university of groningen

faculty of science and engineering

Testing the Validity of Run 3 LHCb Data Utilising the $\Lambda_b^0 \to \Lambda^0 J/\psi$ Decay

Author:
Emma VAN OIJEN
(s4959841)

Supervisor:
Dr. Mick MULDER
Second examiner:
Prof. Dr. Steven HOEKSTRA

July 9, 2024

Abstract

During the last long shutdown of the LHC from 2019 to 2022, the LHCb detector had a major upgrade of both detector components and event trigger software which should allow a significant increase in the amount of available statistics [1]. Data belonging to the $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay has been analysed to determine whether the detector operates as anticipated such that the rare $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ decay can be investigated in the search for beyond the standard model physics [2]. The data taken in 2023 is statistically very limited and shows an unexpected angular dependence as a result of the compromised position of the VELO subdetector. A comparison by means of the Kolmogorov-Smirnov test between MC simulation and the first 2024 measurements as well as between matter and antimatter has yielded no issues in the 2024 experimental data. The former method is statistically different with mitigating factors yielding no definitive conclusion and the latter is supportive of the validity and usability of the data. The yield per integrated luminosity with a cut-based signal selection is in agreement with the expected value, while a deviation occurs once sideband subtraction is included. Possible explanations include that too much signal is caught in the sideband or that the efficiency determined by MC simulation is overestimated.

Acknowledgments

Special appreciation is in place for Dr. Mick Mulder who provided data and sources and guided by suggesting possibilities to research and potential solutions to questions that arose. Moreover, the input from the LHCb research group in Groningen improved the quality by pointing out important aspects to look at. Finally, the author is grateful to all people people at or related to CERN involved in the design and operation of the LHC and the LHCb detector as well as the production of the software required to obtain the data and previous results this thesis is based on.

Contents

1 Introduction

The Large Hadron Collider (LHC) at CERN is the most powerful particle accelerator currently available [3]. Proton bunches are accelerated and collide at one of the four detectors. Of particular interest is the LHC beauty (LHCb) detector, where heavy-flavour physics is the primary reason for the existence of this forward spectrometer [4]. The main goal is the investigation of the asymmetry between matter and antimatter in search for an explanation to the dominance of matter in the universe [5]. For this investigation new physics needs to be discovered in the form of more extensive CP violation for instance [6]. Furthermore, new physics is investigated through studies on lepton flavour universality, dark matter and lepton flavour violation. However, the capabilities of the detector are beneficial to electroweak, heavy ion and fixed target physics as well, and it is currently perceived as a comprehensive experiment covering the forward direction [4].

During the last long shutdown (LS2) of the LHC in 2019 to 2022, the LHCb detector had a major upgrade [1]. The central improvement is to allow an increment in the luminosity. This is achieved by an increased rate of event detection and storage in combination with a fully software-based trigger allowing for a more sophisticated selection of potentially useful events [4]. Analyses involving rare decays are expected to benefit from this upgrade, as they suffer from a lack of sufficient data. Furthermore, due to multiple system failures on 10 January 2023, there was an accumulation of pressure between the vertex locator (VELO) and beam volumes causing a plastic deformation of the thin aluminium RF foil separating the volumes [1]. As such, the VELO could not be fully closed during the 2023 data collection, compromising the detector, which was resolved before the 2024 data collection period. Consequently, it is important to ascertain the validity of the recorded data before analysing it in more detail.

The $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay, antimatter counterparts such as the $\overline{\Lambda_b^0} \to \overline{\Lambda^0} J/\psi$ decay are implied to be included unless stated otherwise, contributes to the current understanding of physics. The decay amplitudes have been determined from the angular distributions which can be described by the polarisation of the Λ_b^0 baryon [7, 8]. Moreover, heavy-quark effective theory (HQET) predicts the conservation of the transverse polarisation of high-energy b quarks when the Λ_b^0 baryon is formed. Finally, perturbative quantum chromodynamics (QCD) can be applied too, providing a systematic approach to study its characteristics [9].

Additionally, the validation of the $\Lambda_b^0 \to \Lambda^0 J/\psi$ data would allow the investigation of the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ decay which could lead to new physics beyond the Standard Model (SM) [2]. The Λ_b^0 decay, similar to other $b \to s\mu^+\mu^-$ decays by B mesons, is a Flavour Changing Neutral Current (FCNC) as there is a flavour change without a change in its charge [10]. The complicated processes through which FCNCs occur potentially have contributions from new physics. Those interactions that are forbidden at tree level provide a good test for the SM through forward-backward asymmetry measurements. Other observables that show a minor deviation from the SM are related to the P_5' anomaly and the branching fraction. As a result, examining the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ decay is a valuable future endeavour and the collected data must be validated before inferring any conclusions. The $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay provides a useful calibration method since the decays are very similar in terms of particles.

The main objective of this thesis is to determine the validity and usability of the 2023 and 2024 LHCb data. As such, a comparison is made between the 2023 and 2024 data and checked for any physical abnormalities. Furthermore, a comparison between the recorded data and a Monte Carlo (MC) simulation can aid in the detection of any inconsistencies which also applies to a separation between matter and antimatter. To conclude, the fraction of signal events over the integrated luminosity will be calculated and compared to the product of the detector efficiency determined by the MC simulations, the branching fractions, the production cross-section and correction factors.

2 Background Literature

2.1 The LHCb Detector

The LHCb detector shown in Figure 1 operates in the pseudorapidity range $2 < \eta < 5$ [4]. Particles travelling perpendicular to the beam axis z have $\eta = 0$ while the pseudorapidity tends towards infinity when the propagation direction is close to the beam axis. This figure can be determined by means of the polar angle θ between the particle track and the beam axis as well as by the total ($\vec{p}$) and longitudinal (p_L) momentum as shown by Formula 1 [11]:

$$\eta = tanh^{-1}\left(\frac{p_L}{|\vec{p}|}\right) = \frac{1}{2}ln\left(\frac{|\vec{p}|+p_L}{|\vec{p}|-p_L}\right) = -ln\left(tan\left(\frac{\theta}{2}\right)\right) \tag{1}$$

The proton beams propagate in both the z and -z direction and collide inside the Vertex Locator (VELO) which consists of hybrid silicon pixel detectors to record produced ionising particles [4]. The tracks are measured which provides information on the location of the interaction vertices, displaced decay vertices and the distances between them. Subsequently, there is a silicon strip upstream tracker (UT) which detects charged particle tracks. In combination with the VELO this can provide a first momentum estimation. Furthermore, it reduces the amount of fake tracks in the VELO and provides measurements of particles which decay outside the VELO such as the Λ^0. Finally, it speeds

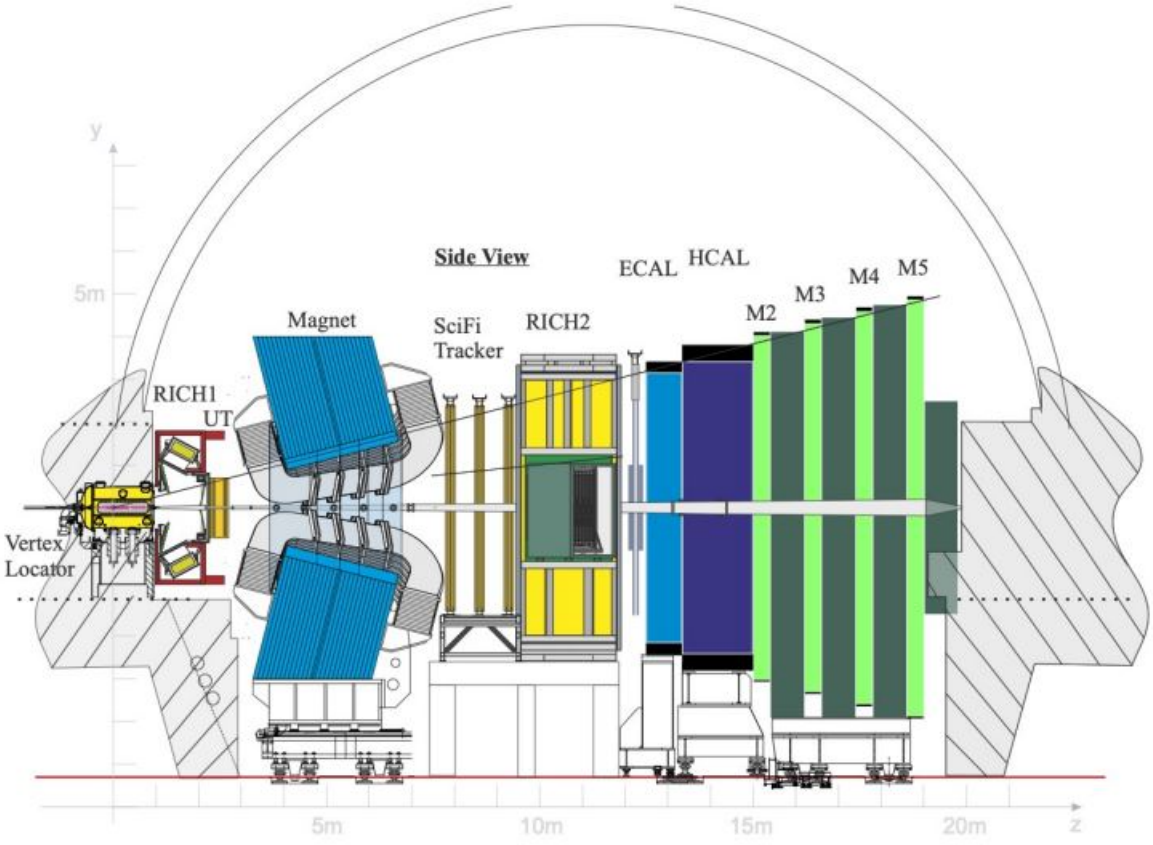

Figure 1: *The LHCb detector after the first major upgrade. Image taken from Ref. [4].*

up matching with the scintillator fibre (SciFi) tracker hits. Since the UT was installed late and was thus still being commissioned in 2023 [1], it is not utilised for detection for the data in this thesis. The three SciFi trackers located after the magnet complete the charged particle tracking after the VELO and the UT [4] and provide the momentum of the particle. The two ring imaging Cherenkov (RICH) detectors, the electromagnetic (ECAL) and hadronic (HCAL) calorimeters and the muon systems (M2-M5) separated by iron shielding provide particle identification. RICH provides hadron discrimination by distinguishing between pions, kaons and protons. The second detector can detect particles with a higher momentum than the first detector. The calorimeters bring the particles to a halt which results in particle showers whose energy and spatial location can be detected [12]. The ECAL detects electrons and photons while the HCAL detects protons, pions, kaons and neutrons. The muons traverse the calorimeters losing small amounts of energy by ionisation and are subsequently detected by the multi-wire proportional chambers (MWPC) in the muon systems [4].

Depending on which subdetectors observe a particle, the tracks are classified as shown in Figure 2. The VELO track belongs to particles with a pseudorapidity below 2 which only have a VELO track segment and then leave the detector. If the pseudorapidity is above 5, the particle will not be detected at all since it propagates close to the beam line. Particles following the upstream track are detected by the first Cherenkov detector and the UT but leave the detector afterwards, often due to deflection by the magnetic field. Downstream and T tracks do not have a VELO track segment but are detected in the subdetectors that follow. For downstream tracks this starts before the magnet in the RICH detector and the UT while the T track is only detected after the magnet. A possible cause is the decay of a particle with a relatively long life time. As such, the decay products will not have any VELO hits. The decaying particle may not be charged and will not be detected by the VELO in that case. Finally, the long track is ideal for observations as all subdetectors which can theoretically measure the particle actually detect it and provide information. This is the only type of track utilised in this thesis, since the data taking including downstream tracks initiated later in 2024 due to the UT as mentioned above. Hence, these tracks could not be incorporated in this investigation.

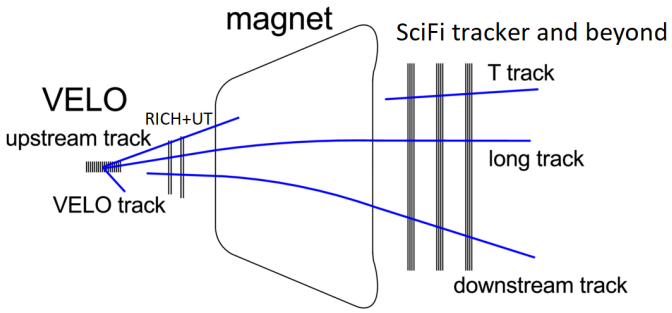

Figure 2: *Defined track types in the LHCb detector. Image adapted from Ref. [13].*

2.2 The Software Trigger

The selection of any decay of interest commences with two software triggers determining what detector information is stored permanently offline [4]. The first High Level Trigger (HLT1) consists of a Global Event Cut (GEC) and an inclusive selection based on a partial event reconstruction. The GEC removes any event with a large number of tracks as the computing time is disproportionate, especially considering the worse detector performance. The following selection is aimed at reducing

the necessary quantity of data storage while combining requirements on tracks or two-track vertices displaced from the primary vertex as well as any lepton detected, but with a particular focus on muons. At the end of this stage, the remaining data is temporarily stored on an online disk buffer awaiting the real-time alignment and calibration. The storage also allows processing to take place in between proton fills allowing more time for HLT2. During the second trigger stage, a full event reconstruction of offline quality takes place such that approximately one thousand selection algorithms based on different physics analyses and/or signal topologies can decide which events go to permanent storage. About half of the events is thrown away in this process, the remaining data reduction is achieved by cutting on the amount of information stored per event [8]. To conclude, software triggers are utilised to reduce the event data to a quantity that can be stored offline indefinitely.

2.3 Monte Carlo Simulations

Monte Carlo simulations are widely utilised to simulate complex systems and are based on random sampling and probability density functions (PDFs) [14]. In general, a system is modelled as a series of PDFs which is sampled repeatedly and subsequently the desired statistic is computed deterministically. Hence, in order to predict the detection properties of LHCb, pp collisions are generated in the GAUSS framework by PYTHIA [15]. Afterwards, the decays of unstable particles are modelled by EVTGEN. Finally, GAUSS is responsible for simulation of the propagation of the generated particles through all sub-detectors and their interactions utilising GEANT4. When similar information as the output of the detector has been obtained, the same algorithms in the software trigger are applied. Ultimately, the data can be examined offline analogous to experimental measurements.

2.4 The $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay

In this thesis, the Λ_b^0 baryon decay to a Λ^0 baryon and a J/ψ meson will be studied, which has a branching fraction of $(3.08 \pm 0.26) \cdot 10^{-4}$ [16]. Subsequently, the J/ψ decays almost immediately into a μ^+ and a μ^- with a branching fraction of $(5.961 \pm 0.033)\%$ [17]. As such, there is no displaced J/ψ vertex [16]. In contrast, the Λ^0 has a much larger mean life of (261.7 ± 1.0) ps [17] causing most particles to decay downstream outside the VELO [16]. The decay products utilised for reconstruction are p^+ and π^- which occur with a branching fraction of $(64.1 \pm 0.5)\%$ [17]. The process is shown in Figure 3.

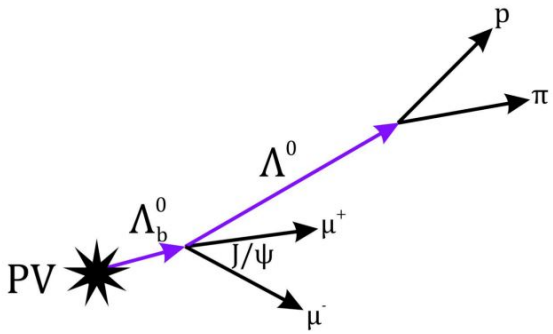

Figure 3: *Graphical representation of the $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay. Image taken from Ref. [16].*

3 Methods

In general, 2024 data sets are utilised with the magnet polarisation down including both the Λ_b^0 and $\overline{\Lambda_b^0}$ decay unless specified otherwise. Furthermore, for every histogram, errors bars are included, with their size given by $\sqrt{N}$ with N the number of events in each bin. Afterwards, the distributions are normalised for the comparison visibility between data sets of different size.

3.1 Signal determination

In order to determine the actual signal in the data and MC files provided for the $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay, several cuts are applied for this study. The fiducial cuts, implemented to ensure a realistic reconstruction, are based on physics requirements. As the momentum of the particles is mostly in the z-direction along the beam line, it is required that the Λ_b^0 decay vertex has a smaller z-coordinate than the Λ^0 vertex. Moreover, the z-coordinate of the Λ^0 vertex needs to be between -30 mm and 650 mm such that it could actually have been detected by the VELO. Furthermore, the Λ^0 lifetime is requested to be larger than zero. Due to detection and processing uncertainties, the recorded lifetime of a real signal could technically become negative, however, as the chance of background by another particle is also very high around a negligible lifetime [18], those events are still excluded. A higher cutoff was considered as it was utilised by The LHCb Collaboration in Ref. [7] for instance. Nevertheless, that seemed to be too non-distinctive between signal and background, thus mostly reducing the statistics. Additionally, the cosine of the angle between the momentum and flight direction is set to be larger than zero for Λ^0 such that the few events where the momentum and flight direction are almost opposite are excluded while all events which have this value close to one, i.e. the angle close to zero degrees, are included. There were no events detected near zero since all events were either close to one or negative one. This was not done for Λ_b^0 since no negative values were visible in the plot while for the J/ψ a negative value could be caused by the resolution due to the almost immediate decay. Finally, the particle identification of the muons had to be larger than 0 in comparison with an electron which could have been misidentified. Even though this does not lead to a major reduction in data which is perceived as signal, it could remove a few background events. On top of the fiducial cuts, restrictions were placed on the mass of the J/ψ, Λ^0 and Λ_b^0 particles as well. Based on the peaks visible in the graphs in combination with the expected value by the Particle Data Group (PDG) [17] the permitted ranges were determined to be as outlined in Table 1:

Particle	PDG mass (MeV/c^2) [17]	Lower bound mass (MeV/c^2)	Upper bound mass (MeV/c^2)
J/ψ	3096.900 ± 0.006	3000	3200
Λ^0	1115.683 ± 0.006	1110	1120
Λ_b^0	5619.60 ± 0.17	5550	5670

Table 1: *Mass ranges utilised for the cuts on the J/ψ, Λ^0 and Λ_b^0 masses for signal selection.*

The cut on the Λ^0 is more strict than the one on J/ψ due to the spread of the peak visible in the data in Figure 4. To conclude, a number of fiducial and mass cuts have been utilised as signal selection cuts.

In order to distinguish the Λ_b^0 and the $\overline{\Lambda_b^0}$ decays, an additional selection is required for which the final decay products are utilised. For the matter decay, the proton should have a positive charge while the pion should be negative. Conversely, the antimatter decay is expected to have a negatively charged

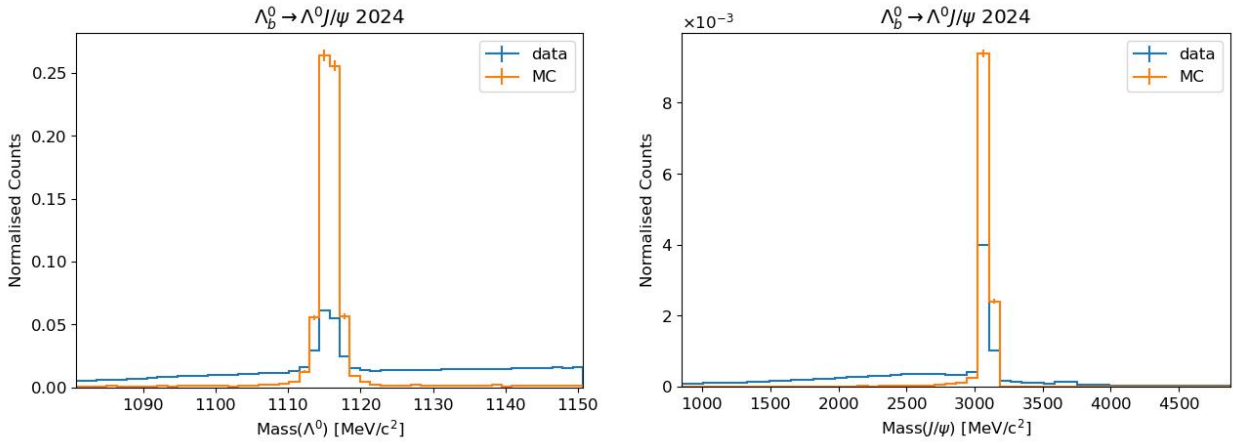

Figure 4: *Normalised histograms of the Λ^0 and J/ψ mass of the 2024 data and MC simulation with only fiducial cuts.*

proton and a positively charged pion. These conditions on the recorded data do not introduce an additional background cut since the decays containing an equal proton and pion charge do not pass the trigger stage for this decay. As such, proton and pion charge are utilised to separate the matter and antimatter decays.

3.2 Kolmogorov-Smirnov Test

The two-sample Kolmogorov-Smirnov (KS) test was utilised to quantitatively establish the overlap between MC simulations and data as well as matter and antimatter. First, the conditions of the test must be met, which are that the samples must have been taken randomly from their population, the samples should be mutually independent and the variable should at least be ordinal, preferably continuous [19]. All variables utilised are ordinal and only few variables such as the number of primary vertices are discontinuous. The continuity of the data is an important argument to utilise this test since it does not rely on binning in contrast to the χ^2-test for instance. The data taking and MC simulation processes do not depend on each other, it is only the detector that has been simulated which thus influences both underlying probability distributions. However, the random sample drawn from that probability distribution does not depend on the other. Furthermore, under the assumption that Λ_b^0 and $\overline{\Lambda_b^0}$ do not tend to be produced simultaneously, there is no reason to believe that the detection of one alters the detection of the other. Finally, even though the cuts eliminate the randomness of the samples, the events can still be considered random but corresponding to a different probability function as the same cuts are applied to all samples. In conclusion, one can use the KS test to determine the similarity of both data and MC along with matter and antimatter.

The procedure of the test contains two mutually excluding hypotheses, a statistical measure and a value to compare it to depending on the significance and the number of data points [19]. The null hypothesis H_0 says the two underlying probability distributions are equal while the alternative hypothesis H_1 states they are unequal. Subsequently, an empirical (cumulative) distribution function (E(C)DF) is created from both data sets. This function starts at zero and at every data point it increases by $\frac{1}{N_{data}}$ thus going to one when the maximum value is reached [20]. The two-sided Kolmogorov-Smirnov statistic is the maximum absolute distance between the two EDFs. This value is calculated by Python's scipy kstest function [21] as well as the p-value which provides a simple metric to deter-

mine whether the null hypothesis should be rejected. For an $\alpha = 0.05$ significance, a p-value below 0.05 means that H_0 should be rejected and the distributions are different with a 95% certainty. If it is higher, no definite conclusion can be made, however, it does suggest similarity between the compared graphs. Hence, the KS test can also serve to see for which variables or particles the data and MC simulation or the matter and antimatter decays are more similar than for others. To conclude, the KS test provides a useful metric to determine and compare the similarity of the data.

3.3 Yield per Integrated Luminosity

The quantity of signal events with respect to the integrated luminosity was determined experimentally from LHCb data. The yield (N) was obtained utilising the selection outlined in Section 3.1 with a $\sqrt{N}$ error. A sideband subtraction on the Λ_b^0 mass was included as well. As the total range is the same as the signal range the background counts can simply be deducted. Due to the large reduction in the yield, the value was obtained both with and without sideband subtraction. The background was taken to be from 5490 MeV/c^2 to 5550 MeV/c^2 on the lower side and 5670 MeV/c^2 to 5730 MeV/c^2 on the upper side as can be observed in Figure 5.

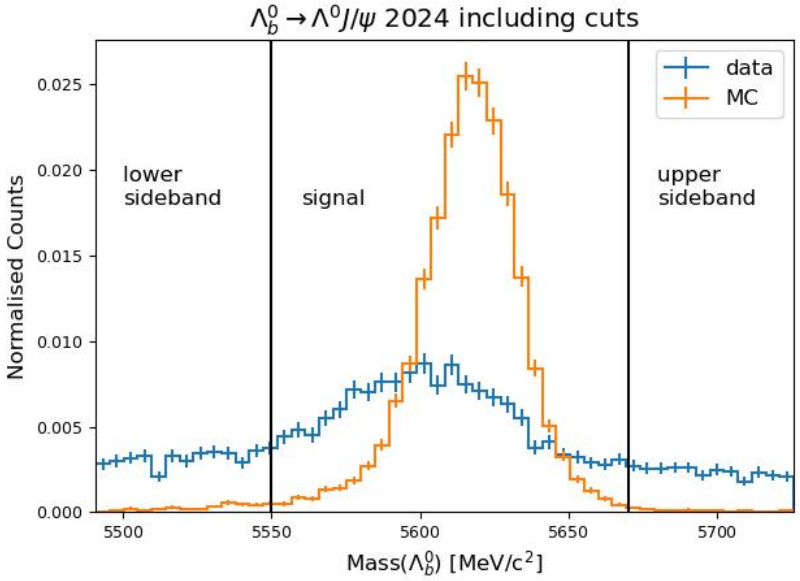

Figure 5: *Normalised histogram of the Λ_b^0 mass of the 2024 data and MC simulation with signal selection cuts except for the Λ_b^0 mass cut to illustrate the sidebands to be subtracted.*

The integrated luminosity ($\mathcal{L}_{int}$) was measured separately by PLUME. The error utilised, which is an upper bound, is 3% [8]. Furthermore, only run numbers for which both the yield and integrated luminosity are available are taken into account. This means that at least one event must pass the trigger for the run to be included even in the case no events are left after signal selection. Although this could lead to a bias in the results since a valid run may simply have a yield of zero. The bias as a result of including them would be much larger since runs are discarded if the data is deemed unfit for physics analyses or runs are simply not investigated in this study while the luminosity is still available. In conclusion, the integrated luminosity and the amount of events are experimentally obtained.

The expected number of events corrected for the integrated luminosity is determined utilising the efficiency, branching fractions, a cross section and correction factors according to Formula 2:

$$\frac{N}{\mathcal{L}_{int}} = \varepsilon_{MC}\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)\mathcal{B}(\Lambda^0 \to p^+\pi^-)\mathcal{B}(J/\psi \to \mu^+\mu^-)\frac{f_{\Lambda_b}}{f_{B^+}}\sigma_{B^+} f_{\sqrt{s}} \tag{2}$$

The efficiency of the detector, trigger system and signal selection is determined by the the fraction of signal left after the MC simulation of 1 million events, hence denoted as ε_{MC}. The signal selection takes place according to the cuts described in Section 3.1 and the value is obtained including and excluding the same symmetric sideband subtraction as for the yield. The error is determined by taking the square root of the signal count and dividing it by 1 million. The MC simulation assumes that every Λ_b^0 decays into $\Lambda^0 J/\psi$ which in turn decays according to $\Lambda^0 \to p^+\pi^-$ and $J/\psi \to \mu^+\mu^-$. Nevertheless, in reality numerous decays are possible and hence the branching fractions ($\mathcal{B}$s) of these three decays should be taken into account. Moreover, in order to link this fraction of detected decays to the number of signal events per integrated luminosity, the Λ_b^0 production cross section needs to be included. However, as that figure is not readily available, while the B^+ production cross section σ_{B^+} as well as the fraction $\frac{f_{\Lambda_b}}{f_{B^+}}$ of b quarks forming a Λ_b^0 baryon compared to B^+ and $\overline{B^0}$ mesons are, those values are utilised. The B^+ cross section was measured at 13 TeV center of mass energy while current measurements were taken at 13.6 TeV. As those values are approximately proportional to each other there is an additional correction factor $f_{\sqrt{s}} = \frac{13.6}{13}$ to consider [8]. To conclude, the yield per integrated luminosity was determined utilising the MC efficiency, branching fractions, the B^+ production cross section and correction factors for the production cross section decay and center of mass energy following Formula 2.

4 Results

4.1 2023 and 2024 Measurements Comparison

In Figure 6 it can be observed that for the data obtained in 2023, not all angles allowed proper detection of the decay while this was resolved in 2024. The distributions are shown for the muon with the same charge as the proton; the distributions for other final state particles looked similar.

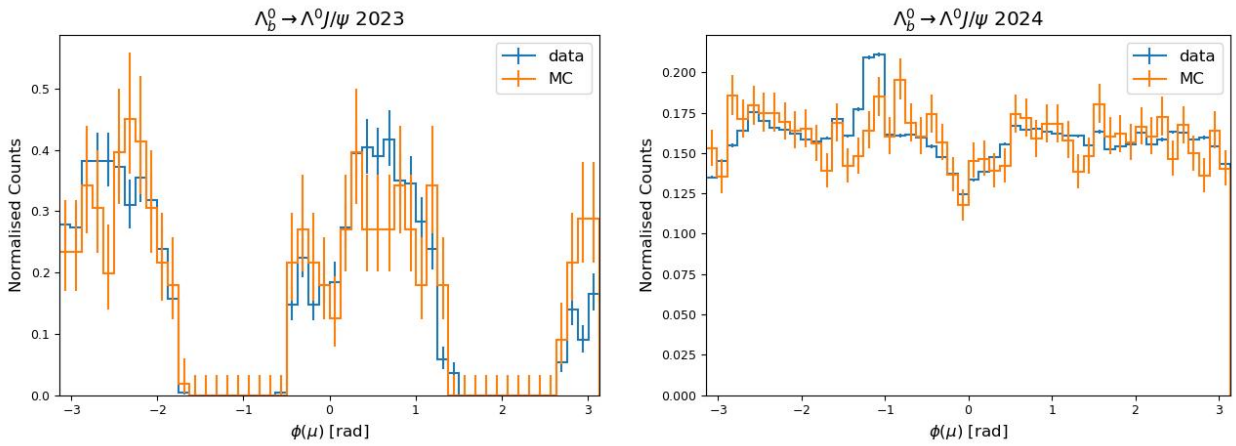

Figure 6: *Normalised histograms of the angle* ϕ *in the x,y-plane for the muon with the same charge as the proton for 2023 (left) and 2024 (right) data and MC simulation. No cuts have been applied other than those present in the trigger.*

The effect can be seen in Figure 7 for the Λ_b^0 particle as well, though at least some events are reconstructed for almost every angle. This is because the decay occurs in the VELO and the decay particles could still travel in the direction of the detectable angles with a limited probability. Another possibility is that the reconstruction is slightly off due to misalignment or imperfect calibration. A similar shape is visible for the J/ψ particle. In contrast, the Λ^0, which is also reconstructed by the final decay particles, has a longer life time and is thus probable to decay a larger distance from the primary vertex. The angular dependence is caused by the malfunctioning of the VELO while a VELO track

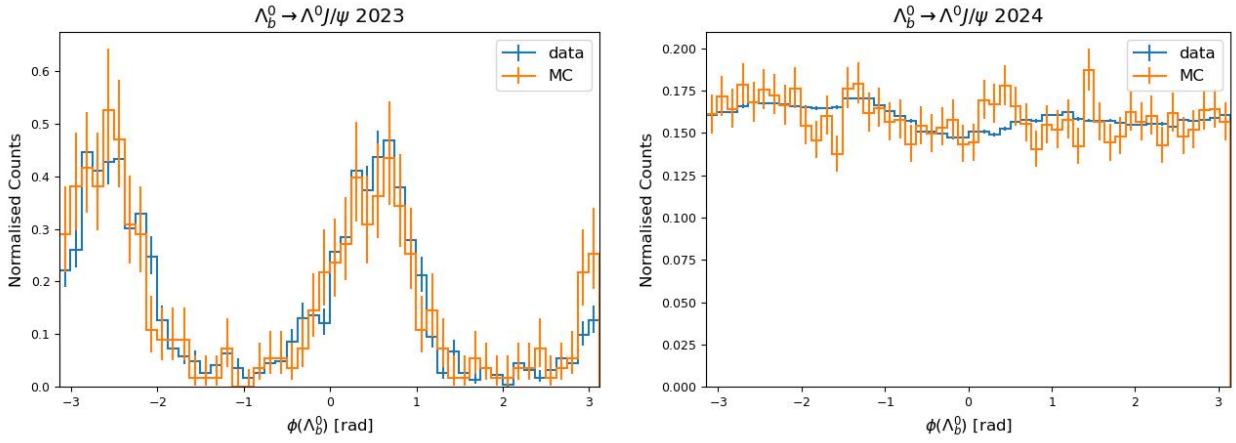

Figure 7: *Normalised histograms of the angle* ϕ *in the x,y-plane for* Λ_b^0 *for 2023 (left) and 2024 (right) data and MC simulation. No cuts have been applied other than those present in the trigger.*

segment of the proton and the pion is required since the UT was not utilised as outlined in Section 2.1. In combination with the small angle between the decay products due to the low mass difference between the Λ^0 and the proton and pion combined [8] as well as the conservation of momentum, the angle for the Λ^0 decay must already be almost correct for the charged decay products to be detectable by the VELO. Therefore, the angular distribution is similar to the final decay products as observed in Figure 6.

Similarly, when considering the x,y-plane transverse to the beam, the x and y momenta of the detected decays show that the rotational symmetry of the detector and hence the data is present in the 2024 data while being absent in the 2023 dataset. For instance, the Λ^0 particle shows this in Figure 8.

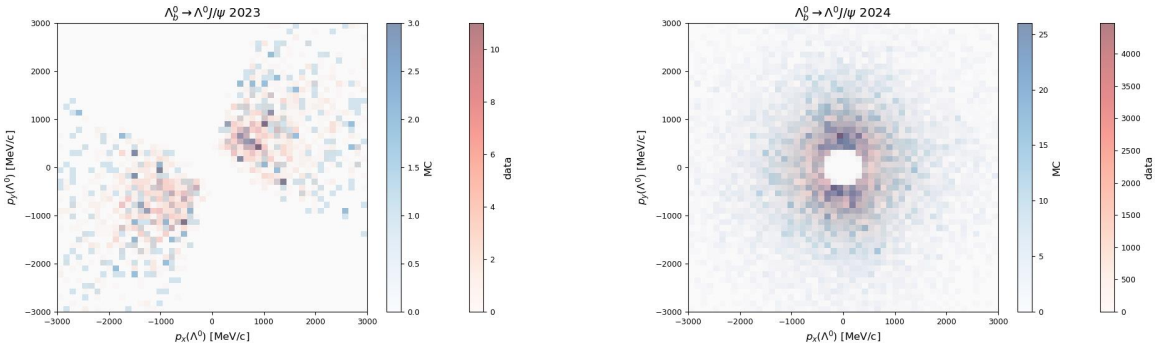

Figure 8: *Two dimensional histograms of the x and y momenta for Λ^0 for 2023 (left) and 2024 (right) data and MC simulation. No cuts have been applied other than those present in the trigger.*

Finally, the usability of the data depends substantially on the amount of statistics available. For the 2023 experimental data, 1786 events were detected and came through the trigger of which only 6 remained after applying the selection as described in Section 3.1. For this reason, only graphs without cuts were utilised in this section. On the other hand, $1,069,940$ events of 2024 data were utilised in this analysis resulting in 3824 being labelled as signal while data taking is still taking place at the time of writing and will thus increase.

4.2 Data and MC Comparison

In order to establish whether there are any unexpected events present in the data and whether the MC simulation provides an accurate description of the experimental data, the two data sets were plotted for all particles for numerous variables such as the mass (including and excluding fit to the primary vertex and J/ψ), momenta, pseudorapidity, the angle in the x,y-plane ϕ and the lifetime with respect to the primary vertex. As such, only a brief selection that can provide additional insights is shown.

Firstly, the Λ_b^0 mass in Figure 9 shows that a lot of background events are present in the experimental data set. The minimal peak around 5619.60 ± 0.17 MeV/c^2 [17] is practically invisible without any selection, while being wider and at a slightly lower mass than MC and PDG prediction. Imperfect calibration could cause the shift while the origin of the increased spread could be a lower than MC prediction resolution or precision. Moreover, more background could be included in the data peak, since the assumption of the presence of the $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay in collisions only holds for MC simulation. The background in the sidebands around the mass peak can be seen in more detail in Figure 5.

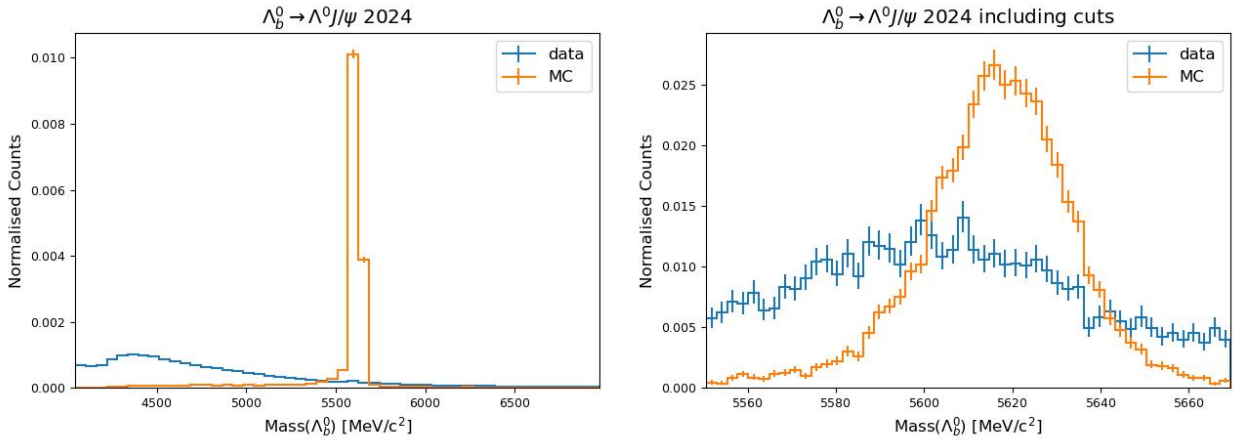

Figure 9: *Normalised histograms of the Λ_b^0 mass of the 2024 data and MC simulation with (right) and without (left) signal selection cuts. The p-values of the KS test are* 0.0 *(left) and* $1.4 \cdot 10^{-216}$ *(right).*

For a quantitative comparison, the Kolmogorov-Smirnov statistics and p-values were calculated. For most variables, both with and without the signal selection cuts, the p-value goes from just below 0.05, to the order of 10^{-321} or even given as 0.0. Hence, the null hypothesis needs to be rejected and the graphs cannot come from the same underlying probability distribution. An exception is the angle ϕ, with only two p-values below 0.05, 0.033 and 0.036 for the proton and the muon of the same charge respectively, both with cuts. The other values range from 0.10 to 0.65 both with and without cuts thus allowing the possibility of the same underlying distribution. The similarity can be observed in Figure 10:

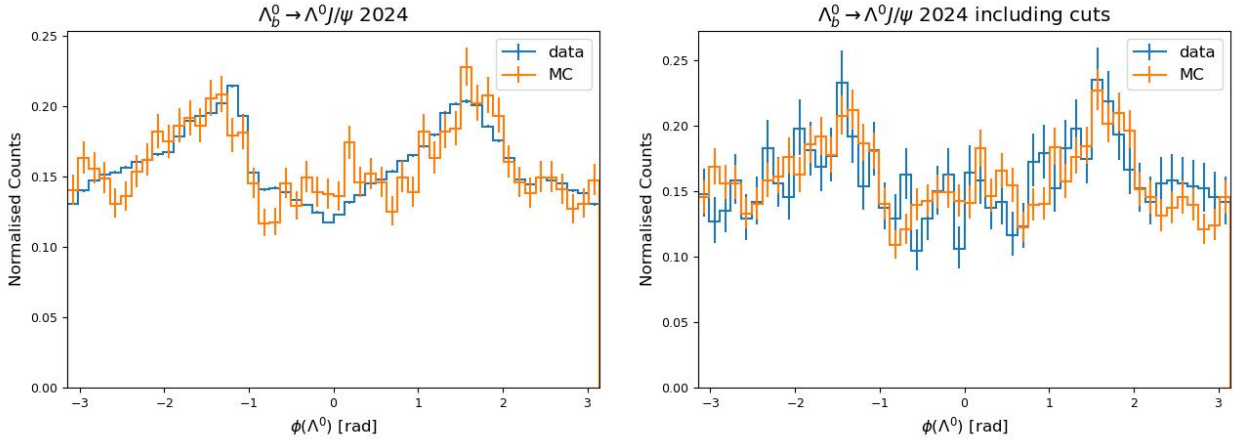

Figure 10: *Normalised histograms of the angle ϕ in the x,y-plane for Λ^0 2024 data and MC simulation with (right) and without (left) signal selection cuts. The p-values of the KS test are* 0.65 *(left) and* 0.23 *(right).*

Additionally, a few single values where cuts have been applied allow for a high similarity which are the mass of the first (0.13) and second (0.23) muon. These are expected to be very similar since the detection and the reconstruction are based on the mass of the final particles. Hence, it is determined precisely and a requirement for being saved as this decay so the only minimal spread can be caused by technical issues such as floating point numbers [8] or a very small mass acceptance window. Even though this is affirmed by the muons after signal selection, the KS test indicates the two distributions

to be different for the proton (0.0069) and pion (0.0021) mass even though the expected similarity is shown in Figure 11:

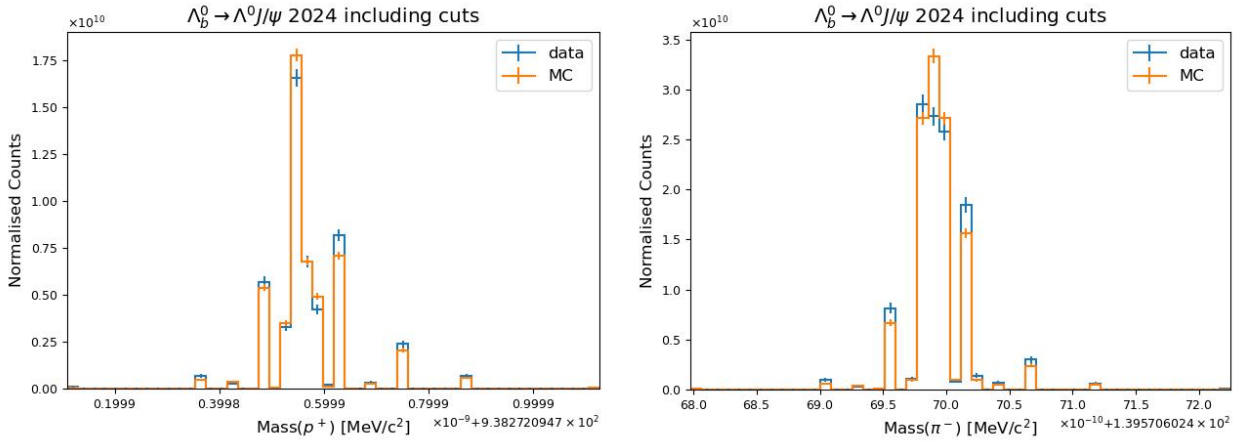

Figure 11: *Normalised histograms of the proton (left) and pion (right) mass utilising 2024 data and MC simulation with signal selection cuts. The p-values of the KS test are* 0.0069 *(left) and* 0.0021 *(right).*

Finally, the p-values seem to increase after the application of the signal selection cuts in general, although Figure 10 shows the opposite for instance.

4.3 Matter Antimatter Comparison

With regard to the occurrence and behaviour of matter and antimatter, it should be similar except for the charge. Nevertheless, some processes such as CP violation do differentiate between them [6]. Furthermore, there could be a small difference in the Λ_b^0 and $\overline{\Lambda_b^0}$ production via pp collisions [8], however, this has no effect on the distributions since all are normalised. The similarity can be

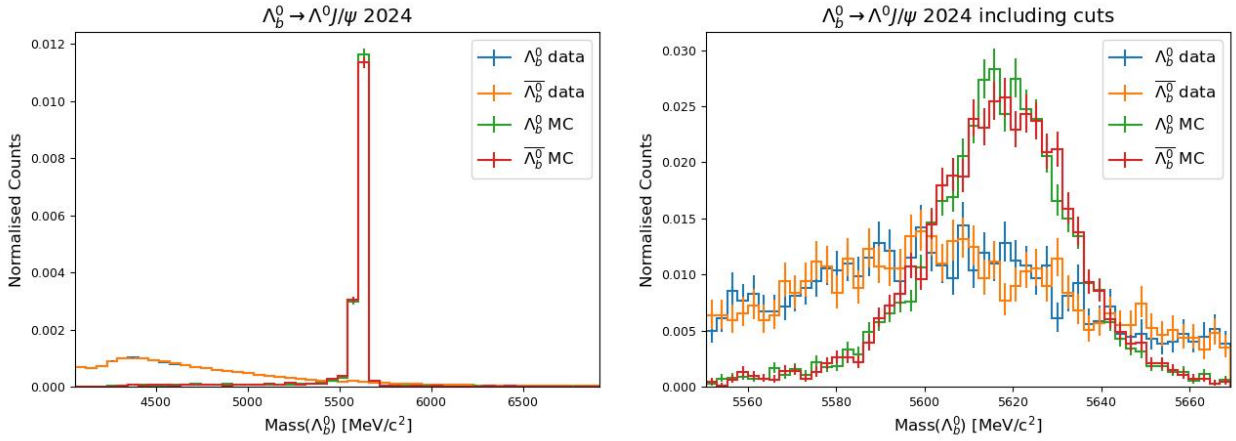

Figure 12: *Normalised histograms of the* Λ_b^0 *and* $\overline{\Lambda_b^0}$ *mass utilising 2024 data and MC simulation with (right) and without (left) signal selection cuts. The p-values of the KS test are* 0.12 *for data and* 0.30 *for MC simulation without cuts (left) while with cuts (right) they are* 0.64 *and* 0.25 *for data and MC respectively.*

observed in most variables for all particles such as the Λ_b^0 and $\overline{\Lambda_b^0}$ mass both before and after signal selection as shown in Figure 12 and by the p-values of the KS test. However, for the p_x and ϕ variables a difference can be observed between matter and antimatter for both experimental measurements and MC simulation such as for the pion in Figure 13.

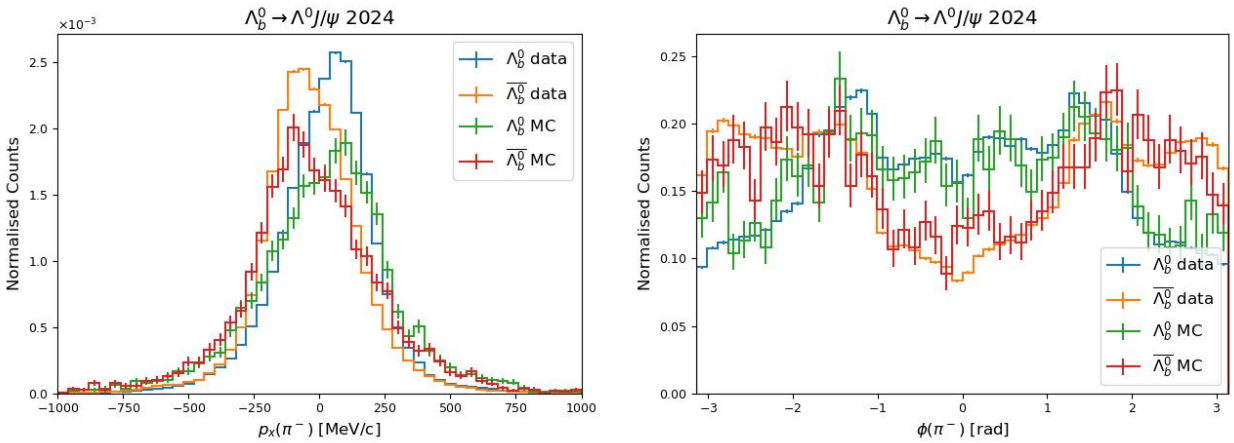

Figure 13: *Normalised histograms of the* π^- *and* π^+ *x momentum (left) and angle* ϕ *(right) utilising 2024 data and MC simulation without signal selection cuts. The p-values of the KS test for* p_x *are* 0.0 *for data and* $5.4 \cdot 10^{-37}$ *for MC simulation while they are* 0.0 *and* $2.3 \cdot 10^{-11}$ *for the* ϕ *data and MC respectively.*

The situation is resolved by considering that the magnetic field polarisation was down in the y-direction for the analysed data. As the main particle momentum is in the z-direction along the beam, a deflection of the charged particles will take place in the x-direction. Since the neutral particles are reconstructed from the charged particles, the effect plays a role there too. Because the experimental data with the magnetic field polarisation up is unavailable at the time of writing, the only comparison option, the MC simulation, indeed shows the opposite effect for both Λ_b^0 and $\overline{\Lambda_b^0}$ in Figure 14.

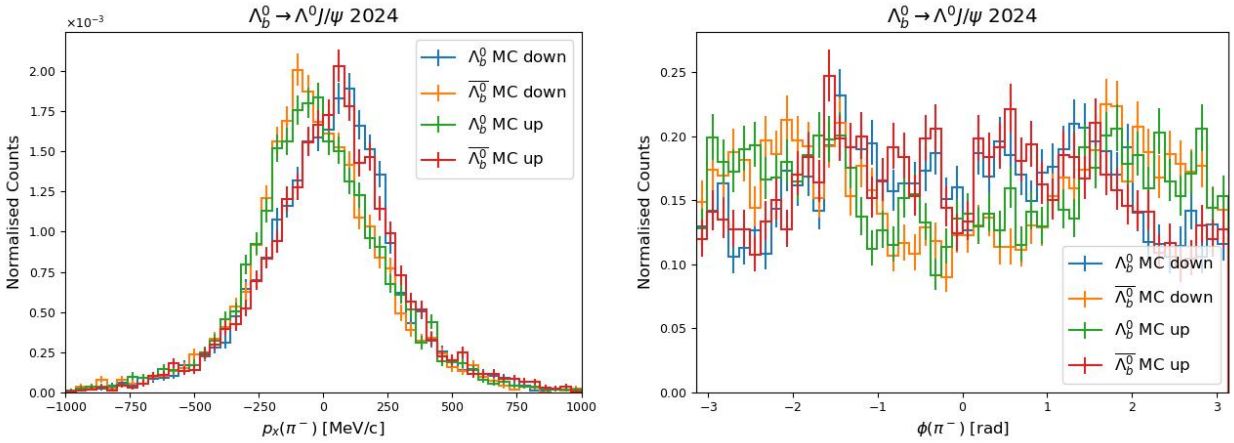

Figure 14: *Normalised histograms of the* π^- *and* π^+ *x momentum (left) and angle* ϕ *(right) utilising 2024 MC simulation without signal selection cuts with both up and down polarisation of the magnetic field. The p-values of the KS test for* p_x *are* 0.31 *for* Λ_b^0 *with up polarisation and* $\overline{\Lambda_b^0}$ *with down polarisation and* 0.20 *with opposite polarisation. The same values for* ϕ *are* 0.72 *and* 0.063 *respectively.*

4.4 Yield per Integrated Luminosity

The fraction of signal events over the integrated luminosity, obtained as outlined in Section 3.3, can be observed in Table 2:

		Excluding sideband subtraction	Including sideband subtraction
Experimental	N	3824 ± 62	1997 ± 75
	$\frac{N}{\mathcal{L}_{int}}$	(4.96 ± 0.17) pb	(2.59 ± 0.24) pb
Expected	ε_{MC}	$(7.896 \pm 0.089) \cdot 10^{-3}$	$(7.728 \pm 0.088) \cdot 10^{-3}$
	$\frac{N}{\mathcal{L}_{int}}$	(4.36 ± 0.67) pb	(4.27 ± 0.66) pb

Table 2: *The yield per integrated luminosity determined experimentally and calculated by means of the efficiency, branching fractions, cross section and correction factors. Both with and without sideband subtraction. The utilised yields and efficiencies are given too.*

For the calculation of these values the integrated luminosity was determined to be (771 ± 23) pb^{-1} discarding any values belonging to runs without any events passed by the software trigger. This provides the exclusion of any runs that were flagged as too bad for data analysis or that have not been included in this investigation. The total integrated luminosity of the data taken in 2024 from 19 March until 8 June was (1985 ± 60) pb^{-1}. The data and MC signal yields and integrated luminosity for runs with a recorded event also imply that MC simulation signal corresponds to an integrated luminosity of (1592 ± 57) pb^{-1} excluding sideband subtraction while it would be (2984 ± 148) pb^{-1} including sideband subtraction. Furthermore, along with the branching fractions mentioned in Section 2.4 and $f_{\sqrt{s}}$ in Section 3.3, a B^+ cross section of $\sigma_{B^+} = (86.6 \pm 0.5 \pm 5.4 \pm 3.4)$ μb [22] was utilised. Since σ_{B^+} is based on B^+, while the value of $\frac{f_{\Lambda_b}}{f_{B^+}}$ is normalised to B^+ and $\overline{B^0}$, the value in Ref. [23] is multiplied by 2 under the common assumption of equal production [8] yielding $\frac{f_{\Lambda_b}}{f_{B^+}} = 0.518 \pm 0.036$. In conclusion, the obtained values are consistent at this order of magnitude, however, a significant deviation of the experimentally determined $\frac{N}{\mathcal{L}_{int}}$ including sideband subtraction can be observed.

5 Discussion

5.1 Results Evaluation and Potential Sources of Error

Firstly, it should be noted that even though the cuts utilised for signal selection are based on physics arguments as well as peaks visible in the data, it remains an arbitrary choice where to set the boundaries. The addition of cuts based on other variables or an increase or decrease in the selected ranges will alter the observed distributions and yields.

Due to the low amount of detected decays as well as the selective angles at which decays could be detected for the data acquired in 2023, it is not useful for further analysis, especially for even rarer decays. The cause can be attributed to the malfunctioning of the VELO as a result of the damage to the RF foil [1]. On the contrary, the first impression of the 2024 data allows for further investigation of the validity as performed in this thesis by the comparison with simulation and of matter and antimatter.

According to the Kolmogorov-Smirnov test, the experimental and simulated data are significantly different such that they cannot have the same underlying probability distribution. However, this does not immediately lead to the conclusion the data must be invalid due to several mitigating factors. Firstly, the MC data is based on events where the Λ_b^0 decay is present. However, in real data numerous decays occur, some of which might be mistakenly detected and hence cause a larger background. For instance, the Λ_b^0 mass in Figure 9 shows a lot of background is present. In addition, the alignment and the calibration of the detector is assumed to be perfect in MC simulations which is considered to be unrealistic for the obtained data. These factors lead the test to conclude that the distributions cannot be from the same probability distribution, especially because the amount of statistics is high for this test. As such, the Kolmogorov-Smirnov test could underestimate the overlap of the distributions studied. Since no physical or unexpected abnormalities were found either, the 2024 dataset cannot be deemed invalid as a result of the KS test.

After a division between Λ_b^0 and $\overline{\Lambda_b^0}$, almost all variables such as the mass in Figure 12 exhibit similar distributions, supported by the high p-values meaning that the null hypothesis cannot be rejected and the possibility exists that the underlying probability distribution is the same. The only exceptions, the x momentum and the angle ϕ, can be explained by the polarisation of the magnet since moving charged particles are deflected in a magnetic field. This is confirmed by the MC simulation. Hence, the validity of the 2024 data is supported by the division into matter and antimatter.

To conclude, when the signal yield of the measurements is determined solely by the trigger and the signal selection cuts, the experimentally obtained yield per integrated luminosity is in agreement with the expected yield per pb^{-1} based on the branching fractions, efficiency and cross section since it is less than one standard deviation higher. In an attempt to decrease the background, sideband subtraction was utilised as well. The effect was small for the expected value since the MC mass peak is very distinguishable as shown in Figure 9 with limited tails which are utilised for the background while the other parameters remain constant. However, that is not the case for the experimental signal yield which has a much more numerous sideband relative to the signal. This is also shown by the large difference in integrated luminosity the signal of the MC simulation would correspond to. A possible cause is the inclusion of some signal in the sidebands which has the double effect of being excluded by the cuts and then being subtracted from the signal amount as though it is background. For instance, the imperfect alignment and calibration could be a reason. Furthermore, Figure 9 shows the background

is not constant. Even though it was attempted to limit the error by choosing symmetric sidebands on both sides of the Λ_b^0 mass peak, it cannot be claimed to have no effect at all. However, the sidebands shown in Figure 5 do not show any sign of the sideband subtraction leading to errors. Moreover, the calculation of the expected yield over the integrated luminosity contains the MC efficiency which is prone to a few errors. Firstly, there could be an overestimation due to the assumption of perfect alignment and calibration in the simulation. Additionally, some bad events are thrown away before entering the MC simulation which are thus not included in the total amount of simulated events [8]. Finally, the x and y momenta of the MC simulation are found to be a bit larger than in data causing a slightly increased chance of detection. If these factors are taken into account, it could significantly lower the expected value of the yield, which is in favour of including the sideband subtraction to reduce background. Therefore, an order of magnitude comparison of the values for the yield per integrated luminosity is in agreement, however, more research is necessary to determine which of the two experimental values is closer to the actual value.

5.2 Suggestions for Future Research

In order to improve the certainty on the validity of the data additional research can be performed. For instance, a similar comparison of data and simulation as well as matter and antimatter can be applied to other decays. Especially the decays that occur frequently as those provide a large amount of statistics. The $\Lambda_b^0 \to \Lambda^0 J/\psi$ decay only occurs 3824 ± 62 times excluding sideband subtraction and including sideband subtraction it is reduced to 1997 ± 75, both for an integrated luminosity of (771 ± 23) pb^{-1}. As such, a more frequent decay would provide more statistics. Unfortunately, that would mean different final particles, which was the reason for researching this decay to determine the validity for the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ decay. Additionally, different track types could be included, mainly the downstream tracks for the Λ^0 particle could provide a significant amount of additional signal events due to the long life time. Moreover, a separation could be made based on the total number of primary vertices in a collision and checked whether it has similar behaviour. Since the detector needs to separate more decays when the number of primary vertices rises, the combinatorial background could increase. Furthermore, the alignment and calibration of the detector could be investigated and improved such that the experimental data becomes more reliable and the MC simulation prediction is closer to reality. Finally, the cuts applied to determine the signal could be optimised such that as much signal as possible with little background remains in the selection. This can be achieved via trial and error as well as a more systematic approach. For instance, one could simulate background as well to train a machine learning algorithm to determine which events are signal and which ones are background. This could also be utilised as a verification for the yield over the integrated luminosity. Unfortunately, this method does rely on the MC simulation which could be too perfect in terms of alignment and calibration for this purpose yielding an argument for the trial and error approach or the production of worse MC data which matches the experimental data better.

Moreover, with the validated data, one can perform further physics analyses. Of particular interest is the much rarer $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ decay which can be investigated in order to potentially discover new physics through the FCNC as well as the P_5' anomaly and the branching fraction. The 2024 data is especially useful in this case since the (1985 ± 60) pb^{-1} collected from 19 March up until 8 June is already close to the 2.19 fb^{-1} [8] collected in the whole current record year 2018. Therefore, the required increment in statistics is provided. Furthermore, the $\Lambda_b^0 \to \Lambda^0 J/\psi$ can be utilised to determine the validity of HQET and perturbative QCD predictions or improve the quality of the estimation of the decay amplitudes for instance with increased availability of statistics.

6 Conclusion

In conclusion, the experimental data gathered in 2023 is deemed to have insufficient statistics, and the compromised operation of the VELO is visible in the results causing the data to be of limited use. On the other hand, data taken in 2024 is more numerous, already (1985 ± 60) pb^{-1} was taken from 19 March until 8 June, and is validated by the comparison of the Λ_b^0 and $\overline{\Lambda_b^0}$ decays via the KS test. A comparison of experimental data with MC simulation yields different results according to the KS test, however, that does not necessarily imply that the results are invalid. The experimental yield per integrated luminosity is in reasonable agreement with the expected value, however, a better estimate of the signal yield in data is required, as the sideband subtraction method by itself has not been validated and gives results at tension with the expected yields. A possibility is the inclusion of too much signal in the sidebands, however, an overestimation of the MC efficiency is likely and reduces the expected value significantly towards the value including sideband subtraction.

Bibliography

[1] P. Albicocco *et al.*, "LHCb/LNF 2023," Tech. Rep. [Online]. Available: https://www.lnf.infn.it/rapatt/2023/LHCb.pdf.

[2] D. Das, "Model independent new physics analysis in $\Lambda_b \to \Lambda\mu^+\mu^-$ decay," *European Physical Journal C*, vol. 78, no. 3, Mar. 2018, ISSN: 14346052. DOI: 10.1140/epjc/s10052-018-5731-2.

[3] CERN, *The Large Hadron Collider*, 2024. [Online]. Available: https://home.cern/science/accelerators/large-hadron-collider (visited on May 20, 2024).

[4] Aaij, R. *et al.* (The LHCb Collaboration), "The LHCb upgrade I," *JINST*, May 2024. DOI: 10.1088/1748-0221/19/05/P05065. arXiv: 2305.10515 [hep-ex].

[5] CERN. "LHCb." (2024), [Online]. Available: https://home.cern/science/experiments/lhcb (visited on Jul. 6, 2024).

[6] The LHCb Collaboration, "Future physics potential of LHCb," Mar. 2022. [Online]. Available: https://cds.cern.ch/record/2806113.

[7] Aaij, R. *et al.* (The LHCb Collaboration), "Measurement of the $\Lambda_b^0 \to J/\psi\Lambda$ angular distribution and the Λ_b^0 polarisation in pp collisions," *Journal of High Energy Physics*, vol. 2020, no. 110, Jun. 2020. DOI: 10.1007/JHEP06(2020)110.

[8] M. Mulder, personal communication, Apr.–Jul. 2024.

[9] Sirunyan, A. M. *et al.* (CMS Collaboration), "Measurement of the Λ_b polarization and angular parameters in $\Lambda_b \to J/\psi\Lambda$ decays from pp collisions at $\sqrt{s} = 7$ and 8 TeV," *Physical Review D*, vol. 97, no. 7, p. 072 010, Apr. 2018, ISSN: 24700029. DOI: 10.1103/PhysRevD.97.072010.

[10] D. Elsby, "The Rare decay of $\Lambda_b \to \Lambda^0\mu^+\mu^-$ with the LHCb detector at CERN," M.S. thesis, The University of Birmingham, Oct. 3, 2012.

[11] A. W. Peck, "Photoproduction of η_c mesons in ultra-peripheral Pb+Pb collisions at $\sqrt{s_{NN}} =$ 5.02 TeV at the LHC," M.S. thesis, Graduate School of Creighton University, Omaha, NE, Jul. 31, 2020. DOI: 10.13140/RG.2.2.35949.87524.

[12] D. Cockerill, *Introduction to Calorimeters*, Lecture slides, Southampton, May 2016. [Online]. Available: https://indico.cern.ch/event/518474/contributions/1198681/attachments/1267581/1877122/Calorimetry-lecture-to-Southampton-students-4May2016-Cockerill-compressed.pdf (visited on Jun. 28, 2024).

[13] Aaij, R. *et al.* (The LHCb Collaboration), "Measurement of the track reconstruction efficiency at LHCb," *JINST*, vol. 10, no. 02, P02007, Feb. 2015. DOI: 10.1088/1748-0221/10/02/P02007.

[14] R. L. Harrison, "Introduction to Monte Carlo simulation," in *AIP Conference Proceedings*, vol. 1204, Seattle: Department of Radiology, University of Washington Medical Center, Jan. 2010, pp. 17–21. DOI: 10.1063/1.3295638.

[15] M. Mazurek, G. Corti, and D. Muller, "New simulation software technologies at the LHCb Experiment at CERN," Dec. 2021. arXiv: 2112.04789 [physics.ins-det].

[16] L. M. Greeven, "Decoding beauty: rare baryonic decays & SciFi detector commissioning," Ph.D. dissertation, Maastricht University, 2024, ISBN: 9789464960563. DOI: 10.26481/dis.20240403lg.

[17] Workman, R. L. *et al.* (Particle Data Group), "Review of Particle Physics," *PTEP*, vol. 2022, p. 083C01, Aug. 2022, and 2023 update. DOI: 10.1093/ptep/ptac097.

[18] K. A. M. de Bruyn, personal communication, May 15, 2024.

[19] Y. Dodge, *The Concise Encyclopedia of Statistics*. Springer New York, NY, Feb. 2008, ISBN: 978-0-387-32833-1. DOI: https://doi.org/10.1007/978-0-387-32833-1.

[20] National Institute of Standards and Technology [NIST], *1.3.5.16. kolmogorov-smirnov goodness-of-fit test*. [Online]. Available: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm (visited on Jun. 12, 2024).

[21] The SciPy community, *Scipy.stats.kstest*, version 1.13.1. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html (visited on Jun. 12, 2024).

[22] Aaij, R. *et al.* (The LHCb Collaboration), "Measurement of the $B^{\pm}$ production cross-section in pp collisions at $\sqrt{s} = 7$ and 13 TeV," *Journal of High Energy Physics*, vol. 2017, no. 26, p. 026, Dec. 2017. DOI: 10.1007/JHEP12(2017)026.

[23] Aaij, R. *et al.* (The LHCb Collaboration), "Measurement of b hadron fractions in 13 TeV pp collisions," *Phys. Rev. D*, vol. 100, no. 3, p. 031 102, Aug. 2019. DOI: 10.1103/PhysRevD.100.031102.