.0	$97.58 \pm 2.48\%$
9	4.0	4.9	$106.43 \pm 3.69\%$

Table 11.7: Hadronic correction factor as data-over-simulation ratio in nine bins of pseudorapidity as presented in Fig. 11.5. The samples are corrected for the track reconstruction efficiency determined on muons for all daughters, and reWeighted in kinematic variables to align the kaon and pion from $D^0 \to K^-\pi^+$ *to* $D^0 \to K^-\pi^-\pi^+\pi^+$*, as well as simulated kaon and pion tracks to recorded kaon and pion tracks.*

Chapter 12

Summary

The LHCb experiment is a highly complex apparatus dedicated to the measurement of particle decays involving b- and c-quarks. At full operation of the LHC, LHCb records proton-proton collisions at an interaction rate of 40 MHz. Each collision can potentially produce hundreds of particles, which must be detected and reconstructed to exploit the physics potential of the recorded data.

The determination and understanding of the *track reconstruction efficiency* is an integral part of many analyses, for example the measurement of particle production cross sections. It describes the probability that a particle crossing through the detector is reconstructed correctly.

A *tag-and-probe method* using muons from $J/\psi \to \mu\mu$ decays has been developed to measure the track reconstruction efficiency of the two main tracking subdetectors of the LHCb experiment during Run 3 of the LHC, the vertex locator (VELO), and the scintillating fibre tracker (SciFi). Employing this method in the same way on both simulated and recorded data samples, correction factors can be developed to describe possible differences between the two samples.

In 2023, with the newly rebuilt LHCb detector still in the process of being commissioned and aligned and only two of the three tracking subdetectors installed, the comparison of the track reconstruction efficiency between simulated and recorded data delivers crucial information for the understanding of the detector. This is used to complete the alignment of the detector and continuously improve the simulation of data, guaranteeing a better start for the data taking in 2024.

With the improved detector and simulation, data taking in 2024 allows for more detailed measurements of the track reconstruction efficiency, using all three tracking subdetectors, which are the VELO, the upstream tracker (UT) and the SciFi. The track reconstruction efficiencies of VELO and SciFi are again measured using the *tag-and-probe method* and their discrepancies between simulated and recorded data are used to compute correction factors. These factors can be used in physics analyses to correct track reconstruction efficiencies measured on different samples. The track reconstruction efficiencies are measured on two different data sets to describe the different detector states.

Thanks to the continuous improvement of the simulation in 2023, the simulated data is very close to the recorded data in 2024, leading to an agreement of the track reconstruction efficiency between recorded and simulated data above 98% in most phase space regions, with an agreement above 99% in the most populated regions. Exceptions are in the outer regions of the detector, where the alignment is less good and the agreement drops to slightly below 96%.

The statistical uncertainty on the track reconstruction efficiency is of the order of ±0.04%. A full analysis of systematic uncertainties on the track reconstruction efficiency has been carried out. The systematic uncertainties are dominated by as-of-yet not fully understood differences between the different methods, which are very likely to be resolved or minimised in 2025. The systematic uncertainties amount to approximately ±0.70%, but are well understood and measures to reduce them are available in 2025.

Sources for the discrepancies and the necessary correction factors for simulated data have been extensively analysed. The most important one of these is the effect of excluded frontend boards in the SciFi subdetector, which are not excluded in the according simulation. Systematic studies of the method have been performed, verifying it on simulated data with an alternative method utilising knowledge of the generated events. The *tag-and-probe method* is tested, and the stability of the track reconstruction efficiency under the influence of different polarities of the LHCb magnet is verified.

While the *tag-and-probe method* provides corrections for the track reconstruction efficiency which are measured using muonic tracks, it cannot correct for hadronic interactions of propagating particles with the detector material. A completely new method is developed to measure the correction of the track reconstruction efficiency due to hadronic interactions of two pions with the detector material. An agreement between the hadronic corrections factors for two pions in simulated and recorded data of about 94% is determined. This can be estimated to correspond to an additional systematic uncertainty on the track reconstruction efficiency correction on the order of a few percent.

However, several tests are still needed to finalise the method and to estimate more precise values for the hadronic correction factor. It is expected that, once these adjustments and additional checks are performed, the hadronic corrections will contribute a systematic uncertainty of 1% or less to the track reconstruction efficiency correction.

Part I

Appendix

Appendix A

Kinematic distributions

In Chapter 8, track reconstruction efficiencies of the SciFi and the VELO as well as long track reconstruction efficiencies are determined as a function of the probe muon momentum, p, and pseudorapidity, η. The corresponding distributions of p and η are shown for recorded and simulated data of the *Downstream method* in Fig. A.1, the *VeloMuon method* in Fig. A.2 and the *MuonUT method* in Fig. A.3. Systematic uncertainties to determine the effect of the visible discrepancies between the distributions in simulated and recorded data are determined by *reWeighting* the simulated data samples also in Chapter 8.

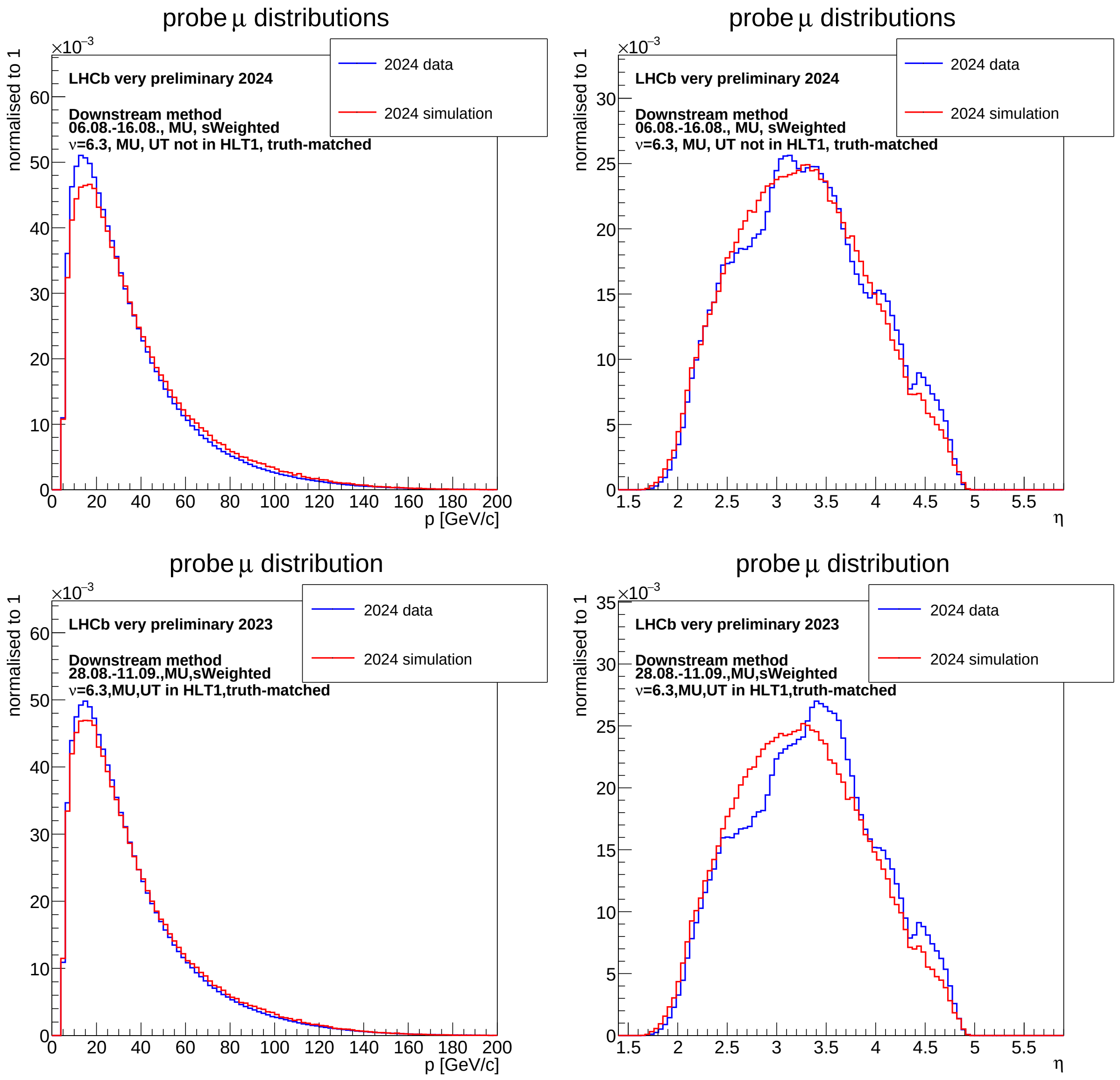


Figure A.1: Distribution of Downstream probe momentum, p, (left) and pseudorapidity of the Downstream probe muon tracks, η, (right) in sWeighted data and truth-matched simulation for data set 1 (above) and data set 2 (below)

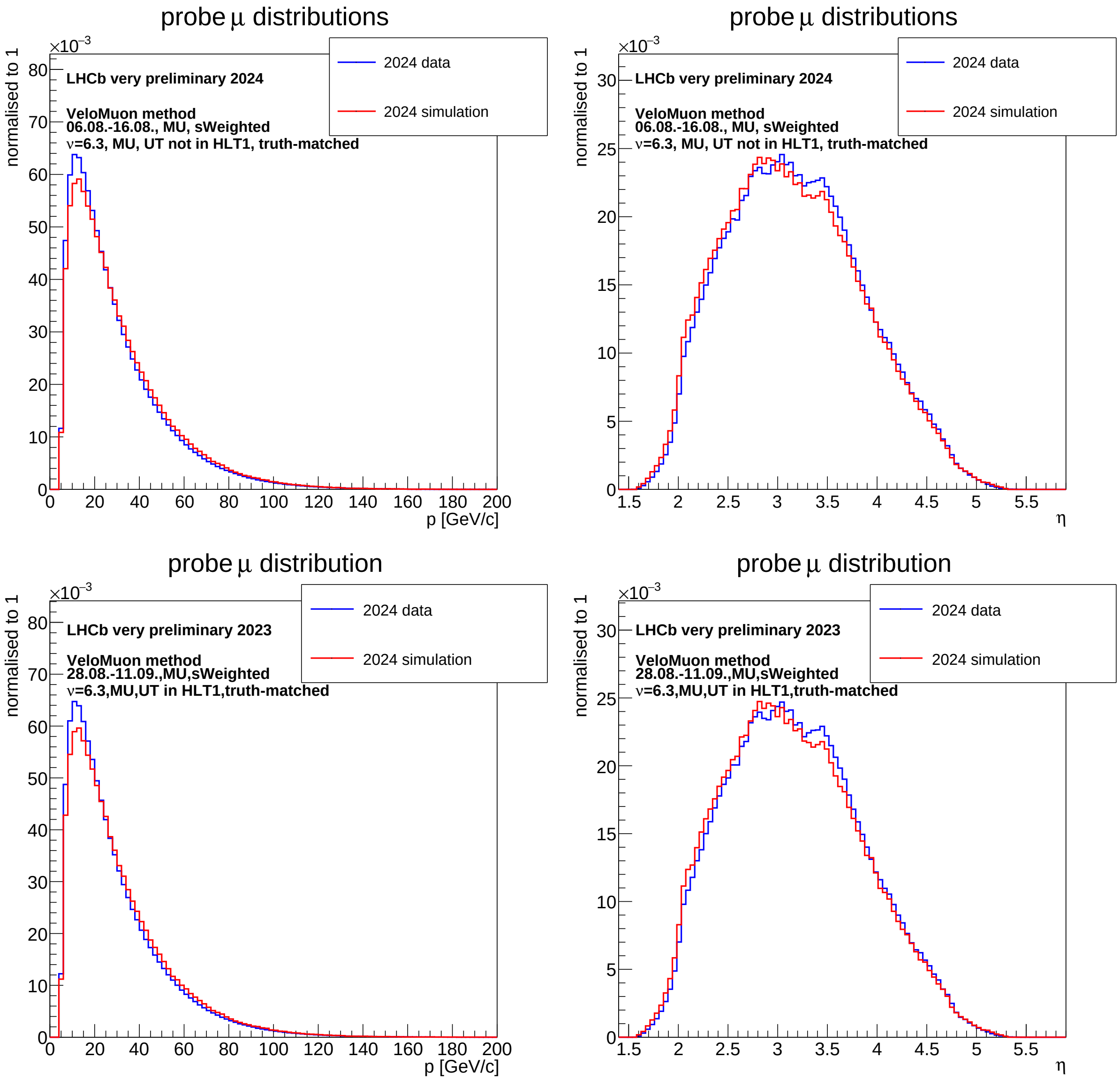


Figure A.2: Distribution of VeloMuon probe momentum, p, (left) and pseudorapidity of the VeloMuon probe muon tracks, η, (right) in sWeighted data and truth-matched simulation for data set 1 (above) and data set 2 (below)

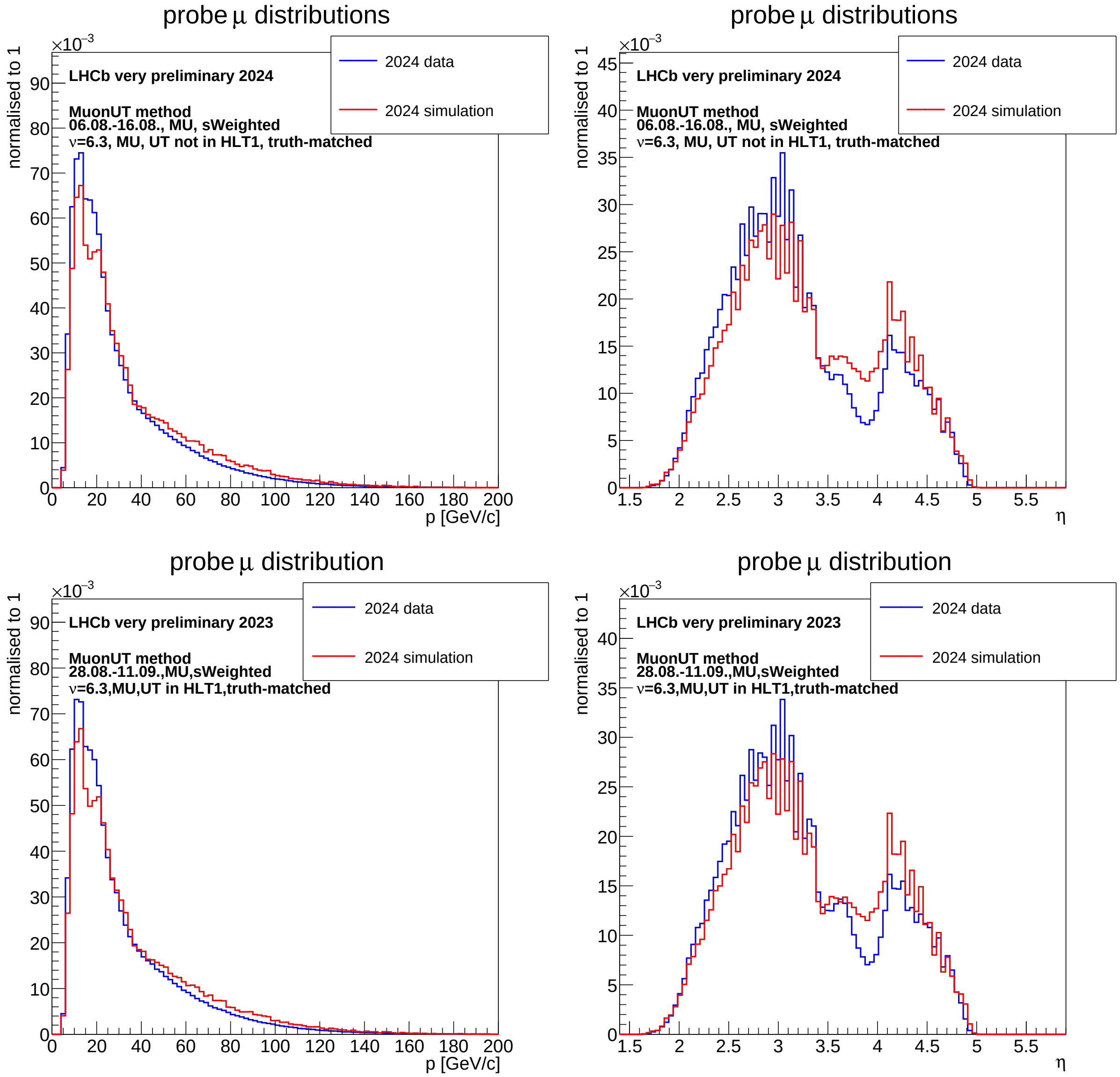


Figure A.3: Distribution of MuonUT probe momentum, p, (left) and pseudorapidity of the VeloMuon probe muon tracks, η, (right) in sWeighted data and truth-matched simulation for data set 1 (above) and data set 2 (below)

Bibliography

[1] David Chapple. *Particle Physics Dark Matter and Dark Energy.* arima publishing, United Kingdom, 2011. ISBN 978-1-84549-477-3. doi: 10.1017/CBO9781139525367.

[2] Mark Thomson. *Modern particle physics.* Cambridge University Press, New York, 2013. ISBN 978-1-107-03426-6. doi: 10.1017/CBO9781139525367.

[3] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024. doi: 10.1103/PhysRevD.110.030001.

[4] Esma Mobs. The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019. 2019. URL `https://cds.cern.ch/record/2684277`. General Photo.

[5] R. Aaij et al. The LHCb Upgrade I. *Journal of Instrumentation*, 19(05):P05065, May 2024. ISSN 1748-0221. doi: 10.1088/1748-0221/19/05/p05065. URL `http://dx.doi.org/10.1088/1748-0221/19/05/P05065`.

[6] The LHCb Collaboration. LHCb Operations Plots Webpage, 2024. URL `https://lbgroups.cern.ch/online/OperationsPlots/index.htm`. Accessed: 2024-12-11.

[7] S. Amato et al. *LHCb magnet: Technical Design Report.* Technical design report. LHCb. CERN, Geneva, 2000. URL `https://cds.cern.ch/record/424338`.

[8] M. Vesterinen. Considerations on the LHCb dipole magnet polarity reversal. Technical report, CERN, Geneva, 2014. URL `https://cds.cern.ch/record/1642153`. On behalf of the LHCb collaboration.

[9] LHCb Collaboration. LHCb VELO Upgrade Technical Design Report. Technical report, 2013. URL `https://cds.cern.ch/record/1624070`.

[10] Maximilien Brice. LHCb's VELO metrology. 2022. URL `https://cds.cern.ch/record/2802195`. General Photo.

[11] R. Ferreira. LHCb VeLo vacuum recovery and vacuum safety system for 2023, 2023. URL `https://indico.cern.ch/event/1261512/#2-lhcb-velo-vacuum-recovery-an`. Accessed: 2024-12-28.

[12] LHCb Collaboration. LHCb Tracker Upgrade Technical Design Report. Technical report, 2014. URL `https://cds.cern.ch/record/1647400`.

[13] LHCb Collaboration. LHCb PID Upgrade Technical Design Report. Technical report, 2013. URL `https://cds.cern.ch/record/1624074`.

[14] RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector . 2020. URL `https://cds.cern.ch/record/2730181`.

[15] Florian Reiss. Real-time alignment procedure at the LHCb experiment for Run 3. 2023. URL `http://cds.cern.ch/record/2846414`.

[16] I. Belyaev, G. Carboni, N. Harnew, C. Matteuzzi, and F. Teubert. The history of LHCb. *The European Physics Journal H*, 46, 2021. doi: 10.1140/epjh/s13129-021-00002-z.

[17] R. Aaij et al. Allen: A High-Level Trigger on GPUs for LHCb. *Computing and Software for Big Science*, 4(1), April 2020. ISSN 2510-2044. doi: 10.1007/s41781-020-00039-7. URL `http://dx.doi.org/10.1007/s41781-020-00039-7`.

[18] Daniel Hugo Cámpora Pérez, Niko Neufeld, and Agustín Riscos Núñez. Search by triplet: An efficient local track reconstruction algorithm for parallel architectures. *Journal of Computational Science*, 54:101422, September 2021. ISSN 1877-7503. doi: 10.1016/j.jocs.2021.101422. URL `http://dx.doi.org/10.1016/j.jocs.2021.101422`.

[19] Run3 Primary Vertex reconstruction description and performance . 2020. URL `https://cds.cern.ch/record/2714851`.

[20] Placido Fernandez Declara, Daniel Hugo Campora Perez, Javier Garcia-Blas, Dorothea Vom Bruch, J. Daniel Garcia, and Niko Neufeld. A Parallel-Computing Algorithm for High-Energy Physics Particle Tracking and Decoding Using GPU Architectures. *IEEE Access*, 7:91612–91626, 2019. ISSN 2169-3536. doi: 10.1109/access.2019.2927261. URL `http://dx.doi.org/10.1109/ACCESS.2019.2927261`.

[21] André Günther. Track Reconstruction Development and Commissioning for LHCb's Run 3 Real-time Analysis Trigger, 2023. URL `https://cds.cern.ch/record/2865000`. Presented 05 Jul 2023.

[22] Alessandro Scarabotto. Tracking on GPU at LHCb's fully software trigger. 2022. URL https://cds.cern.ch/record/2823783.

[23] P. Billoir, M. De Cian, P.A. Günther, and S. Stemmle. A parametrized Kalman filter for fast track fitting at LHCb. *Computer Physics Communications*, 265:108026, August 2021. ISSN 0010-4655. doi: 10.1016/j.cpc.2021.108026. URL http://dx.doi.org/10.1016/j.cpc.2021.108026.

[24] L. Anderlini, F. Archilli, A. Cardini, V. Cogoni, M. Fontana, G. Graziani, N. Kazeev, H. Kuindersma, R. Oldeman, M. Palutan, M. Santimaria, B. Sciascia, P. De Simone, and R. Vazquez Gomez. Muon identification for LHCb Run 3. *Journal of Instrumentation*, 15(12):T12005–T12005, December 2020. ISSN 1748-0221. doi: 10.1088/1748-0221/15/12/t12005. URL http://dx.doi.org/10.1088/1748-0221/15/12/T12005.

[25] Núria Valls Canudas, Míriam Calvo Gómez, Xavier Vilasís-Cardona, and Elisabet Golobardes Ribé. Graph Clustering: a graph-based clustering algorithm for the electromagnetic calorimeter in LHCb. *Eur. Phys. J. C*, 83(2):179, 2023. doi: 10.1140/epjc/s10052-023-11332-1. URL https://cds.cern.ch/record/2846012. 12 pages, 9 figures, submitted to EPJ C.

[26] S. Miglioranzi, M. Clemencic, C. Corti, S. Easo, C.R. Jones, M. Pappagallo, and P. Robbe. The LHCb Simulation Application, Gauss: Design, Evolution and Experience. Technical report, CERN, Geneva, 2011. URL https://cds.cern.ch/record/1322402.

[27] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to PYTHIA 8.1. *Computer Physics Communications*, 178(11):852–867, 2008. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2008.01.036. URL https://www.sciencedirect.com/science/article/pii/S0010465508000441.

[28] David J. Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152–155, 2001. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(01)00089-4. URL https://www.sciencedirect.com/science/article/pii/S0168900201000894. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[29] J. Allison. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006. doi: 10.1109/TNS.2006.869826. URL https://cds.cern.ch/record/1035669.

[30] R Antunes-Nobrega et al. *LHCb computing: Technical Design Report*. Technical design report. LHCb. CERN, Geneva, 2005. URL `https://cds.cern.ch/record/835156`. Submitted on 11 May 2005.

[31] W. Verkerke and D. Kirkby. RooFit Users Manual v2.91, 2019. URL `https://root.cern/download/doc/RooFit_Users_Manual_2.91-33.pdf`. Document version 2.91-33, revised September 18th, 2019.

[32] M. Pivk and F.R. Le Diberder. sPlot: A statistical tool to unfold data distributions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1–2):356–369, December 2005. ISSN 0168-9002. doi: 10.1016/j.nima.2005.08.106. URL `http://dx.doi.org/10.1016/j.nima.2005.08.106`.

[33] Alex Rogozhnikov. Reweighting with Boosted Decision Trees. *Journal of Physics: Conference Series*, 762:012036, October 2016. ISSN 1742-6596. doi: 10.1088/1742-6596/762/1/012036. URL `http://dx.doi.org/10.1088/1742-6596/762/1/012036`.

[34] N. L. Johnson. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika*, 36(1/2):149–176, 1949. ISSN 00063444, 14643510. URL `http://www.jstor.org/stable/2332539`.

[35] F. James. MINUIT - Function Minimization and Error Analysis Reference Manual (PDF) (Report). URL `https://root.cern.ch/download/minuit.pdf`. downloaded 27.10.2024.

[36] Peilian Li, Eduardo Rodrigues, and Sascha Stahl. Tracking Definitions and Conventions for Run 3 and Beyond. Technical report, CERN, Geneva, 2021. URL `https://cds.cern.ch/record/2752971`.

[37] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino. Data driven trigger efficiency determination at LHCb. Technical report, CERN, Geneva, 2014. URL `https://cds.cern.ch/record/1701134`.