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

Chishuai Wang

born in Zhejiang, China

Oral examination: 06.02.2026

The time alignment of the Scintillating Fibre Tracker and a LFU study with $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ at LHCb

Referees:
Prof. Dr. Ulrich Uwer
Prof. Dr. Monica Dunford

Abstract

This thesis describes two different scientific topics: the time synchronisation of the Scintillating Fibre (SciFi) Tracker and a Lepton Flavour Universality (LFU) study with the rare baryonic decay $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ ($\ell = e, \mu$) at the LHCb experiment.

The SciFi Tracker is a part of the LHCb Upgrade I. In the context of its commissioning and early operation, the time synchronisation procedure, hereinafter referred to as the time alignment, was developed to align the detector signal integration window to the incoming physics signal and thus to optimise the hit detection efficiency. The strategy of the time alignment is introduced and the system timing characteristics study based on the early data (2023-2024) is discussed.

LFU is a feature of the Standard Model (SM) of particle physics, which refers to the equal couplings of the electroweak interaction to the 3 generations of leptons. Assuming LFU, the relative branching fraction between the baryon decays $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ is expected to be close to 1.0. A deviation from this expectation can be caused by possible new physics. Using the data collected by the LHCb experiment from 2011 to 2012 (Run 1) and from 2015 to 2018 (Run 2), which corresponds to an integrated luminosity of $9\,\text{fb}^{-1}$, this relative ratio can be probed. The procedure of this LFU study is described and, depending on the different kinematic regions, the preliminary expected sensitivity, either for the ratio observable or for the observation of $\Lambda_b^0 \to \Lambda^0 e^+ e^-$, is reported.

Zusammenfassung

Diese Dissertation beschreibt zwei verschiedene wissenschaftliche Themen: Die Synchronisation des Scintillating Fibre (SciFi) Trackers und einen Test der Lepton-Flavor-Universalität (LFU) in der elektroschwachen Wechselwirkung mit dem seltenen baryonischen Zerfall $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ $(\ell = e, \mu)$ mit Daten des LHCb-Experiments.

Der SciFi Tracker ist Teil des LHCb Upgrade I. Im Rahmen der Inbetriebnahme und des ersten Betriebs, wurde ein Zeitsynchronisationverfahren, im Folgenden als „time alignment“ bezeichnet, entwickelt. Das ist notwendig, um die Integrationsfenster der Elektronik an die eingehenden Detektorsignale anzupassen, um so die Effizienz des Detektors zu optimieren. Die Strategie des „time alignment“ wird vorgestellt und die durchgeführten systematischen Studien zum Verständnis der Detektorsynchronisation werden diskutiert.

LFU ist ein Merkmal des Standardmodells (SM) der Teilchenphysik und bezieht sich auf die gleichen Kopplungen der elektroschwachen Wechselwirkung an die Leptonen der 3 Generationen. Unter der Annahme von LFU erwartet man, dass das Verhältnis der Verzweigungsverhältnisse der beiden Λ_b^0-Baryonzerfälle $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ und $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ nahe bei 1 liegt. Eine Abweichung von dieser Erwartung kann durch mögliche neue physikalische Effekte verursacht werden. Mit Hilfe der vom LHCb-Experiment von 2011 bis 2012 (Run 1) und von 2015 bis 2018 (Run 2) aufgezeichneten Daten, die einer integrierten Luminosität von 9 fb^{-1} entsprechen, kann dieses relative Verhältnis untersucht werden. Diese LFU-Studie wird beschrieben und die erwartete Sensitivität für das beobachtbare Verhältnis bzw. für die erste Beobachtung von $\Lambda_b^0 \to \Lambda^0 e^+ e^-$, abhängig von den verschiedenen kinematischen Regionen, wird angegeben.

Contents

Preface

Thanks to the collective effort of many previous and current collaborators, the Scintillating Fibre (SciFi) Tracker, which is a part of the LHCb Upgrade I, has been commissioned and plays an indispensable role in the LHCb Run 3 operation. During its commissioning and early operation, the author has made significant contributions, especially to the "fine" time synchronisation of the detector, the meaning of which should become clear over the course of this thesis. The author has studied special data collected for this purpose. The acquisition of the data was the result of the effort of numerous experts and the commissioning/operation team, to which the author has contributed by providing feedbacks in terms of operation and data inspection. Although the technical details of this special data acquisition mode are not fully discussed in this thesis, their complexity and importance should not be underestimated.

Apart from the commissioning of the SciFi Tracker, the author also worked on the test of Lepton Flavour Universality (LFU) by analysing the decay $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ where the leptons $\ell^\pm$ can be either electrons or muons. The analysis has been in development since before the author joined the effort. However, the author is the main contributor to the current implementation of this analysis. The previous analysis framework has been essentially rewritten. The data and simulation samples were produced by collaborators. The author inherited a preliminary baseline of selections and simulation corrections (used to improve its compatibility with data). Over the course of his studies, the author has made adjustments/additions to this baseline and conducted several validation checks. Several simulation corrections were reworked. The author has implemented the fitting framework which is used to determine the observables (as will be discussed in this thesis). The author has also conducted crosschecks and systematic studies for these observables, which is still an ongoing effort. For the electron channel, the analysis is blinded, *i.e.* the signal region is hidden and unknown to the author. Since the analysis has not passed all necessary LHCb internal review steps, the author has not got an unblinding permission for these regions. Therefore only expected sensitivity is shown in this thesis.

The work shown in this thesis was conducted within the LHCb collaboration. It would not have been possible without the foundation laid by numerous former and current collaborators.

1 Introduction

Over the course of the 20th and early 21st centuries, following the developments of the theory of special relativity and the quantum theory, physicists have made huge progresses in understanding the microscopic structure of the nature. Summarising various theories (*e.g.* the electroweak theory [1–3] the strong interaction and the quark model [4]), the so-called Standard Model (SM) of particle physics was developed. It was supported by various experimental discoveries such as the $W^{\pm}, Z$ gauge bosons [5–8] and the Higgs boson [9, 10]. Withstanding numerous experimental tests, it provides so far a very successful description of the fundamental (as in our current understanding) particles and their interactions, albeit not including gravity.

Further tests of the SM and the searching for the new physics beyond its description constitute essential parts of the physics research today. In this endeavor, the Large Hadron Collider beauty (LHCb) experiment [11] plays an important role, especially in the sector of heavy flavour physics. The LHCb experiment is one of the four large experiments at the Large Hadron Collider (LHC) [12] operated by the European Organisation for Nuclear Research (CERN). Benefiting from the high energy particle collisions provided by the LHC, the LHCb experiment took valuable data during its initial operation of Run 1 (from 2011 to 2012) and Run 2 (from 2015 to 2018). Afterwards, the LHCb detector went through the so-called Upgrade I [13] and resumed its data-taking at a higher nominal instantaneous luminosity in Run 3 (started in 2022).

The first topic of the thesis is related to the Upgrade I of the LHCb detector. As a part of it, a new tracking detector namely the Scintillating Fibre (SciFi) Tracker was installed and commissioned. The detector apparatus of the LHCb experiment is discussed in Chapter 2, with Section 2.4 focusing on the SciFi Tracker. An important part of the detector commissioning and operation is the time synchronisation, *i.e.* aligning the detector signal integration window to the arrival of the physics signal in terms of timing, hereinafter referred to as the time alignment. The method to perform such an alignment is discussed in Chapter 3. Using the data collected during the detector commissioning and early operation (2023-2024), the timing characteristics of the detector system can be studied, which is discussed in Chapter 4.

The second topic of the thesis concerns the study of the Λ_b^0 baryon decay $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ ($\ell = e, \mu$) using the data collected with the LHCb experiment during its Run 1 and Run 2 operation, which corresponds to an integrated luminosity of $9\,\text{fb}^{-1}$. In the SM, the electroweak interaction couplings are universal for the three lepton flavours, *i.e.* e, μ and τ. This feature is known as

Lepton Flavour Universality (LFU). However, this is not necessarily obeyed by possible new physics. An experimental test of LFU will serve to test the SM and probe new physics. The motivation to perform such a test with the Λ_b^0 baryon decay $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ ($\ell = e, \mu$) is discussed in Chapter 5. The analysis itself is dicsussed from Chapter 6 to Chapter 10.

2 The experimental apparatus

2.1 The LHC

With a circumference of 26.7km, the LHC near Geneva is currently the largest collider in the world. It collides two counter-rotating proton beams at 4 interaction points, which are allocated to 4 large experiments, *i.e.* CMS [14], ATLAS [15], ALICE [16] and LHCb [11].

The proton-proton (pp) operation of the LHC was designed with the goal to achieve the center-of-mass energy up to 14 TeV and an instantaneous luminosity of $\mathcal{L} \sim 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for the two high luminosity experiments (CMS and ATLAS) [12]. During 2011(2012), the LHC operated with the center-of-mass energy of pp collisions of 7(8) TeV [17]. During 2015-2018, this energy was increased to 13 TeV [18]. At the start of Run 3 (2022), it was further increased to 13.6 TeV.

Due to different physics programs, the nominal operation luminosity of the LHCb experiment is lower than that of CMS and ATLAS. During its Run 1 and Run 2, the LHCb experiment operated with $\mathcal{L} \sim 4 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ (twice its design value) [13, 17]. Different luminosity at the different interaction points can be realised via various techniques, *e.g.* different beam focuses and transverse beam offsets [19].

The LHC beams are divided into so-called bunches [1], which can either be filled or empty. The actual configuration of which bunches being empty can vary over the different so-called fills [2]. Usually, during pp operation, the bunches of the two beams meet each other (bunch crossings) at the interaction points with a frequency of 40 MHz, which corresponds to intervals of 25 ns.

2.2 The LHCb experiment in Run 1 and Run 2

2.2.1 The LHCb detector

The LHCb detector was built with a strong emphasis on the study of CP violation and the rare decays of heavy flavour hadrons [11, 22, 23]. Given that, at the energy level of LHC, the $b\bar{b}$ pairs are produced mostly in the same forward or backward cone, the LHCb detector is arranged as a single-arm forward spectrometer covering an angular region of about $\pm(10 - 250)$ mrad

[1] The bunches are focused in the so-called Radio-Frequency (RF) bucket. The RF system of LHC has a RF bucket length of 2.5 ns, while, in reality, a larger bunch spacing corresponding to 25 ns, *i.e.* one in every 10 RF buckets, is used [20].

[2] A fill refers to the period between LHC reaching stable beam and dumping the beam [21].

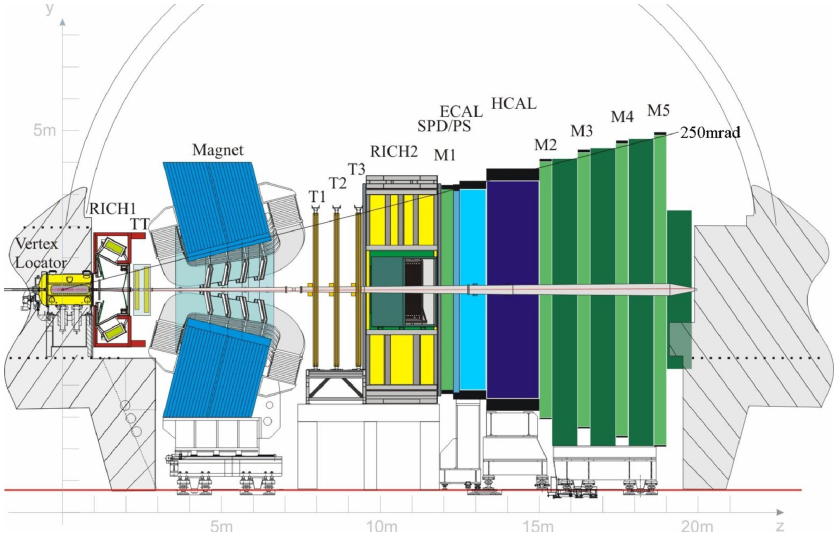


Figure 2.1: An illustration of the LHCb Run 1 and Run 2 detector. Plot taken from Ref. [17].

vertically and about $\pm(10-300)$ mrad horizontally [11]. The detector during Run 1 and Run 2 is illustrated in Figure 2.1, which shows the magnet and several subdetectors, *i.e.* the VErtex LOcator (VELO), the Tracker Turicensis (TT), the tracking stations downstream of the magnet (T1-3), two Ring Imaging CHerencov (RICH1 and RICH2) detectors, a Scintillator Pad Detector (SPD) and PreShower (PS), the Electromagnetic CALorimeter (ECAL), the Hadronic CALorimeter (HCAL) and the muon stations M1-5.

A right-hand coordinate system is used, with the z-axis along the beam pipe and the y-axis along the vertical direction [11]. The direction perpendicular to the z-axis is usually referred to as the transverse direction in the definition of variables, *e.g.* the transverse momentum p_T.

Supported by its subsystems, various functionalities are realised for the LHCb detector, *e.g.* the track reconstruction, Particle IDentification (PID) and the hardware level (L0) triggering.

The magnet provides a magnetic field along the vertical direction, which allows charged particle momentum measurements when combined with the tracking system. The magnet field can be configured in two directions, *i.e.* downwards and upwards. The magnet was designed to provide an bending power (magnet field strength integrated over length) of about 4Tm along the z-direction from $0-10$m [24].

Located at the lower end of the z-axis, the VELO is the tracking subdetector encompassing the pp interaction point, *i.e.* the so-called Primary

Vertex (PV). It provides tracking measurements close to the PV, which contribute to the vertex reconstruction. This is crucial, especially considering that displaced decay vertices are important features of the long-lived b and c-hadron decays [11]. To improve the impact parameter (distance between the PV to a track) measurement precision, it is desirable to place the instrument close to the interaction point [25]. However, due to safety requirements of the LHC operation, a distance to the beam axis is required. This requirement is even larger during the beam injection [17]. Considering the above condition, the VELO is constructed from two halves, which are retractable during the beam injection. The VELO sensors utilise silicon microstrips, which are arranged in two configurations to measure the azimuth angle ϕ coordinate and the radial r coordinate of a track [26]. The geometric layout of VELO is illustrated in Figure 2.2.

Apart from the VELO, the tracking system also contains the TT and downstream stations T1-T3. Based on Ref. [11], they are introduced as follows. The TT uses the silicon micro-strip sensors. It has one station including 4 layers in the so-called x-u-v-x stereo angle layout. The two x layers consist of vertically arranged silicon strips, while the silicon strips in the u/v layer are rotated by ± 5 degrees with respect to the vertical direction. Such a layout provide the tracking algorithm with the necessary information in both the x and y coordinates. The 3 tracking stations downstream of the magnet are separated into the Inner Tracker (TT) and the Outer Tracker (OT). The IT covers regions close to the beam-pipe, which exhibit a high particle occupancy. The OT covers the outer region. The IT uses silicon micro-strip sensors while the OT uses straw-tube technology. Similar as the TT, each station of IT and OT contains 4 layers following the x-u-v-x layout. An illustration of OT and IT in one station is shown in Figure 2.3.

The RICH detectors are a part of the PID system. For a charged particle, the Cherencov light emission angle depends on its velocity and thus, for a certain momentum, depends on the species of the particle. By reconstructing the Cherencov angle, the RICH detectors utilise this feature to provide separation power for various charged particles. Figure 2.4 illustrates the Cherencov angles of several different particles with different momenta in the gas radiator C_4F_{10} of RICH1. As discussed in Ref. [28], the two RICH detectors with different acceptance coverages perform the particle identification in different momentum ranges: The momentum distributions of particles are correlated to the polar angle. Covering the acceptance 25-300 mrad, the upstream RICH1 is mainly responsible for relatively low momentum particles (2-40 GeV/c). The downstream RICH2 covers a narrower acceptance 15-120 mrad, and is mainly responsible for relatively high momentum particles (15-100 GeV/c).

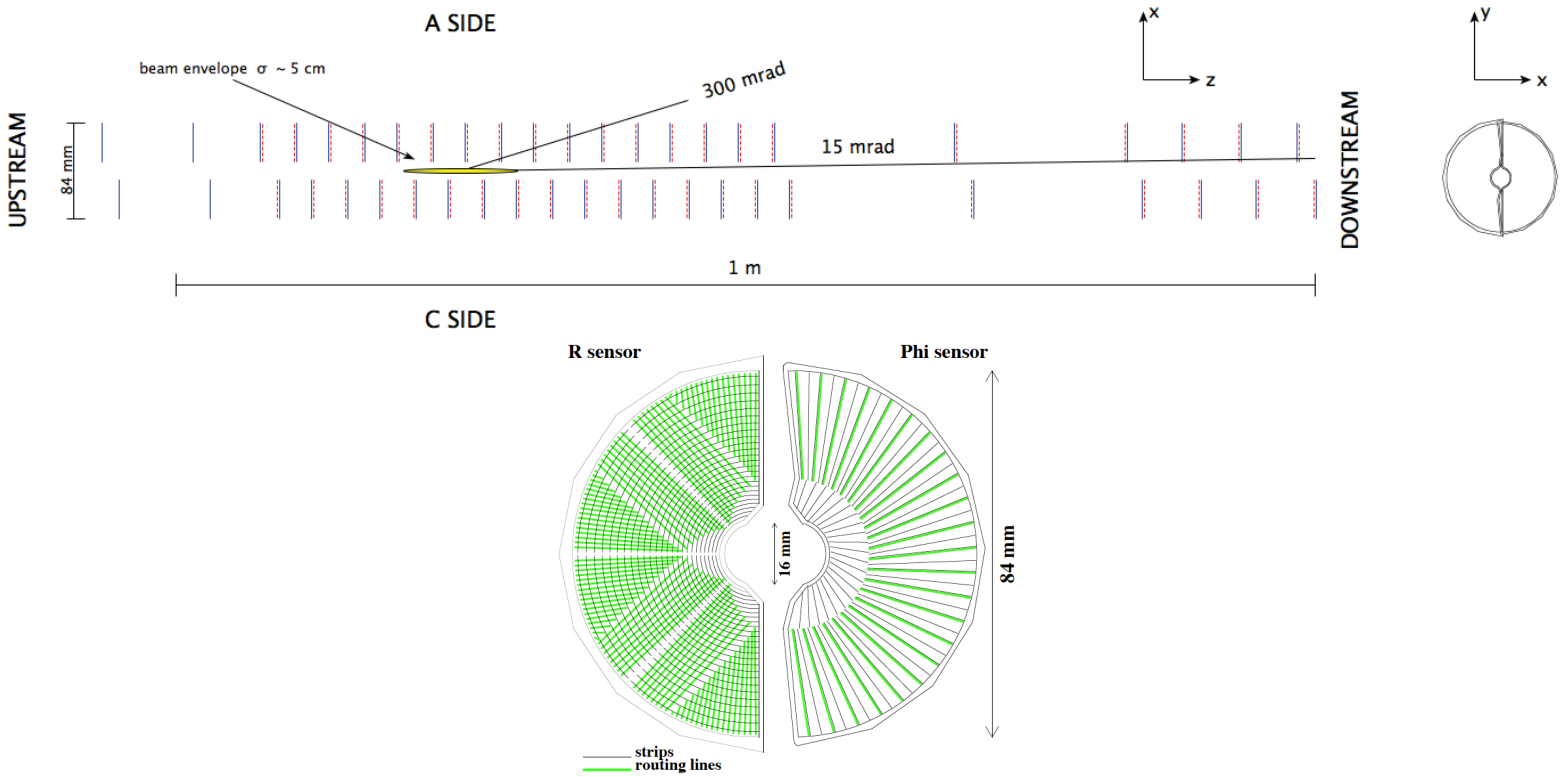


Figure 2.2: Illustrations of VELO layout. The top left plot illustrates a cross-section view in the x-z plane, where the blue solid lines indicate the R sensors and the red dashed lines indicate the ϕ sensors. The top right plot illustrates a view in the x-y plane. The bottom plot illustrates the silicon strip and routing lines (for readout) arrangements in the R and ϕ sensors. Plots taken from Ref. [26].

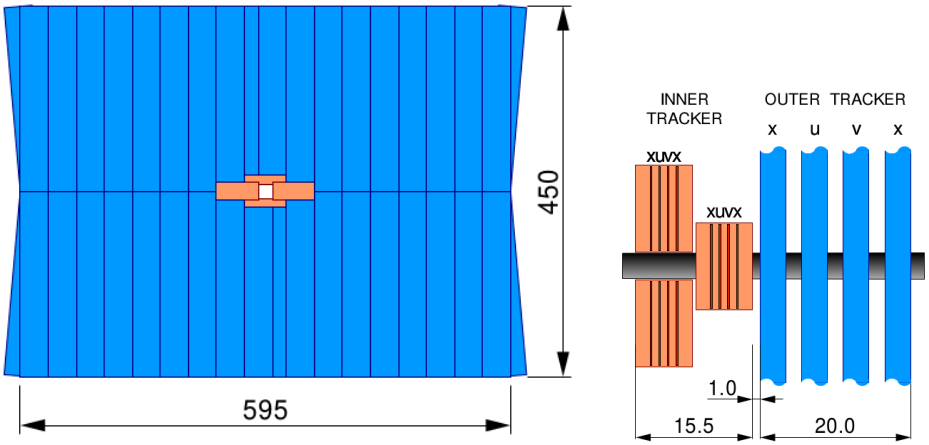


Figure 2.3: Illustrations for the OT (blue) and IT (red) in one station. The left plot is a frontal view. The right plot is a top view. The length numbers are given in the unit of centimetre. Plots taken from Ref. [27].

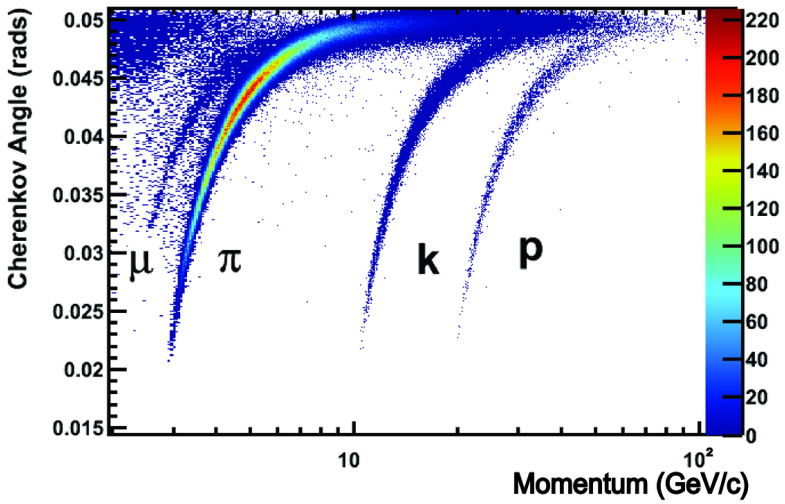


Figure 2.4: The Cherencov angles for different particles with different momenta in the gas radiator C_4F_{10} of RICH1. Plot taken from Ref. [28].

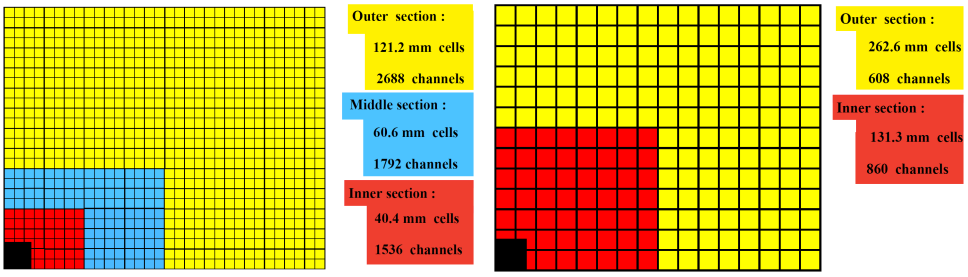


Figure 2.5: An illustration of the ECAL and HCAL cell divisions (in one quarter view). The left side shows the division used for SPD, PS and ECAL. The right side shows the division for HCAL. Plot taken from Refs. [13, 29].

The calorimeter system contributes to PID and to the hardware trigger of LHCb. The calorimeter system consists of, in the sequence of increasing z-coordinate, SPD, PS, ECAL and HCAL. These components share the similar working principle, *i.e.* the scintillator material (working together with absorbers) emits scintillating lights, which are readout by the wavelength-shifting fibres and received by the photomultipliers [29]. The exact structures of them differ [30]: The SPD and PS are made of scintillator pads and are separated by a lead layer. The ECAL features a shashlik structure with scintillator and lead layers. The HCAL consists of iron and scintillating tiles. The calorimeters consist of cells in the x-y plane. Considering the occupancy gradient in the x-y plane, the calorimeters are divided into different zones with different granularity of cells. The division of the granularity zones is illustrated in Figure 2.5. The SPD, PS and ECAL have the same division of three zones (shown in the left plot of Figure 2.5). Considering the different dimension for the hadronic shower, the HCAL has two zones with different cell sizes (shown in the right plot of Figure 2.5). The combined information from the SPD, PS, ECAL and HCAL provides input for the hardware trigger with separation power of several particle types. The SPD contributes to the separation between neutral and charged particles while the PS contributes to the separation between electrons and charged pions [29]. The hardware trigger will be discussed further in Section 2.2.2.

The muon stations include 5 stations (M1-5). They are based on the multi-wire-proportional-chamber technology except for the inner most part of M1, where triple-GEM chambers are used due to the high radiation [11]. The muon stations contribute to the identification of muons and provide information for the hardware trigger. One example for the muon identification is the so-called `IsMuon` boolean variable. As explained in Refs. [31, 32], given

Table 2.1: The station requirement for the `IsMuon` response. Table taken from Refs. [31, 32].

Momentum range	muon stations
$3 < p < 6\,\text{GeV}/c$	M2+M3
$6 < p < 10\,\text{GeV}/c$	M2+M3+(M4 or M5)
$p > 10\,\text{GeV}/c$	M2+M3+M4+M5

a reconstructed track, hits are searched in the Field-Of-Interest (FOI) of the muon stations. The `IsMuon` gives a positive response if at least one hit is found in the FOI of each required muon stations. The station requirement depends on the momentum of the track, as shown in Table 2.1.

The PID information from the RICH system, the calorimeter system and the muon system can be combined to provide more powerful PID distinguishing variables [17].

2.2.2 Trigger system

During the LHCb data-taking, since the data storage and processing resources are not unlimited, not all bunch crossing events can be saved. The decision of whether or not to save an event is taken by the trigger system. During Run 1 and Run 2, the LHCb trigger system consists of three stages (in sequence): the hardware level (L0) trigger, the software High-Level-Trigger-1 (HLT1) and the software High-Level-Trigger-2 (HLT2). The system was operated in the following fashion [21, 33] [3]: The L0 trigger (using dedicated hardware system) reduces the event rates from the full LHC rate of 40 MHz to around 1 MHz. The detector is read out at this level and the information from the detector is provided to the HLT system. At the HLT levels, online reconstruction is performed. The HLT1, which uses partial event reconstruction information, reduces the event rates further to $\sim$ 110 kHz. The HLT2, which uses more complete event reconstruction information [4], reduces further the event rates to $\sim$ 12.5 kHz. The events passing all three trigger levels are saved to the storage. Afterwards, offline reconstruction and selection can be performed. During Run 2, an additional option became available to save, from the HLT2 output, the *compact* event information reconstructed online (full offline reconstruction cannot no longer be performed using only these information) [34, 35]. In the LFU study with $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$, the data based

[3] These are Run 2 event rates. In Run 1, the event rates at various stages are different [33].

[4] In Run 2, full event reconstruction was performed at HLT2 [21].

on the offline procedure are used for both Run 1 and Run 2.

The requirements of the trigger at each level are grouped into the so-called trigger lines. Based on Refs. [21, 33], several L0 trigger lines, which use the information from the calorimeter system and the muon stations, are explained as follows. In the ECAL and HCAL, a cluster is reconstructed from 4 neighouring cells (in the same zone) with its transverse energy calculated as

$$E_T = \sum_{i=1}^{4} E_i \sin(\theta_i) \ , \tag{2.1}$$

where the angle θ_i is determined by the z-axis and the direction pointing from the mean position of the pp interaction region to the center of the calorimeter cell. These clusters, together with the SPD and PS information, can be used to look for electron, photon or hadron signatures. The following trigger responses are provided.

- `L0Electron`: a cluster in ECAL with certain numbers (depending on the zones) of PS cells hit and at least one SPD cell hit in front of it. The E_T of the cluster is required to be larger than a certain threshold.
- `L0Photon` : similar as the `L0Electron` but requiring no corresponding SPD cell hit.
- `L0Hadron`: a cluster in HCAL. If a corresponding ECAL cluster is present, the E_T is calculated combining the ECAL and HCAL clusters.

The muon L0 trigger uses the information of high p_T muon candidates:

- `L0Muon`: require the muon with highest p_T has p_T larger than a certain threshold.
- `L0DiMuon`: the threshold is set on the product of the highest p_T^{highest} and the second highest $p_T^{\text{2nd highest}}$.

In addition, a requirement on the number of SPD hits (nSPDHits) can be used to filter away the events with very high multiplicity, which take a long time to process in the HLT.

In general, the thresholds of these L0 trigger lines vary over different time periods of data-taking. In the LFU study of $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$, the `L0Electron`, `L0Muon` lines are used in the nominal signal selection, which will be discussed in Chapter 7. For these two lines, the typical thresholds are around $2-3\,\text{GeV}$ for E_T (`L0Electron`) and around $1.5-3\,\text{GeV}/c$ for p_T (`L0Muon`) [36].

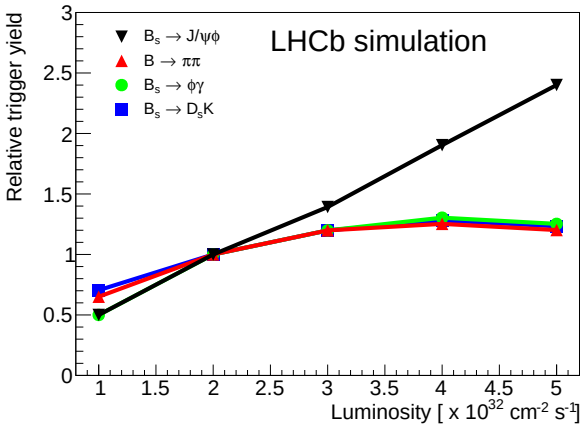


Figure 2.6: An illustration on the limitation of the L0 trigger when the luminosity increases. Plot taken from Refs. [13, 37]

2.3 The LHCb Upgrade I

While the original LHCb detector was a success, an upgrade was needed in order to operate the experiment at a considerably higher instantaneous luminosity ($\mathcal{L}$), which would give a significant increase in statistics. Comparing with the Run 1 and Run 2 operation with $\mathcal{L} \sim 4 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the LHCb Upgrade I detector was designed to run at $\mathcal{L} \sim 2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [13].

At the increased luminosity, the previous L0 trigger output rate would become a limitation, which is especially severe for the full hadronic decays [37]. This is illustrated in Figure 2.6, which shows the trigger yields for several hadronic decays saturate around an instantaneous luminosity of $3 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In addition, comparing with the simple inclusive criteria used in L0 trigger (like E_T), it is more favourable to use the exclusive criteria dedicated to certain signals of interest [38]. Due to these considerations, as one of the significant updates of the Upgrade I, the L0 level trigger was removed. The upgraded LHCb features a full software trigger system with the electronic system designed to support a 40 MHz trigger-less readout [39].

The upgraded detector is illustrated in Figure 2.7. Besides the trigger system, significant upgrades were also made to several subdetectors, which are introduced as follows [13, 40]. A new VELO was installed. Comparing with the previous design with silicon strips measuring r-ϕ coordinates, the new VELO is based on silicon pixel sensors. It also features a smaller inner radius, which is expected to benefit the resolution of the impact parameter. The TT was replaced by the Upstream Tracker (UT), which uses the silicon microstrip technology. Its geometric arrangements have been designed to,

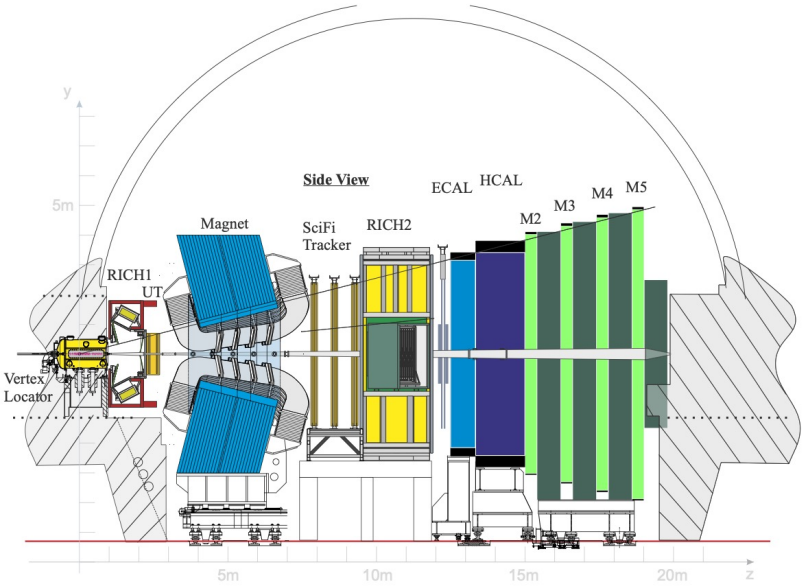


Figure 2.7: An illustration of the LHCb Upgrade I detector. Plot taken from Ref. [13].

comparing with its predecessor TT, significantly reduce gaps in the acceptance. The OT and the IT were replaced by the SciFi Tracker, which will be discussed in Section 2.4. For the RICH detectors, the photon detection chain was replaced to cope with the 40 MHz readout requirement. The optical system of RICH1 was upgraded to cope with the occupancy caused by the increased instantaneous luminosity. The electronics of the ECAL, HCAL and the muon stations were upgraded to fullfill the 40 MHz readout requirement. The SPD, PS and the first muon station M1 were removed due to the consideration that, without the L0 trigger, their purposes would be diminished. The supporting structure of M1 is used to host a neutron shielding for the SciFi Tracker.

2.4 The SciFi Tracker

2.4.1 Detection principle

The SciFi Tracker uses scintillating fibres as active material. The scintillating fibres feature a round cross section with a diameter of $\sim 250\mu$m and utilise Polystyrene based scintillator material [41]. As charged particles traverse the scintillating fibres, scintillating lights are emitted and transmitted to be

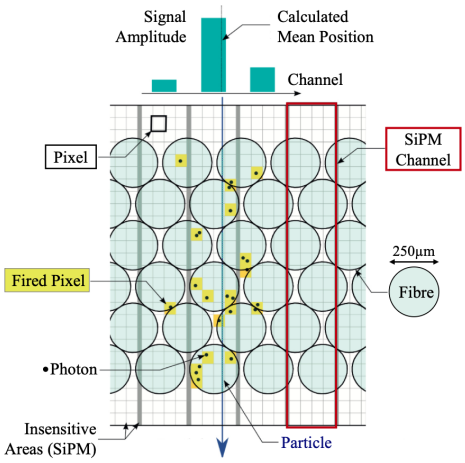


Figure 2.8: An illustration of the particle traversing the scintillating fibres. The red rectangle indicates a single SiPM channel. The squares correspond to the pixels. The circles correspond to the scintillating fibres. Plot taken from Ref. [42].

received by the Silicon Photomultiplier (SiPM) array. The SiPM array is divided into pixels. These pixels are reverse-biased avalanche photodiodes in Geiger mode, which means an initial photoelectron can cause a self-sustaining avalanche, which is quenched by quench resistors [42]. The breakdown voltages of these SiPMs are around $\sim$ 51.75V (with variations) and the SciFi Tracker nominally operates at 3.5V above the breakdown voltage [13]. These pixels are grouped into channels. The signal amplitude of a channel is proportional to the number of fired pixels [43]. The above detection process is illustrated in Figure 2.8. A single SiPM array used by the SciFi Tracker consists of 128 channels. The SiPMs are affected by dark count noise caused by thermal excitation of charge carriers of the photodiodes, especially after being irradiated. To reduce the dark count rate, the SiPMs are cooled to around $-50 \sim -40$°C.

2.4.2 Geometric layout

The SciFi Tracker, located downstream of the magnet, covers an acceptance region of about 6 meters in x and about 5 meters in y [13]. It consists of 3 stations, which are labelled as T1,T2 and T3 (following the increasing z-coordinate). Figure 2.9 illustrates the geometric layout of the SciFi Tracker. As shown in the figure (the side view in the right plot), each of the stations consists of 4 layers in the x-u-v-x stereo angle arrangements. The scintillating

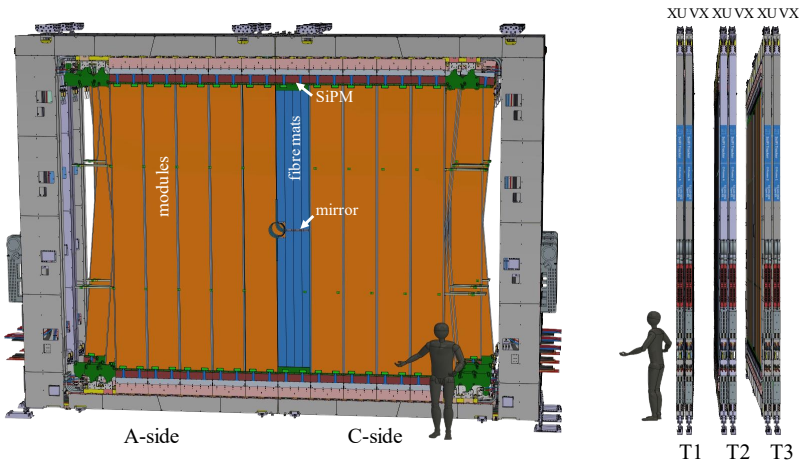


Figure 2.9: Illustrative views of the SciFi Tracker. Plot taken from Ref. [13].

fibres are oriented vertically in the x layers while they are tilted by ± 5 degrees in the u and v layers. Each layer in T1 and T2 consists of 10 modules and each layer in T3 has 12 modules, corresponding to a larger coverage area. In the center, the inner most modules are cut to allow the beam pipe to pass through, as shown in the left plot of Figure 2.9 (front view). The fibres in a module are separated in the middle, where a mirror is installed to reflect scintillating lights. This arrangement aims to increase the photon yields at the ends of the module (top/bottom), where the SiPM arrays are located. The two parts of a module separated by the mirror are referred to as the half-modules. If a charged particle traverses the active area in a point in the x-y plane, scintillating lights emitted are transmitted to the top (for the upper half-module) or the bottom end (for the lower half-module), where they are received by the SiPM arrays. The SiPMs convert the light signals into the electronic signals, which are processed by the Frontend Electronics (FE). For each half-module, at the side opposite to the mirror, there are 16 SiPM arrays and one FE ReadOut Box (ROB), which is the hardware device hosting the FE. In total, the SciFi Tracker consists of 4096 SiPM arrays and 256 ROBs.

The SciFi Tracker is separated horizontally into two sides, which are retractable during interventions. The neighouring x-u or v-x layer pairs of the same station on the same side are grouped into a common mechanical structure called C-Frame. Based on the separation of the sides and the mirrors, a layer can be divided into 4 quarters in the x-y plane.

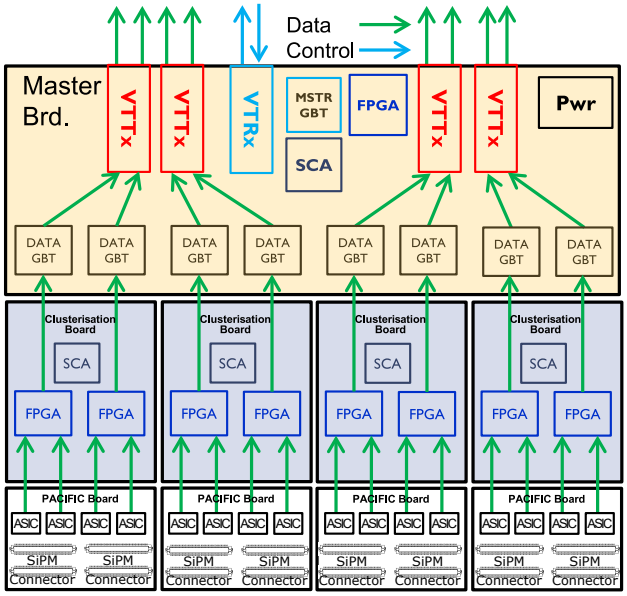


Figure 2.10: An illustration of the data flow in the SciFi FE. Plot taken from Ref. [13].

2.4.3 Frontend elecronics

The signals from the SiPMs are received and processed by the FE. An illustration of the data flow in the FE is shown in Figure 2.10. The unit shown in Figure 2.10, including 4 so-called PACIFIC [5] boards, 4 clusterisation boards and one master board, is responsible for 8 SiPMs. Two master boards, with their corresponding clusterisation boards and PACIFIC boards, are installed in a ROB. Sometimes, the so-called half-ROB is used to refer to the unit associated with a single master board.

The signals from the SiPMs are processed in the Application Specific Integrated Circuit (ASIC) chip named PACIFIC [44], which is hosted on the PACIFIC board. A single PACIFIC chip is responsible for 64 channels of the SiPM array. Figure 2.11 shows the several working blocks of the PACIFIC chip [44]:

- The pre-amplifier converts the input from the SiPM into a voltage signal.
- The shaper is used to reduce the extended tail of the signal.

[5] The name stands for low-power ASIC for the Scintillating Fibre Tracker [42].

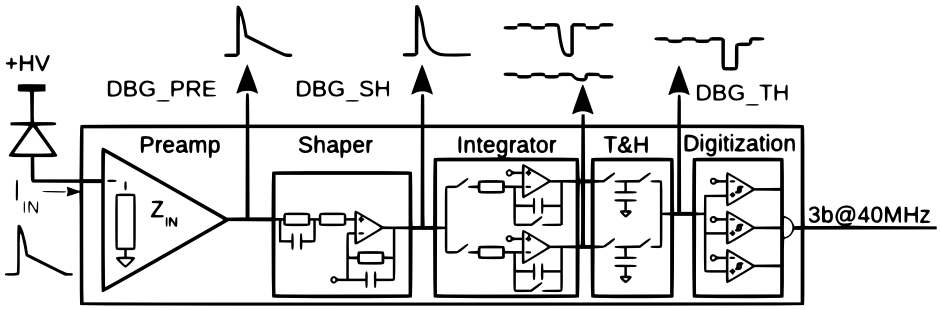


Figure 2.11: The channel block diagram of the PACIFIC chip (version r5q). Plot taken from Ref. [13].

- Two interleaved integrators are responsible for one channel. They switch for each bunch crossing (25 ns) in order to eliminate the dead time associated with the integrators.
- The track-and-hold (T&H) unit is used to combine the output of the two integrators to provide a stable output for the subsequent electronics.
- The digitisation block utilises three comparators to convert the analog signal into a digital one, depending on the thresholds passed.

The stages discussed above are followed by an encoding and serialisation procedure, where manipulation of the data for data-transmission is performed [44].

The analog part of the PACIFIC signal processing is closely related to the signal timing. As discussed in Ref. [13], Figure 2.12 shows the simulated time-dependent responses given by the shaper and the T&H for a 10 photoelectron signal arriving at time $t = 0$ ns. The amplitude of the signal depends on the relative time between the sampling point and the signal arrival. Thus, for the benefit of the detector performance in terms of hit detection efficiency, it is important to align timely the PACIFIC integration window with respect to the signal arrival. It is worth noting that the signal (as shown in Figure 2.12) extends more than 25 ns, which contributes to the spillover in the following bunch crossing.

The digitisation part of the PACIFIC utilises the thresholds set by the three comparators. They require calibrations [45, 46], *i.e.* finding the relationship between the threshold settings in terms of DACs (Digital-to-Analog Converter) and the input signal amplitude from SiPMs in terms of photoelectrons. A light injection system was designed to provide light signals to the SiPMs, which can be used for the calibration.

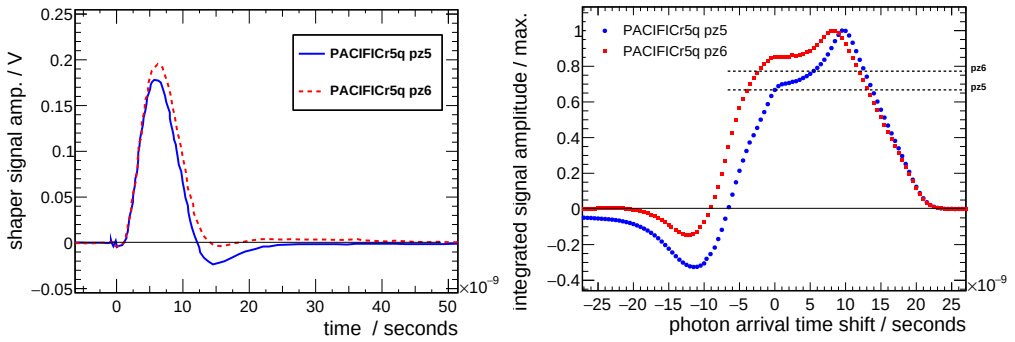


Figure 2.12: The left plot shows a simulated response after the shaper produced by a signal of 10 photoelectron arriving at $t = 0$ ns. The right plot shows the corresponding T&H output. The two curves in each plot are for two different pole-zero settings of the shaper (pz5 and pz6). Plots taken from Ref. [13].

The output data of the PACIFIC are sent to the clusterisation board, where, among other procedures, the clusterisation of the data is performed. Such a procedure is used to reduce the data bandwidth and suppress the thermal noise from the SiPMs. An explanation of the clusterisation algorithm was provided in Ref. [13], which can be summarised as follows. According to the thresholds set by the three comparators (in a certain hierarchy), the channels with different signal amplitudes are assigned weights between 1 and 6, *i.e.* 6 for the signal over the highest threshold, 2 for signal over the second highest threshold and 1 for signal over the lowest threshold. A cluster is formed when the sum of weights of several neighouring channels is larger than 2 or when a single channel has the weight of 6. Several examples are illustrated in Figure 2.13, where clusters are formed in the cases (a), (b), (c), (d). The two instances in (e) do not form clusters. In one instance of (e), the single channel does not have the weight of 6 while, in the other instance of (e), the sum of weights of the two channels is not larger than 2. The case (d), which extends more than 4 channels, is flagged as a large cluster. Different methods for asigning the center of the clusters are applied for the normal and the large clusters, *i.e.* the weighted average over channels is used for the normal cluster while the unweighted center is used for the large cluster. The clusterisation is realised by the Field Programmable Gate Array (FPGA) chips hosted by the clusterisation board, which are referred to as the Clustering FPGAs. A GigaBit Transceiver - Slow Control Adapt (GBT-SCA) ASIC chip is also hosted by the clusterisation board, which is

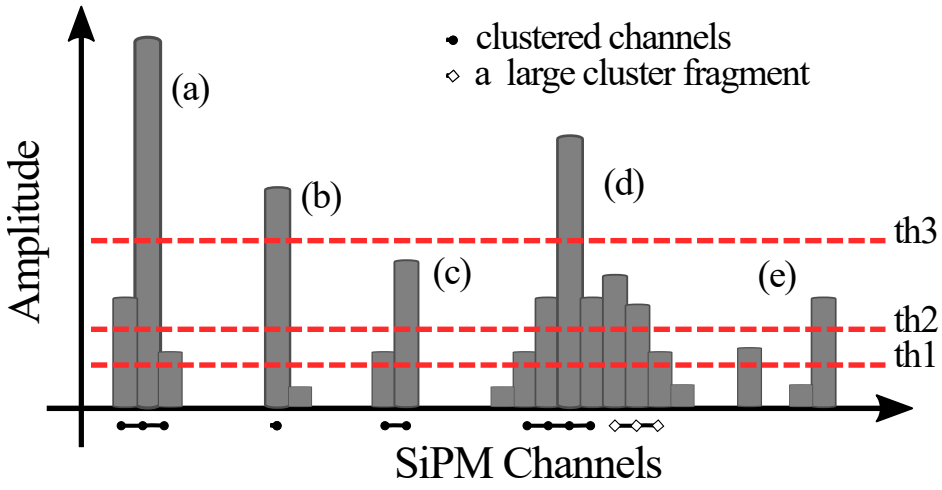


Figure 2.13: An illustration of the clusterisation algorithm. Plot taken from Ref. [13].

capable of providing control and monitoring functionalities [47].

The output data from the clusterisation board are forwarded to the master board. In the master board, the data are further processed by the so-called data GBTx ASIC chip, which encodes the data based on the GBT protocol [42, 48]. Each data GBTx corresponds to one SiPM array. Afterwards, the data are sent by the Versatile Twin Transmitter (VTTx) [49] through the optical link to the backend Data AcQuisition (DAQ) system. Besides handling the data, the master board is also responsible for the communication with the backend control system, which is transmitted by the bidirectional Versatile Transceiver (VTRx) [49] and managed by the master GBTx ASIC chip. The master GBTx and the data GBTx chips are also involved in the distribution of clocks, which will be discussed further in Chapter 3. The so-called housekeeping FPGA is also hosted by the master board, which controls the timing of the light injection system.

2.4.4 Infrastructures

As discussed in Section 2.4.1, the operation of the SiPMs require a bias voltage and cooling [13]: The bias voltage is provided by CAEN power supplies. The cooling system utilises a single phase thermal transfer fluid (fluoroketone or C_6H_{14}), which provide coolant temperature between -50 and $30°C$. To prevent condensation on cold surfaces, dry gas of a dew point of $-70°C$ is flushed through the system. The structure to support the cooling surrounding the SiPMs is illustrated in Figure 2.14.

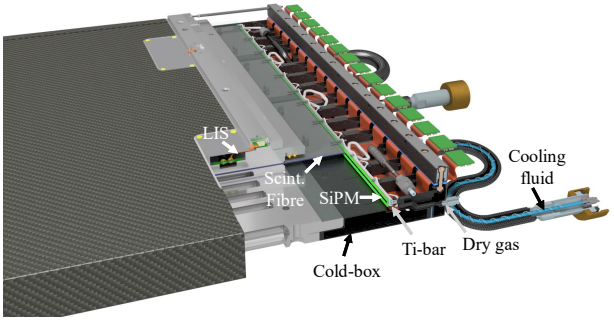


Figure 2.14: An illustration of the cooling support structure surrounding the SiPMs. Plot taken from Ref. [13].

The operation of the FE requires power and cooling [13, 43]: The FE is powered using the Wiener^{TM} MAgnetic field and RAdiation TOleraNt (MARATON) power supply. The cooling of the FE uses chilled water, which limits the FE temperature up to around 50°C (at the location of PACIFIC).

3 Time alignment strategy

The time alignment of SciFi concerns the synchronisation of the FE operation with respect to the physics signal induced by the LHC bunch crossings. The FE collects the signal within the 25 ns interval and produces the data stream sent to the backend DAQ system. The data stream contains the so-called Bunch Crossing IDentifier (BCID or BXID), which uniquely identifies a bunch crossing within an orbit [6]. The nominal LHC proton beam contains 3564 bunch places [12]. This corresponds to a BXID range of $0 - 3563$ (including 0 and 3563). For SciFi, no timing information of finer granularity than the BXID is saved in the data stream.

Following the convention used during the commissioning of the SciFi, the procedure to fine tune the timing of the FE in order to optimise the hit detection efficiency is named as "fine time alignment". One also needs to ensure the correct assignment of the BXID for the data, the procedure of which is named as "coarse time alignment".

Two aspects contribute to the time misalignment. One of them is the system that controls the functioning of FE, which will be introduced in Section 3.1. When time-sensitive commands propagate through the system, delays can be accumulated. For these delays, by examining the data (as will be discussed in Section 3.3 and Section 3.4.1), the time alignment procedure treats the combined effects without tracing them back to the individual components of the system. It is, however, of interest to study the *changes* of the combined delay induced by the modifications of individual components. One such example is discussed in Section 4.3. Apart from the control system, another contributing aspect is the physics process which affects the signal arrival. This includes, *e.g.* the time-of-flight of the particles and the propagation time of the scintillating lights in the scintillating fibres. This aspect can be studied using Monte Carlo (MC) simulation, which will be discussed in Section 3.4.2.

To faciliate the discussion, a notation of referring to different parts of the FE is used. The three stations are referred to as T1, T2 and T3. The four layers within one station are referred to as L0-3. The four quarters within one layer are referred to as Q0-3. The modules are referred to as M0-4 and M5 (only in T3), with the smaller number indicating the modules being closer to the beam pipe in x direction. The two half-ROBs within a half-module are referred as H0 and H1. The 8 SiPM arrays within a half-ROB are referred to as D0-7.

[6] Refers to the beam completing the circle of the LHC ring.

3.1 Timing control system

The LHCb detector is controlled with two types of control commands, *i.e.* the fast control and the slow control. The fast control (Timing and Fast Control, TFC) handles commands with stringent timing requirements (*e.g.* triggers and bunch crossing counter reset). The slow control handles commands without such a timing requirements (*e.g.* changing the setting of FE thresholds). The objective of the time alignment concerns the fast control.

The system for distributing the control commands and receiving data from the FE is briefly introduced as follows [13]. The system utilises the PCI-express modules named PCIe40 boards [50], which host FPGA chips. These powerful electronic devices are used to perform various functionalities. The so-called SODIN [51] board acts as the readout supervisor with the interface for LHC and handles the central TFC. The SOL40 boards act as a control interface, which fans out the TFC. The TELL40 boards are used to process the data transmitted back from the FE and prepare them for the transmission to the event-builder. In order to perform these different functionalities, different firmwares are loaded onto these boards. These boards are located in the data center on the surface. The SOL40 and the TELL40 boards are connected with the FE components, which is located in the experiment cavern, through optical fibres with lengths of about 250m. The connections between SOL40 and FE are referred to as control links and a bidirectional communication between SOL40 and FE is supported. The connections from FE to TELL40 are referred to as data links. For them, only unidirectional transmission from FE to TELL40 is supported. This system is illustrated in Figure 3.1.

During the data-taking operation, LHCb receives the LHC clock information. This includes the LHC bunch clock and the orbit signal [52]: They are distributed to the FE. The orbit signal is used to reset the bunch crossing counters of the FE at every turn of the LHC bunch cycle to synchronise the system. The bunch clock provides the time reference for the sampling of the physics signal and the sampling of the TFC commands.

The LHC clock reaches the LHCb location through approximately 14km long fibres buried in shallow ground ($<$ 1m), which makes the clock transmission susceptible to the environmental temperature variation [53, 54]. As discussed in Refs. [53, 54], the beam phases are monitored for LHCb by the beam pickup devices located $\sim$ 150m away from the interaction point, which can provide feedbacks to the LHCb system in order to compensate the potential drift of the received LHC clock. Usually, for physics data-taking, the experts of LHCb adjust the clock globally [7] to compensate for a drift larger

[7] In LHCb, the so-called "global clock" can be used to shift timing for all subdetectors.

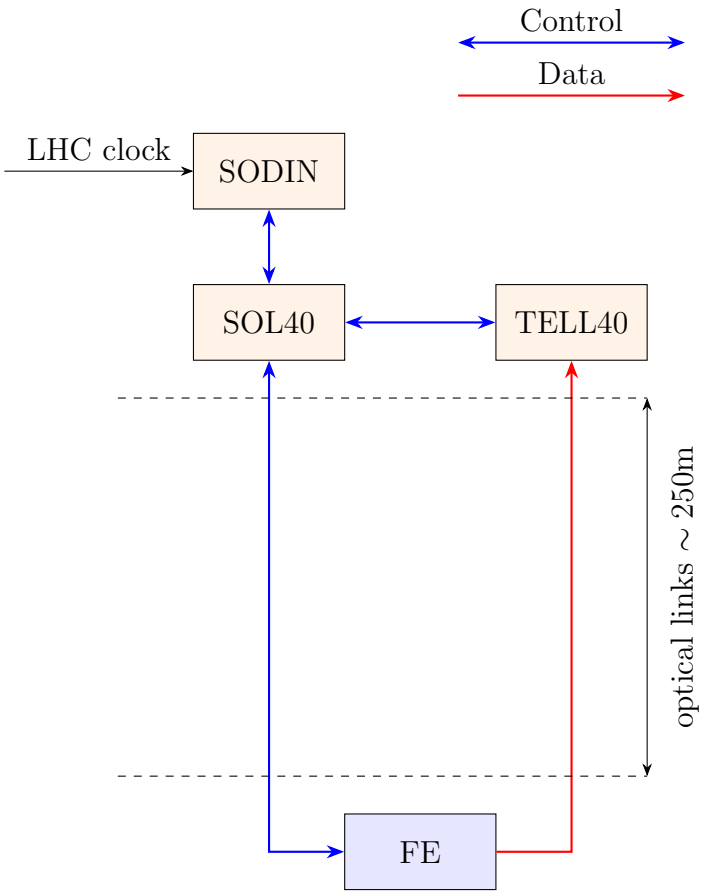


Figure 3.1: A very simplified illustration of the LHCb control and data distribution system. The plot was inspired by Ref. [13].

than 0.5 ns. However, time misalignments can still be generated within the LHCb system for the SciFi Tracker, which need to be treated by the aforementioned time alignment procedure.

Within the LHCb system, the components which contribute to the potential time misalignments include, for example, the SOL40 electronics and the optic transmission through the fibres connecting FE and SOL40. Since the digitised and transmitted data have encoded the BXID information, the data transmission path from FE to TELL40 does not contribute.

The clocks originated from the LHC clock are further distributed in the SciFi FE. Given the clock recevied, each master GBTx of the SciFi FE fans out this clock into the clocks for the 8 data GBTx chips corresponding to 8 data links. The data GBTx further distributes the clock to various electronic components where they can be used for, *e.g.* sampling in clustering FPGA

and controlling the light injection. Offsets can be added at the level of the master GBTx or of the data GBTx by tunable clock delays. The clock distribution in the FE is illustrated in Figure 3.2.

In systems described above, tunable BXID offsets can be introduced in SOL40 to adjust the timing by $n \times 25$ ns (n is an integer). These BXID offsets are used in the aforementioned "coarse time alignment" procedure. These offsets are adjustable separately for each control link, which corresponds to each half-ROB. The delay setting of the FE master GBTx clocks can be tuned for each data link separately to provide a finer granularity adjustment within $0 - 25$ ns. This corresponds to the aforementioned "fine time alignment" procedure. The delay setting of the FE master GBTx clocks is tunable in the granularity of 512 DACs (Digital-to-Analog Converter) covering 25 ns, *i.e.* adjustable at the finest granularity (single DAC) of about 48.8 ps. The data GBTx clock delay setting is determined by FE internal requirements, *e.g.* the required delay of the light injection and the robust internal data sampling. The data GBTx clock delay setting is not intended to be used for the previously mentioned fine or coarse time alignment.

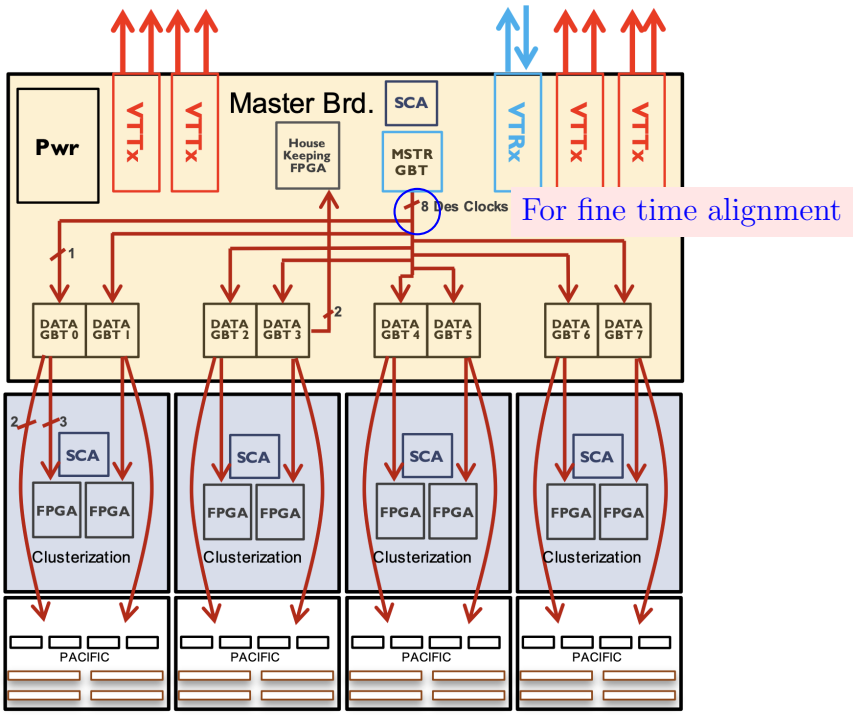


Figure 3.2: The clock distribution in the SciFi FE. The figure was adapted from Ref. [55]. One master GBTx chip fans out the clock into 8 data GBTx chips. The data GBTx further directs the clocks into clustering FPGA and the PACIFIC. The data GBTx number 3 additionally propagates the clock to the housekeeping FPGA, which is used to control the light injection. For the fine time alignment, the master GBTx clocks are adjusted.

3.2 Time alignment event

With the bunch crossings happening every 25 ns, a beam-beam collision can leave imprints beyond the 25 ns interval corresponding to its own bunch crossing, which is often referred to as spill-over. This can be caused by various reasons. In the case of the SciFi Tracker, the contributing factors include, *e.g.* the late arrival of the particles, the delayed arrival of the scintillating lights due to the reflection of the mirror and the extended shape of the PACIFIC response (as discussed in Section 2.4.3).

Thus, to perform the time alignment, the beam-beam collisions from the so-called isolated bunch crossings are needed. An isolated bunch crossing means that, in several neighouring bunch crossings around the bunch crossing of interest, there are no other beam-beam collisions. This is different from the bunch crossings usually used for physics production, which have beam-beam collisions in several consecutive bunch crossings to increase the luminosity.

During the data-taking for the time alignment, the data of several consecutive bunch crossings around an isolated bunch crossing are recorded. The sequence of several consecutive events around the bunch crossing of interest is referred to as the Time Alignment Event (TAE). This window around the bunch crossing of interest is referred to as the TAE window.

Within the TAE window, one can study the cluster occupancy in different BXIDs relative to the central bunch crossing. This information is essential for both the coarse and fine time alignment, which will be discussed in the following sections.

3.3 Coarse time alignment

While the author is not the main contributor for the coarse time alignment, it is closely related to the fine time alignment. It is thus introduced in this section.

The purpose of the coarse time alignment is to ensure the correct BXID assignment in the data. This can be achieved by examining, within the TAE window, the cluster occupancy in the relative BXIDs with respect to the central bunch crossing. Figure 3.3 illustrates an example of the distribution of the cluster occupancy in the relative BXIDs and for the 4096 different data links. The BXID with highest cluster count in the TAE window corresponds to the isolated bunch crossings, where the actual beam-beam interaction can happen. As shown in the plot, the BXID with highest cluster count is aligned to the TAE window center, *i.e.* the BXID is correctly assigned. If that is not the case, for example, if BXID with the highest cluster count is ± 1 off from the center, then the adjustment of the offsets in the SOL40 would be needed

to shift the BXID with the highest cluster count to the center.

Besides the correction assignment of the BXID, Figure 3.3 also shows several other features. The cluster count variation in the center BXID for different data links originates from the geometric distribution of the clusters. The data links arranged along the x-axis of the plot are grouped by quarters, which have higher cluster occupancy near the center and lower cluster occupancy in the outer region. Another feature is that the effect of the beam-beam interaction in the central BXID extends to the subsequent +1(2) BXIDs (spill-over).

The coarse time alignment is closely related to the fine time alignment. The eventual timing working point is a combined result of both the coarse and fine time alignment. This can be illustrated by the following example. Assuming initially the FE operates at the best timing working point of 49.9 ns, which is composed of a coarse time offset and a fine time offset:

$$49.9\,\text{ns} = 1 \times 25\,\text{ns}(\text{coarse offset}) + 24.9\,\text{ns}(\text{fine offset}) \ , \tag{3.1}$$

then, for some reasons, the best working point is shifted to 50.1 ns [8]. Even though this is only a 0.2 ns shift, both the coarse and the fine time offsets need to be adjusted in order to align the FE to the new working point, *i.e.*

$$50.1\,\text{ns} = 2 \times 25\,\text{ns}(\text{coarse offset}) + 0.1\,\text{ns}(\text{fine offset}) \ . \tag{3.2}$$

Another demonstration of the connection between the coarse and fine time alignment can be seen in Section 4.2.

[8] For simplicity, this illustrative example assumes the fine time alignment can be done to an arbitrary precison.

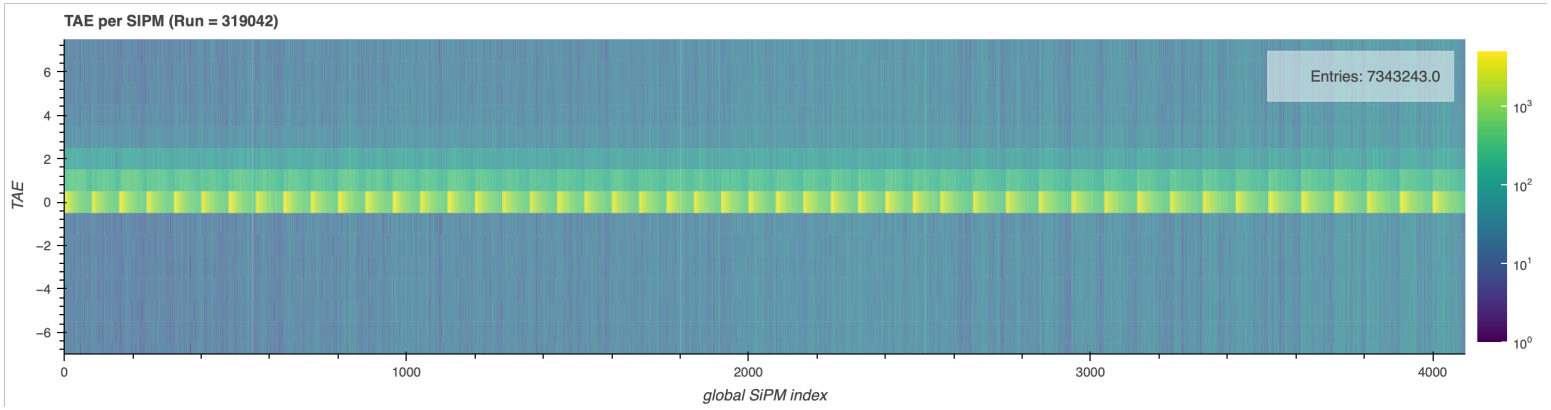


Figure 3.3: An example of the cluster occupancy (indicated by the color) distribution in the TAE window (± 7 around the central BXID) for the 4096 data links (corresponding to 4096 SiPM arrays) of the SciFi Tracker. The relative BXIDs within the TAE window are arranged along the y-axis of the plot. The different data links are arranged along the x-axis of the plot. The figure was produced using the LHCb monitoring software through its web interface [56].

3.4 Fine time alignment

3.4.1 Beam timing scan

The main objective of the fine time alignment is to find the optimised delay settings of the FE master GBTx clocks. Considering that no time information with a finer granularity than the BXID is contained in the data, fine timing information is extracted by examining the relationship between the cluster occupancy in relative BXIDs within the TAE window and the delay setting of the FE master GBTx clocks. To this end, the setting of the master GBTx clocks is scanned and data are taken at different settings [9]. Each setting corresponds to a so-called step in the scan. Such a procedure is called the beam timing scan. At the finest granularity, the scan can be conducted with 1 DAC per step ($\sim 48.8\,\text{ps}$ per step). In practice, a shift of $\sim 48.8\,\text{ps}$ near the optimised point is not expected to significantly change the hit efficiency while a coarser granularity can help to reduce the amount of time to collect and process the data. During 2023-2025, for the fine time alignment, the step granularity of 16 DACs was used ($\sim 0.78\,\text{ns}$ per step).

When the beam timing scan data are taken, they are labelled with a so-called run number. To faciliate the discussion, these run numbers will be used to refer to their corresponding data [10].

Figure 3.4 shows an example of the cluster occupancy distribution for one data link in the relative BXIDs within a TAE window and the step number of the scan. As the step number increases, the delay introduced by the master GBTx clock increases, which leads to a gradual shift of the cluster distribution into the next BXID. An one dimensional delay-occupancy curve can be constructed by examining the cluster count as a function of the time coordinate (t) assigned to the clusters according the following formula:

$$t = (\text{Relative BXID}) \times 25\,\text{ns} + (\text{Delay from FE})\ . \tag{3.3}$$

As an example, the resulting curve of a single data link is shown in Figure 3.5 (top). The central peak corresponds to the beam-beam interactions in the isolated bunch crossings. The tail of the peak extends beyond 25 ns, which corresponds to the spill-over effect.

From the delay-occupancy curve, one can define a reference point by using the so-called bunch crossing asymmetry method. The asymmetry is defined

[9] This is a special mode of operating the detector, *i.e.* the so-called Step Run. At each step, the FE setting can be changed according to the specification.

[10] In practice, sometimes several consecutive runs are needed to complete the scan. In this thesis, such a case is referred to as a single run with the run number being the first one in the group.

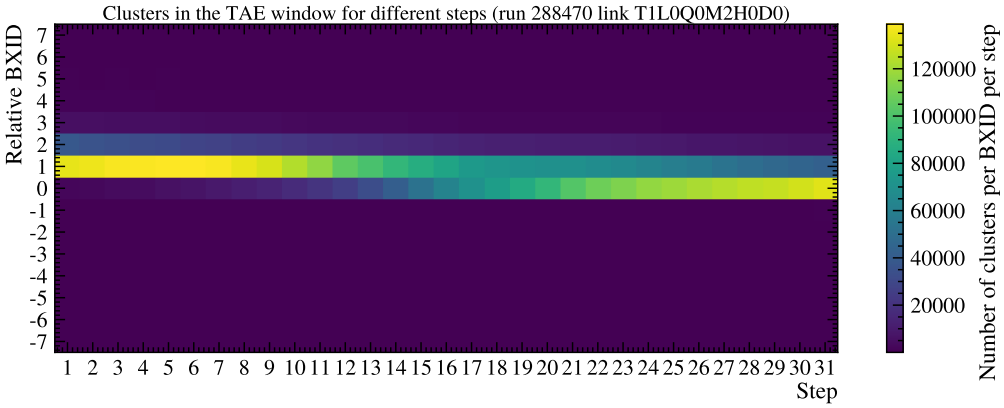


Figure 3.4: An example of cluster occupancy versus relative BXID and scanning step. The plot shows the clusters for one data link (T1L0Q0M2H0D0).

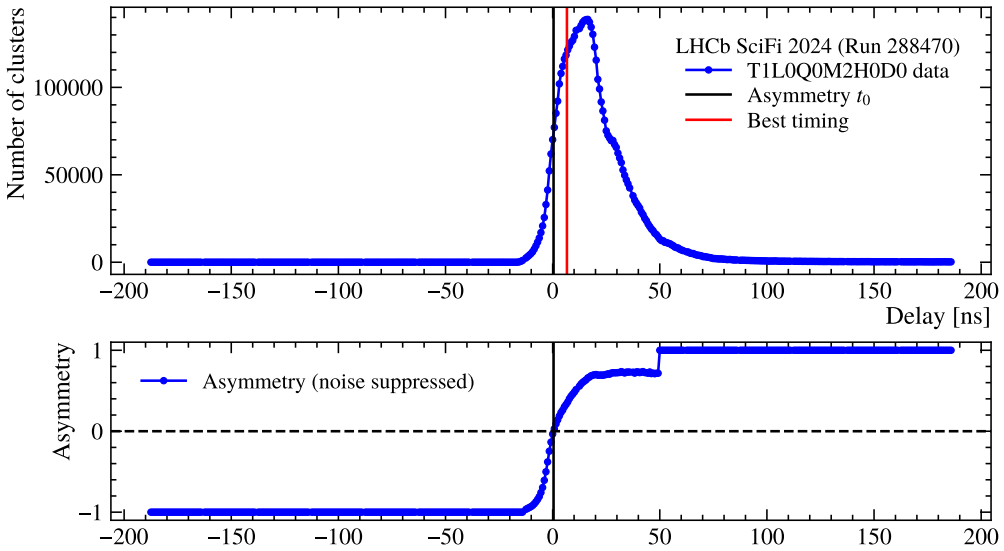


Figure 3.5: An example of the delay-occupancy curve (top) and delay-asymmetry curve (bottom) constructed from beam timing scan data. The plot shows the curve for one data link (T1L0Q0M2H0D0). The plot was also used in Ref. [57].

as

$$\text{asymmetry} = \frac{O_i - O_{i+1}}{O_i + O_{i+1}} \ , \tag{3.4}$$

where O_i stands for the occupancy in the current bunch crossing and O_{i+1} stands for the occupancy in the next bunch crossing, *i.e.* 25 ns later. The delay-asymmetry curve derived from the aforementioned delay-occupancy curve is shown in Figure 3.5 (bottom). The timing point where asymmetry = 0 can be used as a reference point, which is named t_0. Such a point takes the advantage of the rising (falling) behaviour of the curve around the mid-height and thus has the advantage of being statistically robust comparing to a point one might choose from the peak region [58, 59]. In general, this reference point is not the working point of the best performance but it characterises the relative position of the signal peak within the TAE window. Based on the reference point t_0, the optimised best working point t_{best} can be calculated as

$$t_{\text{best}} = t_0 + t_{\text{offset}} \ , \tag{3.5}$$

where the offset t_{offset} should be chosen according to the physics objectives. In the next section, the determination of t_{offset} is discussed.

3.4.2 Timing working point optimisation

In the SciFi Tracker, not all hits are induced by the particles of interest from the perspective of physics studies. Particles which travel through the SciFi layers can originate from sources of interest (*e.g.* b-hadron decays) or from other sources (*e.g.* the particles produced by the material interaction). In the following discussion, the former are referred to as the primary particles while the latter are referred to as the secondary particles. The clusters induced by primary (secondary) particles are referred to as primary (secondary) clusters. Due to the different production mechanisms and kinematic distributions, the arrival time of the primary particles and the secondary particles can be very different. For the interest of physics studies, one would usually aim to optimise for the detection of the primary particles during the time alignment. Since both the primary and the secondary clusters contribute to the delay-occupancy curve (discussed in the previous section), t_{offset} is not chosen to maximise the number of clusters at t_{best}.

In the data collected by the beam timing scan, there is no easy way to clearly separate the primary clusters from the secondary clusters. This separation, however, can be achieved in the simulation, where the truth information about the sources of the particles are available.

Using the simulation [11], the timing working point $t_{\text{best}}^{\text{MC}}$ can be determined to optimise for the primary particles (once the actual method is defined). In the same fashion as in the data, the asymmetry zero point t_0^{MC} in the simulation can also be determined using the combined occupancy of both the primary and secondary clusters. The difference between the two points can then be used as t_{offset} (as in Equation 3.5):

$$t_{\text{offset}} = t_{\text{best}}^{\text{MC}} - t_0^{\text{MC}} \ . \tag{3.6}$$

In the determination of $t_{\text{best}}^{\text{MC}}$, two important factors to consider are the arrival time of the charged particles and the propagation time of the scintillating lights in the scintillating fibres.

Figure 3.6 illustrates the distribution of the simulated hit creation time, *i.e.* the arrival time of the charged particles. The three peaks correspond to the spatially separated three stations of SciFi. As shown in the plot, comparing with the primary clusters, the secondary clusters have a considerable part of late arrivals.

When a charged particle traverses the scintillating fibres, scintillating light is emitted. The scintillating light propagates through the scintillating fibre and reach the SiPMs, where the scintillating photons are converted into electronic signals. Given the refractive index of the core material $n = 1.59$ [41], the speed of light in the scintillating fibre is about 1m/5.3 ns. This leads to about 13 ns for the light to travel through the 2.5m long half-module. Thus, considerable differences exist in the arrival time at the SiPMs for lights generated by hits in different y-positions of the SciFi. To study this dependence, one can divide the half-modules into 5 y-regions in the interval of 500 mm. Figure 3.7 illustrates this idea.

Figure 3.8 (using T1L0M0 as an example) shows the delay-occupancy curve (normalised) in different y-regions. The notation T1L0M0 is used to refer to the result merging the corresponding half-modules in 4 quarters, *i.e.* the cluster counts of T1L0Q0M0, T1L0Q1M0, T1L0Q2M0 and T1L0Q3M0 are merged. Such a treatment was applied because, at the time, it was believed that the 4 quarters should be equivalent considering the geometric symmetry (ignoring the stereo angle in the case of u, v layers). It is not entirely the case in reality, which will be discussed later. As shown in the plot, for the primary clusters, each y-region has a plateau, within which the cluster count is close to the maximum. In order to optimise hit efficiency for all 5 y-regions, the optimised working point should be chosen in the

[11] In the following discussion, minimum bias simulation (general pp interaction without exclusive requirements of certain signal decays) samples were used, which were produced by collaborators [60].

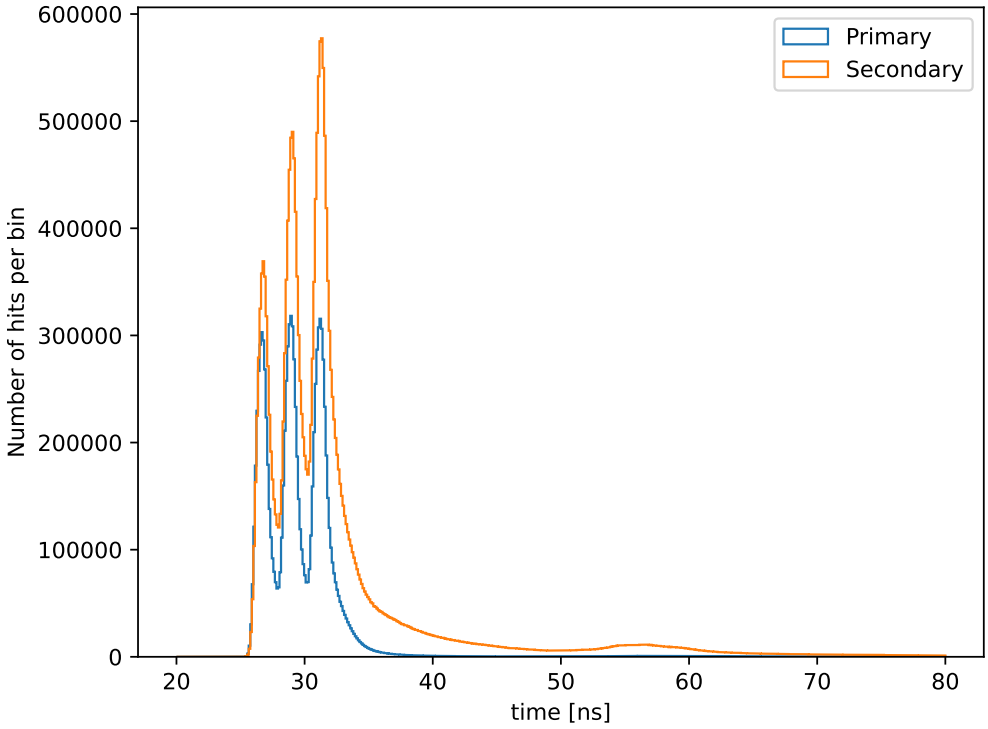


Figure 3.6: The distribution of hit creation time in the simulated sample. The three peaks correspond to the three SciFi stations. As can be seen from the region around 35 ns, the distribution corresponding to the secondary particles has a longer tail.

overlapping region of these plateaus. The relationship between the delay and the cluster count vary for different parts of the detector. The counterparts of Figure 3.8 in several other parts of the detector are shown in Appendix A. It is worth noting that, comparing to the inner (smaller module number) half-modules, the overlappings between the plateaus in several outer (larger module number) half-modules are reduced. For these outer half-modules, while a common overlapping plateau region in a strict sense for all 5 y-regions cannot be found, the plateaus are still close to each other. Although the exact mechanism behind this difference between the inner half-modules and outer half-modules is not fully understood, it is likely due to the variation in the arrival time of particles with different momenta and different travel distances. Given the magnetic field along the y direction, the charged particles with lower momentum are expected to have larger drifts in the x direction and more likely to reach the outer half-modules.

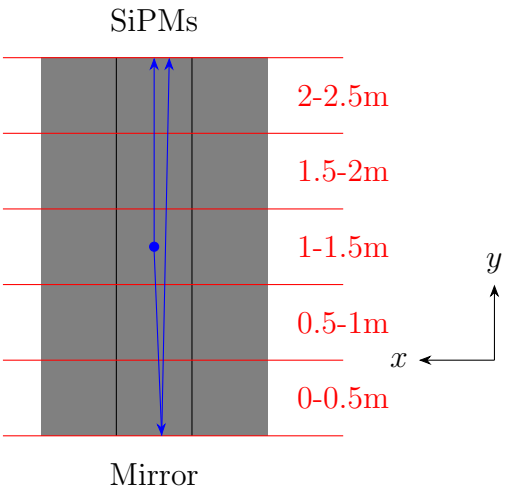


Figure 3.7: An illsutration of the y region division. As an example, the grey block indicates the active area of 3 half-modules. The bottom end corresponds to the mirror position. The top end corresponds to the position of SiPMs. The blue dot indicates a point where a charged particle traverses the x-y plane. The blue lines indicate the scintillating lights, which transmit directly or via mirror reflection to the SiPMs.

The working point optimisation method discussed in the following paragraphs is based on the idea to optimise hit efficiency for as many y-regions as possible. Naively, one could try to find an optimised working point by examining directly primary cluster delay-occupancy curves. However, due to the large statistics in the central region (y near 0), this would lead to a plateau largely dominated by the central region. Instead, one can average the aforementioned normalised delay-occupancy curves over the 5 y-regions, optionally with weights.

As an example, Figure 3.9 shows such averaged curves with two different y-region weights, *i.e.* even weights $1:1:1:1:1$ and weights $3:3:3:3:1$ (from the center to the outer region). The optimised working point can then be found by finding the high points of the averaged curves. In practice, to be statistically robust, the optimised working point is chosen as the median of the interval defined by the 0.96 line [12] (as illustrated in Figure 3.9).

The $1:1:1:1:1$ configuration emphasizes that all 5 y-regions should have good hit efficiency. In principle, it is a more stringent requirement than

[12] In principle, one can check the implications of choosing different values at the similar level. However, it is not expected to significantly affect the extracted median point.

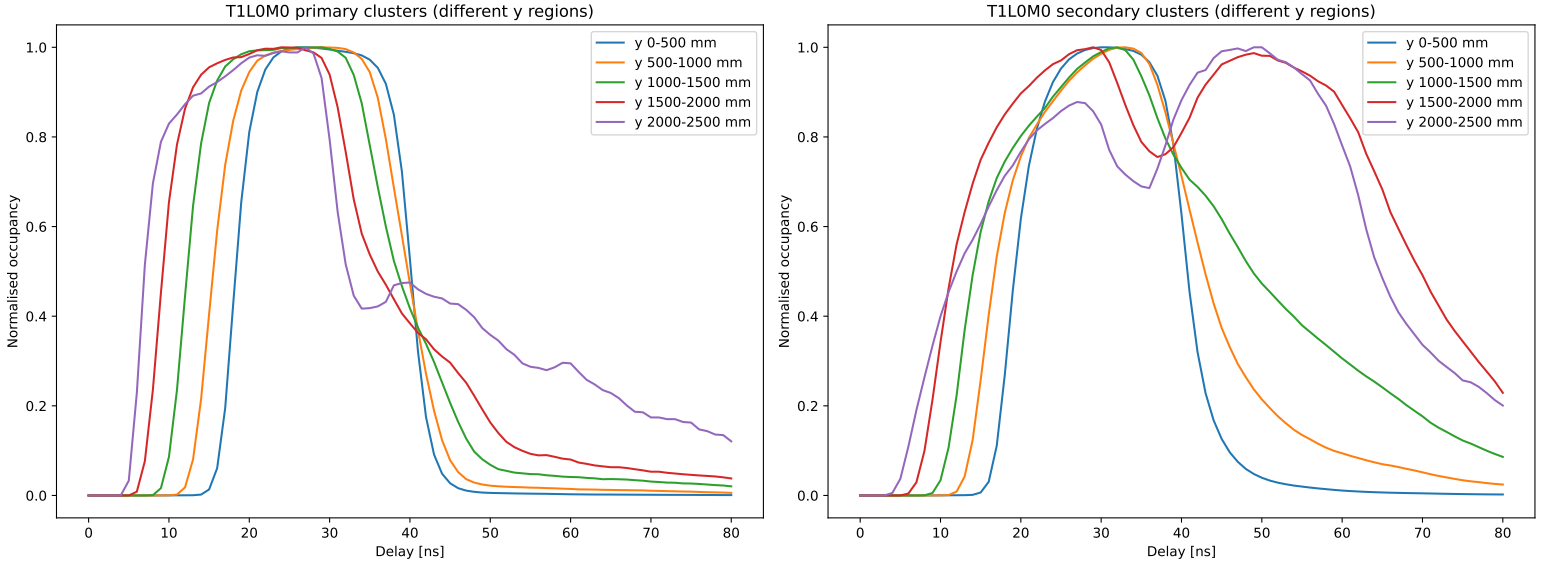


Figure 3.8: An example (T1L0M0) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y direction (*of the plot*) such that their peak values are normalised to 1. The corresponding half-modules in the 4 quarters (T1L0Q0M0, T1L0Q1M0, T1L0Q2M0 and T1L0Q3M0) are merged. In the outer y region, there is a second peak structure. This can be caused by the delayed arrivals of the scintillating photons reflected from the mirror separating the two half-modules and by the shape of the extended tail of the PACIFIC (analog part) response.

the 3 : 3 : 3 : 3 : 1 configuration. This is reflected by the difference of the peak widths at the 0.96 line shown in Figure 3.9 (the width is larger for the 3 : 3 : 3 : 3 : 1 configuration). The 3 : 3 : 3 : 3 : 1 configuration is based on a more realistic consideration: From the perspective of the data-taking, the outermost y-region is less important due to the lower chance to have a hit. From the perspective of the simulation t_{offset} study, it contributes a higher statistical fluctuation to the determination of t_{offset}. In 2024, the t_{offset} result derived from the 1 : 1 : 1 : 1 : 1 configuration was used to perform the fine time alignment. Due to the aforementioned consideration for the lower occupancy in the outermost y-region, an updated version of t_{offset} derived from the 3 : 3 : 3 : 3 : 1 configuration was used in 2025 for the fine time alignment.

As described by Equation 3.6, the offsets t_{offset} can be extracted using the optimised working points and the asymmetry t_0^{MC} points in the simulation. Figure 3.10 illustrates an example of the optimised working point and the t_0^{MC} in the overlay with two cluster delay-occupancy curves (one case with only primary clusters and the other case with primary and secondary clusters combined). Considering that the t_0 point in data is determined using the delay-occupancy curve based on the total cluster counts, the t_0^{MC} point is determined using the delay-occupancy curve combining primary and secondary clusters. It is worth noting that, in Figure 3.10, the optimised working point is displaced from the peak position of the "Primary + Secondary clusters" delay-occupancy curve. Indeed, the goal of the working point optimisation is not to maximise the total cluster count. Another feature shown in Figure 3.10 is that the primary cluster delay-occupancy curve has a large plateau. This is dominated by the inner regions (closer to the $y = 0$ position) due to the higher cluster occupancy comparing to the outer regions. Indeed, the aforementioned y-weighting treatment "magnifies" the contribution of the outer regions.

The counterparts of Figure 3.10 (T1L0M0) in T1L0M3 and T1L0M4 are shown in Appendix A Figure A.7 (T1L0M3) and Figure A.8 (T1L0M4). They exhibit considerably different delay-occupancy curve shapes comparing to Figure 3.10. This is not surprising considering the difference in the particle arrival time between the inner and outer half-modules (as discussed before).

Figure 3.11 shows one version of the t_{offset} determined in different parts of the detector. In this version, the 5 y-regions are evenly weighted and the 4 quarters of each layer are merged when determining the optimised working points. This version was used for the 2024 time alignment.

The layout in Figure 3.11 is a compact way of displaying the values of a quantity associated with different parts of the detector. The layout can be separated into the left side and the right side from the middle of x axis.

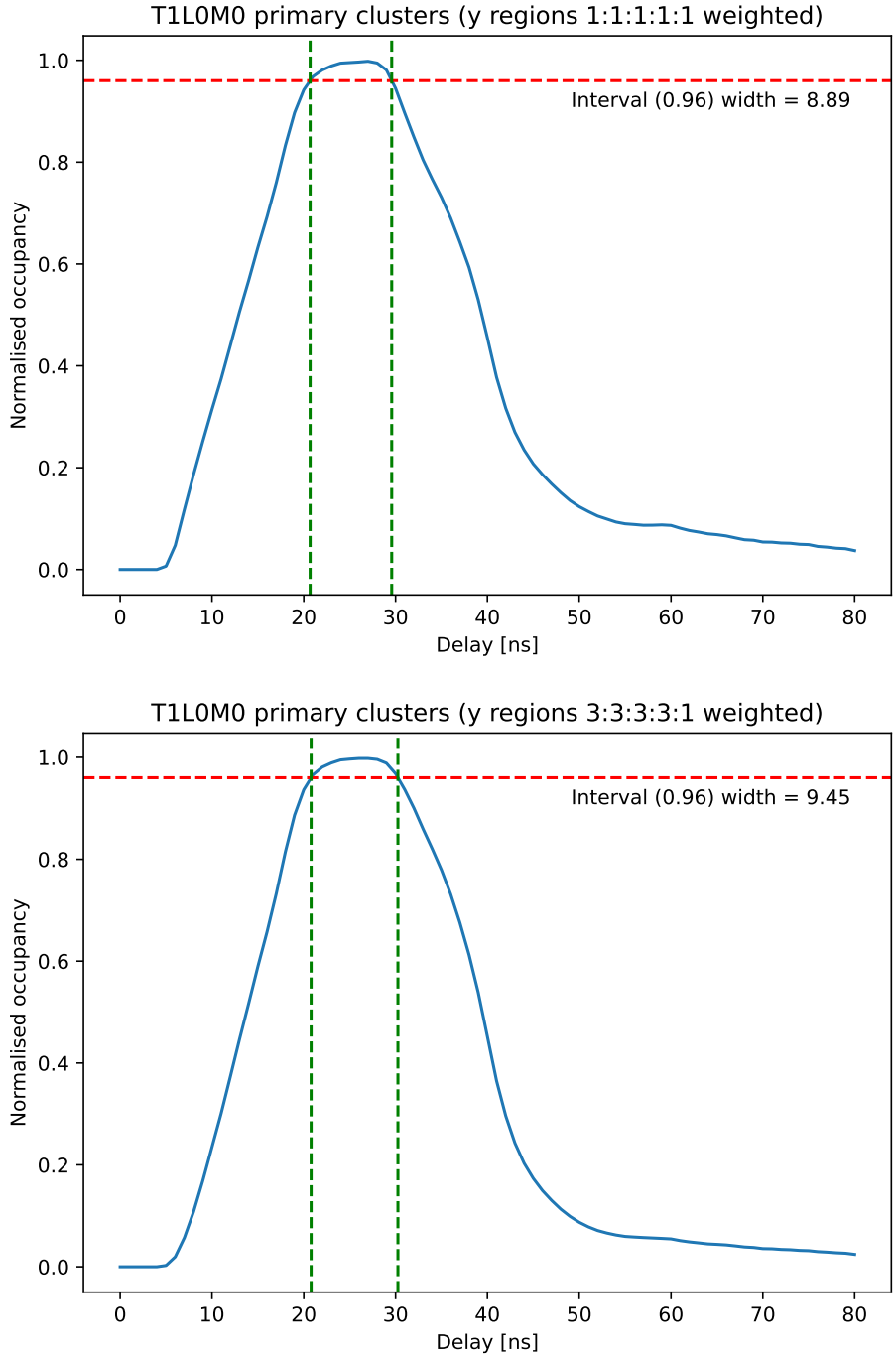


Figure 3.9: The y-weighted delay-occupancy curves of the four-quarters-merged half-module T1L0M0 from the simulation. The top plot corresponds to the even weights. The bottom plot corresponds to the $3:3:3:3:1$ weights.

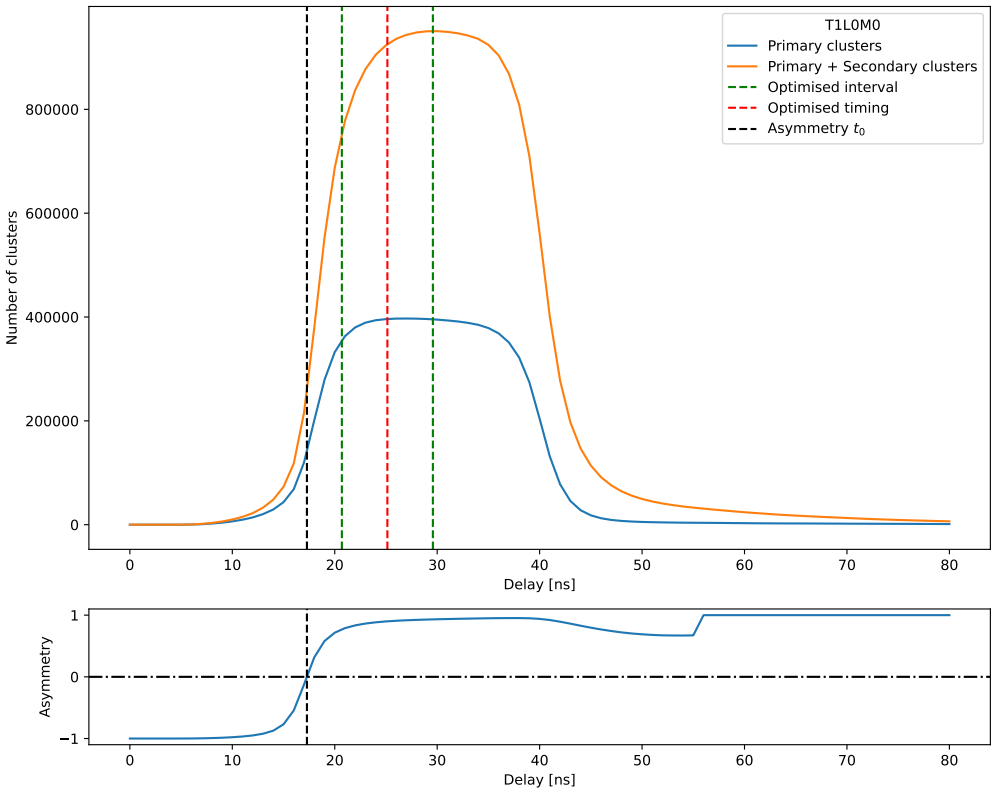


Figure 3.10: An illsutration (T1L0M0) of the asymmetry t_0^{MC} and the optimised working point in the overlay with the delay-occupancy curves ("Primary clusters" and "Primary + Secondary clusters") based on the MC samples. The optimised interval corresponds to the aforementioned interval defined by the 0.96 line. In this plot, the optimised working point was derived with the $1:1:1:1:1$ weight configuration and the corresponding half-modules in the 4 quarters are merged.

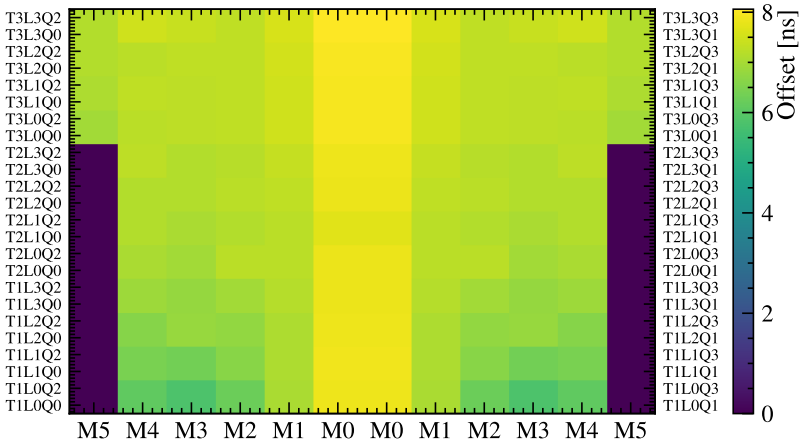


Figure 3.11: One version of t_{offset} extracted from the simulation. In this version, the 4 quarters are merged when determining the optimised working points. The 5 y-regions are evenly weighted. The M5 in T1/T2 are non-existing locations, which are showing here only for the convenience of ploting.

Within each side, a single slot in the y axis corresponds to one quarter in one layer of the detector (as indicated by the y axis labels). Along the x axis, different data links (or half-ROBs/half-modules) within a single quarter are arranged following the same sequence in the geometric layout of the detector, with the inner (outer) parts of the detector assigned to be around the middle (two ends) of the x axis.

Figure 3.11 shows a variation of t_{offset} across the different parts of the detector, which can be explained with the variation in the delay-occupancy curves. As discussed before, although the exact mechanism behind it is not fully understood, it is likely due to the variation in the arrival time of particles with different momenta and different travel distances.

The above treatment of merging 4 quarters assumes the 4 quarters to be equivalent. However, in the late 2024, it was found that the LHCb magnet polarity has a quarter-dependent effect on the t_0, which leads to a revisit of the t_{offset} determination. Figure 3.12 illustrates the difference of t_0 in various data links from two sets of beam timing scan runs with different magnet polarities.

In the simulation sample (with the downwards magnet polarity), the similar pattern is seen in the difference between quarters (Q0 versus Q3 and Q1 versus Q2). From the geometric perspective, the magnet polarity flip is equiv-

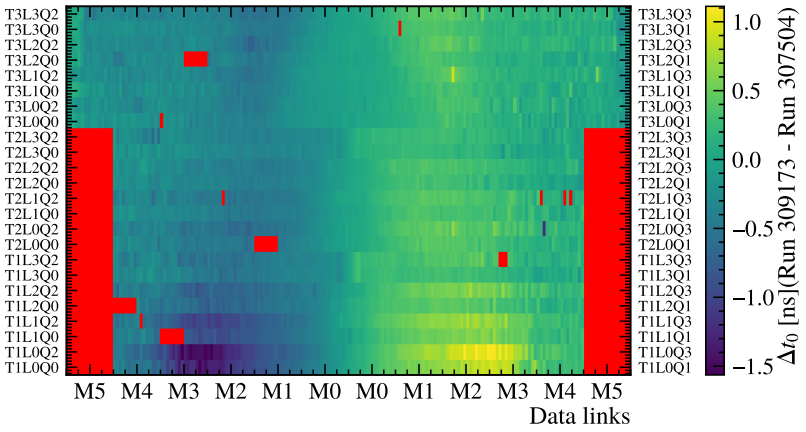


Figure 3.12: The difference of t_0 from two sets of beam timing scan runs with different magnet polarity. The dataset labelled with run number 307504 was taken with the magnet-down configuration. The dataset labelled with run number 309173 was taken with the magnet-up configuration. The red color indicates the non-existing link (M5 of T1/T2) and some masked problematic links (*e.g.* too noisy or excluded during the data-taking due to other issues).

alent to a rotation of 180 degrees along the beam pipe direction, *i.e.* a Q0-Q3 switching and a Q1-Q2 switching. Figure 3.13 illustrates such a difference between quarters in the simulation. The quantity shown, for a given location (loc.), is calcualted as the t_0 at the location minus the average t_0 between itself and its counterpart in the other corresponding quarter (Q0 versus Q3, Q1 versus Q2):

$$
\begin{aligned}
&\Delta t_0(\text{loc. in } Q_0) \\
&= t_0(\text{loc. in } Q_0) - \frac{t_0(\text{loc. in } Q_0) + t_0(\text{corresponding loc. in } Q_3)}{2} \,, \\
&= \frac{t_0(\text{loc. in } Q_0) - t_0(\text{corresponding loc. in } Q_3)}{2} \,, \\
&\Delta t_0(\text{loc. in } Q_3) = \frac{t_0(\text{loc. in } Q_3) - t_0(\text{corresponding loc. in } Q_0)}{2} \,, \\
&\Delta t_0(\text{loc. in } Q_1) = \frac{t_0(\text{loc. in } Q_1) - t_0(\text{corresponding loc. in } Q_2)}{2} \,, \\
&\Delta t_0(\text{loc. in } Q_2) = \frac{t_0(\text{loc. in } Q_2) - t_0(\text{corresponding loc. in } Q_1)}{2} \,.
\end{aligned}
\tag{3.7}
$$

The pattern of such a quarter difference resembles the pattern observed in

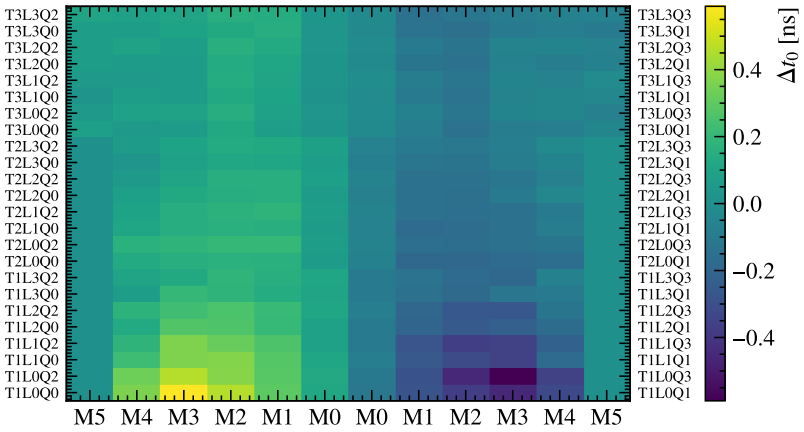


Figure 3.13: An illustration of the t_0 difference between quarters (as described in Equation 3.7) in the simulation. The simulation samples used has the magnet-down configuration. The M5 links in T1/T2 are non-existing locations.

data (Figure 3.12).

The physics reason behind such a magnet effect was not fully understood. However, it is likely that the material interaction effect is asymmetric under the magnet polarity flip.

To take into account the magnet effect, the t_{offset} needs to be determined for each quarter separately. Used in 2025, an updated version of the t_{offset} are determined with each quarter treated separately and using the y-region weights $3:3:3:3:1$ (as discussed).

Figure 3.14 shows the version of t_{offset} with these updates and extracted from the magnet-down simulation samples.

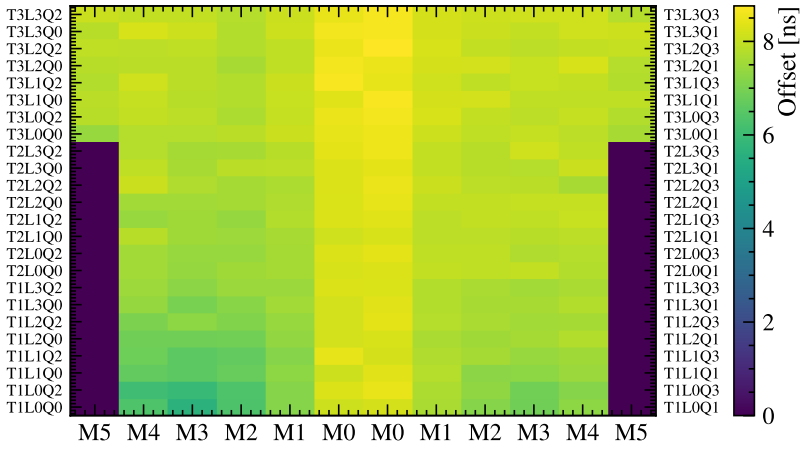


Figure 3.14: One version of t_{offset} extracted from the simulation (with the magnet-down configuration). In this version, the quarters are treated separately. The 5 y-regions are weighted by $3:3:3:3:1$. The M5 links in T1/T2 are non-existing locations.

4 Time alignment studies

During the commissioning and the early operation of the SciFi Tracker, several studies have been carried out for the time alignment. These studies provided significant insights into the timing characteristics of the related system, *e.g.* its stability and variation in different parts of the detector. In the following, some studies using the 2023-2024 data are discussed.

4.1 Statistical fluctuation

As discussed in Section 3.4.1, the extraction of t_0 with the asymmetry method depends on the cluster counts. Thus, the determination of t_0 is affected by statistical fluctuations. The statistics of the data depends on many factors, *e.g.* the number of events collected and the so-called μ value, which refers to the average number of visible pp interactions in a single bunch crossing.

To quantify the effect of statistical fluctuations and check the reproducibility of the method, two runs taken shortly one after the other are compared. Figure 4.1 shows the distribution of the difference of the per-data-link t_0 values extracted from the two runs. Each of the runs has 2.5 million events per step. They were taken with the pp collisions at the center-of-mass energy 13.6 TeV in the same LHC fill, which has an average $\mu \sim 3.95$ and a peak $\mu \sim 5.33$. The per-data-link Δt_0 in different parts of the detector is shown in Figure 4.2. Some problematic links (*e.g.* links with very high noise) are masked. One important feature is that the outer modules have fewer clusters and this leads to a larger statistical fluctuation, which is visible in the plot.

In Figure 4.1, the Δt_0 for the majority of the data links are within ± 0.2 ns.

4.2 Variation across the detector

Due to the variation of the delay in the electronic system and the lengths of the control link fibres, the determined t_0 can vary across different parts of the detector. As discussed in Section 3.1 and Section 3.3, there are BXID offsets in the SOL40, which are adjustable per control link. Depending on the BXID offsets applied at the time of the data-taking, the t_0 with respect to the center of the TAE window can shift by the granularity of 25 ns. In other words, a different BXID offset would have resulting in a shift of $n \times 25$ ns (n is an integer) for the delay-occupancy curve constructed in the way described in Section 3.4.1, which would mean a shift of $n \times 25$ ns for t_0 (using the TAE

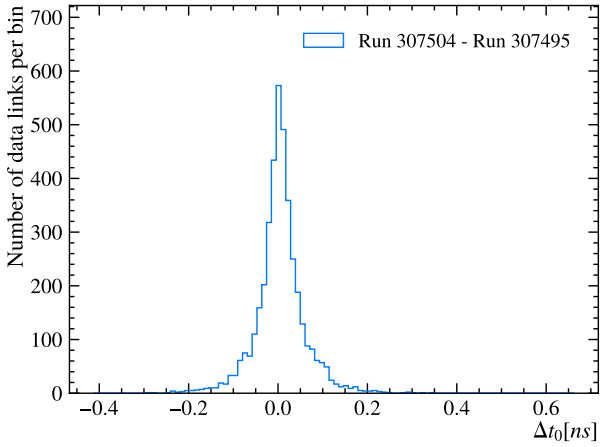


Figure 4.1: The distribution of per-data-link Δt_0 (between two runs taken shortly one after the other) for different data links in the detector. Some problematic data links are not included. A similar plot was used in Ref. [57].

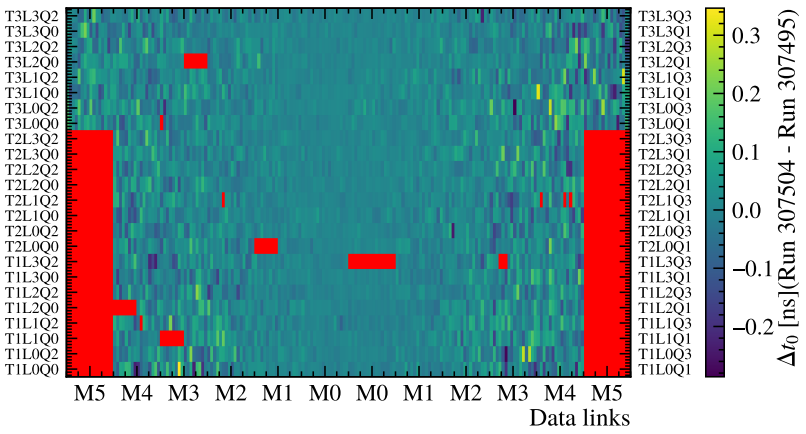


Figure 4.2: The Δt_0 in different data links for two runs taken shortly one after the other. The red color indicates the non-existing (*i.e.* M5 for T1/T2) and masked links.

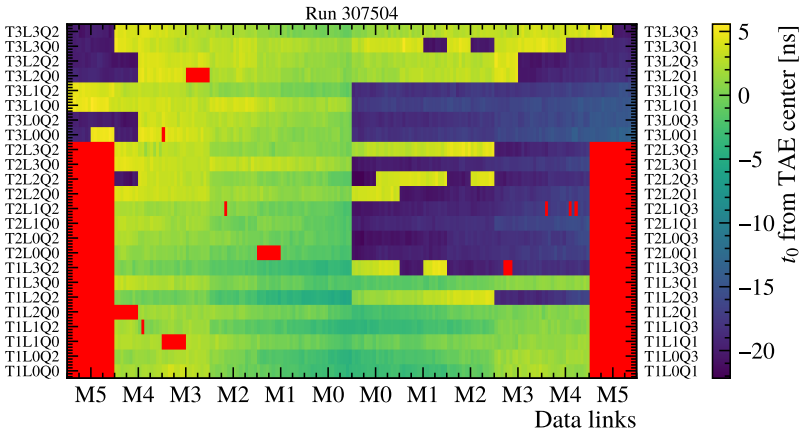


Figure 4.3: The t_0 in different data links. The time coordinate uses the center of the central bunch crossing as the zero point. Along the x-axis, the data links are grouped in such a way that the range within a half-module Mi label ± 2 ticks indicates the data links belong to that half-module. The data links between the Mi tick and 2 ticks to it left (right) belongs to a single half-ROB. The red color indicates the non-existing and masked links.

window center as the 0 point of the time coordinate) [13].

As an example, given the BXID offsets applied at the time of data-taking, the t_0 variation across the detector is illustrated in Figure 4.3. A large difference across the detector, especially in different half-ROBs (corresponding to different control links), can be seen.

Within a single half-ROB, the t_0 difference is expected to originate from, for example, the internal electronic differences in the FE and the differences of the geometric positions of the data links, which can cause slight differences in the delay-occupancy curve shapes and particle arrival time. To check variation within a single half-ROB, one would compare the t_0 between the 8 different data-links within a single half-ROB. Since the variation of t_0 with respect to the TAE window center is dominated by the difference between different half-ROBs (Figure 4.3), to study the t_0 variation within each half-ROB, one needs to find a reference time point for each half-ROB separately, which can be used as a baseline to compare t_0. By summing up the delay-occupancy curves for the 8 data links within a half-ROB and applying the

[13] If n is too large (severe coarse time misalignment), the peak of the delay-occupancy curve might even reside outside of the TAE window, which would prevent a successful fine time alignment.

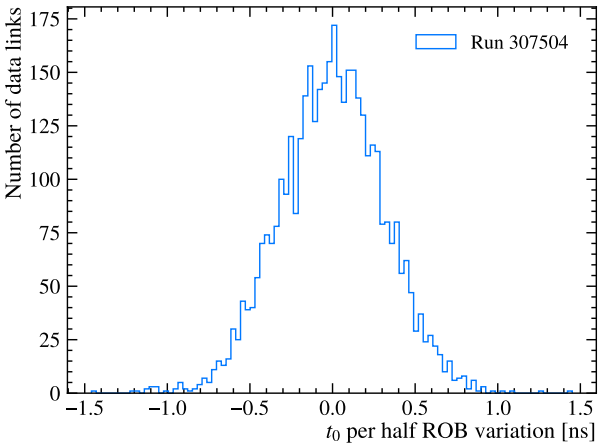


Figure 4.4: The t_0 per half-ROB variation.

asymmetry method on such a combined curve, one can derive a per-half-ROB t_0. This can serve as the aforementioned reference time point. The difference between the per-data-link and per-half-ROB t_0 can be used to study the t_0 variation within half-ROBs. This is demonstrated in Figure 4.4. Considering that the same run 307504 is used in Figure 4.1 and Figure 4.4, the variations shown in Figure 4.4 are beyond the statistical fluctuations. Figure 4.4 indicates that, for majority of the data links, the t_0 per-half-ROB variation is within ± 1 ns.

4.3 Effects of SOL40 reprogramming

As discussed in Section 3.1, the SOL40 boards act as interfaces in the control command and clock distribution. A change in the electronic system could cause a clock phase shift. One such possible change is the reprogramming of the SOL40 FPGA firmware. Since a reprogramming affects the logical gate configuration on the FPGA, it also has timing effects. During the detector operation, a reprogramming of SOL40 might be needed as a part of the infrastructure maintenance. It is thus useful to check the timing effect of such an intervention.

Figure 4.5 shows the difference of t_0 from runs taken before and after a test of SOL40 reprogramming. The run 265388 (taken before the reprogramming) is used as the baseline for the calculation of Δt_0. Figure 4.6 shows another such instance of test with run 268875 (taken before the reprogramming) as the baseline. As shown in the histogram plots, the post-reprogramming Δt_0 distribution clearly deviates from the prior-reprogramming distribution. In

the detector layout plots, it is visible that a structure appears in the group of 4 quarters, which corresponds to the C-Frame structure introduced in Section 2.4.2. In the electronic system, such a group (*e.g.* T2L0Q0, T2L0Q2, T2L1Q0 and T2L1Q2) is connected to the same SOL40 board. In both instances (Figure 4.5 and Figure 4.6), most of the data links show t_0 shifts at the sub-nanosecond level.

Although a single instance of SOL40 reprogramming does not necessarily lead to a timing shift large enough to cause a serious deterioration of detector hit efficiency, the exact shifts in different parts of the detector are usually unpredictable. During the detector operation, it is thus recommended to check the fine timing after a SOL40 reprogramming, subject to the availability of the isolated bunch crossings [14].

[14]The isolated bunch crossings are costly from the perspective of luminosity production. They are not always available.

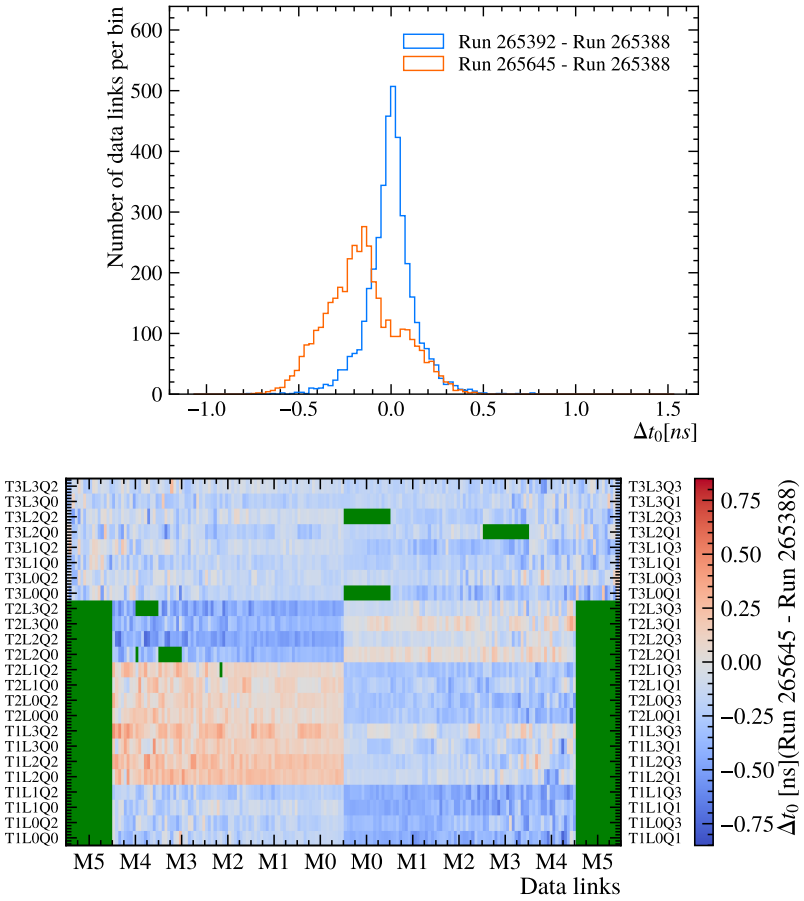


Figure 4.5: The distribution of the difference of t_0 before and after a test SOL40 reprogramming conducted on June 1st 2023. The run 265388 and 265392 were taken before the reprogramming while run 265645 was taken after the reprogramming. The top plot shows the per-data-link Δt_0 in histograms. The bottom plot shows the Δt_0 between run 265388 and 265645 in the detector layout. The green color in the bottom plot indicates the non-existing and masked links.

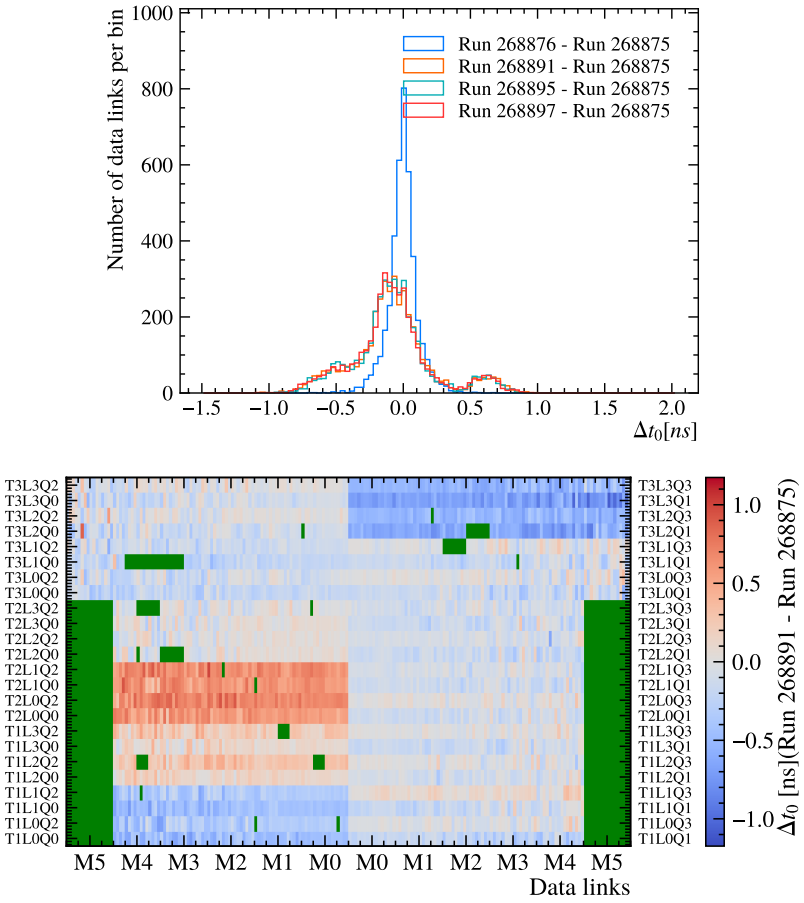


Figure 4.6: The distribution of the difference of t_0 before and after a test SOL40 reprogramming conducted on July 4th 2023. The run 268875 and 268876 were taken before the reprogramming while run 268891, 268895 and 268897 are taken after the reprogramming. The top plot shows the per-data-link Δt_0 in histograms. The bottom plot plot shows the Δt_0 between 268891 and 268875 in the detector layout. The green color in the bottom plot indicates the non-existing and masked links.

4.4 Timing stability

Numerous runs of time alignment data were taken during the course of the detector commissioning and early operation. A comparison of t_0 from a group of runs taken during 2023 is shown in Figure 4.7. As discussed in Section 4.2, the t_0 are largely different across different data links. The comparison is thus done by taking a baseline run (the leftmost run in the plot) and calculating the difference of t_0 for each data link. The BXID offsets in the SOL40 were adjusted when the coarse time alignment changes, which can happen even if the actual time displacement has only a minor change (as discussed in the illustrative example in Section 3.3). This can lead to shifts of t_0 (relative to the TAE window center) with the granularity of 25 ns. Thus, the comparison was performed with Δt_0 subtracting $n \times 25\,\text{ns}$, where n is the integer when $n \times 25\,\text{ns}$ is closest to Δt_0. In Figure 4.7, the points of the curve represent mean value of Δt_0 (with the treatment discussed above) across different data links. The errorbar represents the standard deviation across different data links. Problematic data links (*e.g.* too noisy or disabled) in any of the runs shown in Figure 4.7 are excluded from the calculation.

Similarly, the comparison of a group of runs from 2024 is shown in Figure 4.8.

In Figure 4.7 and Figure 4.8, the central points of the curves shifts within about $\sim 2\,\text{ns}$, which indicate that the system has reasonably good timing stability.

It is worth noting that, in the time period corresponding to an early part of the 2023 curve (a region of run number from around 261000 to around 264000), a drift of LHC clock of about 1 ns was observed by the LHCb global clock monitoring [15]. Comparing with the shift of about 1 ns observed by the SciFi fine time alignment procedure, this is a strong indication of the good sensitivity of the fine time alignment method. Although a similar comparison cannot always be achieved due to many reasons: For example, during physics data-taking period, the global clock is adjusted by experts when the drift of LHC clock is too large (as discussed in Section 3.1), which means there would be only small variations induced by the LHC clock which *could* be observed. In addition, an intervention in the system (*e.g.* SOL40 reprogramming) can potentially change the relative phases between clocks within the LHCb system. It should be noted that the SciFi time alignment concerns more about treating the time misalignments generated within the LHCb system than treating the drift of LHC clock.

[15] Beam phases are monitored for LHCb as discussed in Section 3.1

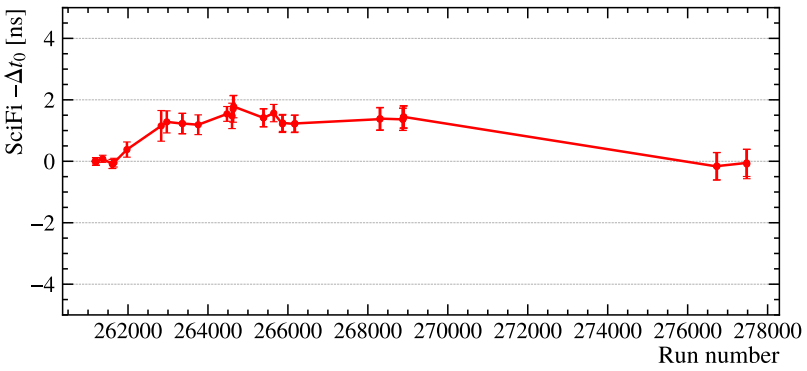


Figure 4.7: The comparison of a group of runs taken during the year 2023. The leftmost run 261179 was taken on April 22nd 2023 and is used here as the baseline. The rightmost run 277479 was taken on September 27th 2023. The meanings of the points and the errorbars are as explained in the text. Problematic data links (*e.g.* too noisy or disabled) in any of the runs shown here are excluded from the calculation. A similar plot was used in Ref. [57].

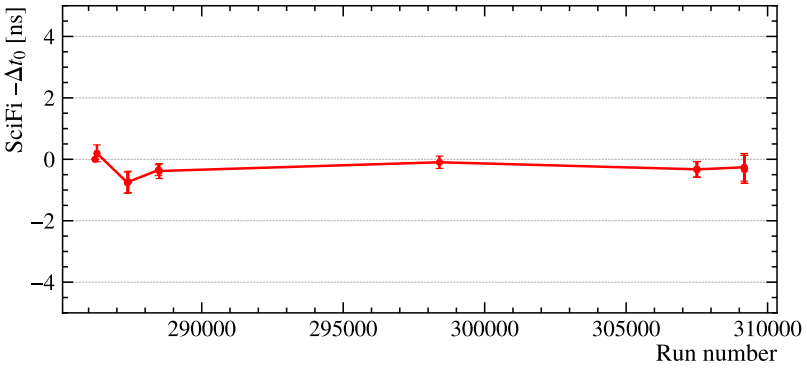


Figure 4.8: The comparison of a group of runs taken during the year 2024. The leftmost run 286229 was taken on March 19th 2024 and is used here as the baseline. The rightmost run 309186 was taken on October 28th 2024. The meanings of the points and the errorbars are as explained in the text. Problematic data links (*e.g.* too noisy or disabled) in any of the runs shown here are excluded from the calculation. A similar plot was used in Ref. [57].

4.5 Remarks on the time alignment and the hit efficiency

Considering the studies discussed in the previous sections, the fine time alignment method has demonstrated a good sensitivity to detect time shifts: With a sufficient amount of data, as in the case discussed in Section 4.1, the statistical fluctuations for t_0 can be suppressed to the level of 0.2 ns. In the cases discussed in Section 4.3, sub-nanosecond level time variations in the electronic readout and clock distribution system can be detected.

However, in the discussion so far, two questions concerning the main objective of the fine time alignment (*i.e.* optimising the hit detection efficiency) remain unanswered. The first question is: what is the hit efficiency of the SciFi detector obtained with the delay settings determined by the fine time alignment procedure. The second question is: what is the size of the time shift which can be tolerated without a severe loss of hit efficiency.

Concerning the first question, dedicated hit efficiency studies have been conducted (and are still being conducted) by collaborators (*e.g.* Ref. [46]). It should be noted that the measured hit efficiency depends on, not only the time alignment, but also various other factors, *e.g.* the setting of the thresholds in the SciFi FE, the spatial alignment and the event multiplicity (affecting accidental matches). Figure 4.9 shows an example of the measured hit efficiency of the 12 detector layers. The content of the plot is explained in Ref. [61]: The efficiencies shown are the averaged efficiencies of all SiPM arrays in each layer (calculated separately for 128 channel indices). A SiPM array has two dies (each die corresponds to 64 channels) and there is a gap in between. The lower efficiencies in the middle and at the two ends correspond to the edges of the dies. The lower efficiencies in layer 10 are due to some data links being excluded during the data-taking. There were also some channels (a few percent) across the detector with non-ideal thresholds, which contribute to the slightly lower hit efficiencies.

Figure 4.9 shows that, excluding these aforementioned known effects, the average hit efficiencies have achieved the detector hit efficiency design value (99 percent [43]). More detailed studies are needed to better understand the different contributing effects and be able to disentangle the effects from potential time misalignment from the other contributing factors.

Concerning the second question, it has to be conceded that a test of intentionally misaligning the detector timing was not performed, considering that such a test would be performed at the expense of the physics data-taking program. In addition, there is a technical difficulty. In order to perform hit efficiency measurement, track reconstruction is needed [46]. Thus, a sufficient part of the detector needs to be time aligned. It is likely that the such a time

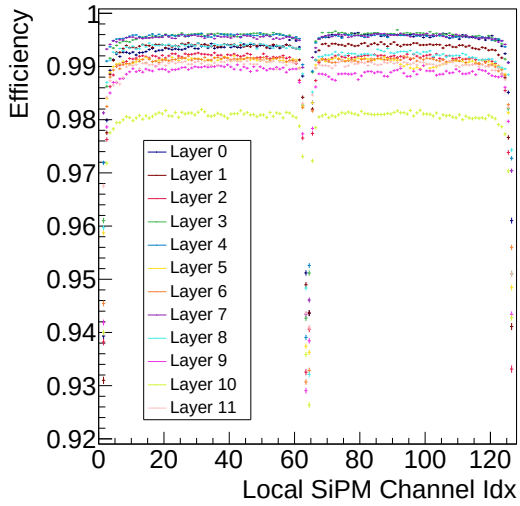


Figure 4.9: The measured hit efficiency based on the data from a LHC fill in July 2024. The y-axis corresponds to the efficiency value. The x-axis corresponds to the index of the 128 channels of a SiPM array (described in Section 2.4.1). Plot taken from Ref. [61].

misalignment test needs to be performed for each layer separately. This is technically challenging and potentially time consuming.

An easier way will be to perform the study with simulation (although there are discrepancies between the simulation and the reality). As discussed in Section 3.4.2, the primary clusters are used to optimise the timing working point. Given a certain primary cluster delay-occupancy curve, the tolerable time shift can be estimated based on the width of the plateau region. However, the delay-occupancy curves vary significantly in different y-regions and between the inner and outer half-modules, In the cases shown in Figure 3.8 and its counterparts in Appendix A, a typical plateau width for a single y-region (covering 0.5m) is about 10 ns (for primary clusters). The relationship between different y-regions vary for inner and outer half-modules (as discussed in 3.4.2).

5 Motivation for the $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ LFU study

5.1 The Standard Model and its LFU

The Standard Model (SM) of particle physics formulates an elegant theory for the known basic building blocks of the matter and their interactions (except for the gravity).

In the fermionic sector of the SM, there are three generations of quarks and leptons. They are described as spinor fields with spin 1/2. The quark sector consists of three up-type quarks (u, c, t) and three down-type quarks (d, s, b). The lepton sector consists of the charged leptons (e, μ, τ) and the neutrinos $(\nu_e, \nu_\mu, \nu_\tau)$.

The basic interactions of the SM is described in the form of the gauge theory. The electroweak interaction follows the $\mathrm{U}(1)_Y \times \mathrm{SU}(2)_L$ gauge symmetry while the strong interaction follows the $\mathrm{SU}(3)_c$ gauge symmetry. These interactions are mediated by vector bosons, which have spins equal to 1. Besides the vector bosons, the SM includes a special scalar boson with spin 0, *i.e.* the Higgs boson. It provides the mechanism for the particles to acquire masses.

The strong interaction is mediated by the gluons. The interaction between the gluons and the fermions is limited to the quark sector. The electroweak interaction, however, involves both the quarks and leptons. Through the Higgs mechanism, an non-zero vacuum expectation is introduced and the electroweak symmetry is broken, resulting in the ElectroMagnetic (EM) interaction mediated by the massless photon, the Neutral Current (NC) weak interaction mediated by the massive Z boson, and Charged Current (CC) weak interaction mediated by the massive $W^\pm$ bosons.

Mathematically, the interaction between these bosons with the leptons are described by their corresponding terms in the Lagrangian. To faciliate the discussion, the photon field is denoted as A_μ, the $Z/W^\pm$ fields are denoted as $Z_\mu/W_\mu^\pm$ and the lepton fields are denoted as ℓ/ν_ℓ, where ℓ stands for the electron (e), the muon (μ) or the tau (τ). The electromagnetic (EM) and weak interaction (NC and CC) terms in the Lagrangian for the leptons can then be written as [62, 63] [16]:

$$\mathcal{L}_{\mathrm{EM}}^{\mathrm{lepton}} = -e \sum_\ell Q_\ell \bar{\ell} \gamma^\mu \ell A_\mu \ , \tag{5.1}$$

$$\mathcal{L}_{\mathrm{NC}}^{\mathrm{lepton}} = -\frac{g}{2\cos\theta_W} \sum_\ell Z_\mu \left(\bar{\ell}\gamma^\mu (g_V^\ell - g_A^\ell \gamma^5)\ell + \bar{\nu}_\ell \gamma^\mu (g_V^{\nu_\ell} - g_A^{\nu_\ell}\gamma^5)\nu_\ell \right) \ , \tag{5.2}$$

[16] The coupling e stands for the electric charge of the positron, the angle θ_W is the weak mixing angle and the coupling g is related to them as $g = e/\sin\theta_W$ [62].

$$\mathcal{L}_{\rm CC}^{\rm lepton} = -\frac{g}{2\sqrt{2}} \sum_{\ell} \left(W_{\mu}^{-} \bar{\ell} \gamma^{\mu} (1-\gamma^5) \nu_{\ell} + W_{\mu}^{+} \bar{\nu}_{\ell} \gamma^{\mu} (1-\gamma^5) \ell \right) \ , \tag{5.3}$$

where the couplings are universal for different lepton generations [62, 63]:

$$Q_{\ell} = -1 \ , \ g_V^{\ell} = 2\sin^2\theta_W - \frac{1}{2} \ , \ g_V^{\nu_{\ell}} = \frac{1}{2} \ , \ g_A^{\ell} = -\frac{1}{2} \ , \ g_A^{\nu_{\ell}} = \frac{1}{2} \ . \tag{5.4}$$

The $\mathcal{L}_{\rm EM}^{\rm lepton}$ term describes the interaction between the photon and the charged leptons. The photon can directly couple with two charged leptons of the same flavour. The neutrinos have zero electric charge and thus do not participate in the electromagnetic interaction. The $\mathcal{L}_{\rm NC}^{\rm lepton}$ term describes the interaction between the Z boson and the leptons. The Z boson can directly couple to either two charged leptons or two the neutrinos of the same flavour. The $\mathcal{L}_{\rm CC}^{\rm lepton}$ term describes the interaction between the $W^{\pm}$ boson and the leptons. The $W^{\pm}$ boson can directly couple to a charged lepton and a neutrino of the same flavour.

The feature that the three generations of lepton have the same couplings in the SM electroweak interaction is known as the Lepton Flavour Universality (LFU). Experimentally, LFU in the Z and $W^{\pm}$ boson decays is well-tested [64] [17]:

$$\begin{aligned}
\frac{\Gamma(Z \to \mu^+\mu^-)}{\Gamma(Z \to e^+e^-)} &= 1.0001 \pm 0.0024 \ , \\
\frac{\Gamma(Z \to \tau^+\tau^-)}{\Gamma(Z \to \mu^+\mu^-)} &= 1.0010 \pm 0.0026 \ , \\
\frac{\Gamma(W^+ \to \mu^+\nu)}{\Gamma(W^+ \to e^+\nu)} &= 1.000 \pm 0.004 \ , \\
\frac{\Gamma(W^+ \to \tau^+\nu)}{\Gamma(W^+ \to \mu^+\nu)} &= 1.002 \pm 0.02 \ .
\end{aligned} \tag{5.5}$$

Another worth noting feature is that the $\mathcal{L}_{\rm EM}^{\rm lepton}$, $\mathcal{L}_{\rm NC}^{\rm lepton}$ and $\mathcal{L}_{\rm CC}^{\rm lepton}$ terms shown above does not allow a direct coupling between the leptons of different flavours. In the quark sector, the structure has an important difference. In the SM, both the up-type and down-type quarks are massive and acquire their masses through the Yukawa couplings with the Higgs boson. Their mass eigenstates, which correspond to the diagonalised Yukawa coupling matrices, do not overlap with the electroweak interaction flavour eigenstates. For the electromagnetic and neutral current interactions, the transformation from one set of eigenstates to the other effectively cancels with its Hermitian conjugate. Thus, no flavour changing mechanism is introduced for the quarks in

[17] The error of $\Gamma(W^+ \to \tau^+\nu)/\Gamma(W^+ \to \mu^+\nu)$ reported by PDG includes a 2.0 scale factor [64].

the electromagnetic and neutral current interaction. For the charged current, however, such a cancellation cannot be achieved. The corresponding flavour changing mechanism is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [65, 66]. In the case of leptons, if the neutrinos are considered massless, the transformation between the mass eigenstates and the electroweak interaction flavour eigenstates can be absorbed through a redefinition of the neutrino eigenstates. In the case of massive neutrino (as known from the neutrino oscillation), a similar mechanism is described by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [67, 68].

5.2 The $b \to s\ell^+\ell^-$ process and the R_{H_s} observable

Possible new physics models, unlike the SM, do not necessarily obey LFU. While being experimentally well-constrained in the $Z/W^\pm$ decays, the potential lepton flavour non-universal effect can become visible when it interfer with SM loop suppressed processes. One of such processes is the $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ decay. The underlying quark-level transition for $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ is $b \to s\ell^+\ell^-$, which is a Flavour-Changing-Neutral-Current (FCNC) process. In the SM, no direct flavour changing mechanism exists in the electromagnetic interaction and the neutral current interaction mediated by the Z boson (as discussed in the previous section). Thus, in the SM, FCNC cannot happen at the tree-level but only via higher-order diagrams. Figure 5.1 shows two examples of these higher-order diagrams in the SM for the $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ decay. Being suppressed in the SM, the $b \to s\ell^+\ell^-$ process provides an excellent testing ground for the new physics, which can have interferences with the SM process. For example, a hypothetical leptoquark which couples with both leptons and quarks can introduce a tree level contribution to the $b \to s\ell^+\ell^-$ process (*e.g.* Ref. [69]). Another example is a new Z' boson with flavour non-universal couplings (*e.g.* Ref. [70]).

The b or s quark is usually part of a hadron, therefore a more general form of the decays like $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ can be written as $H_b \to H_s\ell^+\ell^-$, where H_b stands for a hadron containing b quark and H_s stands for a hadron containing s quark. To study LFU (in this case, specifically between the muon and the electron) of these decays, a ratio observable, *i.e.* the relative branching fraction between the muon and the electron channels, can be defined:

$$R_{H_s} = \frac{\mathcal{B}(H_b \to H_s\mu^+\mu^-)}{\mathcal{B}(H_b \to H_s e^+e^-)} \ . \tag{5.6}$$

Using this observable has a certain advantage from the theoretical perspective. Usually, the non-perturbative effects from the strong interaction can complicate theoretical studies for the hadron decays. However, in the case of

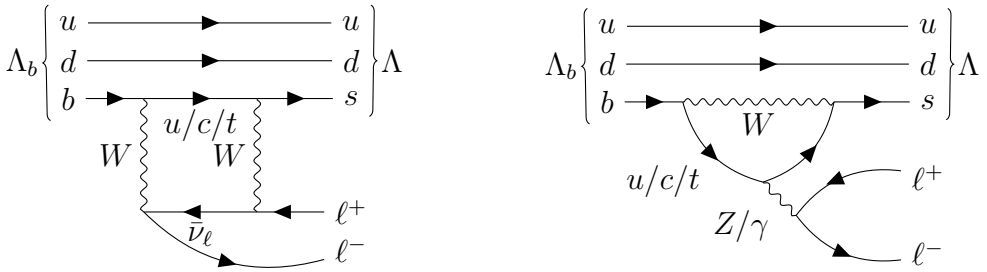


Figure 5.1: Two examples of the Feynman diagrams for the $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ decay in the SM.

the R_{H_s} type observables in the SM, these strong interaction effects largely cancel between the muon and the electron channels (to the extent of the kinematic difference) [71–73]. As a result, in the SM, R_{H_s} can be predicted with high precision to be ~ 1. Should an experimental measurement show incompatibility with this expectation, it would be a hint for the potential new physics.

In the study of the $b \to s\ell^+\ell^-$ type decays, there is an extra complication, namely the decays via the charmonium resonances. The tree-level $b \to sc\bar{c}$ transition, with the charmonium decay $c\bar{c} \to \ell^+\ell^-$, can lead to the same final state particles as the FCNC $b \to s\ell^+\ell^-$ transition. These charmonium mode decays can have large branching fractions comparing to the FCNC $b \to s\ell^+\ell^-$ process. However, they have distinctive signatures in terms of the dilepton invariant mass, the square of which is usually denoted as q^2. The charmonium contributions focus around the two prominent charmonium resonances, *i.e.* the J/ψ ($q^2 \sim 9.6\,\mathrm{GeV}^2/c^4$) and the $\psi(2S)$ ($q^2 \sim 13.6\,\mathrm{GeV}^2/c^4$) [64]. Figure 5.2 illustrates the overlapping of the $b \to s\ell^+\ell^-$ type rare decay and the charmonium peaks in the dilepton q^2 spectrum. Taking this complication into consideration, the observable R_{H_s} is usually studied in restricted q^2 regions to avoid the charmonium peaks:

$$R_{H_s}|_{q_1^2<q^2<q_2^2} = \frac{\int_{q_1^2}^{q_2^2} \mathrm{d}q^2 \frac{\mathrm{d}\mathcal{B}(H_b \to H_s \mu^+\mu^-)}{\mathrm{d}q^2}}{\int_{q_1^2}^{q_2^2} \mathrm{d}q^2 \frac{\mathrm{d}\mathcal{B}(H_b \to H_s e^+e^-)}{\mathrm{d}q^2}} \,. \tag{5.7}$$

5.3 Related experimental research

LFU of muon and electron for the $b \to s\ell^+\ell^-$ transition has been extensively studied at LHCb. Among the relatively statistically abundant decays of this type are the $B^+ \to K^+\ell^+\ell^-$ decay and the $B^0 \to K^{*0}\ell^+\ell^-$ decay. The

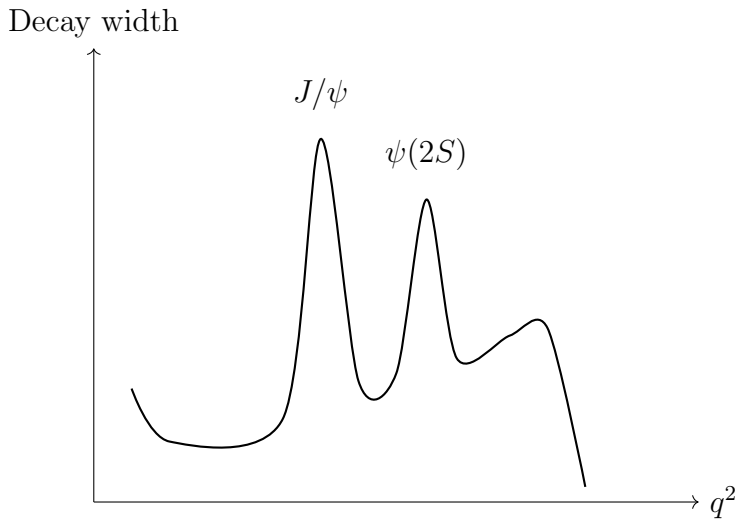


Figure 5.2: An illustration of the dilepton q^2 spectrum for the $b \to s\ell^+\ell^-$ type rare decay overlapping with the charmonium peaks. The scales in the decay width and the q^2 are only for the illustrative purpose. The actual scale of the amplitudes at different q^2 can vary for different decays.

corresponding observables are R_K and R_{K^*}, *i.e.* substituting the H_s with K, K^* in the aforementioned R_{H_s}. The so-called R_X ($X = K, K^*$) analysis gave the latest result on the LFU test of these two observables for q^2 below the charmonium resonances [74, 75]:

$$\begin{aligned} R_K|_{0.1<q^2<1.1\,\mathrm{GeV}^2/c^4} &= 0.994^{+0.094}_{-0.087}\ , \\ R_K|_{1.1<q^2<6.0\,\mathrm{GeV}^2/c^4} &= 0.949^{+0.048}_{-0.047}\ , \\ R_{K^*}|_{0.1<q^2<1.1\,\mathrm{GeV}^2/c^4} &= 0.927^{+0.099}_{-0.093}\ , \\ R_{K^*}|_{1.1<q^2<6.0\,\mathrm{GeV}^2/c^4} &= 1.027^{+0.077}_{-0.073}\ . \end{aligned} \tag{5.8}$$

The R_K was also measured in the q^2 region above the charmonium resonances [76]:

$$R_K|_{q^2>14.3\,\mathrm{GeV}^2/c^4} = 1.08^{+0.11}_{-0.09}(\mathrm{stat.})^{+0.04}_{-0.04}(\mathrm{syst.})\ . \tag{5.9}$$

Another example is the decay $B_s \to \phi\ell^+\ell^-$, with the corresponding observable R_ϕ^{-1} measured to be [77]:

$$\begin{aligned} R_\phi^{-1}|_{0.1<q^2<1.1\,\mathrm{GeV}^2/c^4} &= 1.57^{+0.28}_{-0.25}(\mathrm{stat.}) \pm 0.05(\mathrm{syst.})\ , \\ R_\phi^{-1}|_{1.1<q^2<6.0\,\mathrm{GeV}^2/c^4} &= 0.91^{+0.20}_{-0.19}(\mathrm{stat.}) \pm 0.05(\mathrm{syst.})\ , \\ R_\phi^{-1}|_{15.0<q^2<19.0\,\mathrm{GeV}^2/c^4} &= 0.85^{+0.24}_{-0.23}(\mathrm{stat.}) \pm 0.10(\mathrm{syst.})\ . \end{aligned} \tag{5.10}$$

Compared with the meson sector, the baryon sector is less explored by LHCb. Only the measurement with $\Lambda_b^0 \to pK^-\ell^+\ell^-$ has been published [78]:

$$R_{pK}^{-1}|_{0.1<q^2<6.0\,\mathrm{GeV}^2/c^4} = 1.17^{+0.18}_{-0.16}(\mathrm{stat.}) \pm 0.07(\mathrm{syst.})\ , \tag{5.11}$$

which has an additional requirement on the combined invariant mass of the proton and the kaon: $m(pK^-) < 2600\,\mathrm{MeV}/c^2$. Considering the tolerance of uncertainties, these measurements are in agreement with the LFU expectations. Nevertheless, considering that the Λ_b^0 and Λ^0 baryons have different hadronic structures than the mesons, a study with $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ can provide additional insights. The $\Lambda_b^0 \to \Lambda^0$ transition also has an attractive feature from the theoretical perspective: the Λ^0 particle is stable under the strong interaction, which makes the calculation of $\Lambda_b^0 \to \Lambda^0$ form factors very suitable for the standard Lattice QCD method [79].

Besides LFU, the $b \to s\ell^+\ell^-$ transition is also extensively studied in other aspects, especially in the differential branching fractions and the angular distributions of the $H_b \to H_s\mu^+\mu^-$ type decays. In some cases, the measurements exhibit intriguing tensions with respect to the SM, For example, the branching fractions of the $B_s \to \phi\mu^+\mu^-$ decay in some q^2 bins below the charmonium resonances are measured to be lower than the SM predications [80], which is shown in the left plot of Figure 5.3. Similar trends were also seen in several other decays: $B^0 \to K^*(892)^0\mu^+\mu^-$ [81], $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^0\mu^+\mu^-$ [82]. Another example is the angular distribution parameter (the so-called P_5') for the $B^0 \to K^{*0}\mu^+\mu^-$ decay [83]. This is illustrated in right plot of Figure 5.3. Unlike the R_{H_s} type observables, the theoretical predications for the differential branching fractions and the angular distributions are, in general, more susceptible to the uncertainties caused by the strong interaction. This warrants further caution to make new physics discovery claims with these observables.

For the decays of interest in this analysis, the measurement for the branching fractions of the $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ decay in various q^2 bins exist, which uses the data collected by LHCb during Run 1 [84] [18]. There is an ongoing effort within LHCb to perform the measurements using combined Run 1 and Run 2 data. Figure 5.4 shows the results of the published Run 1 measurement. Based on this measurement, the main contribution of the $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ decay resides in the q^2 region above the charmonium resonances. Given the significance of the signal, the $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ is considered observed (significance $> 5\sigma$) in the q^2 region above the charmonium resonances while, for the q^2

[18] The measurement was done using the normalisation channel $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$. The branching fraction $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)$ used in Ref. [84] to convert the result into $\mathcal{B}(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)$ has been superseded by the measurement in Ref. [85].

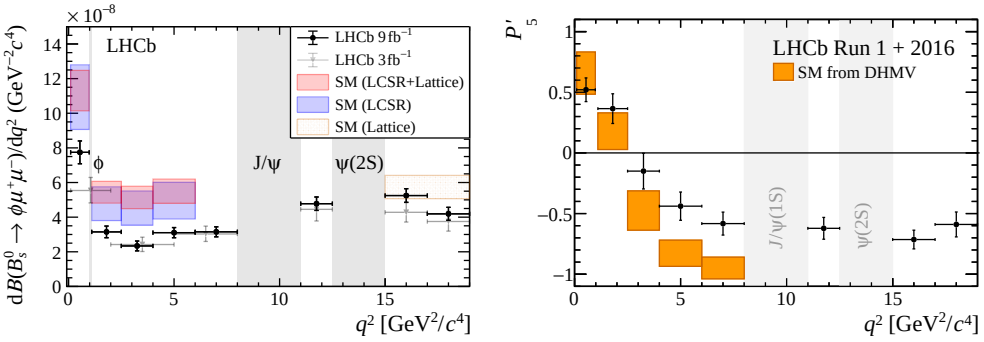


Figure 5.3: The left plot (taken from Ref. [80]) shows the comparison between the measurement and several SM predications for the branching fractions of the $B_s \to \phi\mu^+\mu^-$ decay. The right plot (taken from Ref. [83]) shows the comparison between the measurement and a SM predication for the angular distribution parameter P_5' for the $B^0 \to K^{*0}\mu^+\mu^-$ decay in different q^2 bins.

region below the charmonium resonances, only evidence of signal was claimed (significance $> 3\sigma$ for $0.1 < q^2 < 2.0\,\mathrm{GeV}^2/c^4$) [84]. For the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ decay, no corresponding measurements exist.

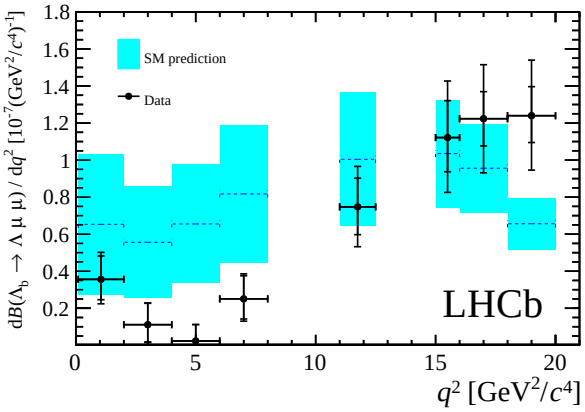


Figure 5.4: The previously measured branching fractions of $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ in different q^2 bins. Plot taken from Ref. [84].

6 Analysis strategy

A LFU test with $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ ($\ell = e$ or μ) decays can be performed by comparing the branching fractions $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)$ and $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$. Experimentally, the branching fraction $\mathcal{B}$ is related to the number of observed signal decay N_{sig}, which is given by:

$$N_{\text{sig}} = \left(\int \mathcal{L} \mathrm{d}t \right) \cdot \sigma(\Lambda_b^0) \cdot \mathcal{B} \cdot \epsilon_{\text{sig}} \ , \tag{6.1}$$

where $\int \mathcal{L} \mathrm{d}t$ stands for the integrated luminosity, $\sigma(\Lambda_b^0)$ stands for the production cross section of Λ_b^0 and ϵ_{sig} stands for the efficiency of observing such a decay.

Since the integrated luminosity and the Λ_b^0 production cross section are common between $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$, the analysis focuses on the evaluation of signal yields N_{sig} and efficiencies ϵ_{sig}. In practice, to better control the systematic uncertainties associated with the efficiencies, two normalisation channels are used, *i.e.* $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+ e^-)$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+ \mu^-)$.

A double ratio observable can be defined:

$$R_\Lambda^{-1} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)/\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+ e^-))}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)/\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+ \mu^-))} \ . \tag{6.2}$$

Given the LHCb detector, larger statistical uncertainty is expected to be associated with the electron channels than the muon channels, due to effects such as smaller efficiencies and worse resolution in some kinematic variables. Thus, the observable is constructed with $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ instead of $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ in the numerator. The ratio between the decays $J/\psi \to \mu^+ \mu^-$ and $J/\psi \to e^+ e^-$ is well-constrained experimentally [64]:

$$\frac{\mathcal{B}(J/\psi \to e^+ e^-)}{\mathcal{B}(J/\psi \to \mu^+ \mu^-)} = 1.0016 \pm 0.0031 \ . \tag{6.3}$$

Thus, in a practical sense, the observable R_Λ defined in Equation 6.2 is essentially the R_{H_s} variable in Equation 5.6 when $H_s = \Lambda$.

Using Equation 6.1, the observable R_Λ^{-1} can be expressed as:

$$R_\Lambda^{-1} = \frac{\frac{N_{\text{sig}}}{\epsilon_{\text{sig}}}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)}{\frac{N_{\text{sig}}}{\epsilon_{\text{sig}}}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)} \times \frac{\frac{N_{\text{sig}}}{\epsilon_{\text{sig}}}(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+ \mu^-))}{\frac{N_{\text{sig}}}{\epsilon_{\text{sig}}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+ e^-))} \ , \tag{6.4}$$

which contains the signal yields N_{sig} and the efficiencies ϵ_{sig} of the rare and the J/ψ resonance channels.

In the strategy of this analysis, the signal yields N_{sig} are extracted by performing fits to the invariant mass distribution of the Λ_b^0 candidates and the efficiencies ϵ_{sig} are evaluated based on simulation samples, possibly with data-driven correction factors (as discussed in Section 7.3). The invariant mass of Λ_b^0 candidates used are reconstructed with the help of the so-called `DecayTreeFitter` (DTF) algorithm, which is a fit to simultaneously extract the parameters of the decay chain, including the decay vertices and the momenta of particles [86]. As dicussed in Ref. [86], in this algorithm, the knowledge of the production vertex of the mother particle and the known masses of the composite particles can be used as constraints to improve the results. In the case of this analysis, the PV (the pp collision point) and the mass of Λ^0 can be constrained. In the following discussion of the thesis, the notation $m_{\text{DTF}}(\Lambda_b^0)$ is used for this DTF constrained mass. Optionally, for the J/ψ and $\psi(2S)$ resonance regions, the dilepton mass can also be constrained to the corresponding charmonium masses. The notations $m_{\text{DTF}}^{J/\psi}(\Lambda_b^0)$ and $m_{\text{DTF}}^{\psi(2S)}(\Lambda_b^0)$ are used for them in the thesis. Further details of the signal yield extraction and the efficiency evaluation will be discussed in Chapter 9.

6.1 Strategy for resonance ratio crosschecks

The resonance channels $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ can, besides serving as the normalisation channels, also be used for crosschecks. Since $\mathcal{B}(J/\psi \to e^+e^-)/\mathcal{B}(J/\psi \to \mu^+\mu^-)$ is well-constrained experimentally (as shown in Equation 6.3), measuring the single ratio

$$r_{J/\psi}^{-1} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))} , \tag{6.5}$$

which is expected to be ~ 1.0, can serve as a powerful crosscheck especially for the evaluation of efficiencies. Following the same argument used by Equation 6.4, this can be measured experimentally as

$$r_{J/\psi}^{-1} = \frac{\frac{N_{\text{sig}}}{\epsilon_{\text{sig}}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))}{\frac{N_{\text{sig}}}{\epsilon_{\text{sig}}}(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))} . \tag{6.6}$$

A further crosscheck can be performed using the resonance $\psi(2S)$ to substitute the rare decay in Equation 6.2. The measurement of a double ratio

$$R_{\psi(2S)}^{-1} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))/\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))/\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))} \tag{6.7}$$

Table 6.1: The q^2 division of the electron and the muon channels.

q^2 regions	electron q^2 [GeV^2/c^4]	muon q^2 [GeV^2/c^4]
Low	0.1 - 1.1	
Central	1.1 - 6.0	
High	15.0 - 22.0[20]	
J/ψ	6.0 - 11.0	8.7 - 10.05
$\psi(2S)$	11.0 - 15.0	12.5 - 14.2

can serve to further crosscheck the method of measuring the double ratio R_Λ^{-1}. The LFU between $\psi(2S) \to e^+e^-$ and $\psi(2S) \to \mu^+\mu^-$ is also relatively well-constrained experimentally [64]: [19]

$$\frac{\mathcal{B}(\psi(2S) \to \mu^+\mu^-)}{\mathcal{B}(\psi(2S) \to e^+e^-)} = 1.00 \pm 0.08 \ . \tag{6.8}$$

Experimentally, the double ratio $R^{-1}_{\psi(2S)}$ can be measured as

$$R^{-1}_{\psi(2S)} = \frac{\frac{N_{\mathrm{sig}}}{\epsilon_{\mathrm{sig}}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))}{\frac{N_{\mathrm{sig}}}{\epsilon_{\mathrm{sig}}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(\mu^+\mu^-))} \times \frac{\frac{N_{\mathrm{sig}}}{\epsilon_{\mathrm{sig}}}(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))}{\frac{N_{\mathrm{sig}}}{\epsilon_{\mathrm{sig}}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))} \ . \tag{6.9}$$

6.2 Division of q^2 regions

The rare decay $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ and the resonance decays $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$, $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ share the same final state particles while having different distributions in the dilepton invariant mass. In order to separate the different channels and provide meaningful measurements, the data samples are divided into different regions using the reconstructed q^2 variable, *i.e.* the low, central, J/ψ, $\psi(2S)$ and high q^2 regions. To faciliate the discussion, the low, central and high q^2 regions are sometimes referred to as rare q^2 regions while the J/ψ and $\psi(2S)$ regions are sometimes referred to as resonance q^2 regions.

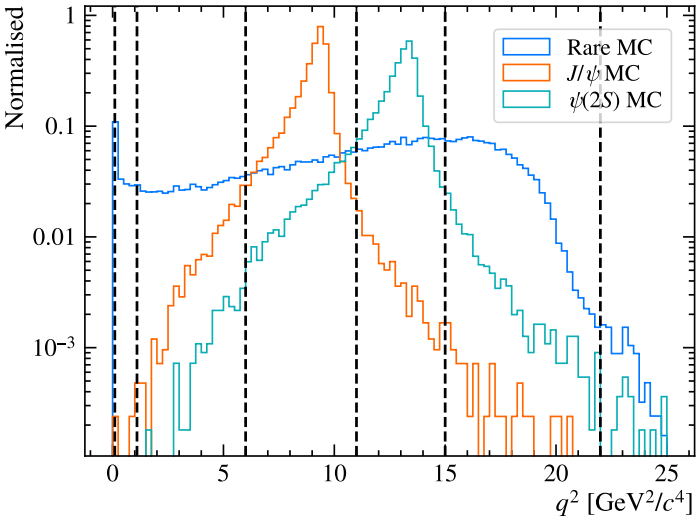
The q^2 division is shown in Table 6.1. The electron q^2 reconstruction is heavily affected by the bremsstrahlung, which will be discussed further

[19] This is based on the value directly reported in PDG [64]. In the R_X ($X = K, K^*$) analysis [74,75], the ratios $R_{\psi(2S)}$ are reported to be 0.987 ± 0.007 ($X = K^+$) and 1.012 ± 0.013 ($X = K^{*0}$). This implies a more stringent experimental constraint.

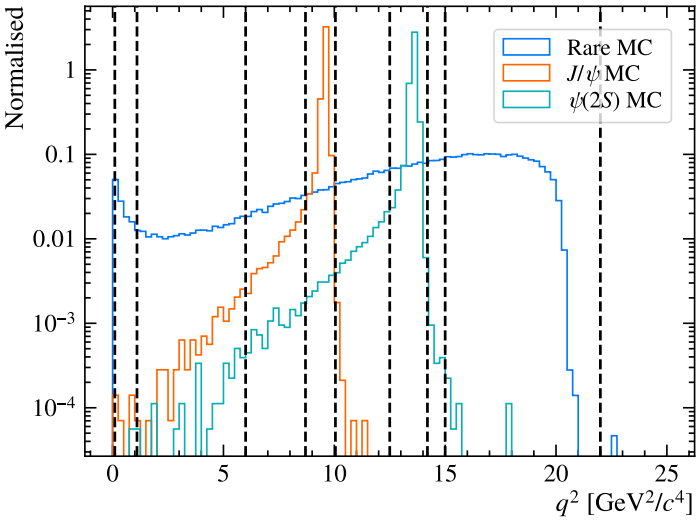
[20] This is higher than the phase-space limit to take the resolution effect into consideration. From the perspective of the theoretical interpretation, this is effectively $q^2 > 15.0\,\mathrm{GeV}^2/c^4$.

in Section 6.4. This leads to a very different q^2 resolution of the electron compared to the muon, which requires a different q^2 division for the resonance regions between the two types of leptons. The resonance J/ψ and $\psi(2S)$ regions are wider for the electron channels due to their worse q^2 resolution. This difference in the q^2 region definition is accounted for in the efficiencies. For the rare q^2 regions, for the benefit of the theoretical interpretation, the definition is aligned between the electron and the muon channels.

The different q^2 regions are illustrated, with the overlay of the reconstructed q^2 distribution of simulated $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ (rare), $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+ \ell^-)$ and $\Lambda_b^0 \to \Lambda^0 \psi(2S)(\ell^+ \ell^-)$ samples, in Figure 6.1. The aforementioned resolution difference between the electron and the muon channels can be seen.



(a) Electron channel



(b) Muon channel

Figure 6.1: The reconstructed q^2 distribution of the electron (top) and muon (bottom) channels in the simulation. The black dashed lines indicate the q^2 region division.

6.3 Analysis objectives and the blinding strategy

As discussed in Chapter 5, the main signal of $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ is expected to reside in the high q^2 region. Thus, the primary objective of the analysis is to measure R_Λ^{-1} in the high q^2 region. For the central and low q^2 regions, under the LFU assumption, it is likely that the significance of the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal cannot constitute an observation. In the case of no observation, the secondary objective of this analysis is to estimate an upper limit for the relative branching fraction between $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+ e^-)$ (denoted as $r_{ee}^{\mathcal{B}}$). The $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+ e^-)$ channel is used to as a normalisation channel to better control the systematic uncertainties associated with the efficiencies. Such a priority is reflected in the trigger strategy (to be discussed in Section 7.2).

As mentioned in Section 5.3, there is an ongoing effort within LHCb to measure $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$ (relative to $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)$) in different q^2 bins. Thus, this analysis does not attempt to give statements concerning $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$ (although it would be implicitly measured in the high q^2 region in the process of measuring R_Λ^{-1}). An important difference between the dedicated $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$ analysis and this analysis is that, while the former emphasizes the sensitivity of measuring $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$, the latter needs to control the systematic uncertainties in the branching fraction ratio between the electron and the muon channel, which affects the analysis optimisation.

To avoid biasing the analysis procedure, *e.g.* the candidate selections, this analysis adopts a blinding strategy. The distribution of $m_{\mathrm{DTF}}(\Lambda_b^0)$ is blinded in an interval around the expected Λ_b^0 peak for the low, central and high q^2 regions of the electron channel data. For the muon channel data, since the high q^2 signal was already observed [84], only the low and central q^2 regions are blinded. The blinding interval for the electron channel is from $5240\,\mathrm{MeV}/c^2$ to $5840\,\mathrm{MeV}/c^2$, while, for the muon channel, it is from $5500\,\mathrm{MeV}/c^2$ to $5700\,\mathrm{MeV}/c^2$. The blinding interval is wider for the electron channel due to its worse resolution in the $m_{\mathrm{DTF}}(\Lambda_b^0)$ variable.

The $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal efficiencies and yields in the high q^2 region are involved in the fit in the high q^2 region. However, due to the ongoing $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$ analysis within LHCb, the author is not allowed to explicitly provide an estimation for $\mathcal{B}\,(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$. Thus, these efficiencies and yields are not shown in this thesis.

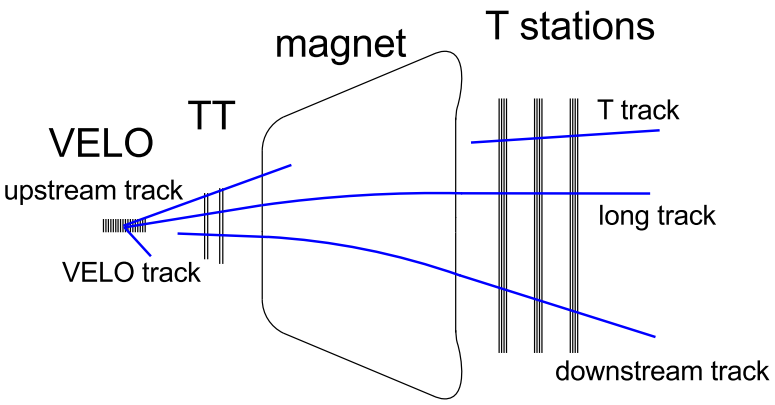


Figure 6.2: The different track types in the LHCb tracking system. Plot taken from Ref. [87].

6.4 Candidate reconstruction

As discussed in Section 2.2, during Run 1 and Run 2, the tracking system of LHCb consists of VELO, TT and the downstream T stations. Depending on the subsystems which a particle traverses, it can be reconstructed as a track of different types. As discussed in Ref. [87] (illustrated in Figure 6.2), the so-called long tracks traverse all the tracking stations while the so-called downstream tracks traverse the TT and the downstream T stations. These two types of tracks are the most important ones for this analysis.

The neutral Λ^0 particle is reconstructed as $\Lambda^0 \to p\pi^-$, *i.e.* with a proton track and a pion track. With a non-negligible lifetime, the decay vertices of the Λ^0 particles have a considerable spread in their z-coordinates. Depending on the location of its decay vertex, two types of reconstructed Λ^0 are used in this analysis. In one of the cases when the Λ^0 decays within the VELO, the proton and the pion have the possibility of being reconstructed as long tracks. In the other case when the Λ^0 decays outside the VELO but before the TT, the proton and the pion can potentially be reconstructed as downstream tracks. In the following discussion, the former is named LL (Long-Long) category while the latter is named DD (Down-Down) category. The leptons of $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ used in this analysis are reconstructed as long tracks, since they are produced from the Λ_b^0 decay, which happens close to the PV.

The reconstruction of the electron requires special treatments due to the bremsstrahlung effect. At LHCb, bremsstrahlung photons reconstructed by the ECAL can be added back to the electron candidate. The practice differs for bremsstrahlung emission before or after the magnet. In the former case,

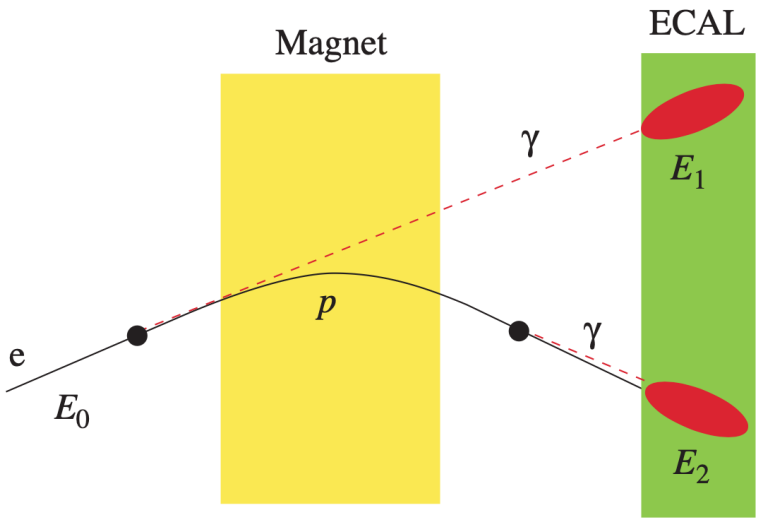


Figure 6.3: An illustration of the electron emitting bremsstrahlung photons before and after the magnet. Plot taken from Ref. [23].

it is needed to extrapolate the electron candidate track before the magnet while in the latter case, the energy deposit in the ECAl already contains the bremsstrahlung component [23]. This is illustrated in Figure 6.3 (taken from Ref. [23]). For the rare and the resonance electron channels of this analysis, depending on the number of the bremsstrahlung photons recovered, the samples can be divided into different categories. These categories sometimes require different treatments. For example, in the Λ_b^0 mass spectrum (without a DTF dilepton mass constraint), the distribution for the dielectrons with different bremsstrahlung recovery scenarios can be very different. Counting how many bremsstrahlung photons are recovered for the two electrons together, the samples can be categorised into G0 (no bremsstrahlung photon recovered), G1 (one bremsstrahlung photon recovered) and G2 (at least two bremsstrahlung photons recovered).

6.5 Data sample category overview

In this analysis, the data samples of two lepton types, *i.e.* the data of the electron channels and the muon channels, are used.

As discussed in Chapter 1, the data samples used in this analysis were taken by LHCb during its Run 1 and Run 2 operation. Based on the similarity in terms of operation condition (*e.g. pp* collision center-of-mass energy and trigger configuration), the data-taking periods are separated into R1

(Run 1, corresponding to 2011-2012), R2p1 (Run 2 part 1, corresponding to 2015-2016) and R2p2 (Run 2 part 2, corresponding to 2017-2018) [74].

Following the discussion in Section 6.4, two types of $p\pi$ tracks are used to reconstruct the Λ^0 candidates, *i.e.* LL and DD.

Combining above categories, the data samples can be separated into $12(2 \times 3 \times 2)$ categories:

- The lepton types: electrons, muons.
- The data-taking periods: R1, R2p1, R2p2.
- The $\Lambda^0 \to p\pi^-$ track types: LL and DD.

Within one category, the data samples can be further divided according to the q^2 division discussed in Section 6.2.

In this analysis, this categorisation is used in various steps. It can be considered as the "primary" data categorisation of the analysis. For example, the single ratio $r^{-1}_{J/\psi}$ and the double ratio $R^{-1}_{\psi(2S)}$ are evaluated in 6 categories, *i.e.* 3 data-taking periods and 2 track types. In some cases, due to limited statistics, some categories are merged. For example, during the rare q^2 region fit, the three data-taking periods of the electron channel are merged, which will be discussed in Section 9.4.

6.6 Background overview

As previously discussed, one of the key aspects of the analysis is the extraction of signal yields based on the invariant mass distribution. Such a procedure is heavily influenced by the presence of various backgrounds.

One of the important backgrounds, which is also common in various other analyses at LHCb, is the so-called combinatorial background. Considering the complicated environment of the pp collision event, random combinations of tracks can resemble the signal candidate. In the case of $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$, this can happen with a random combination of $\ell^+, \ell^-, \Lambda^0$ or ℓ^+, ℓ^-, p, π (the Λ^0 candidate itself is of combinatorial origin). A dedicated MultiVariate Analysis (MVA) procedure is developed to suppress this background, which will be discussed in Chapter 8. Empirically, it is usual that the combinatorial background exhibits an exponential behaviour in the invariant mass spectrum. However, the phase-space imposes the restriction

$$m(\Lambda^0 \ell\ell) = \sqrt{(E_{\Lambda^0} + E_{\ell\ell})^2 - (\vec{p}_{\Lambda^0} + \vec{p}_{\ell\ell})^2} \geq m_{\Lambda^0} + m_{\ell\ell} \ , \tag{6.10}$$

which causes the sculpting of the distribution in the high q^2 region (large $m_{\ell\ell}$). In order to model the combinatorial background in the high q^2 region,

the so-called Same-Sign (SS) data sample, *i.e.* the sample reconstructed with Λ^0 and the same-charged lepton pairs ($\ell^\pm\ell^\pm$), is used. Correspondingly, the normal data with the opposite-charged lepton pairs are referred to as the Opposite-Sign (OS) data. The SS data proxy is used during both the selection optimisation (Section 8.3) and the signal extraction (Section 9.4.2).

Due to the relative abundance of the resonance decays and the limited resolution of the electron kinematic variables, the q^2 division (described in Section 6.2) is not able to completely separate the resonance decays from the rare q^2 regions for the electron channels. To faciliate the discussion, this type of behaviour is referred to as the leakage, *i.e.* the decay candidates from a certain q^2 region can be reconstructed in another q^2 region due to the limited resolution of the q^2 variable. In the high q^2 region, a certain amount of the $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ leakage can be present. These leaked candidates are smeared by the reconstruction effect towards a higher q^2 region. This also reflects on the Λ_b^0 mass distribution, which will also be smeared towards a higher region as a result. Thus, the leaked $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ in the high q^2 region is expected to appear in the upper sideband of the Λ_b^0 mass distribution. In contrast, the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ leakage in the central q^2 region, which is smeared towards a lower q^2 region, is expected to appear in the lower sideband of the Λ_b^0 mass distribution. These leakage backgrounds will be considered in the invariant mass fit, which will be discussed in Chapter 9.

The presence of the Λ^0 particle, which has a displaced decay vertex, helps to distinguish the signal from various prompt-decay backgrounds (*e.g.* $\Lambda_b^0 \to pK^-\ell^+\ell^-$). However, the K_S^0 particle, with its non-negligible lifetime, can still imitate the Λ^0 particle. This can happen if one of the pion from $K_S^0 \to \pi^+\pi^-$ is misidentified as the proton. The K_S^0 can often be generated from the B meson decay $B^0 \to K_S^0 J/\psi(\ell^+\ell^-)$, $B^0 \to K_S^0\psi(2S)(\ell^+\ell^-)$ and $B^0 \to K_S^0\ell^+\ell^-$. These background decays are relatively abundant. Based on Ref. [88], the production fraction of the Λ_b^0 at LHCb relative to that of the B^- and $\bar{B}^0$ (together) is measured to be 0.259 ± 0.018 (averaged over the b-hadron transverse momentum p_T from 4 to 24 GeV/c and pseudorapidity η from 2 to 5). Assuming the isospin symmetry for the charged and neutral B meson production, this gives a relative production fraction ~ 0.518 between Λ_b^0 and B^0. In Ref. [85], the relative branching fraction between $\Lambda_b^0 \to \Lambda^0 J/\psi$ and $B^0 \to K_S^0 J/\psi$ is measured to be

$$\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)}{\mathcal{B}(B^0 \to K_S^0 J/\psi)} = 0.750 \pm 0.005 \pm 0.022 \pm 0.005 \pm 0.062 \ . \tag{6.11}$$

These give the relative abundance between the two decays:

$$\frac{\text{produced } \Lambda_b^0 \to \Lambda^0 J/\psi}{\text{produced } B^0 \to K_S^0 J/\psi} \sim 0.39 \ . \tag{6.12}$$

To suppress these K^0_S backgrounds, a veto selection is applied, which will be discussed in Section 7.1.2.

Another type of the background is the partially reconstructed background which shares the same final state particles as the signal but with extra decay products not reconstructed. For example, the decay $\Lambda_b^0 \to \Lambda^* J/\psi(\ell^+\ell^-)$ can be reconstructed as the resonance signal when the Λ^* decays into Λ^0 and some missing products (*e.g.* $\Lambda^* \to \Sigma^0(\to \Lambda^0\gamma)\pi^0$). During a previous study for the isospin amplitudes of $\Lambda_b^0 \to J/\psi\Lambda^0(\Sigma^0)$ and $\Xi_b^0 \to J/\psi\Xi^0(\Lambda^0)$ [89], the partially reconstructed backgrounds for the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ decay were studied. Taken from Ref. [89], Figure 6.4 illustrates the results of their investigation of those partially reconstructed backgrounds. In their fit, a Gaussian shape (shown in Figure 6.4 as "Other part. rec. bkgs", *i.e.* other partially reconstructed backgrounds) is used to model several Λ^* (heavier than $\Lambda^0(1600)$) components and the decay $\Lambda_b^0 \to \Lambda^0\psi(2S)(\to J/\psi\pi\pi)$. The modelling of the partially reconstructed backgrounds in this analysis is inspired by their analysis. In contrast to their interest in the decays such as $\Lambda_b^0 \to J/\psi\Sigma^0$, the objective of this analysis concerns the signal yield of $\Lambda_b^0 \to \Lambda^0 J/\psi$. Thus, only several major distinguishable partially reconstructed backgrounds are explicitly modelled while several minor ones are ignored. Further details concerning the modelling of these backgrounds are discussed in Section 9.3. For the partially reconstructed rare decays, *e.g.* $\Lambda_b^0 \to \Lambda^*\ell^+\ell^-$, very little is known. They are not included in the nominal approach for the signal extraction of $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$. But, in principle, the possibility exists that they can have an impact. A systematic uncertainty study for this is discussed in Section 10.3.

The semileptonic decay $\Lambda_b^0 \to \Lambda_c^+\ell^-\bar{\nu}_\ell$, with its relatively large branching fraction $\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+\ell^-\bar{\nu}_\ell) = 6.2^{+1.4}_{-1.3} \times 10^{-2}$ ($\ell = e$ or μ) [64], is potentially another background. The charm-baryon Λ_c can decay into Λ^0, *e.g.* $\mathcal{B}(\Lambda_c^+ \to \Lambda^0\mu^+\nu_\mu) = (3.48 \pm 0.17) \times 10^{-2}$, $\mathcal{B}(\Lambda_c^+ \to \Lambda^0 e^+\nu_e) = (3.56 \pm 0.13) \times 10^{-2}$ and $\mathcal{B}(\Lambda_c^+ \to \Lambda^0\pi^+) = (1.31 \pm 0.05) \times 10^{-2}$ [64]. Due to the missing neutrinos, when reconstructed as the signal $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$, these semileptonic decays tend to reside at lower reconstructed Λ_b^0 invariant mass and lower q^2 regions. The treatment for this semileptonic background will be discussed in Section 7.1.3.

During the LFU study for the R_X ($X = K, K^*$) measurement [74, 75], it was found that the electron rare channels are susceptible to the b-hadron decay backgrounds containing one or two hadrons (pions or kaons) which are misidentified as the electrons. They developed a data-driven method namely the so-called `Fail-To-Pass` method to model these backgrounds, the advantage of which is to avoid the need to investigate individually each possible channels and their decay kinematics. In the case of $\Lambda_b^0 \to \Lambda^0 e^+e^-$, similar

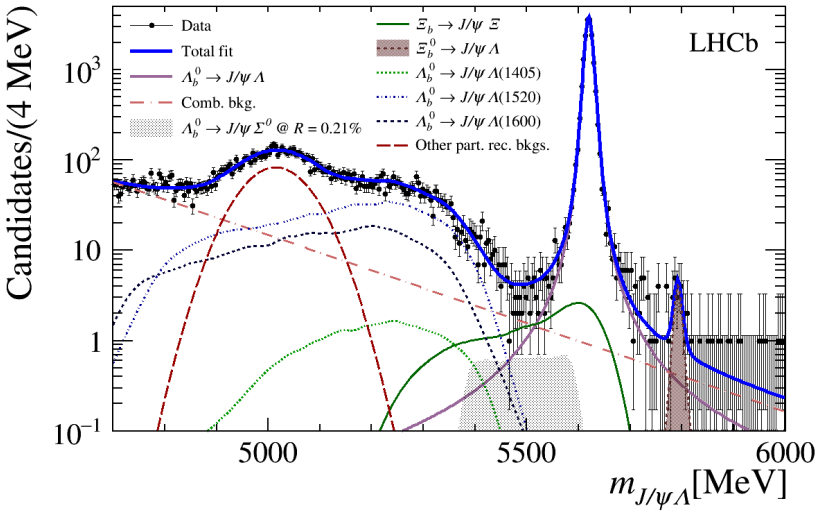


Figure 6.4: The $J/\psi\,\Lambda^0$ invariant mass fit plot taken from Ref. [89]. Several partially reconstructed backgrounds appearing on the lower side of the signal peak are shown. The "Other part. rec. bkgs" component includes $\Lambda_b^0 \to \Lambda^* J/\psi$ (Λ^* heavier than $\Lambda^0(1600)$) and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\to J/\psi\pi\pi)$.

misidentified backgrounds (*e.g.* $\Lambda_b^0 \to \Lambda^0 h_1 h_2$) can potentially affect the signal extraction. Table 6.2 shows several possible origins (non-exhaustive) of such misidentified backgrounds with their branching fractions. In this analysis, the aforementioned `Fail-To-Pass` method is used to investigate the combined effects of this type of misidentified backgrounds, which will be discussed in Section 9.4.1 and Section 10.2.

6.7 The *sPlot* technique

In this analysis, the so-called *sPlot* technique [90] is used in several occasions such as the correction of the simulation (to be discussed in Section 7.3).

Very often, the knowledge about the behaviour of certain variables not only in the simulation but also in the real data is desired. Considering the blinding of the rare signal regions, if the variables of interest have similar behaviours between the rare and the resonance channels, such a knowledge can be extracted from resonance channels. The data samples, however, consist of a mixture of different components including the signal and various backgrounds. It is often the behaviour of a certain component (the signal or a type of background) that is of interest. The *sPlot* technique is a method to

Table 6.2: Possible hadronic decays (non-exhaustive) contributing as misidentified backgrounds to $\Lambda_b^0 \to \Lambda^0 e^+ e^-$. The branching fractions are based on PDG [64].

Possible origins	Branching fraction
$\mathcal{B}(\Lambda_b^0 \to \Lambda^0 \pi^+ \pi^-)$	$(4.6 \pm 1.9) \times 10^{-6}$
$\mathcal{B}(\Lambda_b^0 \to \Lambda^0 K^+ \pi^-)$	$(5.7 \pm 1.2) \times 10^{-6}$
$\mathcal{B}(\Lambda_b^0 \to \Lambda^0 K^+ K^-)$	$(1.61 \pm 0.22) \times 10^{-5}$
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \pi^-) \times \mathcal{B}(\Lambda_c^+ \to \Lambda^0 \pi^+)$	$(6.42 \pm 0.58) \times 10^{-5}$
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ \pi^-) \times \mathcal{B}(\Lambda_c^+ \to \Lambda^0 \pi^+ \pi^0)$	$(3.48 \pm 0.33) \times 10^{-4}$
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ K^-) \times \mathcal{B}(\Lambda_c^+ \to \Lambda^0 \pi^+)$	$(4.66 \pm 0.41) \times 10^{-6}$
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+ K^-) \times \mathcal{B}(\Lambda_c^+ \to \Lambda^0 \pi^+ \pi^0)$	$(2.53 \pm 0.23) \times 10^{-5}$

extract the distribution of variables for a certain component from a sample with mixed components.

Based on Ref. [90], the idea of the *sPlot* technique is introduced as follows. Given a sample with mixed components and two independent variables, *i.e.* a variable of interest (x) and another variable (y) with good distinguishing ability for the different components, the *sPlot* endeavors to extract the information on x for the individual component using the knowledge on y. Usually, one can perform a maximum-likelihood fit of the y variable to extract the y-distributions of the individual components and their yields. After these are acquired, for a given component, a weight (sWeight) can be constructed for each entry in the sample. The x-distribution of the weighted (sWeight) sample can then serve an estimation of the real distribution of x for the corresponding component. In the case of this analysis, the reconstructed invariant mass of the Λ_b^0 candidate often serves as the aforementioned y variable with distinguishing ability.

7 Analysis sample processing

As discussed in Section 2.2.2, during the LHCb Run 1 and Run 2, the triggered events were recorded and offline reconstruction was performed. The reconstructed events then went through the centralised selection procedure known as the Stripping selection [91]. In addition to the centralised Stripping, further customised offline selections are applied for this analysis. The Stripping and the customised offline selections are discussed in Section 7.1.

During the data acquisition, only triggerred data can be saved. However, in the trigger system, there are various different trigger lines. To better control the samples, especially in terms of efficiency, only a subset of the available trigger lines are selected for the analysis. In practice, this is done via an offline selection of the decision flags of various trigger lines. Such a selection is described in Section 7.2.

In addition to the data samples, this analysis also uses simulation samples, *i.e.* Monte Carlo (MC) samples. The simulation samples went through similar procedures as the real data with the difference that the inputs were generated via a series of software algorithms instead of being provided by the detector. These softwares include, *e.g.* Pythia for the b-hadron generation [92, 93], EvtGen for the particle decays [94] and Geant4 for the particle propagation through the detector material [95, 96]. The simulation usually does not represent the reality perfectly and thus corrections are needed to improve the compatibility between simulation and data. These corrections are discussed in Section 7.3.

The Stripping selection is the collective work of the collaboration [91]. The customised offline selections and the trigger strategy were modified based on the previous work by collaborators [36] and the work in Refs. [74, 75]. For the simulation correction, while a preliminary baseline existed in the previous work by collaborators [36], it was largely reworked with adjustments and/or checks (as discussed in Section 7.3).

7.1 Offline selection

The aforementioned Stripping selection is organised in various so-called Stripping lines. In this analysis, the Stripping lines named `Bu2LLK_eeLine2` (electrons) and `Bu2LLK_mmLine` (muons) are used for the normal data and simulation samples. These Stripping lines combine the dilepton pair with various hadron system (*e.g.* Λ^0, ϕ, K^+, K^0_S), and were used by several previous LFU studies (*e.g.* Refs. [74, 75, 78, 97]). In addition, as discussed in Section 6.6, this analysis also uses the SS data. They are selected with the Stripping lines named `Bu2LLK_eeSSLine2` (electrons) and `Bu2LLK_mmSSLine` (muons).

These Stripping lines set requirements on the reconstruction quality of the candidates using criteria such as the decay vertex χ^2 and the distance between tracks which are assumed to originate from the same vertex. They also set requirements on the decay topology using flight distance and impact parameters, *e.g.* Λ_b^0 should originate from the PV with some displacements before decaying and the leptons/proton/pion should not originate from PV. The details of the Stripping selections [98–101] are shown in Appendix B.

The samples passing the Stripping selection are further selected before they are used for more sophisticated analysis procedures. An overview of these customised selections are shown in Appendix B. A few important ones are emphasized in the following sections.

7.1.1 Lepton identification

The identification of the leptons utilises the LHCb PID system discussed in Section 2.2. The PID selections applied to the lepton candidates are summarised in Table 7.1. The variable `CombDLL`e refers to the Combined Differential Log-Likelihood (CombDLL) based on the likelihood information from the RICH, the calorimeter and the muon systems, which gives a measure of how likely the candidate belongs to the hypothesized particle species [17, 102]. The variable `ProbNN`μ (or `ProbNN`e) is an Artificial Neural Network (ANN) based PID response, which combines the information from the PID system and the tracking information [17, 102]. The `IsMuon` boolean variable discussed in Section 2.2 is also used.

The efficiencies of the PID selection are often not well simulated and thus they are treated with a data-driven calibration procedure [103]. This calibration will be discussed in Section 7.3.4. In order for the calibration to function properly, certain basic requirements are applied. In the case of electrons, the corresponding tracks are required to have hits registered by the calorimeter. For both electrons and muons, the corresponding tracks are required to be registered in the RICH. The momentum and the transverse momentum of the leptons are limited to a certain range in order for them to be within the coverage of the calibration samples with good statistics:

- Electron: $p_T > 500\,\text{MeV}/c$ and $3 < p < 200\,\text{GeV}/c$.
- Muon: $p_T > 800\,\text{MeV}/c$ and $3 < p < 300\,\text{GeV}/c$.

7.1.2 Veto against K_S^0 background

As discussed in Section 6.6, one of the main backgrounds in this study comes from the decay $K_S^0 \to \pi^+\pi^-$ where the π^+ is misidentified as the proton.

Table 7.1: An overview of the PID requirements.

Leptons	Stripping	Customised selection
Electron	`CombDLL`$e > 0.0$	`ProbNN`$e > 0.2$
Muon	`IsMuon`	`ProbNN`$\mu > 0.2$

For the muon channel, such a misidentified background leads to a distinguishable shape in the reconstructed Λ_b^0 mass spectrum, the situation is worse in the electron channels due to the worse resolution of the electron momenta. The situation is especially complicated in the case of the high q^2 region of the electron channel, where not only the rare decay $B^0 \to K_S^0 e^+ e^-$ needs to be considered but also the resonance decay $B^0 \to K_S^0 \psi(2S)(e^+e^-)$ can leak into the signal region.

By using the momenta of the possibly misidentified proton and pion tracks, one can recalculate the combined invariant mass. This is done by keeping the momenta while replacing the proton mass hypothesis with the pion mass hypothesis. Such a recalculated invariant mass is denoted as $M(\pi(p \to \pi))$. A veto against the K_S^0 can be performed using such a variable. The distributions of the recalculated mass for the $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ and $B^0 \to K_S^0 \ell^+ \ell^-$ MC samples of the R2p2 period are shown in Figure 7.1, where the veto cut

$$M(\pi(p \to \pi)) < 475 \text{ or } > 525\,\text{MeV}/c^2 \tag{7.1}$$

is also illustrated. As indicated in the plots, the veto cut removes almost all $B^0 \to K_S^0 \ell^+ \ell^-$ background while retaining about 80 percent of the $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ signal.

7.1.3 Veto against semileptonic background

As discussed in Section 6.6, the semileptonic decay $\Lambda_b^0 \to \Lambda_c^+ \ell^- \bar{\nu}_\ell$, with the subsequent decay $\Lambda_c^+ \to \Lambda^0 \ell^+ \nu_\ell$, can contribute to the background. In principle, this type of background can be vetoed using the combined invariant mass of the Λ^0 and one of the lepton candidate. For example, as the decay products of $\Lambda_c^+ \to \Lambda^0 \ell^+ \nu_\ell$, the combined invariant mass of Λ^0 and ℓ^+ (denoted as $M(\Lambda^0 \ell^+)$) should be smaller than that of Λ_c^+ (without considering the reconstruction resolution effects). When the decay $\Lambda_b^0 \to \Lambda_c^+ \ell^- \bar{\nu}_\ell$ ($\Lambda_c^+ \to \Lambda^0 \ell^+ \nu_\ell$) is reconstructed as $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$, the lepton from the Λ_c^+ decay can be identified using the charge sign. The lepton associated with the Λ_c^+ will have the same charge as the proton.

Figure 7.2 shows the truth level distribution for the $\Lambda_b^0 \to \Lambda_c^+ e^- \bar{\nu}_e$ (with the subsequent $\Lambda_c^+ \to \Lambda^0 \ell^+ \nu_\ell$) MC sample in $q^2(\ell^+\ell^-)$ versus the visible mass

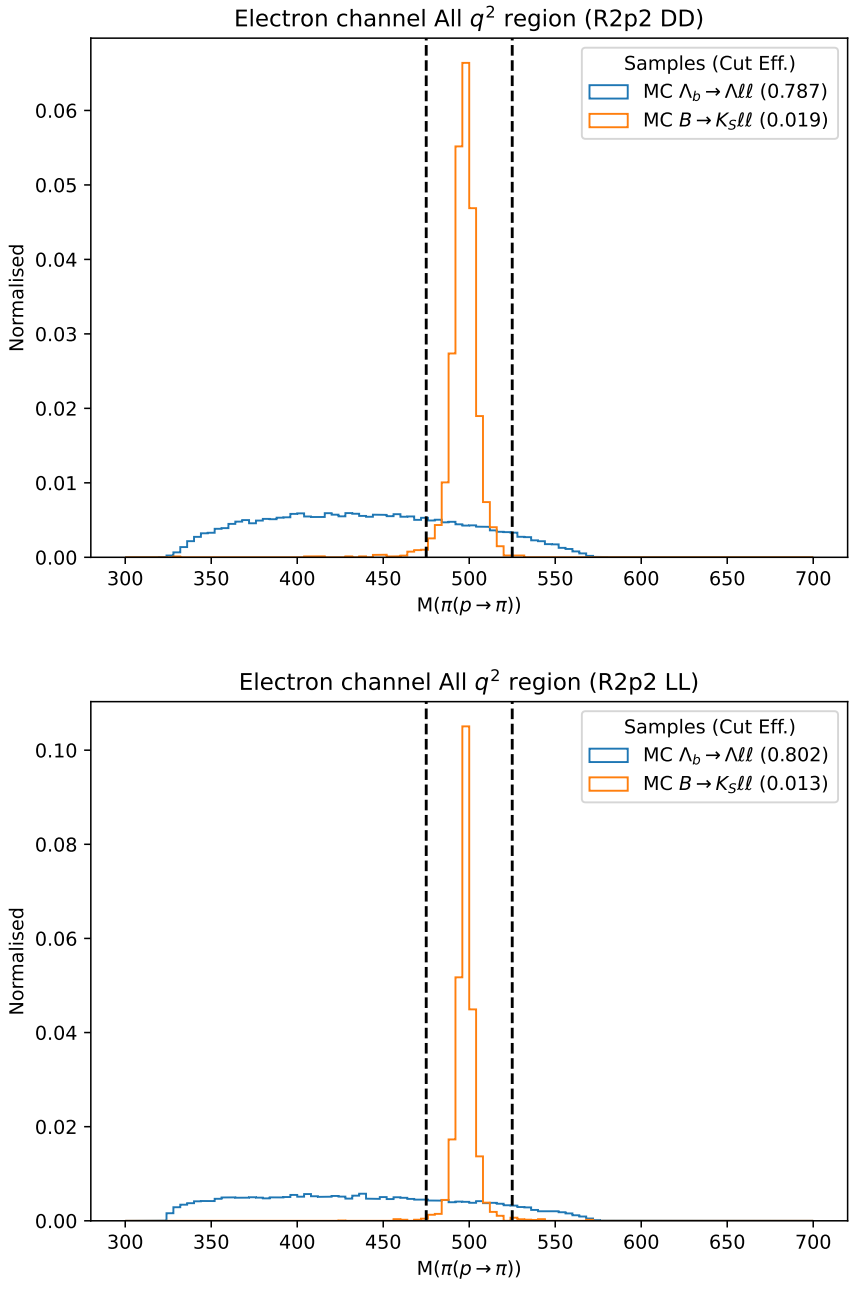


Figure 7.1: The recalculated invariant mass $M(\pi(p \to \pi))$ of the simulation samples $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ (signal) and $B^0 \to K_S^0 \ell^+ \ell^-$ (misidentified background). The veto cut is shown as the black dashed line. The top plot shows for the R2p2-DD category while the bottom plot shows for the R2p2-LL category.

(the invariant mass of the decay products excluding the neutrinos). Due to the momentum carried away by the neutrinos, the reconstructible part of the decay, as shown in Figure 7.2, resides mainly in the lower dilepton q^2 and lower visible invariant mass regions. In the plot, the visible mass does not extend below $4500\,\mathrm{MeV}/c^2$. This is due to a requirement in the MC production, *i.e.* the generated decays with the visible mass $< 4500\,\mathrm{MeV}/c^2$ were removed before performing the full simulation in order to avoid spending too much time for those candidates in the extremely low invariant mass region. This implies, should the simulation be conducted with the full kinematic range of the decay, there would be even more candidates reside in the lower mass region.

Figure 7.3, Figure 7.4 and Figure 7.5 show examples of the combined invariant mass $M(\Lambda^0\ell^+)$ distributions for the signal MC, the $\Lambda_b^0 \to \Lambda_c^+ e^- \bar{\nu}_e$ ($\Lambda_c^+ \to \Lambda^0 e^+ \nu_e$) background MC and the blinded data (*i.e.* without the candidates inside the blinding window) in the three rare q^2 regions. The blinded data in the low and central q^2 regions show structures which are likely caused by the semileptonic decays while no obvious similar structure is seen for the high q^2 region. As shown in the plot, the loss of the simulated signal caused by the veto selection $M(\Lambda^0\ell^+) > 2300\,\mathrm{MeV}/c^2$ is considerably large in the high q^2 region (losing about 34 percent) while, for the low and central q^2 regions, such losses are considerably smaller (losing about 4 percent for the low q^2 region and about 10 percent for the central q^2 region).

Due to the above considerations, in the nominal selection, the veto selection is used only for the low and central q^2 regions.

For the possibility of $\Lambda_c^+ \to \Lambda^0\pi^+$, a similar veto selection, with the lepton candidate using the pion mass hypothesis (similar technique as discussed in Section 7.1.2), is applied for the low and central q^2 regions: $M(\Lambda^0(\ell^+ \to \pi^+)) > 2400\,\mathrm{MeV}/c^2$.

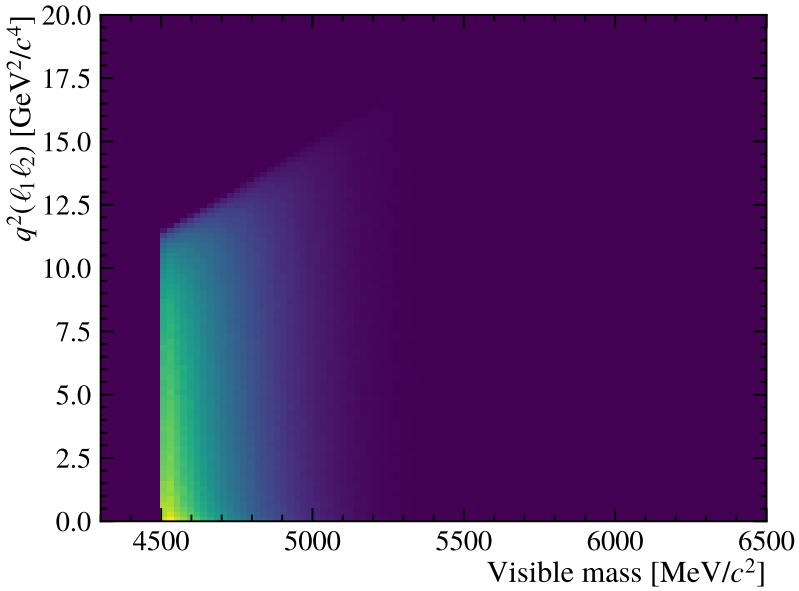


Figure 7.2: The distribution in the dilepton q^2 versus the visible mass (combined invariant mass except for the neutrinos) of the truth level simulation for $\Lambda_b^0 \to \Lambda_c^+ e^- \bar{\nu}_e$ ($\Lambda_c^+ \to \Lambda^0 e^+ \nu_e$).

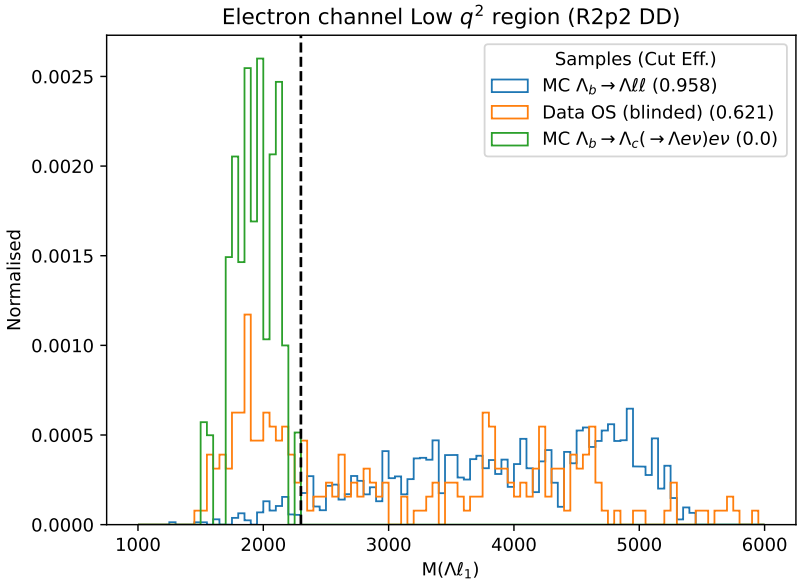


Figure 7.3: The combined invariant mass $M(\Lambda^0\ell^+)$ distributions in the low q^2 region (electron-R2p2-DD category).

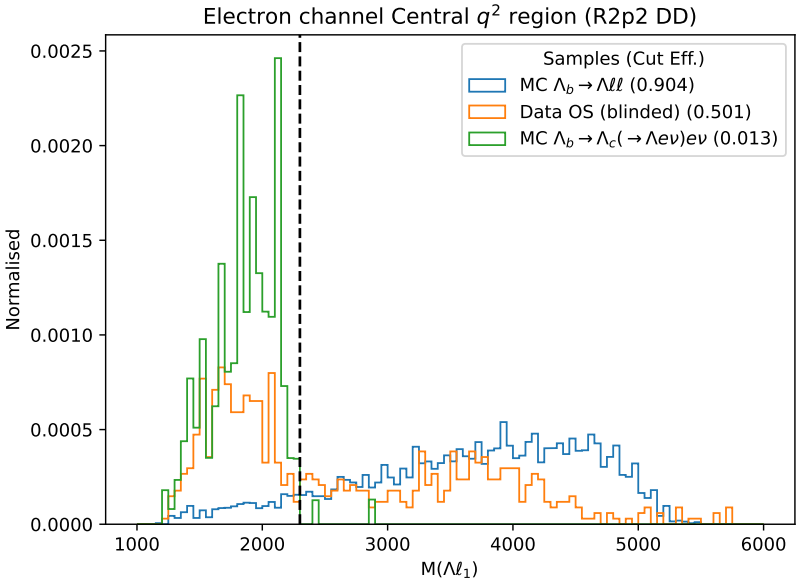


Figure 7.4: The combined invariant mass $M(\Lambda^0\ell^+)$ distributions in the central q^2 region (electron-R2p2-DD category).

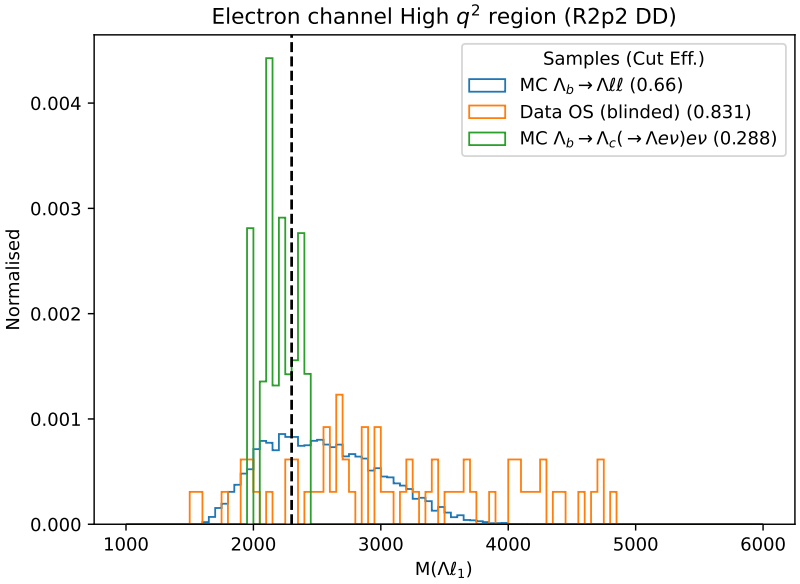


Figure 7.5: The combined invariant mass $M(\Lambda^0\ell^+)$ distributions in the high q^2 region (electron-R2p2-DD category).

Table 7.2: Overview of the used L0 trigger lines

Leptons	L0 trigger lines
Electron	`L0Electron` (ℓ_1) or `L0Electron` (ℓ_2)
Muon	`L0Muon` (ℓ_1) or `L0Muon` (ℓ_2)

7.2 Trigger strategy

As discussed in Section 2.2.2, during Run 1 and Run 2, the LHCb trigger system consists of three levels, *i.e.* L0, HLT1 and HLT2.

In this analysis, the leptons are required to have triggered the event at the L0 level (as shown in Table 7.2). For the two leptons, at least one of them is required to has a positive trigger response for the corresponding trigger line.

As discussed in Section 2.2.2, the trigger line `L0Electron` sets the threshold for the transverse component of the electron energy deposit (E_T) in the ECAL, while the trigger line `L0Muon` requires the muon transverse momentum (p_T) to be larger than a certain threshold. In addition, both L0 trigger lines impose requirements on the SPD multiplicity to remove events with very high occupancy (time consuming for HLT), *i.e.* requiring number of SPD hits $< 600(450)$ for Run 1 (Run 2) [21, 33]. In this analysis, the same cut on the SPD multiplicity is applied as a part of the offline selection so that it does not complicate the trigger efficiency evaluation. The calibration of the L0 trigger efficiency will be discussed in Section 7.3.5.

The trigger lines `L0Electron` and `L0Muon` favour the energetic lepton candidates. This strategy is justified considering the main signal of $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ resides in the high q^2 region. These trigger lines are less ideal for the central and low q^2 regions [21].

At the HLT1 level, the lepton tracks are used for the trigger. For Run 1, the so-named `TrackAllL0` line is used while the `TrackMVA` line is used for Run 2. These trigger lines set requirements on the single track features. They look for tracks with criteria such as good fit quality (using track χ^2 per degrees-of-freedom) and displacement from the PV (using impact parameter information). The details of their requirements are shown in Appendix B (Table B.4). At least one of the leptons is required to have a positive trigger response.

[21] Several LFU studies (*e.g.* Refs. [74, 75, 78]) focusing on the q^2 region below the charmonium resonances employ different L0 trigger strategies, which use, in combination with the lepton triggers, also the remaining part of the event excluding the signal decay (require a separate efficiency calibration). The exact details vary for different analyses.

At the HLT2 level, a group of topological trigger lines [104,105] are used, which sets requirements using the signal decay topology. As discussed in Refs. [21, 33], such topological trigger lines look for displaced vertices with multiple tracks using multivariate techniques.

7.3 Simulation correction

In many cases, the simulation plays an important role in the estimation of parameters, especially for the efficiencies. However, the simulation does not describe the reality perfectly. Therefore, corrections are applied to the MC samples to improve the compatibility between the data and simulation. In practice, these corrections are applied in the form of per-decay-candidate weights, which will be discussed in the following sections. The effects of these corrections on the single ratio $r^{-1}_{J/\psi}$ and the double ratio $R^{-1}_{\psi(2S)}$ will be discussed in Chapter 10.

7.3.1 The lifetime of Λ_b^0

In the MC samples, the Λ_b^0 particle is simulated with the mean lifetime of $\tau = 1.451\,\text{ps}$, which differs from the world average measurement given by HFLAV [106] $\tau = 1.471 \pm 0.009\,\text{ps}$. The correction weights are calculated as the ratio of two exponential probability density functions:

$$w_i^\tau = \frac{1.451}{1.471} \times \exp\left(-t_i \left(\frac{1}{1.471} - \frac{1}{1.451}\right)\right) \ , \tag{7.2}$$

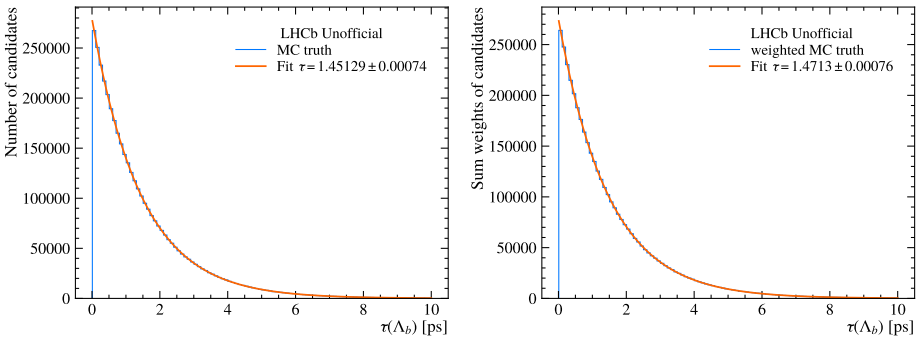
where t_i stands for the sampled lifetime of the simulated Λ_b^0 (i stands for the candidate index).

As an example, the effect of the weighting is shown in Figure 7.6, where the lifetime distributions without and with the τ weight are shown. As shown in the plot, the weighting is able to correct the mean lifetime as intended.

7.3.2 Angular distributions

The angular structures of the resonance decay channels $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ are not well described in their simulations, where the phase-space model is used. The rare decay $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ simulation, on the other hand, uses the Lattice QCD model [79], which is supposed to describe the decay kinematics.

For the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ decay, existing measurements of the angular moments [107] are used to weight the phase-space MC. Following the



Without τ weight With τ weight

Figure 7.6: Comparison between the MC truth level τ distribution without and with the τ weight. They are fitted with an exponential shape using the least square method. The fitted average lifetime is also shown. The sample displayed here is the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ 2018 MC.

same convention used in the measurements, the angular structure is characterised by five angular variables, which is illustrated in Figure 7.7 (taken from Ref. [107]). The angle θ represents the angle between the direction of Λ^0 momentum and the normal vector $\hat{n}$ (perpendicular to the plane defined by the proton momentum and the Λ_b^0 momentum) in the rest frame of the Λ_b^0 particle. The angles ϕ_b and θ_b describe the direction of the proton momentum in the Λ^0 rest frame. The angles ϕ_l and θ_l describe the direction of the μ^+ in the J/ψ rest frame.

Following Ref. [107], the differential decay width (Γ) in terms of the five angles $\vec{\Omega} = (\cos\theta, \cos\theta_b, \phi_b, \cos\theta_l, \phi_l)$ is expressed as

$$\frac{\mathrm{d}^5\Gamma}{\mathrm{d}\vec{\Omega}} = \frac{3}{32\pi^2} \sum_{k=1}^{34} J_k f_k(\vec{\Omega}) \ , \tag{7.3}$$

where the f_k are different functions of the angular variable $\vec{\Omega}$. The factor J_k encodes the information of the decay physics and the Λ_b^0 production polarisation. The integration of $\mathrm{d}\vec{\Omega}$ over full phase space can be written as

$$\int \mathrm{d}\vec{\Omega} = \int_{-1}^{1} \mathrm{d}\cos\theta \int_{-1}^{1} \mathrm{d}\cos\theta_b \int_{0}^{2\pi} \mathrm{d}\phi_b \int_{-1}^{1} \mathrm{d}\cos\theta_l \int_{0}^{2\pi} \mathrm{d}\phi_l \ . \tag{7.4}$$

Given the definition of f_k following Ref. [107], only $f_1(\vec{\Omega}) = \sin^2\theta_l$ and $f_2(\vec{\Omega}) = \cos^2\theta_l$ do not vanish under the integration of $\mathrm{d}\vec{\Omega}$ over full phase

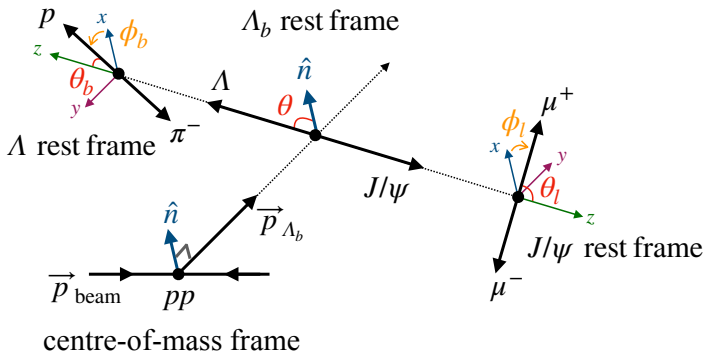


Figure 7.7: An illustration of the angular variables of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ decay. Plot taken from [107].

space:

$$\begin{aligned} \int \mathrm{d}\vec{\Omega} \cdot \sin^2\theta_l &= \frac{64\pi^2}{3} \ , \\ \int \mathrm{d}\vec{\Omega} \cdot \cos^2\theta_l &= \frac{32\pi^2}{3} \ , \end{aligned} \tag{7.5}$$

which, combining with Equation 7.3, gives the total decay width $2J_1 + J_2$.

The results of Ref. [107] were given in terms of the angular moments $M_k = J_k/(2J_1 + J_2)$ where J_k is normalised to the total decay width.

In the phase-space model, the $\vec{\Omega}$ follows an uniform distribution at generation (before effects such as the detector acceptance and reconstruction). The per-decay-candidate angular weight can then be calculated as

$$w_i^{\text{angle}} = \sum_{k=1}^{34} 3M_k f_k(\vec{\Omega}_i) \ , \tag{7.6}$$

where i stands for the candidate index. The factor 3 tunes the average weight scale to unity, considering:

$$\frac{\int \mathrm{d}\vec{\Omega} \sum_{k_1}^{34} 3M_k f_k(\vec{\Omega})}{\int \mathrm{d}\vec{\Omega}} = 2M_1 + M_2 = 1 \ . \tag{7.7}$$

To validate the angular weights, the J/ψ mode data are fitted and the distributions of the five angular variables for the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ signal component are extracted with the *sPlot* technique. They are then compared

to the simulation with and without the angular weights. Figure 7.8 shows an example of such comparisons. As shown in the figure, the angular weights are able to improve the compatibility between data and MC in these distributions.

For the other resonance mode $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$, no angular measurements equivalent to that of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ are available. As substitutes, two alternative methods are used. The first method is simply to apply the angular moments measured for $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$. The second method is based on the theory calculation from Ref. [108], which characterises the angular structure with only θ, θ_b and θ_l. To faciliate the discussion, the weights from the second method are referred to as the three-fold weights. Similar validation checks as the J/ψ mode can also be performed. Figure 7.9 shows the corresponding comparisons. As shown in the plot, both methods give good data-MC compatibility in these angular distributions. In this analysis, the J/ψ mode angular moments weighting is used as the nominal method for the $\psi(2S)$ mode angular weights.

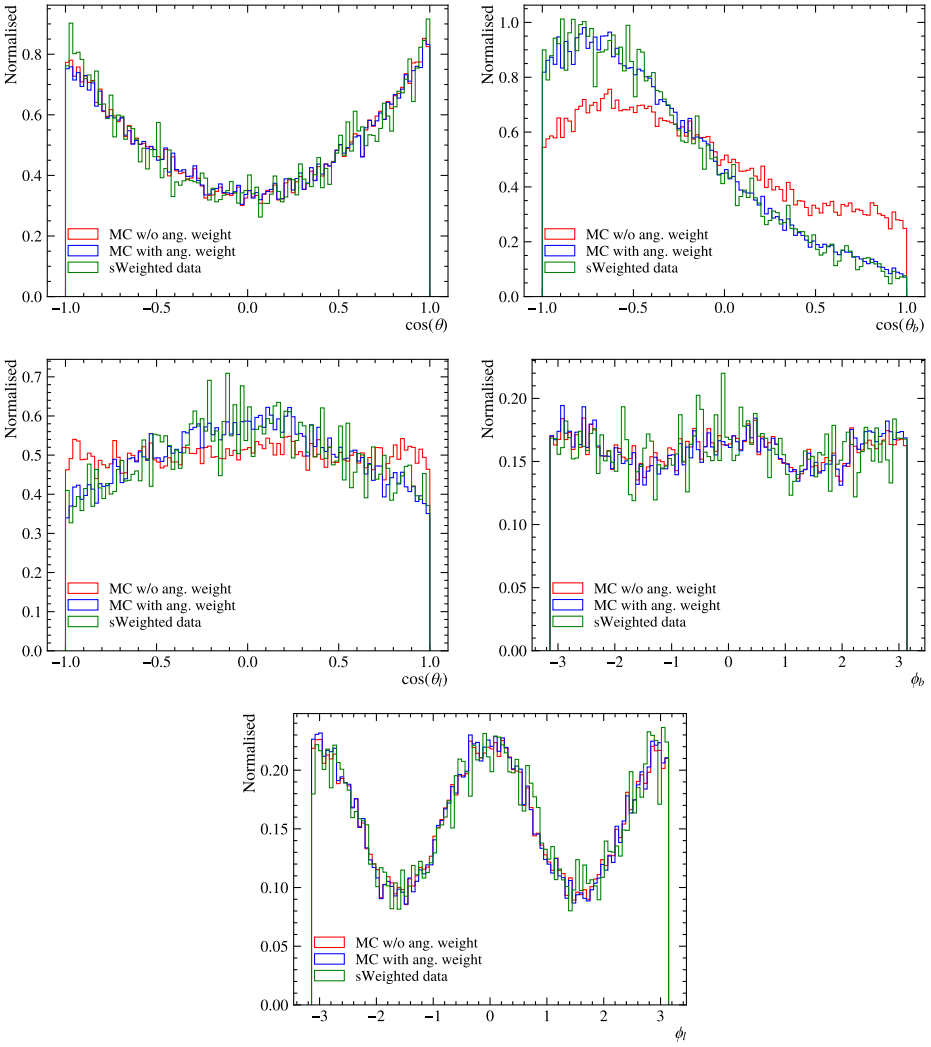


Figure 7.8: The comparisons of the angular distributions for the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ mode between the *sPlot* based data (labelled as "sWeighted data") and the MC with/without angular weights. The samples shown here are of the muon-R2p2-DD category.

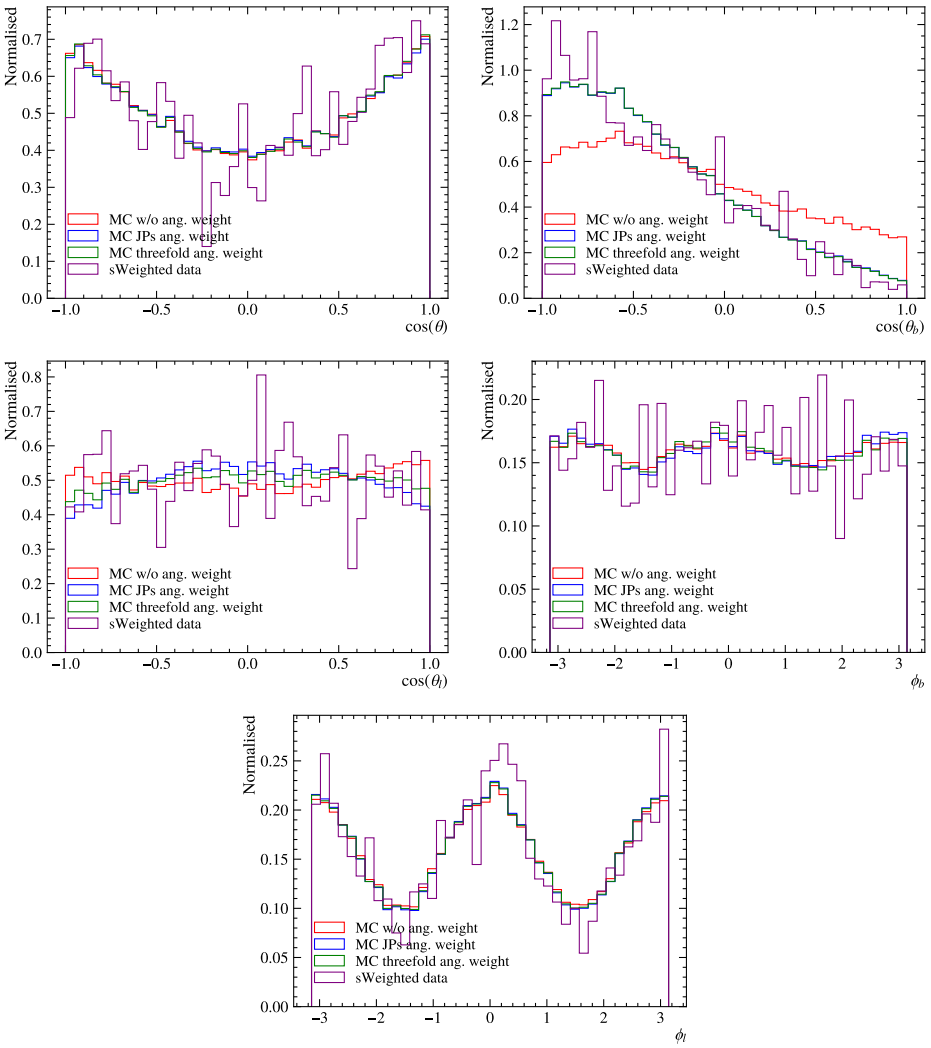


Figure 7.9: The comparisons of the angular distributions for the $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ mode between the *sPlot* based data (labelled as "sWeighted data"), the MC with J/ψ mode angular weights, the MC with three-fold weights and MC without angular weights. The sample shown here are of the muon-R2p2-DD category.

7.3.3 Tracking efficiency

The tracking efficiency in the simulation does not necessarily match the reality. In this analysis, this effect on the tracking efficiencies of the proton and pion tracks are expected to cancel when taking the electron to muon channel ratios. For the muon, the analysis work from a previous LFU study [74, 75] suggested the tracking efficiency for the muons is well-described. Due to these considerations, no tracking corrections are applied for the proton, the pion and the muon tracks.

For the tracking effects of the electrons, the studies conducted in Ref. [109] are considered. As discussed in Ref. [109], the electron tracking efficiency is affected by its momentum and the material it traverses, which leads to the data/MC tracking efficiency correction factor maps being given in bins of transverse momentum p_T, pseudorapidity η and the azimuthal angle ϕ. Using the given maps, a correction factor is assigned for each electron track which represents the ratio of tracking efficiencies between data and MC. The per-decay-candidate weight is then calculated as the product of the correction factors of the two electron tracks.

Mathematically, the tracking weights, as described above, are expressed as:

$$w_i^{\text{track}} = C^{\text{track}}(p_T^{e_1}(i), \eta^{e_1}(i), \phi^{e_1}(i)) \times C^{\text{track}}(p_T^{e_2}(i), \eta^{e_2}(i), \phi^{e_2}(i)) \ , \qquad (7.8)$$

where i stands for the candidate index, C^{track} stands for the correction factor in bins of (p_T, η, ϕ) and $p_T^{e_{1(2)}}, \eta^{e_{1(2)}}, \phi^{e_{1(2)}}$ stand for the transverse momentum, the pseudorapidity, azimuthal angle for the electron $e_{1(2)}$ respectively.

7.3.4 Particle identification efficiency

Due to many factors, *e.g.* the temperature and the gas pressure inside the RICH detector, which might affect the response of the PID subdetectors, the PID efficiency is not described precisely in the simulation [103]. The data-driven method, namely the so-called PIDCalib [102, 103], is used to calibrate the PID efficiencies. The efficiencies of the PID selection are evaluated on the data calibration samples. Technically, the PIDCalib uses the *sPlot* method to realise the background subtraction in the data calibration samples, which enables the estimation of the number of signal candidates before and after the PID selection [103]. The resulting PID efficiencies for given particle tracks are then assigned to the MC samples as weights. For the two leptons, a factorization of the efficiencies is assumed, *i.e.* the efficiency of the whole signal decay candidate passing the PID selection is assumed to be the product of the corresponding efficiencies for each leptons. Since the weights represent

the PID efficiency, the corresponding PID selections are no longer applied for the weighted MC samples. By construction, such a procedure does not preserve the knowledge of the PID response of a given candidate but only corrects the efficiency of the total sample. Thus, these PID weights are only used when an accurate efficiency estimation is required. In other cases, *e.g.* during the MVA training (to be discussed in Section 8.2), the PID selections are used instead of the PID weights.

Among the PID requirements shown in Table 7.1, the `IsMuon` requirement is treated differently. In contrast to the other PID variables, which are susceptible to the environmental conditions during the operation (as mentioned before), the `IsMuon` variable is based on the information of the track and muon station hits. It is known to be well-described in the simulation. Thus, it is applied as a cut on the MC samples instead of using the calibration weights.

The PID likelihood response of the calorimeter system is affected by the recovery of bremsstrahlung photons [102] (the recovery discussed in Section 6.4). Thus, the PID efficiency calibration is performed separately for the electron with and without the bremsstrahlung recovered.

As the PID efficiencies are known to be dependent on the kinematics of the particles [102, 103], the efficiency maps are produced in the bins of lepton momentum (p) and pseudorapidity (η). Technically, this is done with two steps. The first step produces the maps in one dimensional bins of each variable with fine binning. The binning optimisation is then performed by merging the fine bins together based on criteria, *e.g.* the efficiency uncertainty of each bin and the difference of efficiency values in neighouring bins relative to their uncertainties. The second step takes optimised binning in each dimension and produce the efficiency maps in the combined two dimensional binning. Figure 7.10 shows examples of the resulting two dimensional maps.

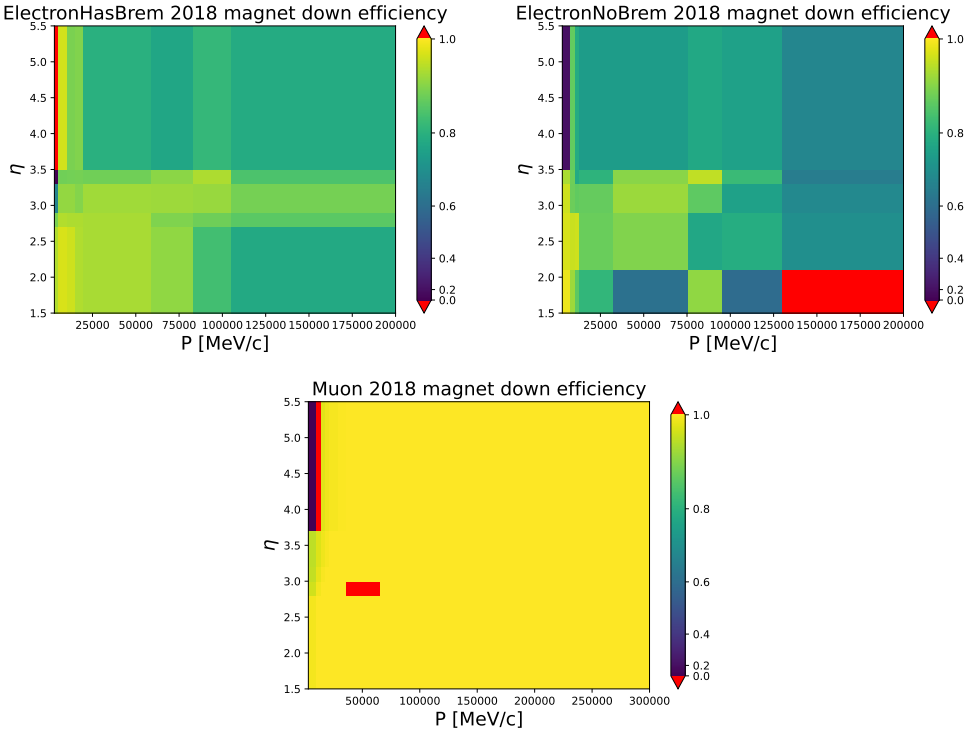


Figure 7.10: Examples (2018, magnet polarity down) of the PID efficiency map in the optimised binning. The top left plot shows for the electron with recovered bremsstrahlung, the top right plot shows for the electron without recovered bremsstrahlung and the bottom plot shows for the muon. The red region indicates the bins with unphysical value, *e.g.* larger than 1.0 or less than 0.0. Such a situation may rise due to the limited statistics in the calibration data sample, considering the background subtraction is realised by the *sPlot* technique. In the nominal approach, these bins are assigned the value 1.0 (in the case of value larger than 1.0), 0.0 (in case of value less than 0.0), or the efficiency calculated by merging all bins (in the case that the bin is missing calibration).

7.3.5 L0 trigger efficiency

As discussed in Section 7.2, the `L0Electron` and `L0Muon` lines are choosen at the hardware trigger level, which set requirements on the E_T or p_T. The hardware response is not easy to perfectly simulate. In addition, the actual thresholds on E_T or p_T vary over the different data-taking periods. Such a variation is not reproduced in the simulation, which has one fixed threshold for each year. Due to these reasons, the L0 trigger efficiency is not well described in the simulation and requires to be calibrated with a data-driven method.

Based on Ref. [110], the data-driven method for trigger calibration, namely the TISTOS method, is explained as follows. The relationship between the trigger response and the signal object (*e.g.* a certain track) can be categorised into three categories, *i.e.* Triggered-On-Signal (TOS), Triggered-Independent-of-Signal (TIS) and Triggered-On-Both (TOB). The TOS category refers to the case when the existance of the signal object alone is enough for a positive trigger response. The TIS category refers to the case when the remaining part of the event (removing the signal object) is sufficient to give a positive trigger response. The TOB category refers to the case which is neither TOS nor TIS, *i.e.* the positive trigger response requires both the signal object and the remaining part of the event. In the TISTOS method, only the first two categories (TIS and TOS) are used. In the data samples, since the events without any positive trigger responses were not saved, the direct trigger efficiency is not accessible. Usually, the accessible data have already certain trigger and selection requirements applied. Instead of using the direct trigger efficiency, one can reorder several stages of the data acquisition without affecting the total efficiency (ϵ_{tot}):

$$\epsilon_{\text{tot}} = \epsilon_{\text{acc}}\epsilon_{\text{trig|acc}}\epsilon_{\text{rec|trig}}\epsilon_{\text{sel|rec}} = \epsilon_{\text{acc}}\epsilon_{\text{rec|acc}}\epsilon_{\text{sel|rec}}\epsilon_{\text{trig|sel}} \ , \tag{7.9}$$

where the notation $\epsilon_{A|B}$ represents the efficiency of A for the events already fulfilling B. The different stages shown are: detector geometric acceptance (acc), trigger filtering (trig), reconstruction (rec) and selection (sel). The efficiency of interest $\epsilon_{\text{trig|sel}}$ can be expressed as

$$\epsilon_{\text{trig|sel}} = \frac{N_{\text{trig}\cap\text{sel}}}{N_{\text{sel}}} \ , \tag{7.10}$$

where $N_{\text{trig}\cap\text{sel}}$ represents the number of candidates passing the trigger and selection while N_{sel} represents the hypothetical number of candidates selected but without trigger requirements. Since the number N_{sel} is not accessible in the data, the efficiency cannot be directly evaluated from the above equation. In case of this analysis, as discussed in Section 7.2, the chosen triggers are

Table 7.3: The TIS trigger lines used for the L0 trigger efficiency study. The trigger is required to have a positive response independent of the signal decay candidate.

Leptons	TIS trigger lines
Electron	`L0Muon` or `L0Hadron`
Muon	`L0Electron` or `L0Hadron`

TOS (trig = TOS in the above equation). Assuming the independence of the TIS and TOS response, the trigger efficiency $\epsilon_{\text{trig|sel}}$ can be calculated as

$$\epsilon_{\text{TOS|sel}} = \frac{N_{\text{TIS}\cap\text{TOS}\cap\text{sel}}}{N_{\text{TIS}\cap\text{sel}}} , \tag{7.11}$$

where $N_{\text{TIS}\cap\text{TOS}\cap\text{sel}}$ represents the number of selected candidates with both the TIS and TOS responses while $N_{\text{TIS}\cap\text{sel}}$ represents the number of selected candidates with the TIS response. Both $N_{\text{TIS}\cap\text{TOS}\cap\text{sel}}$ and $N_{\text{TIS}\cap\text{sel}}$ can be evaluated from the data.

In principle, the TISTOS method efficiency can be extracted based on the reference channel data (with the *sPlot* technique used for background subtraction) and then compared with its counterpart in the simulation. The ratio between the two can then be considered as a correction weight. In practice, it was found that the reference channels $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ ($\ell = e, \mu$) do not have sufficient statistics for the calibration, especially in the case of the electron. Instead, the calibration results based on $B^+ \to K^+ J/\psi(\ell^+\ell^-)$ from the R_X ($X = K, K^*$) study [74,75] are used. The results from $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ are compared with the R_X results. Since the L0 triggers used in this analysis act on the individual lepton, the portability between the two is expected.

With the `L0Electron` (on the electron) and `L0Muon` (on the muon) being the nominal L0 trigger requirements for this analysis, the TIS trigger lines used for the trigger efficiency study are shown in Table 7.3.

With the L0 trigger acting on $E_T(e)$ or $p_T(\mu)$ calculated at the L0 level, the trigger efficiency is evaluated as a function of the reconstructed level variable E_T or p_T. Considering the resolution of the variables calculated at the L0 level, the L0 trigger efficiencies are expected to behave as sigmoid-like curves with respect to the reconstructed E_T (or p_T). Following the R_X ($X = K, K^*$) study [74, 75], the low electron E_T region poses a certain problem. Considering the low statistics for the electron, the calibration in the low E_T regions can be difficult since the efficiency is close to zero. This is especially problematic if the trigger threshold is tighter in the simulation than in the data. To treat this issue, the so-called Trigger Configuration Key (TCK)

Table 7.4: The TCK alignment cuts for electrons used in the R_X ($X = K, K^*$) study [74, 75], which requires a minimum E_T.

Year	Cut on E_T [MeV]
2011	2500
2012	3000
2015	3000
2016	2700
2017	2700
2018	2400

alignment cut is applied to the electron, which is shown in Table 7.4. When an electron gives a positive `L0Electron` trigger response, it is additionally required to have higher E_T than the threshold shown in Table 7.4. For the muon, no equivalent cut is applied. However, the offline selection includes the cut $p_T > 800\,\text{MeV}/c$ for muons regardless of their trigger responses.

As examples (2018), the single-lepton L0 trigger efficiency curves are shown in Figure 7.11 (for electrons) and Figure 7.12 (for muons). The comparison between the R_X result from $B^+ \to K^+ J/\psi(\ell^+\ell^-)$ and the result of this study from $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ is also shown. Albeit the large statistical uncertainties on the part of $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$, especially in the case of the electrons, the agreement between the basic trends of the two can be seen.

The calibration maps (based on the $B^+ \to K^+ J/\psi(\ell^+\ell^-)$ channel) from the R_X study [74, 75] provide the single lepton L0 trigger efficiency for data and MC in bins (E_T and ECAL region bins for the electron, p_T and pseudorapidity η bins for the muon). For the electron, the binning in the ECAL region is justified considering that the different granularities of the ECAL cells in different regions (discussed in Section 2.2) are likely to lead to different hardware responses. For the muons, the pseudorapidity η bins are justified considering the segmentations of the muon stations in the x-y plane [111],

Using these maps, the L0 efficiency factors are assigned to the lepton candidates in the simulated samples of the $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ analysis, *i.e.* a data L0 efficiency factor and a MC L0 efficiency factor for each of the two leptons (ℓ_1, ℓ_2) in one decay candidate:

$$\epsilon^{\text{trig,data}}_{\ell_1,i} \;,\; \epsilon^{\text{trig,data}}_{\ell_2,i} \;,\; \epsilon^{\text{trig,MC}}_{\ell_1,i} \;,\; \epsilon^{\text{trig,MC}}_{\ell_2,i} \;, \tag{7.12}$$

where i represents the decay candidate index.

Since the L0 trigger decision for a decay candidate is made as the logical `OR` of the two single lepton trigger responses, the single lepton efficiency factor

(as discussed above) can be combined into per-decay-candidate efficiency as

$$\begin{aligned} \epsilon^{\text{trig,data}}_{\text{decay},i} &= 1 - (1 - \epsilon^{\text{trig,data}}_{\ell_1,i})(1 - \epsilon^{\text{trig,data}}_{\ell_2,i}) \ , \\ \epsilon^{\text{trig,MC}}_{\text{decay},i} &= 1 - (1 - \epsilon^{\text{trig,MC}}_{\ell_1,i})(1 - \epsilon^{\text{trig,MC}}_{\ell_2,i}) \ . \end{aligned} \tag{7.13}$$

For the electrons, where the TCK alignment cuts (as discussed in Section 7.2) are applied, the efficiency factor is set to 0.0 when the E_T falls below the alignment threshold. For example, in the case that one electron ℓ_1 passes the alignment threshold while the other ℓ_2 does not, this effectively sets the $\epsilon^{\text{trig,data}}_{\text{decay},i}$ to $\epsilon^{\text{trig,data}}_{\ell_1,i}$ ($\epsilon^{\text{trig,MC}}_{\text{decay},i}$ to $\epsilon^{\text{trig,MC}}_{\ell_1,i}$) as desired.

As the main output of this calibration procedure, the per-decay-candidate calibration weight is then calculated as the data over MC ratio

$$w^{\text{trig}}_i = \frac{1 - (1 - \epsilon^{\text{trig,data}}_{\ell_1,i})(1 - \epsilon^{\text{trig,data}}_{\ell_2,i})}{1 - (1 - \epsilon^{\text{trig,MC}}_{\ell_1,i})(1 - \epsilon^{\text{trig,MC}}_{\ell_2,i})} \ . \tag{7.14}$$

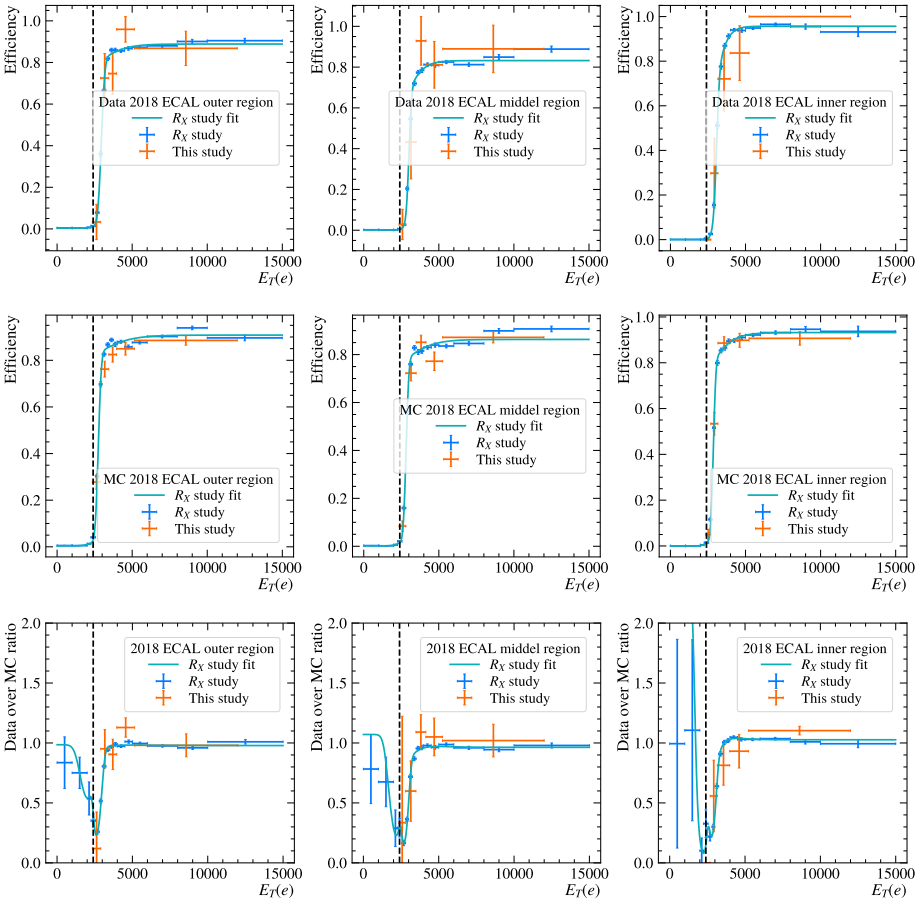


Figure 7.11: An example (2018) of the L0 trigger efficiency results for electrons. The top row shows the data efficiencies. The middle row shows the MC efficiencies. The bottom row shows the data over MC ratio. Each of the three columns shows for one of the three zones of ECAL (zones with different cell granularities). The black dashed lines indicate the TCK alignment cuts.

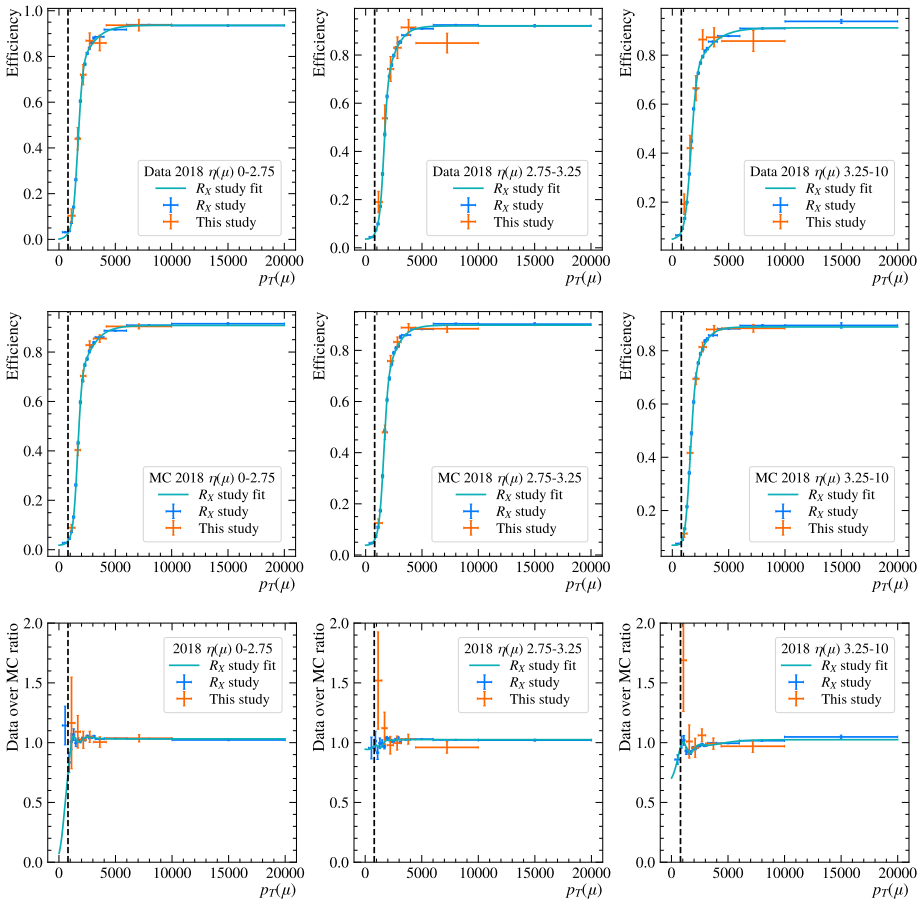


Figure 7.12: An example (2018) of the L0 trigger efficiency results for muons. The top row shows the data efficiencies. The middle row shows the MC efficiencies. The bottom row shows the data over MC ratio. The three columns show for different pseudorapidity (η) ranges of the muon. The black dashed lines indicate the $p_T > 800\,\mathrm{MeV}/c$ cut.

7.3.6 Production kinematics and track multiplicity

The production kinematics of Λ_b^0 and the overall event multiplicity are not well described in the PYTHIA based simulation. They are corrected via a reweighting procedure using the so-called GBReweighter [112]. This technique provides the framework for reweighting multiple variables simultaneously. It is a machine learning technique which performs the training with the original and the target distributions and gives per-decay-candidate weights as output such that the weighted distributions match the target distributions.

In this analysis, the transverse momentum $p_T(\Lambda_b^0)$, the pseudorapidity $\eta(\Lambda_b^0)$ and the number of tracks (nTracks) in the event are used as proxies for the Λ_b^0 kinematics and multiplicity for the training. The distributions of these variables in data are extracted using the *sPlot* technique on the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ sample. The GBReweighters are trained to match the reconstructed $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ MC sample to the data (the *sPlot* technique applied). The effects of the weighting are illustrated in Figure 7.13 ($\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ samples, R2p2-DD category). As shown in the plot, the reweighting procedure, as intended, improves the compatibility between data and MC in the distributions of $p_T(\Lambda_b^0)$, $\eta(\Lambda_b^0)$ and nTracks.

The trained reweighters are also applied to the electron channels. This is based on two considerations. Firstly, the electron channel is more statistically limited than the muon channel and thus not ideal to be used for the training of the reweighters. Secondly, the mismodelling of the Λ_b^0 kinematics is expected to be the similar between the electron and the muon channel since it mainly concerns the Λ_b^0 generation.

The purpose of the reweighting of Λ_b^0 kinematics is to correct for its mismodelling in the MC generation. Thus, the trained GBReweighters are applied to both the reconstructed MC samples (the samples containing reconstructed decay candidates) and the generated truth level MC samples (the samples containing the decay candidates after detector geometric acceptance but before the reconstruction). As to be discussed in Section 9.2, the MC samples of both levels are needed for the efficiency evaluation. The reconstructed MC samples contain both the reconstruction level information and the associated truth level information. However, for the data, only reconstructed variables are accessible. The training of the GBReweighters is thus performed with the reconstructed variables. The trained GBReweighters are applied to reconstructed MC samples based on the reconstructed variables. However, for the Λ_b^0 kinematic variables in truth level MC samples, only the truth level information are available [22]. For these MC samples, the

[22] A generated Λ_b^0 decay is not necessarily reconstructed. However, certain reconstructed variables which are associated with the whole event (*e.g.* nTracks), can be matched to a

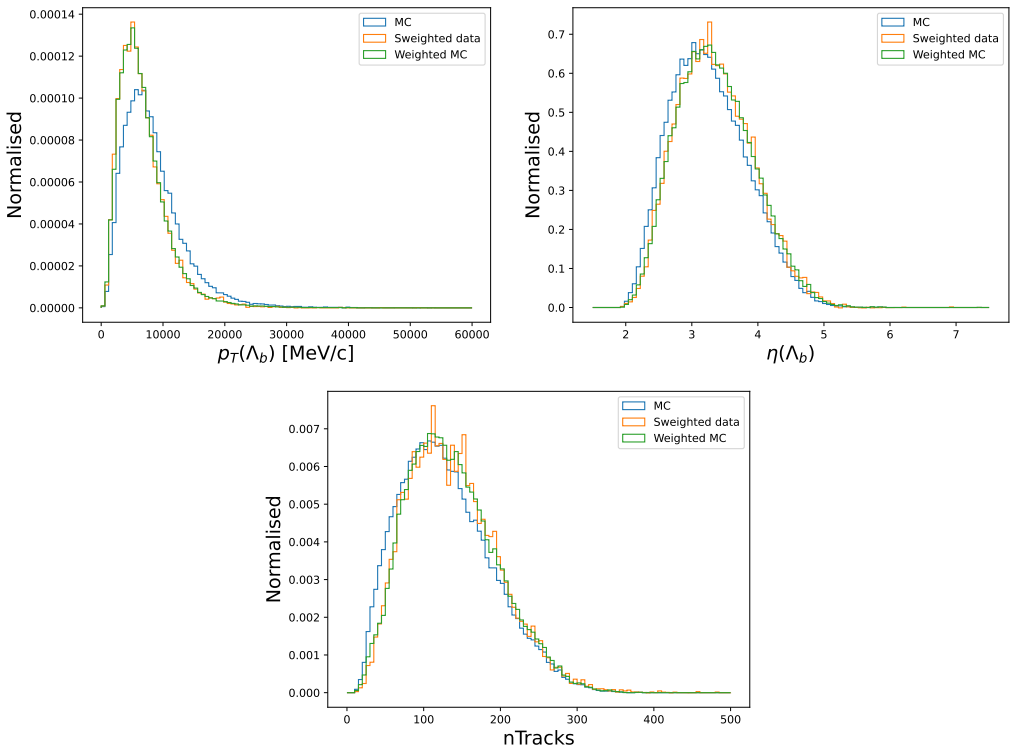


Figure 7.13: A comparison of the distributions of the reconstructed weighting variables between the *sPlot* data (labelled as "Sweighted data"), MC with and without weights. The weighted MC shows better agreement with the data than the MC without weights. The samples shown here are $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ in the R2p2-DD category.

reweighters are thus applied with the truth level Λ_b^0 kinematic variables. This is justified, because the effects of the reconstruction resolution are expected to be small compared to the effects of kinematic mismodelling of Λ_b^0 in the MC generation.

Figure 7.14 shows the comparison of the $p_T(\Lambda_b^0)$, $\eta(\Lambda_b^0)$ and nTracks distributions between the reconstructed MC sample and the *sPlot* data for the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ decay channel (R2p2-DD). Besides the reconstructed $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ variables, the truth level $p_T(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ associated with these reconstructed MC candidates are also shown. As shown in the plots, the weights improve the compatibility of MC with respect to the data. Compared to the difference caused by the weights, the agreement between the truth level and the reconstruction level Λ_b^0 p_T, η is good both before and after the weighting. This serves to illustrate two points. The first point is that the reweighter extracted from the muon channel is also able to reweight the electron channel. The second point is that the reconstruction resolution effect is, indeed, small comparing with the MC mismodelling.

generated Λ_b^0 decay. The truth level MC samples contain the variable nTracks.

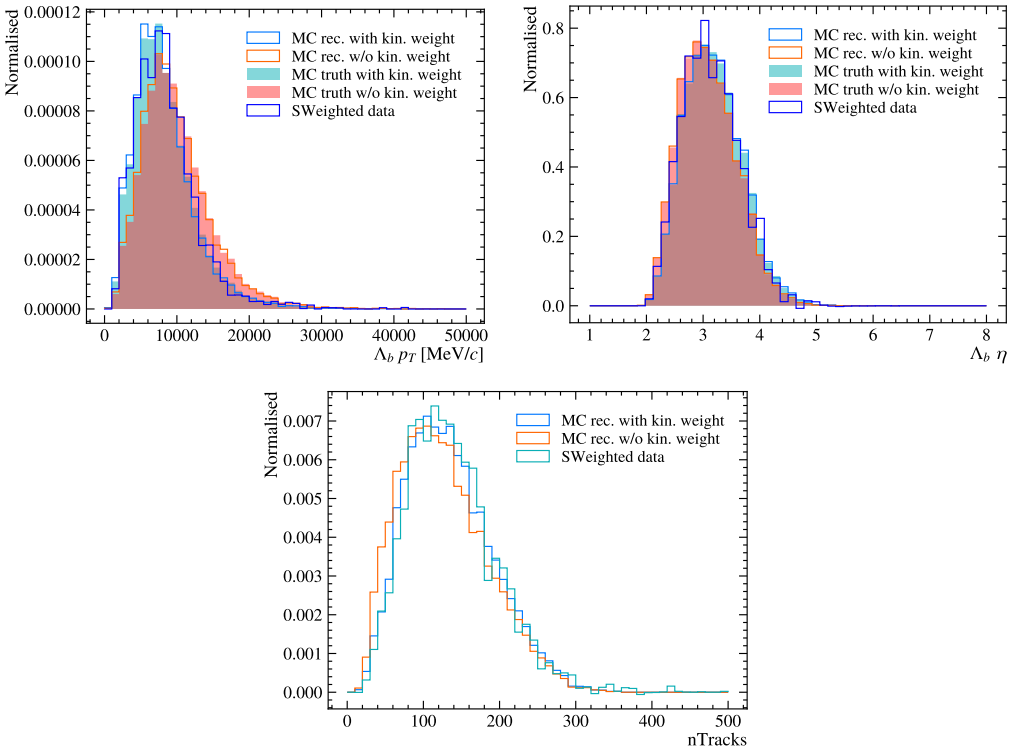


Figure 7.14: The weighting variables of the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ MC and the *sPlot* data. For the Λ_b^0 p_T and η, the reconstructed level and truth level distributions are shown. The samples shown here are in the R2p2-DD category.

8 Combinatorial background suppression

One of the important backgrounds in the analysis is the combinatorial background. In the signal extraction, which is to be discussed in Chapter 9, such a background is differentiated from the signal via different shapes of the Λ_b^0 invariant mass distribution. Nevertheless, due to the complexity of the pp collision events, the amount of the combinatorial background can be large. In this analysis, the method of MultiVariate Analysis (MVA) is used to help to distinguish the signal from the combinatorial background.

8.1 Technique introduction

One of the key concepts of the MVA is that of the classifier. A classifier is essentially a map from the space of multiple input variables to an output variable space. Such a map usually contains lots of adjustable parameters, the adjustment of which is called the training of the classifier. In the MVA procedure, the classifiers are trained and then applied to the samples of interest to evaluate an output score for each candidate, which indicates how likely a certain candidate is a signal candidate. The training of the classifiers requires a labelled training sample, *i.e.* a collection of candidates whose signal/background classification is known. Various features (input variables) and the classification label of the training candidates are feed into the classifier model. The parameters of the classifier model are adjusted such that the predicted scores of these training candidates are well-matched to their known labels. The criterion for how well the predicted scores match the known labels is the so-called loss function.

In this analysis, the classifier algorithm Gradient Boosted Decision Tree (GBDT) [113] implemented by the package scikit-learn [114] is used.

Based on Refs. [115, 116], the basic idea of the GBDT is introduced as follows. The GBDT is an ensemble of the basic classifiers called the basic decision trees. A basic decision tree recursively splits the training sample according to a series of boolean decisions, which are referred to as nodes. Given a training sample with input variables $\vec{x} = (x_1, x_2, \cdots)$ and the known label y (y being either the signal or the background), the boolean decision at the node m is made by comparing the value of a certain input variable x_i to a threshold t_m. The parameters $t_m (m = 1, 2, \cdots)$ are tunable parameters during the training of the decision tree. In the training process, these nodes are adjusted such that the entries in the sample with the same known label are grouped together. The quantitative criterion for this is the loss function, which can be defined in various ways depending on the specific cases. Figure 8.1 illustrates the sample-spliting process of a basic decision tree. Such a

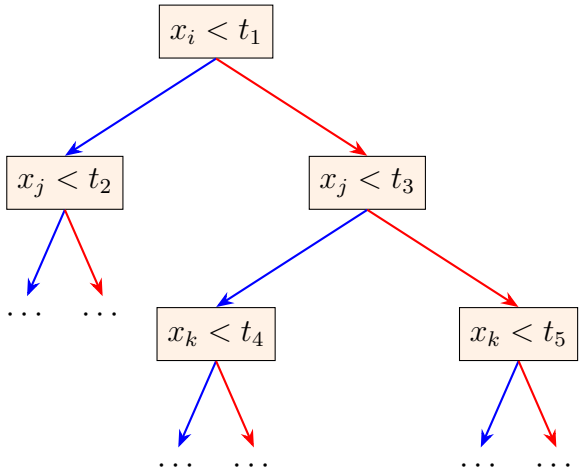


Figure 8.1: An illustration of the basic decision tree. Each box indicates a decision node. The red lines indicate positive decisions. The blue lines indicate negative decisions. Plot based on Refs. [115, 116].

basic decision tree has several weaknesses, *e.g.* being susceptible to the fluctuation in the training samples or being locally instead of globally optimised. The GBDT classifier provides a more robust classifier by using an ensemble of the basic decision trees to mitigate these weaknesses.

One of the issues which can happen during the training of the classifiers is the so-called over-training, which exhibits as a tendency for the classifiers to perform better in the training samples themselves than other samples they are applied to. The over-training can be caused by the training model having too many tunable parameters or the limited statistics in the training sample [116]. Due to the potential over-training, it entails risks to apply the trained classifiers on their own training samples. In this analysis, the samples used during MVA training are also used in the later steps of the analysis [23]. The potential over-training can thus cause biases, especially in terms of efficiency estimation. In order to avoid the bias caused by the potential over-training, the so-called k-folding technique [117] is used. In using this technique, the training samples are divided into k subsets. By combining $k-1$ subsets each time for the training, *i.e.* keeping one subset out, k classifiers are trained. Each classifier can then be evaluated on the subset kept out during its training. Figure 8.2 illustrates such an idea.

[23] The available statistics should be fully utilised.

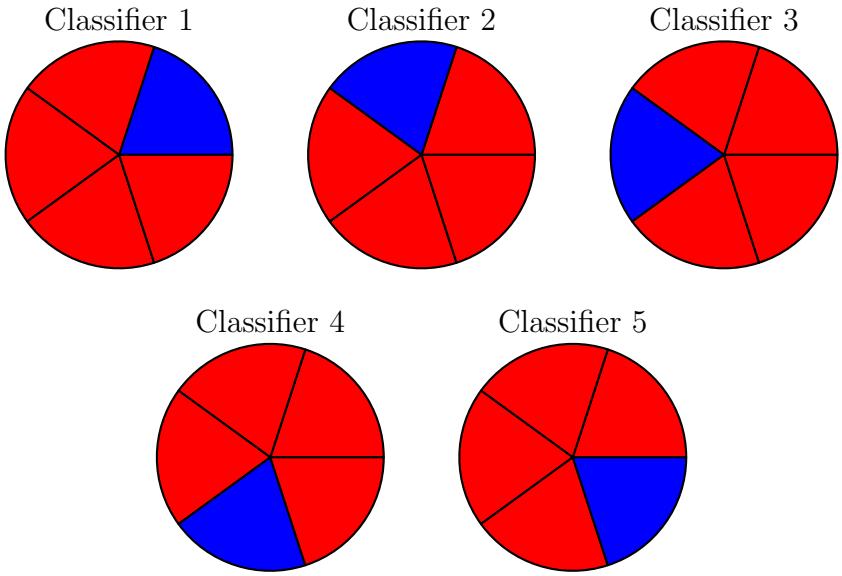


Figure 8.2: An illustration of the k-folding technique (in the case k= 5). The 5 pie charts correspond to 5 classifiers. The training sample is divided into 5 subsets. Each classifier is trained using 4 subsets (red) and evaluated on the remaining subset (blue). The subsets are iterated for the 5 classifiers as illustrated. The relative size of the pieces in the pie charts is illustrative. The plot was inspired by Ref. [117].

8.2 Classifier training and validation

As discussed in the previous section, to perform the training of the classifiers, a labelled training sample is needed. The rare signal MC samples (*i.e.* $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ MC) are used as the signal proxy. With the known Λ_b^0 mass $\sim 5620\,\mathrm{MeV}/c^2$ [64], the data upper sidebands, which are defined as $5900(5750) < m_{\mathrm{DTF}}(\Lambda_b^0) < 7120\,\mathrm{MeV}/c^2$ for the electron(muon) channel , are used as the background proxy. The upper sidebands reside above the expected Λ_b^0 mass, which are expected to mainly consist of the combinatorial background. Due to the potential contribution from the partially reconstructed background, the lower sidebands are not used for the MVA training.

Due to the limited statistics, the MVA training is not conducted separately in different q^2 regions. The merged q^2 range $0.1-22.0\,\mathrm{GeV}^2/c^4$ is used with some exclusions as discussed in the following. Considering the large amount of J/ψ and $\psi(2S)$ resonances in the pp collision environment (not necessarily from Λ_b^0 decay), a considerable part of the combinatorial background in the corresponding resonance q^2 regions originates from the random combination of J/ψ or $\psi(2S)$ with another Λ^0 particle or a randomly combined $p\pi^-$ pair. However, outside the resonance q^2 regions, the dilepton pairs are likely random combinations themselves. Since the main interest of this analysis is the rare decay $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$, the resonance q^2 regions as defined in Table 6.1 are excluded from the training samples. In summary, the q^2 regions used for the MVA training are as follows.

- Electron channel: $0.1-22.0\,\mathrm{GeV}^2/c^4$ excluding $6.0-11.0\,\mathrm{GeV}^2/c^4$ (J/ψ region) and $11.0-15.0\,\mathrm{GeV}^2/c^4$ ($\psi(2S)$ region).
- Muon channel: $0.1-22.0\,\mathrm{GeV}^2/c^4$ excluding $8.7-10.05\,\mathrm{GeV}^2/c^4$ (J/ψ region) and $12.5-14.2\,\mathrm{GeV}^2/c^4$ ($\psi(2S)$ region).

Given the $m_{\mathrm{DTF}}(\Lambda_b^0)$ mass region and the q^2 region as discussed above, the available statistics in the signal proxy ($\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ MC) and background proxy (data upper sideband) are shown in Table 8.1. It can be seen that the available statistics is especially limited in the electron channel background proxy. As discussed in Section 8.1, limited statistics in the training sample can lead to over-training. The aforementioned k-folding technique is used to protect the classifier evaluation against the potential over-training. Nevertheless, the limited statistics in the training samples can still limit the performance of the classifiers. Thus, another action taken is to merge the R2p1 and R2p2 samples during the training, which will be further explained later. To faciliate the discussion, the merged R2p1 and R2p2 samples are labelled as R2.

Table 8.1: The available statistics (number of candidates) in the signal (sig.) proxy and the background (bkg.) proxy for the electron and the muon channels.

	Muon sig.	Muon bkg.	Electron sig.	Electron bkg.
R1 DD	23770	10630	5823	995
R1 LL	13406	3013	2991	237
R2p1 DD	33736	23529	9785	3232
R2p1 LL	19022	6900	5219	951
R2p2 DD	68239	46101	21769	6469
R2p2 LL	39623	12614	11749	1566

In summary, the MVA training is performed separately in 8 sample categories, *i.e.* muon-R1-DD, muon-R1-LL, muon-R2-DD, muon-R2-LL, electron-R1-DD, electron-R1-LL, electron-R2-DD, electron-R2-LL.

When choosing the features for training (*i.e.* the input variables), good signal/background distinguishing ability as well as good data/MC compatibility are desired. The variables used for training are listed and explained as follows.

- $\log(\chi^2_{\rm FD})(\Lambda_b^0)$: $\chi^2_{\rm FD}$ refers to the Flight Distance (FD) significance of the Λ_b^0 particle from the PV.
- $\chi^2_{\rm IP}(\Lambda_b^0)$: the Impact Parameter (IP) χ^2 of the Λ_b^0 relative to the PV. The $\chi^2_{\rm IP}$ of a track relative to a vertex refers to the difference of the vertex fit χ^2 with or without the track.
- $\log(\chi^2_{\rm FD})(\Lambda^0)$: the FD significance of the Λ^0 from the PV.
- $\chi^2_{\rm FD}(\ell\ell)$: the FD significance of the dilepton vertex from the PV.
- $\max(\log(\chi^2_{\rm IP})(\ell_1), \log(\chi^2_{\rm IP})(\ell_2))$: the larger one of the $\log(\chi^2_{\rm IP})$ of the two leptons relative to the PV.
- $\min(\log(\chi^2_{\rm IP})(\ell_1), \log(\chi^2_{\rm IP})(\ell_2))$: the smaller one of the $\log(\chi^2_{\rm IP})$ of the two leptons relative to the PV.
- DIRA(Λ_b^0): the cosine of the angle between the momentum of Λ_b^0 and the direction from PV to the Λ_b^0 decay vertex.
- $m(\Lambda^0)$ (only for LL): the invariant mass of the reconstructed Λ^0.
- $\chi^2_{\rm endvtx}(\Lambda_b^0)$ (only for LL): the χ^2 of the Λ_b^0 decay vertex.

Table 8.2: The input variables for the MVA training

LL	DD
$\log(\chi^2_{\rm FD})(\Lambda^0_b)$	
$\chi^2_{\rm IP}(\Lambda^0_b)$	
$\log(\chi^2_{\rm FD})(\Lambda^0_b)$	
$(\chi^2_{\rm FD})(\ell\ell)$	
$\max(\log(\chi^2_{\rm IP})(\ell_1), \log(\chi^2_{\rm IP})(\ell_2))$	
$\min(\log(\chi^2_{\rm IP})(\ell_1), \log(\chi^2_{\rm IP})(\ell_2))$	
$\mathrm{DIRA}(\Lambda^0_b)$	
$m(\Lambda^0)$	-
$\chi^2_{\rm endvtx}(\Lambda^0_b)$	-
$\mathrm{DIRA}(\Lambda^0)$	-

- $\mathrm{DIRA}(\Lambda^0)$ (only for LL): the cosine of the angle between the momentum of Λ^0 and the direction from the PV to the Λ^0 decay vertex.

The Λ^0_b should originate from the PV with a displacement before it decays, which is reflected by $\chi^2_{\rm FD}$ (Λ^0_b), $\chi^2_{\rm IP}$ (Λ^0_b) and $\mathrm{DIRA}(\Lambda^0_b)$. The Λ^0 and the dileptons should not originate from the PV, which is reflected by $\chi^2_{\rm FD}$ (Λ^0), $\mathrm{DIRA}(\Lambda^0)$, $\chi^2_{\rm FD}(\ell\ell)$, $\chi^2_{\rm IP}(\ell_1)$ and $\chi^2(\ell_2)$. The Λ^0 and the dileptons should form a good vertex, which is reflected by the Λ^0_b decay vertex quality $(\chi^2_{\rm endvtx})$. The invariant mass of Λ^0 can contribute to distinguish between a genuine Λ^0 and a random combination of $p\pi$ tracks.

The training variables are summarised in Table 8.2.

As an example, two of the input variables for the muon-LL category are shown in Figure 8.3. Distributions for more variables can be found in Appendix C. The distributions are shown for the signal and the background in three different data-taking periods, *i.e.* R1, R2p1 and R2p2. As can be seen in Figure 8.3, the distributions for R2p1 and R2p2 are close to each other while being different to R1. Although no detailed study was conducted to find the exact reason, there are various possible reasons due to the differences between Run 1 and Run 2, *e.g.* different pp collision center-of-mass energy and different HLT triggers. Considering the similarity between R2p1 and R2p2 as well as the limited statistics in the electron channel, the training was performed by merging R2p1 and R2p2 [24].

The aforementioned k-folding technique is used for training. The signal

[24]It has to be conceded that this still leaves Run 1 training samples to be quite statistically limited, which can limit the performance of the classifiers. However, the potential bias induced by over-training is prevented using the k-folding technique.

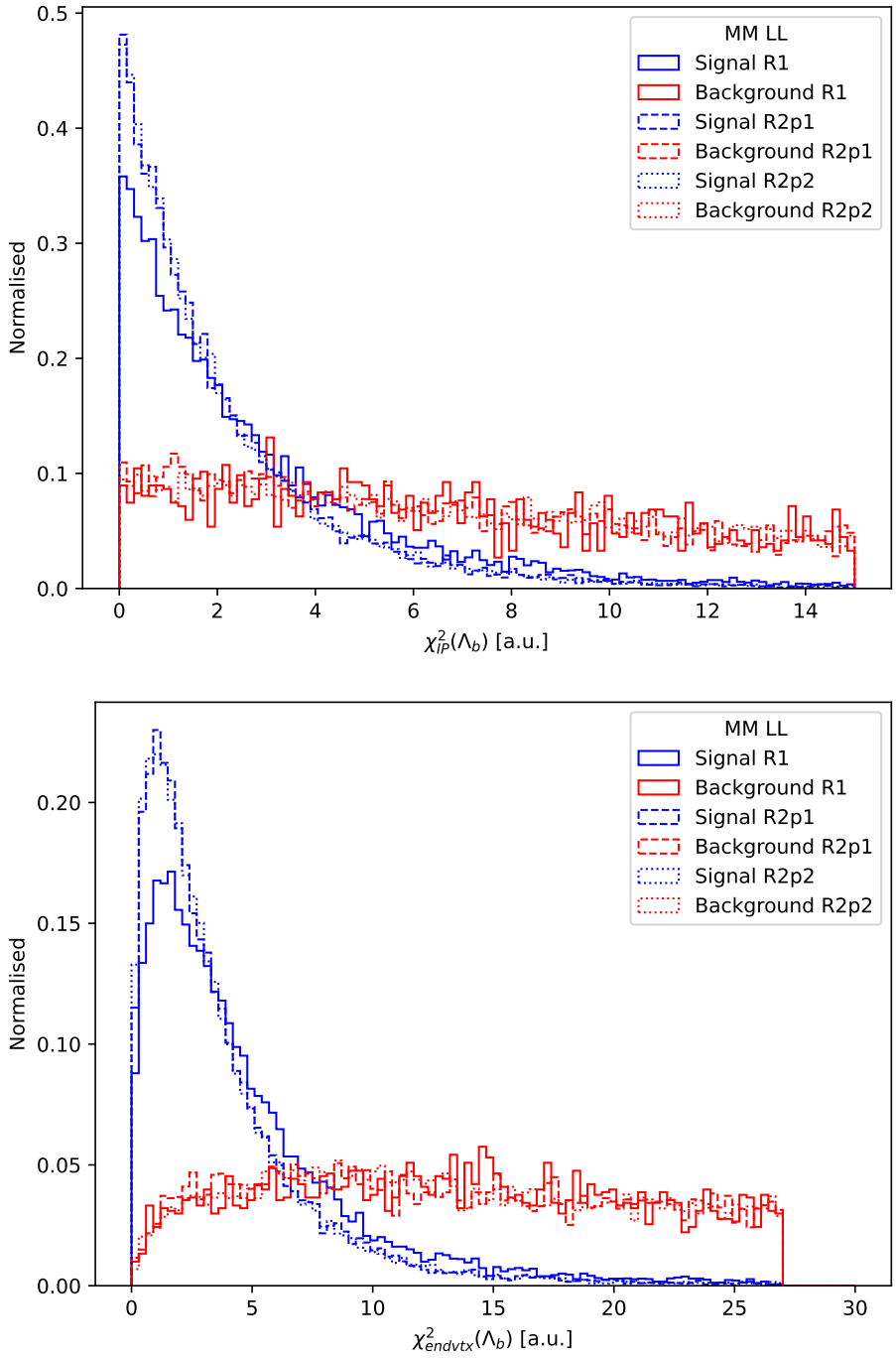


Figure 8.3: Two of the input variables used for the MVA training in the muon-LL category.

and the background proxy samples are divided into 5 subsets. Each time the classifier is trained on 4 of them and tested on the remaining fold. In the later steps of the analysis, all 5 classifiers are used. For the candidates involved in the training and testing, their classification outputs are evaluated with the classifier which they are not trained with. Technically, this can be achieved utilising the so-called `eventNumber`[25] modulo 5 as a label to divide the training samples and to assign the classifiers.

As an example, the outputs of one of the classifiers on the signal and the background proxy samples are shown in the left plot of Figure 8.4. The right plot of Figure 8.4 shows, for the 5 classifiers of the electron channel Run 2 DD category, the Receiver-Operating-Characteristic (ROC) curves, *i.e.* the signal versus the background retention rates when different cuts on the MVA output variable are applied. Further classifier outputs are shown in Appendix C.

For the MVA classifier responses, the compatibility between data and simulation is checked. Since the MVA selection will contribute to the efficiency, which will be evaluated on the simulation samples (with the corrections discussed in Section. 7.3), a good data/simulation compatibility of the MVA response is desired to avoid potential biases. The *sPlot* technique [90] (as discussed in Section 6.7) is used to extract the distribution of the MVA response for the data of $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ channel. The extracted response is compared with the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ MC sample. Although the training was performed merging R2p1 and R2p2, in several later steps of the analysis (*e.g.* efficiency evaluation and resonance fit as to be discussed in Chapter 9), the R2p1 and R2p2 are still treated separately. Thus, this crosscheck is performed on the R2p1 and R2p2 samples separately, *i.e.* the R2p1 and R2p2 samples shown separately but both have the R2 classifiers applied. As shown in Figure 8.5 (electron channels) and Figure 8.6 (muon channels), the compatibility between the MC and the *sPlot* data for the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ channels is good.

[25] A number identifying individual event in the samples.

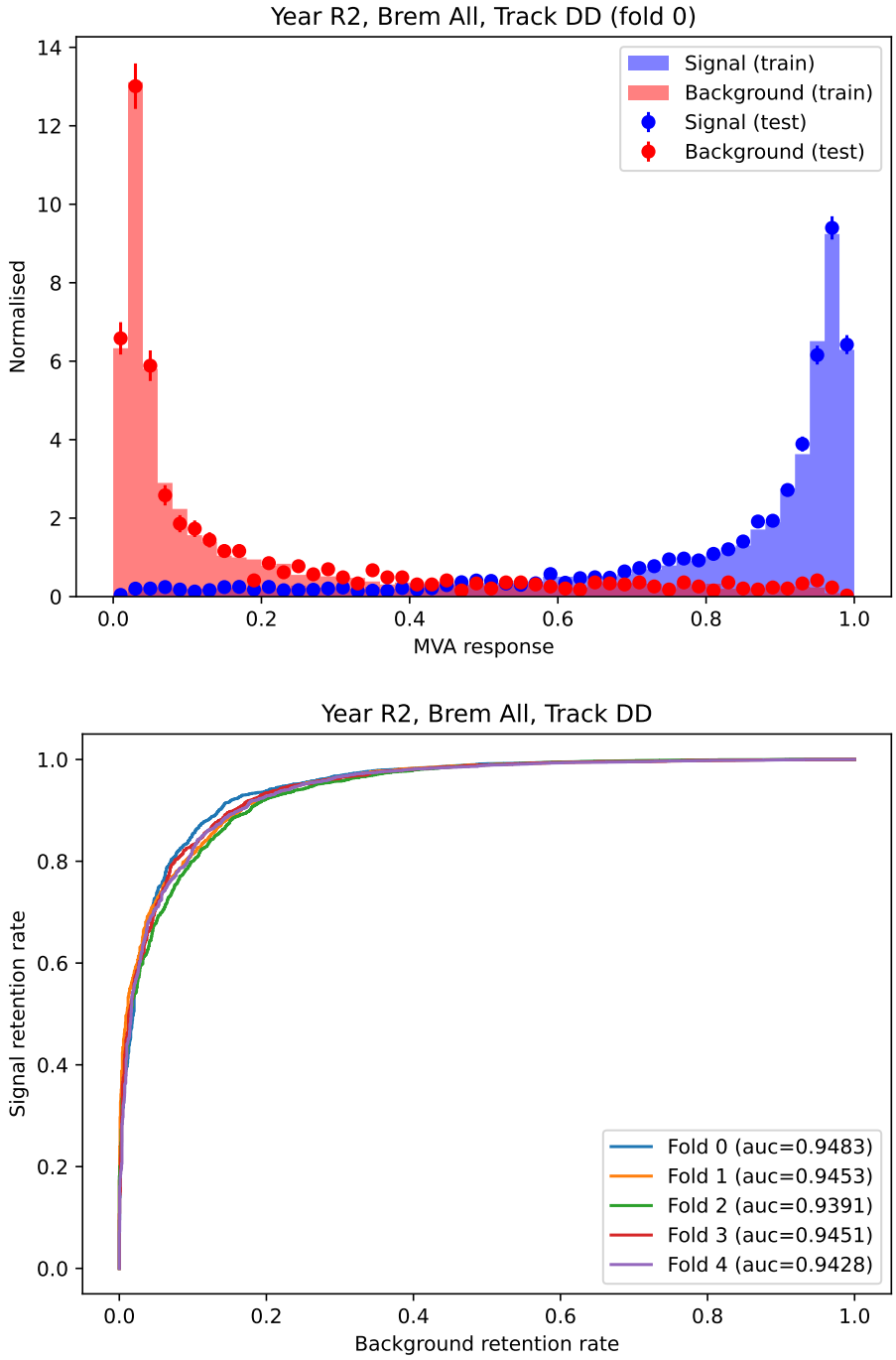
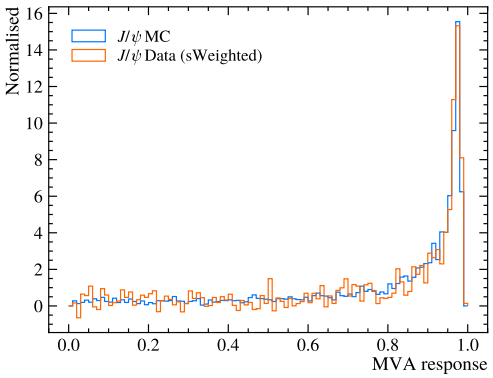
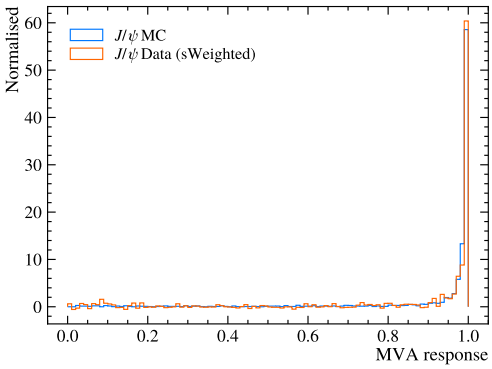


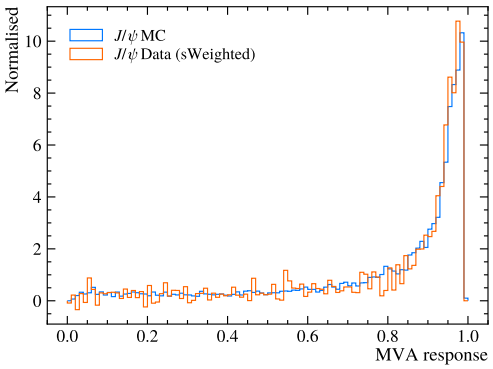
Figure 8.4: The top plot shows the output distributions of one of the five classifiers corresponding to the electron-R2-DD category. The bottom plot shows the ROC curves of those 5 classifiers. The Area-Under-Curve (AUC) serves to gauge the classifier performance.



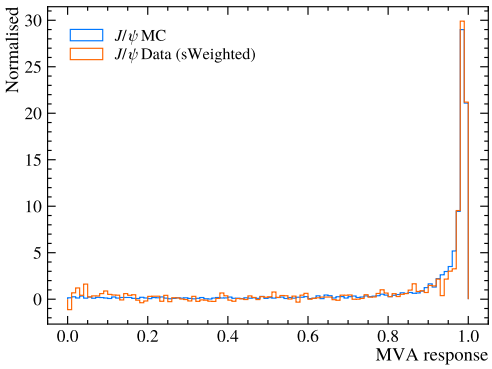
(a) R1-DD



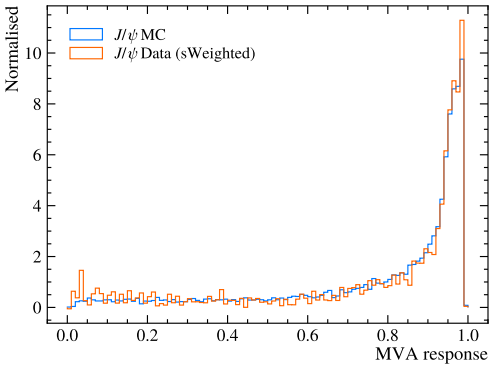
(b) R1-LL



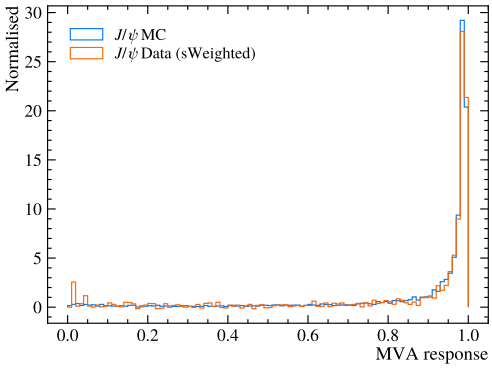
(c) R2p1-DD



(d) R2p1-LL

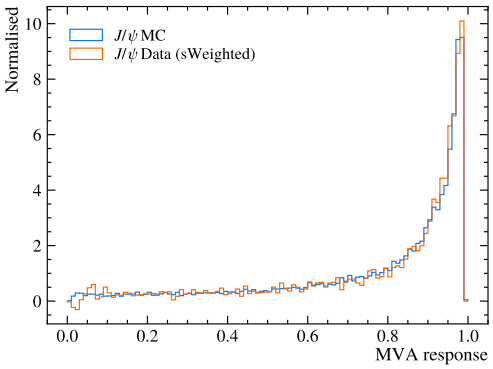


(e) R2p2-DD

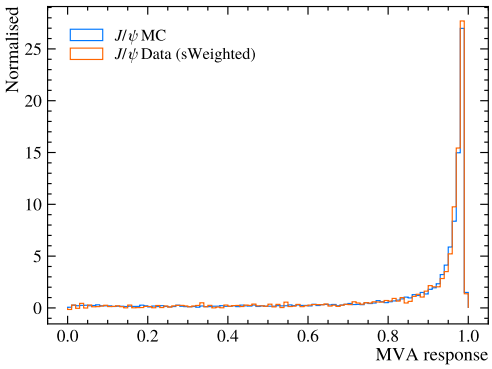


(f) R2p2-LL

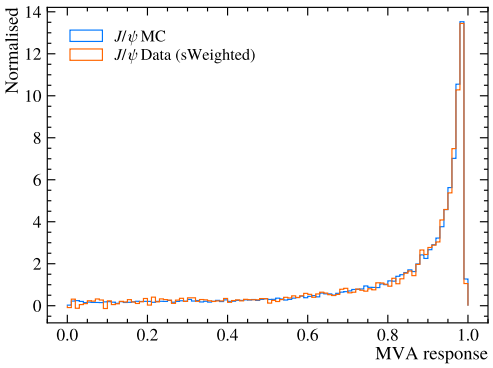
Figure 8.5: The comparison of the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ channel MVA responses between the simulation and data (*sPlot* technique applied).



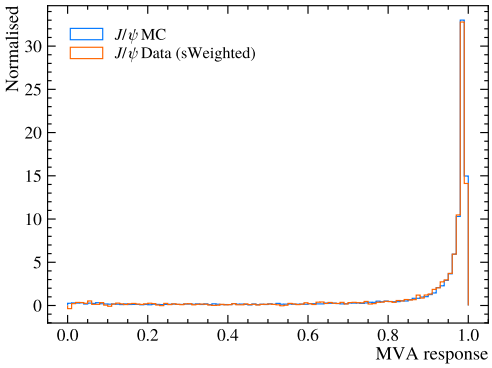
(a) R1-DD



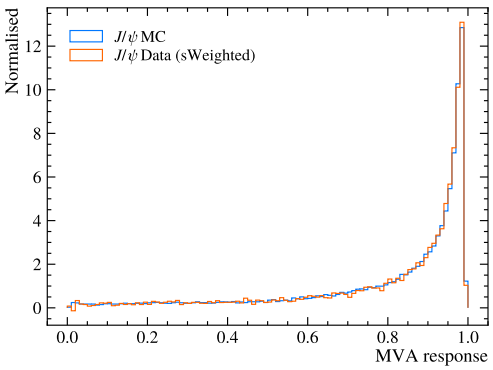
(b) R1-LL



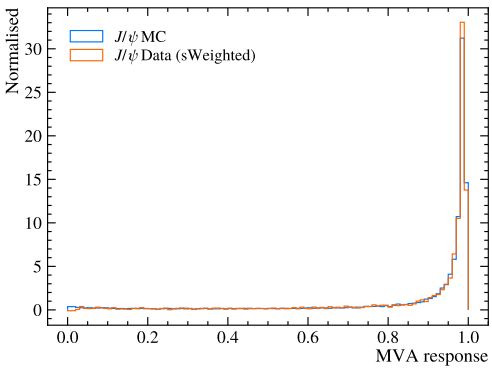
(c) R2p1-DD



(d) R2p1-LL



(e) R2p2-DD



(f) R2p2-LL

Figure 8.6: The comparison of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ channel MVA responses between the simulation and data (*sPlot* technique applied).

8.3 Working point optimisation

In order to optimise the sensitivity of the measurement, the working points of the MVA classifiers, *i.e.* the selection points based on the MVA output variables, need to be optimised. Such an optimisation is usually conducted by evaluating a certain Figure-of-Merit (FoM) at different selection points. In this analysis, two types of FoM are used, *i.e.* the signal significance and the Punzi FoM [118]. The signal significance is defined as

$$\mathrm{FoM}_{\mathrm{Sig}} = \frac{S}{\sqrt{S+B}} \ , \tag{8.1}$$

where $S(B)$ stands for the expected signal(background) yields after the MVA selection, aims to optimise for the statistical uncertainty of the signal yields assuming certain amplitudes of signal and background before the MVA cut. It can be calculated as

$$\mathrm{FoM}_{\mathrm{Sig}} = \frac{S_0\epsilon_{\mathrm{sig}}}{S_0\epsilon_{\mathrm{sig}} + B_0\epsilon_{\mathrm{bkg}}} \ , \tag{8.2}$$

where $S_0(B_0)$ stands for the expected signal(background) yields before the MVA selection and the $\epsilon_{\mathrm{sig}}(\epsilon_{\mathrm{bkg}})$ stands for the signal(background) efficiency of the MVA cut. The Punzi FoM [118], assuming the background-only hypothesis as the null hypothesis, is defined as

$$\mathrm{FoM}_{\mathrm{Punzi}} = \frac{\epsilon_{\mathrm{sig}}}{a/2 + \sqrt{B}} \ . \tag{8.3}$$

It aims to optimise, for the signal, the sensitive limit corresponding to a discovery claim with a times the standard deviation. It can be calculated as

$$\mathrm{FoM}_{\mathrm{Punzi}} = \frac{\epsilon_{\mathrm{sig}}}{a/2 + \sqrt{B_0\epsilon_{\mathrm{bkg}}}} \ . \tag{8.4}$$

In practice, in this analysis, the S_0 and B_0 are estimated for a certain signal region, which is an interval of $m_{\mathrm{DTF}}(\Lambda_b^0)$ (or $m^{J/\psi}_{\mathrm{DTF}}(\Lambda_b^0)$, $m^{\psi(2S)}_{\mathrm{DTF}}(\Lambda_b^0)$) largely overlapping with the expected signal peak (known invariant mass of $\Lambda_b^0 \sim 5620\,\mathrm{MeV}/c^2$ [64]):

- Electron channel rare(low,central,high) q^2 region: $5300 < m_{\mathrm{DTF}}(\Lambda_b^0) < 5700\,\mathrm{MeV}/c^2$.
- Electron channel J/ψ q^2 region: $5570 < m^{J/\psi}_{\mathrm{DTF}}(\Lambda_b^0) < 5670\,\mathrm{MeV}/c^2$.
- Electron channel $\psi(2S)$ q^2 region: $5570 < m^{\psi(2S)}_{\mathrm{DTF}}(\Lambda_b^0) < 5670\,\mathrm{MeV}/c^2$.

- Muon channel rare(low,central,high) q^2 region: $5570 < m_{\mathrm{DTF}}(\Lambda_b^0) < 5670\,\mathrm{MeV}/c^2$.
- Muon channel J/ψ q^2 region: $5595 < m^{J/\psi}_{\mathrm{DTF}}(\Lambda_b^0) < 5645\,\mathrm{MeV}/c^2$.
- Muon channel $\psi(2S)$ q^2 region: $5595 < m^{\psi(2S)}_{\mathrm{DTF}}(\Lambda_b^0) < 5645\,\mathrm{MeV}/c^2$.

The differences in range widths are due to the different mass resolutions between the electron and the muon channel as well as the DTF reconstructed mass with and without the dilepton mass constraints.

The efficiencies ϵ_{sig} and ϵ_{bkg} are evaluated from the simulated signal and the data upper sidebands respectively.

Although the R2p1 and R2p2 are merged during the training (considering the limited statistics in the training sample), the optimisation is done for them separately. The expected relative ratios between the signal and the background are quite different in the different q^2 regions, considering the q^2 dependence of the previously measured $\mathcal{B}$ $(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)$ [84] and the relatively large branching fraction of the resonance channels. Thus, the MVA working points are optimised in the different q^2 regions separately.

- For the resonance regions, due to the considerable branching fractions of $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$, the signal extraction, *i.e.* the fit of the mass spectrum, can even already be performed before the MVA selection. The signal significance FoM is chosen for them to optimise for their statistical uncertainties after the MVA selection. Their input yields S_0 and B_0 are taken from their fits prior to the MVA.
- In the high q^2 region, the goal is to measure R_Λ^{-1}, for which the signal significance FoM is chosen in order to optimise for the statistical uncertainty on R_Λ^{-1}. The measurement in Ref. [84] is used to estimate the muon channel signal yields S_0 while the electron channel signal yields are estimated with the LFU assumption. In practice, this is done by scaling the signal yields of the J/ψ mode with the efficiency factors calculated from the simulation. The background yield B_0 is extrapolated from a reference region in the data upper sideband. As discussed in Section 6.6, the mass spectrum of Λ_b^0 is sculpted in the high q^2 region and the conventional exponential shape is not suitable to describe the combinatorial background shape. Thus, the Same Sign (SS) data sample is used as a proxy for the shape of the combinatorial background in the normal Opposite Sign (OS) data. The background yield B_0 is

extracted by

$$B_0(\text{signal region}) = \frac{N_{\text{OS}}(\text{reference region})}{N_{\text{SS}}(\text{reference region})} \times N_{\text{SS}}(\text{signal region}) \ , \tag{8.5}$$

where $N_{\text{OS}}(N_{\text{SS}})$ stands for the number of candidates in the OS(SS) data samples in the corresponding regions. This procedure is illustrated in Figure 8.7.

- In the low and central q^2 regions, considering the relatively low branching fraction of $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$, the Punzi FoM [118] is chosen to optimise for the chance of the discovery of $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ with 3 standard deviations ($a = 3$ in Equation 8.3). Since the conventional exponential shape for the combinatorial background is applicable for the low and central q^2 regions, the background yield B_0 is estimated by fitting a reference region in the OS upper sideband with an exponential shape and then extrapolating to the signal region.

Following the above procedure, the FoM can be calculated at different MVA cut points, which gives the FoM curves. In order to mitigate the statistical fluctuation, the FoM curves are smoothed by averaging over ± 2 neighouring points (limited to the first or the last scanned point). Figure 8.8 and Figure 8.9 shows examples of the optimisation curves, *i.e.* the signal efficiency, background efficiency and the FoM at various MVA cut points. The optimised working points obtained for the different sample categories and the different q^2 regions are summarised in Table 8.3. While the exact working points vary in different cases, there is a tendency that the optimised MVA selection is looser in the resonance and the muon high q^2 region. Intuitively and qualitatively, this can be understood as that, when the signal is more prominent, the MVA selection should be looser to avoid losing too much signal.

Table 8.3: A summary of the working points in the different sample categories and the different q^2 regions.

q^2 regions	high	central	low	J/ψ	$\psi(2S)$
Electron-R1-LL	0.91	0.88	0.92	0.44	0.66
Electron-R1-DD	0.67	0.90	0.73	0.36	0.29
Electron-R2p1-LL	0.95	0.96	0.88	0.41	0.61
Electron-R2p1-DD	0.85	0.92	0.91	0.33	0.42
Electron-R2p2-LL	0.87	0.96	0.96	0.32	0.41
Electron-R2p2-DD	0.87	0.93	0.87	0.24	0.50
Muon-R1-LL	0.66	0.91	0.87	0.12	0.08
Muon-R1-DD	0.61	0.80	0.81	0.15	0.15
Muon-R2p1-LL	0.71	0.94	0.88	0.20	0.12
Muon-R2p1-DD	0.62	0.91	0.82	0.18	0.20
Muon-R2p2-LL	0.76	0.95	0.95	0.14	0.12
Muon-R2p2-DD	0.67	0.94	0.90	0.16	0.15

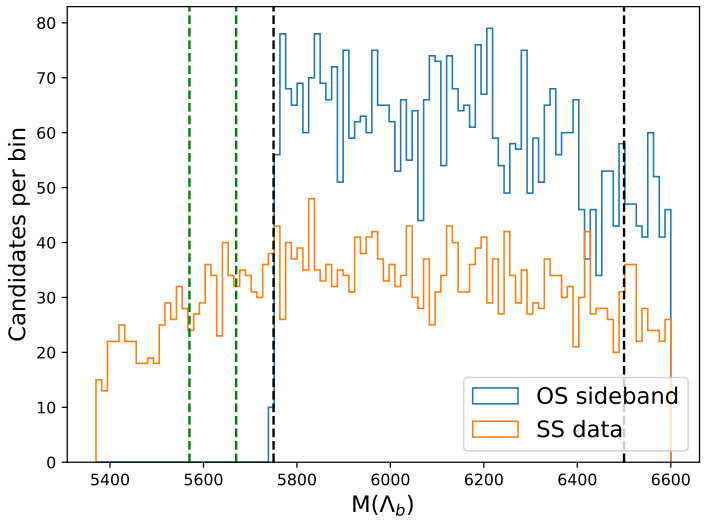
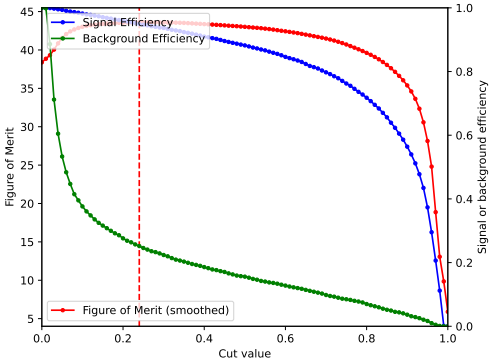
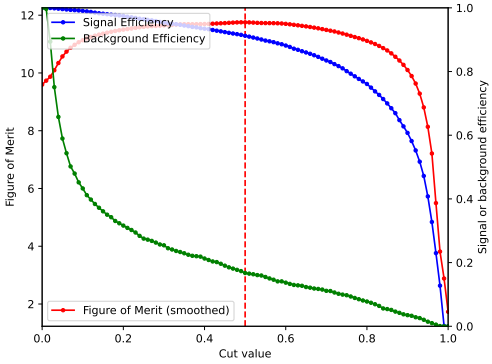


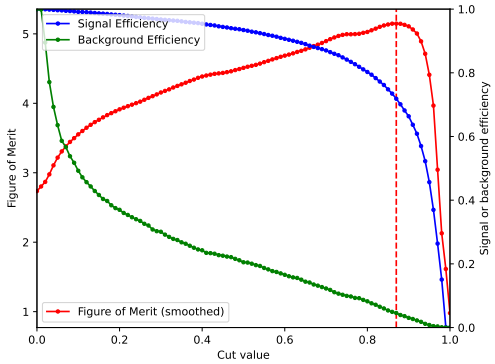
Figure 8.7: An illustration of the procedure to use the SS data as a proxy for the OS combinatorial background in the high q^2 region. The signal region and the reference region (as in Equation 8.5) are indicated with the green and the black dashed lines respectively. The samples shown here belong to the muon-R2p2-DD category.



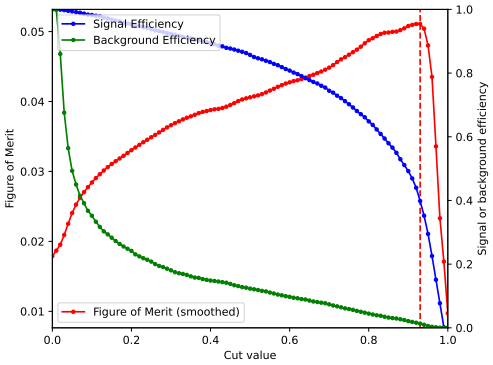
(a) J/ψ region



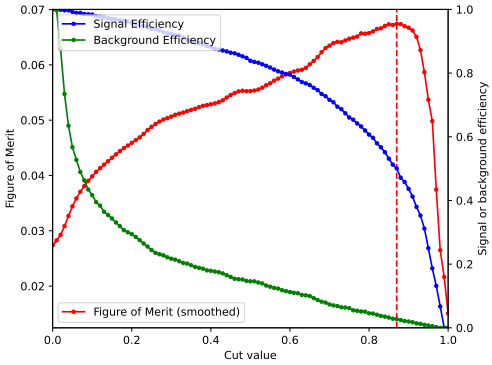
(b) $\psi(2S)$ region



(c) High q^2 region

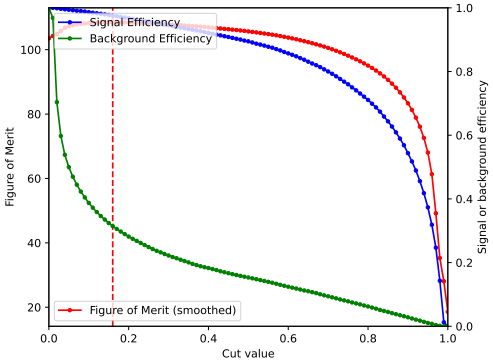


(d) Central q^2 region

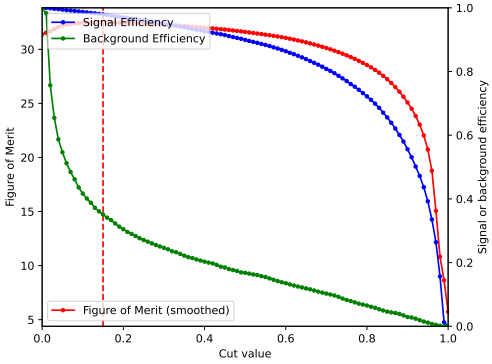


(e) Low q^2 region

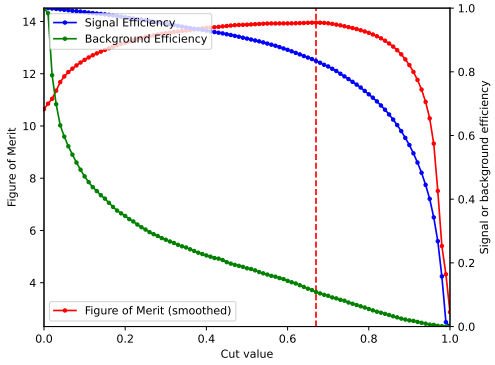
Figure 8.8: Examples of the MVA optimisation curves. The plots correspond to the electron-R2p2-DD category. The FoM curves are smoothed by averaging over several neighouring points. The red dashed lines indicate the chosen working points.



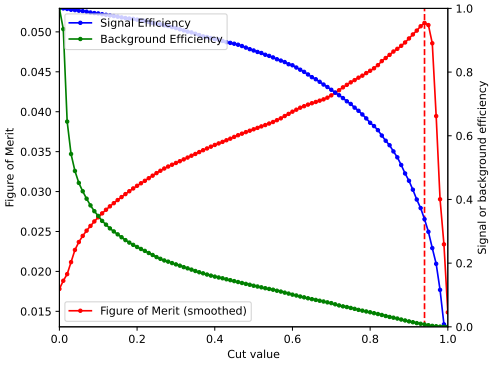
(a) J/ψ region



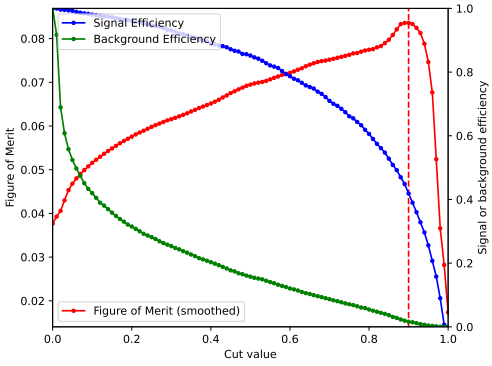
(b) $\psi(2S)$ region



(c) High q^2 region



(d) Central q^2 region



(e) Low q^2 region

Figure 8.9: Examples of the MVA optimisation curves. The plots correspond to the muon-R2p2-DD category. The FoM curves are smoothed by averaging over several neighouring points. The red dashed lines indicate the chosen working points.

9 Ratio observable determination

As discussed in Chapter 6, the determination of the observables R_{Λ}^{-1}, $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ depends on the evaluation of signal yields and the efficiencies. In practice, the Λ_b^0 mass spectra of different channels are fitted simultanouesly with the observables of interest included as fit parameters. These observables are evaluated from the simultaneous fits with the efficiencies and, when applicable, other variables used as nuisances parameters.

Section 9.1 discusses several related mathematical techniques, *i.e.* a technique involved in the estimation of the efficiency uncertainty (Section 9.1.1), the fitting techniques (Section 9.1.2) and the so-called CL_s method (Section 9.1.3). The evaluation of the efficiencies, which provides the information needed in order to constrain them as nuisance parameters during the fit, is described in Section 9.2. The fitting of the resonance regions to determine $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ is described in Section 9.3. With the signal region for the rare electron channel blinded, Section 9.4 discusses the method for the fits in the rare q^2 regions, *i.e.* the high, central and low q^2 regions.

9.1 Mathematical techniques

9.1.1 Uncertainty estimation for efficiencies

As to be discussed in Section 9.2, weighted MC samples are used in the procedure of efficiency estimation. A generalised form of efficiency calculation using two samples (labelled as A and B, which can have overlapping) can be expressed as

$$z = \frac{\sum_{i\in A} w_i^A}{\sum_{j\in B} w_j^B} \ , \tag{9.1}$$

where z stands for the generalised efficiency [26] and $w_{i(j)}^{A(B)}$ stands for the weight of entry $i(j)$ in the sample $A(B)$.

For the variable z, the limited statistics in the samples contributes to its uncertainty. In principle, the weights w_i^A and w_j^B can also have uncertainties. In this section, only the uncertainty induced by the sample statistics is discussed while the weights w_i^A and w_j^B are assumed to be accurate. The method for estimating such an uncertainty is explained as follows.

Given a group of independent random variables (denoted as $X_1, X_2, \cdots, X_n$) following Poisson distributions of mean values $\lambda_1, \lambda_2, \cdots, \lambda_n$:

$$X_i \sim \text{Poisson}(\lambda_i) \ , \tag{9.2}$$

[26] Usually, for a conventional selection efficiency, the A and B samples have the relationship $A \subset B$.

the sum of them would follow the Poisson distribution of mean value $\sum_{i=1}^{n} \lambda_i$:

$$\sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right) \ . \tag{9.3}$$

In this discussion, it is sufficient to limit $\lambda_i = 1$. When considering the statistical fluctuation associated with a given sample of n entries, its total number of entries can be considered as a random variable following a Poisson distribution of mean value n. Based on Equation 9.3, it can be equivalently considered as a sum of n independent random variables, each of which follows a Poisson distribution of mean value 1 (Equation 9.3 when $\lambda_i = 1$). When the sample is weighted, the sum of weights (denoted as y) is expressed as

$$y = \sum_{i=1}^{n} w_i \ , \tag{9.4}$$

where w_i stands for the weight for the entry i in the sample. A corresponding random variable (denoted as Y) can be constructed as a linear combination of n independent Poisson distributed random variables:

$$Y = \sum_{i=1}^{n} w_i X_i \ , \tag{9.5}$$

where $X_i \sim \text{Poisson}(1)$ and w_i is the weight for entry i. The uncertainty of y can be estimated as $\sqrt{\text{Var}(Y)}$, where $\text{Var}(Y)$ stands for the variance of Y. The variance of Y can be calculated as:

$$\text{Var}(Y) = \sum_{i=1}^{n} w_i^2 \text{Var}(X_i) = \sum_{i=1}^{n} w_i^2 \ . \tag{9.6}$$

This corresponds to the well-known statistical uncertainty of a weighted sum.

For the variable z as defined in Equation 9.1, a corresponding random variable Z can be constructed:

$$Z = \frac{\sum_{i \in A} w_i^A X_i}{\sum_{j \in B} w_j^B X_j} \ . \tag{9.7}$$

The uncertainty of z can be estimated as $\sqrt{\text{Var}(Z)}$, where $\text{Var}(Z)$ stands for the variance of Z. The variance of Z can be estimated via the Toy MC method, where the random variables X_i are sampled from their corresponding Poisson distributions.

In the degenerated case where $A \subset B$ and $w_i^A = w_j^B = 1$, using Taylor expansion, the variance of Z can be estimated as:

$$\begin{aligned}\mathrm{Var}(Z) &\approx \sum_{i\in A\cap B}\left(\frac{1}{n_B}-\frac{n_A}{n_B^2}\right)^2\mathrm{Var}(X_i)+\sum_{j\in B, j\notin A}\left(\frac{n_A}{n_B^2}\right)^2\mathrm{Var}(X_j)\\ &= \frac{(n_B-n_A)^2}{n_B^4}n_A+\frac{n_A^2}{n_B^4}(n_B-n_A)\\ &= \frac{1}{n_B}\times\frac{n_A}{n_B}\left(1-\frac{n_A}{n_B}\right)=\frac{1}{n_B}z(1-z)\ ,\end{aligned} \tag{9.8}$$

where $n_{A(B)}$ stands for the number of entries in $A(B)$. This corresponds to the well-known binomial error of a conventional selection efficiency.

9.1.2 Fitting techniques

The real data is a mixture of different components, *i.e.* the signal and various backgrounds. Each component (labelled with c) is characterised by its yield (denoted as N_c) and shape in the reconstructed Λ_b^0 mass spectrum, which is denoted as the Probability Density Function (PDF) $f_c(m)$. In principle [27], the information on N_c and f_c can be extracted via maximum likelihood fits. The combined PDF (denoted as f_{total}) of these components can be expressed as

$$f_{\mathrm{total}}(m)=\frac{\sum_c N_c f_c(m)}{\sum_c N_c}\ . \tag{9.9}$$

For a certain dataset, which consists of a set of reconstructed Λ_b^0 mass entries $(m_1,\cdots,m_n)$, a likelihood term can be constructed as

$$\mathcal{L}_{\mathrm{PDF}}\propto\prod_{i=1}^{n}f_{\mathrm{total}}(m_i)=\prod_{i=1}^{n}\frac{\sum_c N_c f_c(m_i)}{\sum_c N_c}\ . \tag{9.10}$$

This is the likelihood of acquiring n independent observations $(m_1,\cdots,m_n)$ from the random variable which follows the distribution with PDF f_{total}. Such a likelihood, however, does not allow the determination of the absolute scale of N_c, since $\mathcal{L}_{\mathrm{PDF}}$ remains unchanged when all component yields N_c are scaled by a common factor. To determine the absolute scale of N_c, an additional Poissonian likelihood term can be constructed as

$$\mathcal{L}_{\mathrm{Poisson}}\propto\frac{(\sum_c N_c)^n}{n!}e^{-\sum_c N_c}\ , \tag{9.11}$$

[27] Assuming: The data have sufficient statistics. The number of components is not too large. There are not too many free parameters in the PDFs.

which corresponds to the probability of acquiring the sampled value n from a random variable following the Poisson distribution of mean value $\sum_c N_c$. Combining the the two likelihood terms, the total likelihood ($\mathcal{L}$) can be constructed as

$$\begin{aligned}\mathcal{L} &\propto \frac{(\sum_c N_c)^n}{n!} e^{-\sum_c N_c} \prod_{i=1}^{n} \frac{\sum_c N_c f_c(m_i)}{\sum_c N_c} \\ &\propto e^{-\sum_c N_c} \prod_{i=1}^{n} \sum_c N_c f_c(m_i) \ .\end{aligned} \tag{9.12}$$

In the standalone fit, *i.e.* using a single dataset instead of fitting several datasets simultaneously, one can maximise such a likelihood to determine yields N_c or certain shape parameters contained in f_c. In the case of only one component, the extended Poissonian term can be dropped and the likelihood is simply

$$\mathcal{L} \propto \prod_{i=1}^{n} f_c(m_i) \ , \tag{9.13}$$

which can be used to determine the shape parameters contained in the PDF f_c. This case is useful for extracting shapes from the simulation samples.

In the simultaneous fit of several datasets, the likelihood of several datasets (*e.g.* several decay channels) can be combined by taking the sum of the logarithmic likelihoods and then the combined likelihood can be maximised to determine parameters. If the yields are parameterised with the ratio observables, then the simultaneous fit gives directly the results for the observables. If some nuisance parameters contained in the likelihood are constrained with additional estimations (values and uncertainties), then the likelihood can be further extended by Gaussian terms based on these additional estimations. In this way, the information of the additional estimations is incorporated into the measurement of the observables with the uncertainties propagated.

Technically, the implementation of the fits utilises the software package RooFit [119], which provides many supports, *e.g.* the implementations of many fit models, the functionality to construct likelihoods and the interface for maximising the likelihood.

9.1.3 The CL_s method

In the low and central q^2 regions of the electron channel, only a very small amount of signal is expected when assuming LFU. In the case of no signal observation, the so-called CL_s method [120] can be used to estimate an upper limit on the relative branching fraction between $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$.

Based on Refs. [120,121], the basic idea of CL_s method is briefly explained as follows.

When given a set of observed data, certain criteria are needed to gauge the (relative) compatibility of various hypotheses against the said data. For this purpose, one can use a so-called test statistic, which is a random variable following different distributions under different hypotheses. In this analysis, the test statistic

$$q_\mu = \begin{cases} -2\ln(L(\mu,\hat{\hat{\theta}})/L(\hat{\mu},\hat{\theta})) \ , \ \hat{\mu} \leq \mu \\ 0 \ , \ \hat{\mu} > \mu \end{cases} \tag{9.14}$$

is used, where μ stands for the parameter of interest corresponding to the hypothesis under test, θ stands for the nuisance parameters and L stands for the likelihood under different conditions. The likelihood $L(\mu,\hat{\hat{\theta}})$ is evaluated under the conditional fit with the parameter of interest fixed to the value μ based on the hypothesis. The nuisance parameters determined in this conditional fit are denoted as $\hat{\hat{\theta}}$. The likelihood $L(\hat{\mu},\hat{\theta})$ is evaluated with both the parameter of interest and the nuisance parameters allowed to be determined from the fit. The determined parameters in this case are denoted as $\hat{\mu}$ and $\hat{\theta}$.

Given the definition of the test statistic, its distribution under the signal-plus-background $(s+b)$ hypothesis and the background-only (b) hypothesis can be estimated. The signal s amplitude in the $s+b$ hypothesis corresponds to the value of μ (the hypothesis under test). The CL_s value for the given observed data can then be defined as

$$CL_s = \frac{1 - \text{c.d.f}(q_\mu^{obs}|s+b)}{1 - \text{c.d.f}(q_\mu^{obs}|b)} \ , \tag{9.15}$$

where c.d.f stands for the cumulative distribution function [28] under the hypothesis $s+b$ (or b) and the q_μ^{obs} stands for the q_μ evaluated with the observed data. By scanning the μ value, one can evaluate the CL_s-μ relationship. The upper limit of μ corresponding to a certain confidence level (C.L.) can be defined by the $CL_s >$ C.L. requirement.

Another concept is the expected CL_s curve for the background-only hypothesis. It is especially important in the absence of the real observed data, such as in the case of blinded data. Such a curve can be determined by replacing q_μ^{obs} in Equation 9.15 with the median value of q_μ under the background-only hypothesis. Similarly, the curve for the median plus or minus several standard deviation (σ) can also be determined.

[28] For a given random variable X, its c.d.f(x) (x being a value) equals the probability of $X \leq x$.

9.2 Efficiency

The efficiencies are calculated from two fractions:

$$\epsilon_{\text{tot}} = \epsilon_{\text{acc}} \times \epsilon_{\text{reco\&sel}} \ , \tag{9.16}$$

where ϵ_{acc} stands for the detector geometric acceptance while $\epsilon_{\text{reco\&sel}}$ stands for the reconstruction and the selection efficiency.

Based on the simulation, the acceptance ϵ_{acc} can be estimated by counting the number of signal candidates before and after applying the detector geometric requirement [29]. Such a geometric requirement can already be applied at the event generation stage before the full detector simulation.

The efficiency $\epsilon_{\text{reco\&sel}}$ is evaluated using the simulation samples with the corrections described in Section 7.3 applied. The MC samples at two simulation stages are used for the calculation. The first MC sample consists of the generated truth level candidates passing the geometric acceptance. The second MC sample consists of the reconstructed candidates passing the nominal selection of this analysis.

$$\epsilon_{\text{reco\&sel}} = \frac{\sum_{i\in\text{reco\&sel}} w_i^{\tau} w_i^{\text{angle}} w_i^{\text{track}} w_i^{\text{trig}} w_i^{\text{PID}} w_i^{\text{kin\&multi}}}{\sum_{j\in\text{gen}} w_j^{\tau} w_j^{\text{angle}} w_j^{\text{kin\&multi}}} \ , \tag{9.17}$$

where the numerator is a sum of weights over the reconstructed and selected MC samples and the denominator is a sum of weights over the generated MC samples (after acceptance but before reconstruction). The weights (as discussed in Section 7.3) are listed as follows.

- w_i^{τ} : the Λ_b^0 lifetime weight.
- w_i^{angle} : the angular weights of the resonance simulation (only applicable for $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0 \psi(2S)(\ell^+\ell^-)$).
- w_i^{track} : the tracking efficiency corrections (for the electrons).
- w_i^{trig} : the trigger efficiency corrections.
- w_i^{PID} : the PID calibration weights.
- $w_i^{\text{kin\&multi}}$: the corrections for the Λ_b^0 kinematics and the track multiplicity.

[29] In practice, it is reported as a part of the centralised simulation production procedure.

Special attention needs to be drawn to the q^2 regions. For the reconstructed and selected MC samples (numerator), the q^2 requirement is part of the selection. The reconstructed q^2 is selected according to the q^2 region under study. For the generated truth level MC samples (denominator), the q^2 requirement concerns the theoretical interpretation.

In the case of the resonance signal efficiency, it is the total branching fraction that is of interest. Thus, only the reconstructed q^2 (numerator) are selected based on the q^2 region division while no q^2 requirement is applied on the truth level (denominator). An efficiency calculated in this way can be interpretated as the efficiency for a generated signal candidate in any q^2 region to be reconstructed in the q^2 region under study.

In the case of the rare signal efficiency, the observables of interest (*e.g.* R_Λ^{-1}), are defined in a certain q^2 region. Thus, q^2 requirements are needed on both the numerator and the denominator samples. An efficiency calculated in this way can be interpretated as the efficiency for a generated signal candidate in the q^2 region corresponding to the definition of the observable to be reconstructed in the q^2 region under study. To match the constructed observables to their intended theoretical interpretations, the pre-Final-State-Radiation (preFSR) truth level q^2 variable is used, which is calculated using the four-momentum difference between the Λ_b^0 and the Λ^0 particles.

For the efficiencies evaluated with Equation 9.17, the limited statistics of the simulation samples contributes to their uncertainties. Following the discussion in Section 9.1.1, the uncertainties on $\epsilon_{\text{reco\&sel}}$ induced by the limited statistics of the MC samples can be estimated via a toy MC method. Such toy MC samples can be generated by assigning a Poissonian sampling weight (w^{Pois}), which is drawn from the Poisson distribution of mean value 1, to each decay candidate:

$$\epsilon_{\text{reco\&sel}}(w^{\text{Pois}}) = \frac{\sum_{i\in\text{reco\&sel}} w_i^\tau w_i^{\text{angle}} w_i^{\text{track}} w_i^{\text{trig}} w_i^{\text{PID}} w_i^{\text{kin\&multi}} w_i^{\text{Pois}}}{\sum_{j\in\text{gen}} w_j^\tau w_j^{\text{angle}} w_j^{\text{kin\&multi}} w_j^{\text{Pois}}} \,. \quad (9.18)$$

Since the reconstructed samples are reconstructed using the generated samples as the input, the numerator and the denominator are not completely independent in the above equation. Due to this consideration, the Poissonian weights are matched between the two samples in such a fashion that the same event appearing in the numerator and in the denominator are assigned the same Poissonian weight. The sample mean and the sample standard deviation of the toy MC sample can be taken as the central value and the error for the efficiency.

As discussed in Section 6.5, several important steps (*e.g.* the $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ determination) of the analysis are conducted in the three data-taking

Table 9.1: Efficiencies of $\Lambda_b^0 \to \Lambda^0 e^+ e^-$. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	High q^2	Central q^2	Low q^2
R1 LL	1.198 ± 0.032	0.505 ± 0.021	0.39 ± 0.031
R1 DD	2.127 ± 0.039	0.453 ± 0.019	0.528 ± 0.037
R2p1 LL	1.967 ± 0.04	0.831 ± 0.031	0.769 ± 0.045
R2p1 DD	3.357 ± 0.049	0.94 ± 0.031	0.732 ± 0.048
R2p2 LL	2.203 ± 0.028	0.764 ± 0.019	0.632 ± 0.03
R2p2 DD	3.093 ± 0.032	0.83 ± 0.02	0.844 ± 0.035

periods (R1,R2p1,R2p2) and two $\Lambda^0 \to p\pi^-$ track type categories. The corresponding efficiencies are required in these categories. In practice, the simulation samples were provided in different years and magnet polarity configurations. To be more compatible with the data, the efficiencies are evaluated in different years and magnet polarity configurations and then merged using weights [30], which are based on the integrated luminosity of the corresponding data samples.

The fits (which will be discussed later) used to determine the observables are conducted within a certain reconstructed Λ_b^0 mass window. For the efficiencies used as nuisance parameters in the fit, the mass window requirements used for the fits should be considered. This mainly concerns the backgrounds such as and the partially reconstructed backgrounds and the background caused by misreconstructing q^2 (leakage background, as discussed in Section 6.6). The reconstructed Λ_b^0 mass distributions of these backgrounds can have considerable displacements from the nominal mass window center. The Λ_b^0 mass windows used in the fit described in Section 9.3, Section 9.4.2 and Section 9.4.3 are referred to as the nominal mass windows.

Following the method described above, the evaluated total efficiencies for the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal decay are shown in Table 9.1. Comparing with the high q^2 region, the efficiencies in the low and the central q^2 regions are significantly lower. A significant contributor to this difference is the L0 trigger strategy (as discussed in Section 7.2), which favours the energetic electrons by requiring a certain E_T threshold. In addition, the electron track reconstruction efficiency, in general, has an increasing trend towards higher p_T regions [109].

The evaluated total efficiencies for the decays $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0 \psi(2S)(\ell^+\ell^-)$ are shown in Appendix D. The efficiencies in the electron channel are significantly lower than their counterparts in the muon

[30] Provided by collaborators.

channel. This is expected considering the lower electron reconstruction efficiency [109] and L0 trigger efficiency [21]. As discussed in Section 6.6, due to the very limited q^2 resolution of the dielectrons, they can be reconstructed in a q^2 region other than their origin q^2 region (leakage background). The efficiencies of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ leaking into the central and the J/ψ q^2 regions as well as the efficiencies of $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ leaking into the J/ψ and the high q^2 regions are also shown in Appendix D. As mentioned in Section 5.3, there is an ongoing effort within LHCb to measure $\mathcal{B}(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)$ in various q^2 bins. In order to prevent indirectly showing $\mathcal{B}(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)$, the efficiencies of $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ are not given in this thesis.

The evaluated total efficiencies for several partially reconstructed background decays ($\Xi_b^- \to \Xi^- J/\psi$, $\Xi_b^- \to \Xi^-\psi(2S)$ and $\Lambda_b^0 \to \Lambda(1520)J/\psi$) are also shown in Appendix D. These background components are used in the fit described in Section 9.3. For these Ξ_b^- backgrounds, the Λ_b^0 lifetime correction, the angular correction and the kinematic(Λ_b^0)-multiplicity correction are not used, since they are not applicable. For $\Lambda_b^0 \to \Lambda(1520)J/\psi(\ell^+\ell^-)$, the angular correction is not applicable and the kinematic(Λ_b^0)-multiplicity correction is not used considering the Λ_b^0 kinematic variables are not correctly reconstructed due to missing decay products. These efficiencies can in principle be affected without the full simulation corrections. However, as will be discussed in Section 9.3, these efficiencies are used to constrain backgrounds in the fits and thus their relationship with the observable $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ is indirect. In addition, they are used in such a way (Section 9.3) that a shared scaling factor between the electron and muon channels can be absorbed in free fit parameters.

9.3 Fits in the resonance q^2 regions

As discussed in Chapter 6, the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ ($\ell = e, \mu$) decay channels are used as normalisation channels to better control the systematic uncertainties associated with the efficiencies. Together with the $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ channels, they also serve to construct the crosscheck observables $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$. The measurements of $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ can be compared with their expected value (1.0), which serve as powerful crosschecks for the analysis. In this section, the procedure to determine $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ based on the fits and their comparison to the expected value (1.0) will be discussed.

To determine the single ratio $r_{J/\psi}^{-1}$ and the double ratio $R_{\psi(2S)}^{-1}$, the Λ_b^0 mass spectra of the four channels $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ ($\ell = e, \mu$) are fitted simultaneously. Taking advantage of the knowledge on the masses of J/ψ and $\psi(2S)$, the DTF [86] constrained mass $m_{\mathrm{DTF}}^{J/\psi}(\Lambda_b^0)$ and

$m^{\psi(2S)}_{\rm DTF}(\Lambda^0_b)$ are used (as discussed in Chapter 6). As discussed in Section 9.1.2, the yields are directly involved in the construction of the likelihoods. When some yields are parameterised with the ratio observables, the fitting procedure can directly give the ratio observable results. In order to determine the observables $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$, the electron channel signal yields are parameterised using the ratio observables, the muon channel signal yields and the efficiencies. Mathematically, they are expressed as follows:

$$\begin{aligned} &N_{\rm sig}(\Lambda^0_b \to \Lambda^0 J/\psi(e^+e^-)) = \\ &r^{-1}_{J/\psi} \times \frac{\epsilon_{\rm sig}(\Lambda^0_b \to \Lambda^0 J/\psi(e^+e^-))}{\epsilon_{\rm sig}(\Lambda^0_b \to \Lambda^0 J/\psi(\mu^+\mu^-))} N_{\rm sig}(\Lambda^0_b \to \Lambda^0 J/\psi(\mu^+\mu^-)) \end{aligned} \tag{9.19}$$

and

$$\begin{aligned} &N_{\rm sig}(\Lambda^0_b \to \Lambda^0 \psi(2S)(e^+e^-)) = R^{-1}_{\psi(2S)} r^{-1}_{J/\psi} \times \\ &\frac{\epsilon_{\rm sig}(\Lambda^0_b \to \Lambda^0 \psi(2S)(e^+e^-))}{\epsilon_{\rm sig}(\Lambda^0_b \to \Lambda^0 \psi(2S)(\mu^+\mu^-))} N_{\rm sig}(\Lambda^0_b \to \Lambda^0 \psi(2S)(\mu^+\mu^-)) \ . \end{aligned} \tag{9.20}$$

The yields of the muon channels and the ratio observables are allowed to be determined from the data fit, while the efficiencies serve as nuisance parameters, which can be constrained with Gaussian likelihood terms based on the evaluation discussed in Section 9.2.

The fitting of the resonances is performed for six separate data categories, *i.e.* three data taking period (R1, R2p1, R2p2) each with LL and DD $\Lambda^0 \to p\pi^-$ track types. Within each category and each channel, the data model is constructed as a combined shape of different components, *i.e.* the signal and various backgrounds.

In each of the four channels, the signal shapes are based on the MC samples but allowing a shift of mean and a scaling of the width to accommodate the possible discrepancy between data and simulation. Fits are performed on the signal MC samples and then the fitted shapes are adapted to incorporate additional parameters to shift the mean value and to scale the width. In order to account for the tails of the signal peak in the reconstructed Λ^0_b mass distribution, the signal shapes are modelled with PDFs which accommodate such structures. For the electron channels, the signal peaks are modelled with the Johnson S_U function [122], which is described by the following PDF

$$\begin{aligned} &f_{\rm Johnson}(m|\mu,\lambda,\gamma,\delta) = \\ &\frac{\delta}{\sqrt{2\pi}\lambda} \frac{1}{\sqrt{1+\left(\frac{m-\mu}{\lambda}\right)^2}} \exp\left(-\frac{1}{2}\left(\gamma + \delta \sinh^{-1}\left(\frac{m-\mu}{\lambda}\right)\right)^2\right) \ . \end{aligned} \tag{9.21}$$

For the signal peak shape of the muon channels, the Asymmetrical Double-Side-Crystal-Ball (AsymDSCB) function [123] is used, whose PDF takes the form

$$
\begin{aligned}
&f_{\text{AsymDSCB}}(m|m_0,\sigma_L,\sigma_R,\alpha_L,n_L,\alpha_R,n_R) = \\
&\text{if } \frac{m-m_0}{\sigma_L} < -\alpha_L : \left(\frac{n_L}{\alpha_L}\right)^{n_L} e^{-\frac{\alpha_L^2}{2}} \left(\frac{n_L}{\alpha_L} - \alpha_L + \frac{m-m_0}{\sigma_L}\right)^{-n_L} \\
&\text{if } -\alpha_L < \frac{m-m_0}{\sigma_L} < 0 : \exp\left(-\frac{1}{2}\left(\frac{m-m_0}{\sigma_L}\right)^2\right) \\
&\text{if } 0 < \frac{m-m_0}{\sigma_R} < \alpha_R : \exp\left(-\frac{1}{2}\left(\frac{m-m_0}{\sigma_R}\right)^2\right) \\
&\text{if } \alpha_R < \frac{m-m_0}{\sigma_R} : \left(\frac{n_R}{\alpha_R}\right)^{n_R} e^{-\frac{\alpha_R^2}{2}} \left(\frac{n_R}{\alpha_R} - \alpha_R + \frac{m-m_0}{\sigma_R}\right)^{-n_R} .
\end{aligned}
\tag{9.22}
$$

For both functions, their implementation in RooFit [119] is used.

Among the above PDF parameters, the tail parameters, *i.e.* the parameters γ, δ for f_{Johnson} and the parameters $\alpha_L, n_L, \alpha_R, n_R$ for f_{AsymDSCB} are fixed from the simulation. The center parameters, *i.e.* μ for f_{Johnson} and m_0 for f_{AsymDSCB} are modified for the data fit with additional shift parameters ($\Delta\mu$ or Δm_0):

$$
\mu^{\text{data}} = \mu^{\text{MC}} + \Delta\mu \ , \ m_0^{\text{data}} = m_0^{\text{MC}} + \Delta m_0 \ . \tag{9.23}
$$

Similarly, the width parameters are scaled, *i.e.*

$$
\lambda^{\text{data}} = \lambda^{\text{MC}} \cdot s_\lambda \ , \ \sigma_{L(R)}^{\text{data}} = \sigma_{L(R)}^{\text{MC}} \cdot s_\sigma \ . \tag{9.24}
$$

The shift $(\Delta\mu, \Delta m_0)$ and scaling parameters (s_λ, s_σ) are separate for the electron and the muon channels, but are shared between J/ψ and $\psi(2S)$ modes when they are determined from the data fit.

As examples, Figure 9.1 shows the plots for the fits of the resonance signal MC samples in the category of R2p2-DD.

The combinatorial background is fitted with an exponential shape, whose slope parameter and yield are both left free in the data fit. The slope parameters and the yields are different for different decay channels, data-taking periods and $\Lambda^0 \to p\pi^-$ track types.

As discussed in Section 6.6, the modelling of the partially reconstructed backgrounds for the J/ψ mode fit is inspired by Ref. [89]. The background component $\Lambda_b^0 \to \Lambda(1520)J/\psi$ (with $\Lambda(1520) \to \Sigma^0(\to \Lambda^0\gamma)\pi^0$) is chosen to be explicitly modelled based on the simulation. Its shape is fixed from the simulation using the kernel estimation technique [124], which is implemented

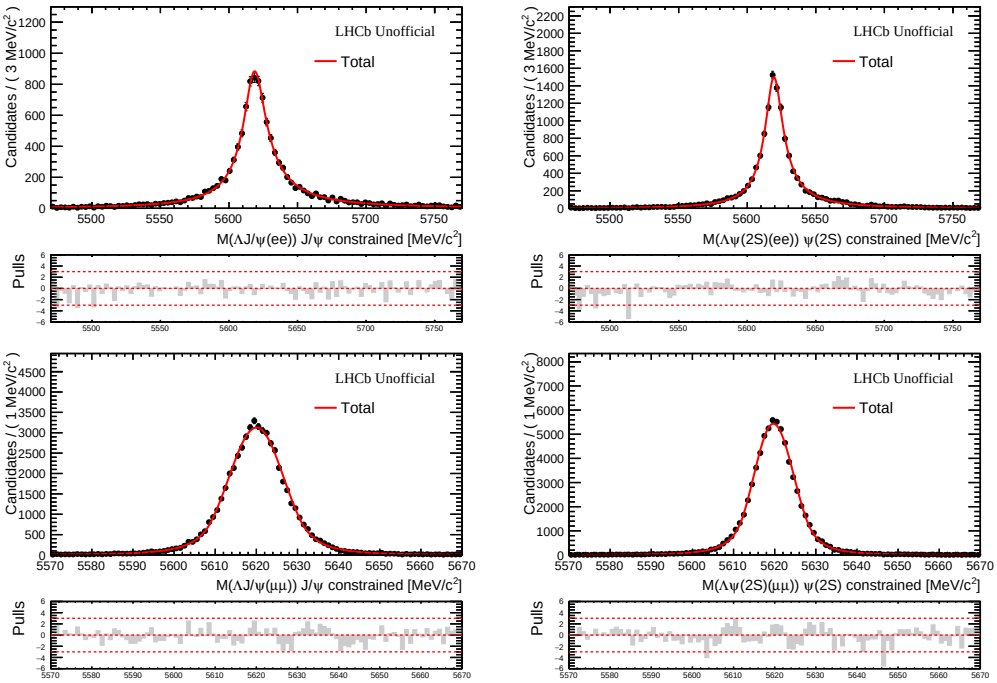


Figure 9.1: The plots for the fits of the resonance signal MC samples in the R2p2-DD category. The top left plot corresponds to the signal peak of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$. The top right plot corresponds to $\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$. The bottom left plot corresponds to $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$. The bottom right plot corresponds to $\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$.

in RooFit [119] as the `RooKeysPdf` class [125]. As discussed in Ref. [124], given a dataset drawn from a certain distribution, the kernel estimation is an unbinned and non-parametric method to provide an estimation of the said distribution. In the case of this analysis, it provides the benefit of avoiding the need to choose specific parametric models for several backgrounds. Figure 9.2 shows such a kernel estimation of the $\Lambda_b^0 \to \Lambda(1520)J/\psi(\ell^+\ell^-)$ shape based on the simulation in the R2p2-DD category. As shown in Figure 6.4 (mass fit from Ref. [89]), the contribution from $\Lambda^0(1600)$ and $\Lambda^0(1405)$ components are relatively small while largely overlapping in the shape comparing with the $\Lambda(1520)$ component. Thus, the $\Lambda^0(1600)$ and $\Lambda^0(1405)$ components are not included in the fit [31]. The contributions from several heavier Λ^* and the decay $\Lambda_b^0 \to \Lambda^0\psi(2S)(\to J/\psi\pi\pi)$ are modelled together using a Gaussian peak, whose mean, width as well as the amplitude (yield) parameters are left free in the data fit. For the $\psi(2S)$ mode, the mass window cannot be extended as low as the J/ψ mode considering the phase-space sculpting effect. A narrower mass window is used for the $\psi(2S)$ mode and thus excluding the corresponding Λ^* peaks.

Another possible source of the partially reconstructed backgrounds is the transition $\Xi_b \to \Xi J/\psi$ (or $\psi(2S)$). As shown in Ref. [89], the Ξ_b background, while relatively small in its amplitude (yield), has a considerably different shape than the Λ^* backgrounds. There are two types of $\Xi_b(\Xi)$, namely the charged $\Xi_b^-(\Xi^-)$ and the neutral $\Xi_b^0(\Xi^0)$. Both Ξ^- and Ξ^0 can decay into Λ^0, *i.e.* $\Xi^- \to \Lambda^0\pi^-$ and $\Xi^0 \to \Lambda^0\pi^0$. The mass differences between $\Xi_b^-(\Xi^-)$ and $\Xi_b^0(\Xi^0)$ are very small [64]:

$$\begin{aligned} m(\Xi_b^0) &= 5791.7 \pm 0.4\,\mathrm{MeV}/c^2\ , \\ m(\Xi_b^-) &= 5797.0 \pm 0.4\,\mathrm{MeV}/c^2\ , \\ m(\Xi^0) &= 1314.82 \pm 0.21\,\mathrm{MeV}/c^2\ , \\ m(\Xi^-) &= 1321.70 \pm 0.09\,\mathrm{MeV}/c^2\ . \end{aligned} \tag{9.25}$$

Considering the small mass differences, the partially reconstructed backgrounds originated from Ξ_b^- and Ξ_b^0 are expected to have similar shapes in the reconstructed Λ_b^0 mass distribution. Thus, the MC samples corresponding to the Ξ_b^- are used as proxies to model the shapes for the backgrounds originated from both Ξ_b^- and Ξ_b^0. It is modelled explicitly with the kernel estimation method [124] based on the MC samples. The modelling of the $\Xi_b^- \to \Xi^- J/\psi(\ell^+\ell^-)$ and the $\Xi_b^- \to \Xi^-\psi(2S)(\ell^+\ell^-)$ backgrounds in the R2p2-DD category is shown in Figure 9.3

[31] It is likely that a systematic study can be performed to check the impact of including these components in the fit. It is not performed at the moment because not all required simulation samples are available.

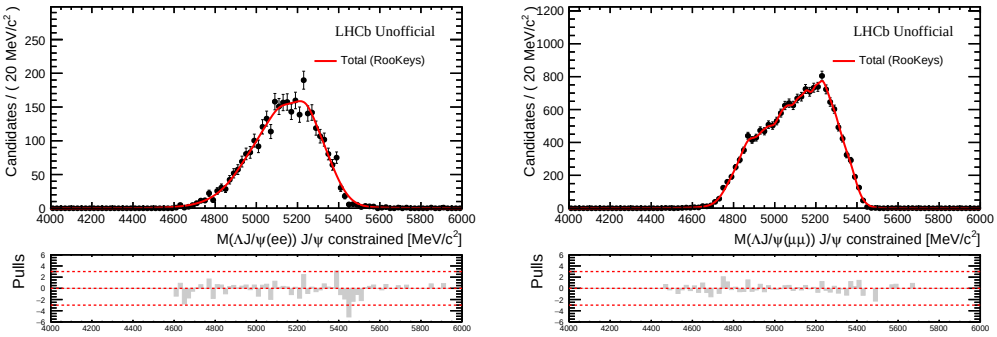


Figure 9.2: The kernel estimation with the $\Lambda_b^0 \to \Lambda(1520)J/\psi(\ell^+\ell^-)$ MC samples in the R2p2-DD category. The left plot corresponds to the electron channel while the right plot corresponds to the muon channel.

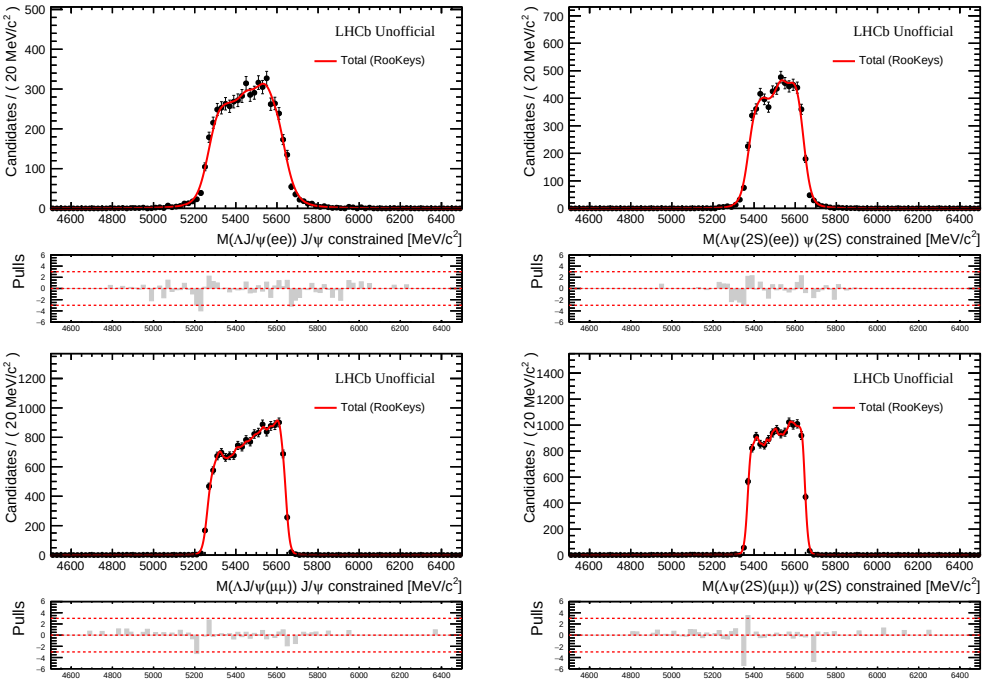


Figure 9.3: The kernel estimation with the $\Xi_b^- \to \Xi^- J/\psi(\ell^+\ell^-)$ and $\Xi_b^- \to \Xi^-\psi(2S)(\ell^+\ell^-)$ MC samples in the R2p2-DD category. The top left plot corresponds to the electron J/ψ mode, the top right plot corresponds to the electron $\psi(2S)$ mode, the bottom left plot corresponds to the muon J/ψ mode and the bottom right plot corresponds to the muon $\psi(2S)$ mode.

For the yields of the $\Lambda_b^0 \to \Lambda(1520)J/\psi(\ell^+\ell^-)$, $\Xi_b^- \to \Xi^- J/\psi(\ell^+\ell^-)$ and $\Xi_b^- \to \Xi^-\psi(2S)(\ell^+\ell^-)$ components, the relationship between the muon and the electron channels can be used to improve the fits. Instead of leaving them separately free to be determined in the data fits, their yields N_{partReco} can be parameterised using the signal yields N_{sig}, efficiencies $\epsilon_{\text{partReco}}$, ϵ_{sig} and a free fraction parameter a_{partReco} as

$$N_{\text{partReco}} = a_{\text{partReco}} \times \frac{\epsilon_{\text{partReco}}}{\epsilon_{\text{sig}}} N_{\text{sig}} \ . \tag{9.26}$$

For the $\Lambda_b^0 \to \Lambda(1520)J/\psi$, the fraction parameter a_{partReco} represents the relative branching fraction between the decay $\Lambda_b^0 \to \Lambda(1520)J/\psi$ (with the subsequent decay $\Lambda(1520) \to \Sigma^0(\to \Lambda^0\gamma)\pi^0$) and the decay $\Lambda_b^0 \to \Lambda^0 J/\psi$, which is expressed as:

$$\begin{aligned} &a_{\text{partReco}}(\Lambda_b^0 \to \Lambda(1520)J/\psi) = \\ &\frac{\mathcal{B}(\Lambda_b^0 \to \Lambda(1520)J/\psi \text{ with } \Lambda(1520) \to \Sigma^0(\to \Lambda^0\gamma)\pi^0)}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)} \ . \end{aligned} \tag{9.27}$$

The fraction parameter $a_{\text{partReco}}(\Lambda_b^0 \to \Lambda(1520)J/\psi)$ can be shared between the electron and muon channels. By doing this, no LFU of the resonance modes is assumed since the branching fractions of $J/\psi \to \ell^+\ell^-$ cancel between the signal and the partially reconstructed background channel. For the $\Xi_b^- \to \Xi^- J/\psi$ and $\Xi_b^- \to \Xi^-\psi(2S)$ components, the meaning of the corresponding a_{partReco} parameter is less straightforward. The shapes of the Ξ_b^- MC samples are used as proxies for the shapes of both the Ξ_b^- and the Ξ_b^0 backgrounds, thus the yields of the corresponding fit components reflect both contributions. In the following discussion, Ξ_b^- (charged) and Ξ_b^0 are used to denote their corresponding decays. The yield of the Ξ_b^- *fit component* (denoted as $N_{\Xi_b^-}$), is the sum of the yields of the Ξ_b^- and the Ξ_b^0 *decays* (denoted as $\tilde{N}_{\Xi_b^-}, \tilde{N}_{\Xi_b^0}$). The ratio between $N_{\Xi_b^-}$ and the signal yield N_{sig} can be expressed as follows:

$$\frac{N_{\Xi_b^-}}{N_{\text{sig}}} = \frac{\tilde{N}_{\Xi_b^-} + \tilde{N}_{\Xi_b^0}}{N_{\text{sig}}} = \frac{\epsilon_{\Xi_b^-}\mathcal{B}_{\Xi_b^-} f_{\Xi_b^-} + \epsilon_{\Xi_b^0}\mathcal{B}_{\Xi_b^0} f_{\Xi_b^0}}{\epsilon_{\text{sig}}\mathcal{B}_{\Lambda_b^0} f_{\Lambda_b^0}} \ , \tag{9.28}$$

where $\mathcal{B}$ stands for the branching fraction of the corresponding mode, ϵ stands for the efficiency of the corresponding mode and f stands for the production fraction of the corresponding mode. Equation 9.28 can be rewritten as:

$$\frac{N_{\Xi_b^-}}{N_{\text{sig}}} = \frac{\epsilon_{\Xi_b^-}}{\epsilon_{\text{sig}}} \frac{\mathcal{B}_{\Xi_b^-} f_{\Xi_b^-} + \frac{\epsilon_{\Xi_b^0}}{\epsilon_{\Xi_b^-}} f_{\Xi_b^0}\mathcal{B}_{\Xi_b^0}}{\mathcal{B}_{\Lambda_b^0} f_{\Lambda_b^0}} = \frac{\epsilon_{\Xi_b^-}}{\epsilon_{\text{sig}}} a_{\text{partReco}} \ . \tag{9.29}$$

Following the same argument as for the $\Lambda_b^0 \to \Lambda(1520)J/\psi$ component, the branching fractions of $J/\psi \to \ell^+\ell^-$ or $\psi(2S) \to \ell^+\ell^-$ (contained in the terms $\mathcal{B}_{\Xi_b^0}, \mathcal{B}_{\Xi_b^-}, \mathcal{B}_{\Lambda_b^0}$) cancel between the partially reconstructed backgrounds and the signal. Assuming the relative efficiencies $\epsilon_{\Xi_b^0}/\epsilon_{\Xi_b^-}$ to be the the same for the electron and the muon channels, the a_{partReco} parameters for the Ξ_b^- components (*i.e.* $\Xi_b^- \to \Xi^- J/\psi$ and $\Xi_b^- \to \Xi^-\psi(2S)$) can then also be shared between the muon the electron channels.

It should be noted that there is a large difference for the Ξ_b backgrounds between the LL and the DD categories. The known lifetime (denoted as τ) of the Ξ^- and Ξ^0 are as follows [64]:

$$\begin{aligned} \tau(\Xi^-) &= (1.639 \pm 0.015) \times 10^{-10}\text{s} \ , \\ \tau(\Xi^0) &= (2.90 \pm 0.09) \times 10^{-10}\text{s} \ . \end{aligned} \tag{9.30}$$

In comparison, the known lifetime of the Λ^0 is $(2.617 \pm 0.010) \times 10^{-10}$s [64]. Considering the displacement from the Ξ_b decay vertex to the Ξ decay vertex and the displacement from the Ξ decay vertex to the Λ^0 decay vertex, the Ξ_b backgrounds are expected to appear much more in the DD category than in the LL category. This is reflected in the efficiencies evaluated from the $\Xi_b^- \to \Xi^- J/\psi$ and the $\Xi_b^- \to \Xi^-\psi(2S)$ MC samples. The corresponding efficiencies are much higher in the DD category than in the LL category (Appendix D Table D.6 and Table D.5).

For the $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ channel, due to the limited resolution of the dielectron q^2, a certain amount of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ candidates can be present. This is an instance of the leakage backgrounds discussed in Section 6.6. The $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ background features a dielectron pair smeared by the resolution effect to a higher q^2 region. This effect will also be reflected on the reconstructed Λ_b^0 mass. As a result, the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ leakage background appears to be on the higher end of the reconstructed Λ_b^0 mass spectrum. The shape of the such a leakage background is extracted based on the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ simulation with the kernel estimation technique [124]. An example (R2p2-DD category) is shown in Figure 9.4. The yield of this leakage background can be parameterised with the efficiencies (denoted as ϵ) and the signal yield $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ in the J/ψ q^2 region as

$$\begin{aligned} &N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-) \text{ in } \psi(2S) \text{ region}) = \\ &\frac{\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-) \text{ in } \psi(2S) \text{ region})}{\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-) \text{ in } J/\psi \text{ region})} N_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)) \ . \end{aligned} \tag{9.31}$$

With the shapes and the yields of the different components configured as in the above discussion, the data samples of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and

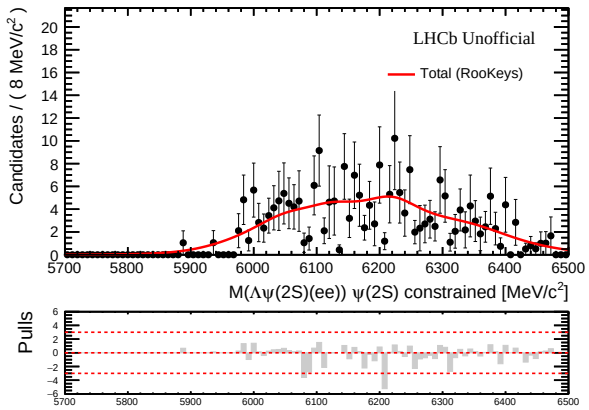


Figure 9.4: The kernel estimation for the leakage of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ in the $\psi(2S)$ region based on the simulation (R2p2-DD category).

$\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ channels can be fitted. As an example, Figure 9.5 shows the fits in the R2p2-DD category. The data fits for the other data-taking periods (R1, R2p1, R2p2) and the $\Lambda^0 \to p\pi^-$ track type (LL, DD) categories are shown in Appendix E.

The determined single ratios $r^{-1}_{J/\psi}$ and double ratios $R^{-1}_{\psi(2S)}$ are shown in Table 9.2 and Figure 9.6. As discussed in Section 9.1.2, with the efficiencies being Gaussian constrained during the fit, their uncertainties (corrresponding to the limited MC statistics, as discussed in Section 9.2) are propagated to the determined ratios. Thus, the determined ratios incorporate the statistical uncertainties of the data and the statistical uncertainties of the MC samples induced in the efficiencies. For the backgrounds, whose shapes are fixed from simulation, the systematic uncertainties associated with their shapes are not included. As shown in Figure 9.6, the determined single ratios $r^{-1}_{J/\psi}$ and double ratios $R^{-1}_{\psi(2S)}$ are compatible to unity considering the tolerance of uncertainties. Comparing with $r^{-1}_{J/\psi}$ in the same category, $R^{-1}_{\psi(2S)}$ has a larger uncertainty, which reflect the difference in the available statistics between the $\psi(2S)$ and the J/ψ channels. The determined signal yields of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ channels are shown in Table 9.3. Since the electron channel signal yields are parameterised using Equation 9.19 and Equation 9.20, they are calculated from the efficiencies, the ratio observables and the muon channel yields based on these equations. The signal statistics are more limited in the electron mode than in the muon mode and more limited in the $\psi(2S)$ mode than in the J/ψ mode.

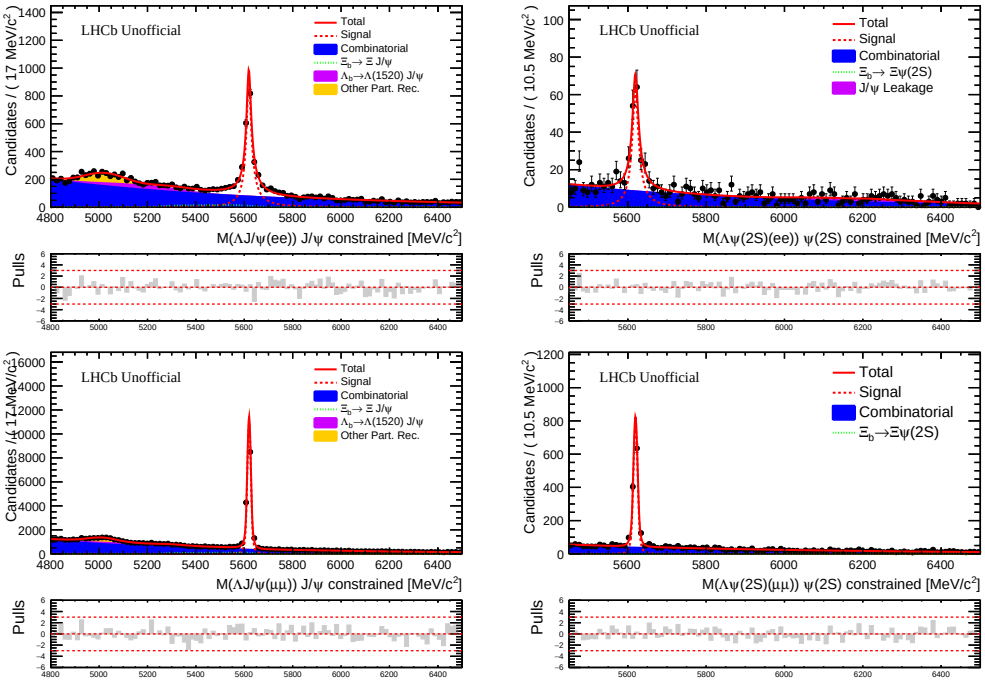


Figure 9.5: The data fit of the R2p2-DD category in the resonance q^2 regions.

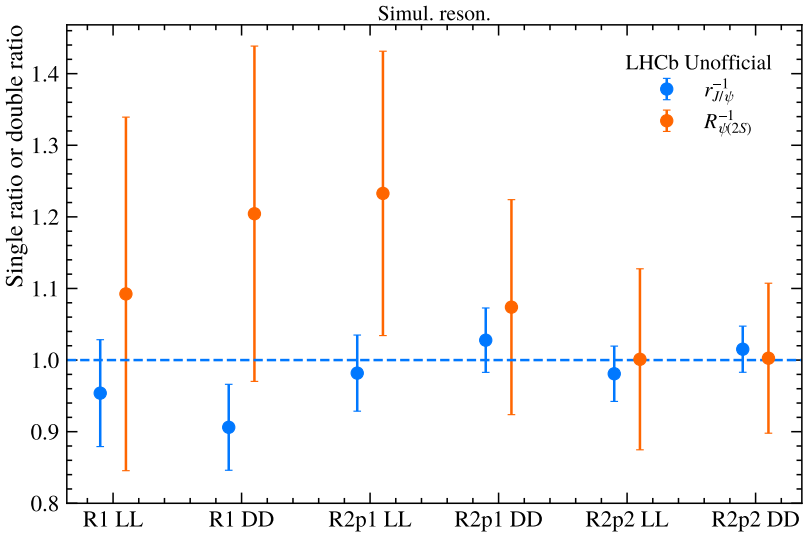


Figure 9.6: The determined $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ in the six data-taking period and Λ^0 track type categories.

Given the existing measurements on the relative branching fraction between $\Lambda_b^0 \to \Lambda^0\psi(2S)$ and $\Lambda_b^0 \to \Lambda^0 J/\psi$ [64], another crosscheck can be performed with the ratios of the efficiency corrected yields ($N_{\rm sig}/\epsilon_{\rm sig}$) between $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$. As shown in Figure 9.7, these ratios are, considering the tolerance of uncertainties, compatible with their expectations. The corresponding ratios in the electron channel have larger uncertainties than the ratios in the muon channel, which reflects the difference in the amount of data statistics between the two.

A few interesting fit parameters are shown in Table 9.4. The aforementioned mean-shift ($\Delta\mu$, Δm_0) and width-scaling (s_λ, s_σ) parameters, which are used to accommodate the data/simulation discrepancies in the signal shapes, are shown. Although, in several cases, the mean-shift parameter values are not compatible with 0.0 considering the uncertainties, the absolute scale of the shifts is small comparing to the scale of the known Λ_b^0 mass ($\sim 5620\,{\rm MeV}/c^2$) [64]. It is worth noting that the central values of the width scaling factor are larger than 1.0 in the various data sample categories except for the electron-R2p2-LL category, where the value is slightly lower than 1.0 but compatible with 1.0 considering the uncertainty. This is an indication that these signal shapes determined from the MC samples have the tendency of being over-optimistic about the resolution. The parameter $a_{\rm partReco}$ for the $\Lambda_b^0 \to \Lambda(1520)J/\psi$ background have similar values across the different data sample categories (although the values in the R2p1-LL and R2p1-DD categories deviate more with respect to the others), which is expected since its physical interpretation is the relative branching fraction between $\Lambda_b^0 \to \Lambda(1520)J/\psi$ (with $\Lambda(1520) \to \Sigma^0(\to \Lambda^0\gamma)\pi^0$) and $\Lambda_b^0 \to \Lambda^0 J/\psi$[32]. For the $a_{\rm partReco}$ parameters of $\Xi_b^- \to \Xi^- J/\psi$ and $\Xi_b^- \to \Xi^-\psi(2S)$, it has to be conceded that in many categories they cannot be considered as meanful estimations. This can be ascribed to two reasons. Firstly, the contributions of the Ξ_b backgrounds are small, especially in the LL category where the suppression from the efficiency is considerably larger than in the DD category. Secondly, the available statistics in data are limited. This is more severe in the $\psi(2S)$ channels than in the J/ψ channels. It is worth noting that Table 9.4 provides meaningful results in the case of $a_{\rm partReco}$ ($\Xi_b^- \to \Xi^- J/\psi$) in the DD category, where the efficiency suppression and the statistical fluctuation are less severe. In other cases, it can be considered that the fits are not sensitive enough to provide meaningful estimations for $a_{\rm partReco}$ ($\Xi_b^- \to \Xi^- J/\psi$) and $a_{\rm partReco}$ ($\Xi_b^- \to \Xi^-\psi(2S)$). In order to check the effects on the $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ observables induced by in-

[32]Considering that the $\Lambda^0(1600)$ and $\Lambda^0(1405)$ components are omitted and have shapes largely overlapping with the $\Lambda^0(1520)$ components, it is likely that they also contribute.

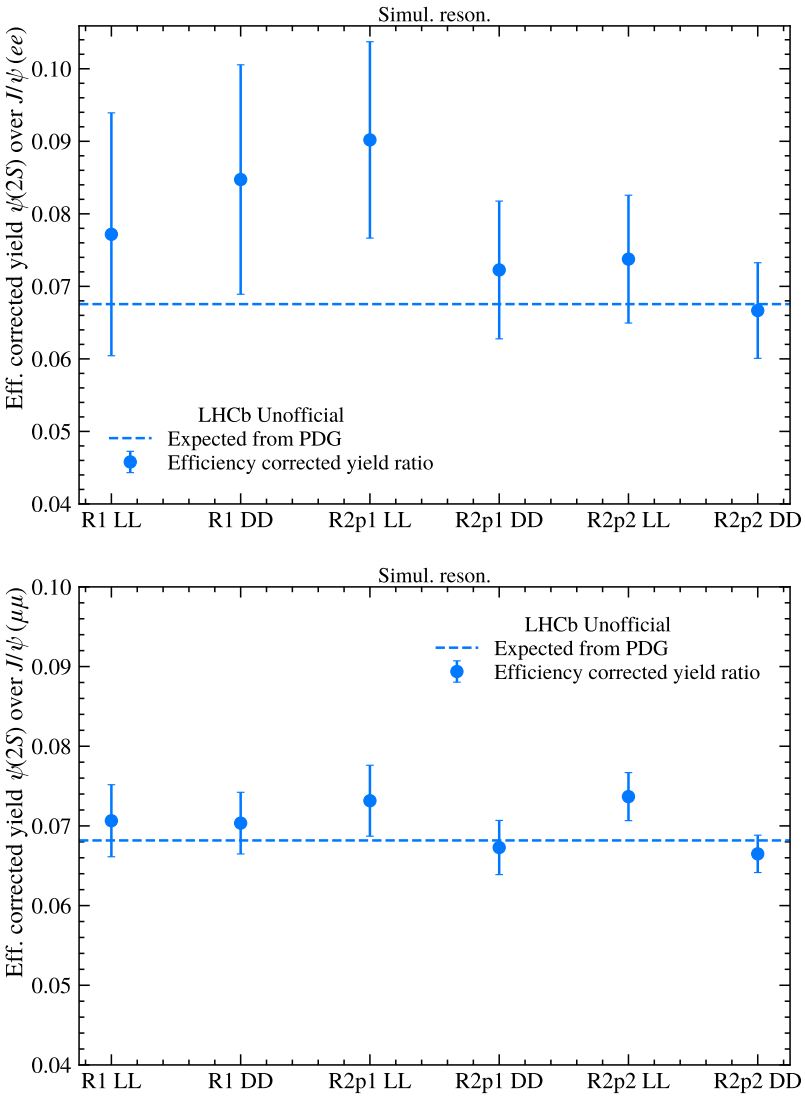


Figure 9.7: The ratio of the efficiency corrected yields between $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ in the electron (top) and muon (bottom) channels.

Table 9.2: Resonance ratios from the fits.

	R1 LL	R1 DD	R2p1 LL	R2p1 DD	R2p2 LL	R2p2 DD
$r^{-1}_{J/\psi}$	0.954 ± 0.075	0.906 ± 0.06	0.982 ± 0.053	1.028 ± 0.045	0.981 ± 0.039	1.015 ± 0.032
$R^{-1}_{\psi(2S)}$	1.092 ± 0.247	1.204 ± 0.234	1.233 ± 0.199	1.074 ± 0.15	1.001 ± 0.126	1.003 ± 0.105

Table 9.3: The resonance signal yields from the data fits.

	R1-LL	R1-DD	R2p1-LL	R2p1-DD	R2p2-LL	R2p2-DD
Muon J/ψ	2971.27 ± 57.24	4305.98 ± 72.6	3666.91 ± 63.47	6089.04 ± 86.0	7728.79 ± 92.4	13230.45 ± 126.04
Muon $\psi(2S)$	325.57 ± 19.36	458.13 ± 23.61	365.32 ± 20.85	531.13 ± 25.35	774.21 ± 29.52	1117.19 ± 37.22
Electron J/ψ	361.58 ± 25.14	479.32 ± 29.49	712.86 ± 33.78	1266.04 ± 50.03	1356.02 ± 47.27	2487.53 ± 71.39
Electron $\psi(2S)$	34.01 ± 7.04	56.42 ± 10.17	73.32 ± 10.54	109.65 ± 13.99	117.75 ± 13.58	192.1 ± 18.45

cluding the a_{partReco} ($\Xi_b^- \to \Xi^- J/\psi$) and the a_{partReco} ($\Xi_b^- \to \Xi^- \psi(2S)$) parameters (which are not meaningfully measured in some cases), crosschecks of removing the $\Xi_b^- \to \Xi^- \psi(2S)$ fit components as well as removing both the $\Xi_b^- \to \Xi^- J/\psi$ and the $\Xi_b^- \to \Xi^- \psi(2S)$ fit components can be performed. They will be discussed in Section 10.1.

More parameters determined during the data fits are shown in Appendix E. The parameters used as constants (fixed based on MC samples) during the data fits are also shown in Appendix E.

Table 9.4: A selected group of the data fit parameters for the fits in the resonance regions.

	R1-LL	R1-DD	R2p1-LL	R2p1-DD	R2p2-LL	R2p2-DD
mean shift ($\mu\mu$)	1.243 ± 0.142	1.559 ± 0.137	-0.044 ± 0.131	-0.146 ± 0.115	-0.2192 ± 0.0877	$(-5.81 \pm 7.74) \times 10^{-2}$
mean shift (ee)	1.9 ± 1.02	1.381 ± 0.932	-0.826 ± 0.724	0.598 ± 0.616	0.352 ± 0.46	-0.143 ± 0.441
width scale ($\mu\mu$)	1.1367 ± 0.0236	1.1345 ± 0.0198	1.1458 ± 0.0211	1.1555 ± 0.017	1.1207 ± 0.0146	1.1364 ± 0.0113
width scale (ee)	1.316 ± 0.152	1.174 ± 0.115	1.1498 ± 0.09	1.1835 ± 0.0783	0.9802 ± 0.0583	1.1846 ± 0.056
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520)J/\psi)$	0.4193 ± 0.0669	0.4382 ± 0.0648	0.5302 ± 0.0599	0.3294 ± 0.0541	0.4444 ± 0.0424	0.4206 ± 0.0355
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.0 ± 0.337	0.3728 ± 0.0564	0.0 ± 0.268	0.3864 ± 0.0438	0.0 ± 0.0755	0.3104 ± 0.0284
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	1.5 ± 1.0	0.0 ± 1.15	0.886 ± 0.972	0.124 ± 0.207	0.0 ± 1.0	0.0 ± 0.231

9.4 Fits in the rare q^2 regions

With the resonance fits, the signal yields of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ channels and the $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ channels can be determined.

Combining the resonance information with the efficiencies, with the previous measurement of $\mathcal{B}(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)$ (relative to $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)$) [84] and with the LFU assumption, the expected yields for the signal $\Lambda_b^0 \to \Lambda^0 e^+e^-$ in the rare q^2 regions (*i.e.* the low, central and high q^2 regions) can be derived. It should be noted that the branching fraction $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)$ used in Ref. [84] to convert $\mathcal{B}(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)/\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi)$ into $\mathcal{B}(\Lambda_b^0 \to \Lambda^0\mu^+\mu^-)$ has been superseded by the measurement in Ref. [85]. The result shown here was calculated using the relative branching fraction between $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ and $\Lambda_b^0 \to \Lambda^0 J/\psi$ as input, and thus is not affected by this supersession. The amount of resonance leakage in the electron central or high q^2 region, *i.e.* the resonance signal decay reconstructed in these q^2 regions, can be estimated using the determined resonance signal yields and the efficiencies.

These expected signal and leakage yields in the electron rare q^2 regions are shown in Table 9.5. The signal expectation is much higher in the high q^2 region than in the central and low q^2 regions. This is partially because the signal branching fraction is expected to be higher in the high q^2 region. Another important reason is the efficiency difference between the high q^2 region and the low/central q^2 regions as shown in Section 9.2. These expected yields are calculated with mass window requirements, *i.e.* the reconstructed decay candidates are require to have reconstructed Λ_b^0 mass within certain windows. These mass window requirements correspond to the mass window used in nominal fits (Section 9.4.2 and Section 9.4.3). This mainly concerns the leakage background, which is considerably displaced from the nominal mass window center and can have a considerable part residing outside of the mass window.

Given these signal expectations, different fitting strategies are applied for the high q^2 region and the central/low q^2 regions. For the high q^2 region, the goal is to extract R_Λ^{-1}, which will be discussed in Section 9.4.2. For the central and low q^2 regions, given the expected signal yields is very limited (under LFU assumption), the objective is to evaluate an upper limit of the branching fraction $\mathcal{B}$ $(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ relative to $\mathcal{B}$ $(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$, which will be discussed in Section 9.4.3.

Table 9.5: The expected values of several yields of interest in the electron rare q^2 regions. The nominal mass window requirements are included. The uncertainty from the relative branching fraction between the rare and resonance channels is not included.

	R1 LL	R1 DD	R2p1 LL	R2p1 DD	R2p2 LL	R2p2 DD
High-q^2 signal	8.31 ± 0.66	11.55 ± 0.77	12.02 ± 0.65	21.06 ± 0.92	25.77 ± 1.0	38.07 ± 1.21
High-q^2 $\psi(2S)$ leakage	0.81 ± 0.21	0.91 ± 0.2	1.5 ± 0.27	2.03 ± 0.31	2.85 ± 0.4	3.45 ± 0.4
Central-q^2 signal	0.25 ± 0.02	0.18 ± 0.01	0.37 ± 0.02	0.43 ± 0.02	0.65 ± 0.03	0.74 ± 0.03
Central-q^2 J/ψ leakage	2.47 ± 0.72	0.47 ± 0.3	2.07 ± 0.76	0.61 ± 0.51	1.98 ± 0.58	2.58 ± 0.72
Low-q^2 signal	0.23 ± 0.03	0.25 ± 0.02	0.4 ± 0.03	0.4 ± 0.03	0.64 ± 0.04	0.89 ± 0.05

9.4.1 Hadronic decays misidentified as $\Lambda_b^0 \to \Lambda^0 e^+ e^-$

As discussed in Section 6.6, hadronic decays with the hadrons misidentified as electrons can potentially contribute to the backgrounds and bias the signal extraction for the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ decay. This section will discuss the treatment of this background in the electron channel, the result of which will be used in the fits (to be discussed in Section 9.4.2 and Section 9.4.3). The method developed in the R_X ($X = K, K^*$) study (which has the advantage of avoiding the need to investigate individually each likely background decays) [74, 75] is used to model this potential background.

Following the work of R_X ($X = K, K^*$) study [74, 75], this so-called `Fail-To-Pass` method is explained as follows. By inverting the nominal PID selection for the electrons, one can intentionally select the data samples in the background-enhanced region (the `Fail` region). The misidentified background in the nominally selected region (the `Pass` region) can then be extrapolated based on the data in the background-enhanced regions. In this analysis, since some PID requirements are already present in the Stripping selection (as discussed in Section 7.1.1), only the customised offline electron PID selection is inverted, which means:

- `Pass` region: `CombDLL`$e > 0.0$ and `ProbNN`$e > 0.2$,
- `Fail` region: `CombDLL`$e > 0.0$ and `ProbNN`$e \leq 0.2$.

Given the two electron candidates, 4 regions can be defined as each of the electron candidate can be in the `Pass` or the `Fail` region. The decay candidates in the `Pass-Pass` (PP) region can be divided, according to the genuine particle species, into 4 possible categories, *i.e.* hadron-hadron (h_1h_2), hadron-electron (h_1e_2), electron-hadron (e_1h_2) and electron-electron (e_1e_2). This can be expressed mathematically as

$$N_{PP} = N_{h_1h_2}\epsilon_P^{h_1}\epsilon_P^{h_2} + N_{h_1e_2}\epsilon_P^{h_1}\epsilon_P^{e_2} + N_{e_1h_2}\epsilon_P^{e_1}\epsilon_P^{h_2} + N_{e_1e_2}\epsilon_P^{e_1}\epsilon_P^{e_2} \ , \tag{9.32}$$

where N_{PP} stands for the number of candidates in the `Pass-Pass` region, $\epsilon_P^{h_{1/2}}(\epsilon_P^{e_{1/2}})$ stands for the efficiency of a hadron(electron) to be selected into the `Pass` region and $N_{h_1(e_1)h_2(e_2)}$ stands for the hypothetical number of candidates of the genuine hadron(electron)-hadron(electron) origin. Similarly, for the `Pass-Fail` (PF), `Fail-Pass` (FP) and `Fail-Fail` (FF) regions, there are

$$N_{PF} = N_{h_1h_2}\epsilon_P^{h_1}\epsilon_F^{h_2} + N_{h_1e_2}\epsilon_P^{h_1}\epsilon_F^{e_2} + N_{e_1h_2}\epsilon_P^{e_1}\epsilon_F^{h_2} + N_{e_1e_2}\epsilon_P^{e_1}\epsilon_F^{e_2} \ , \tag{9.33}$$

$$N_{FP} = N_{h_1h_2}\epsilon_F^{h_1}\epsilon_P^{h_2} + N_{h_1e_2}\epsilon_F^{h_1}\epsilon_P^{e_2} + N_{e_1h_2}\epsilon_F^{e_1}\epsilon_P^{h_2} + N_{e_1e_2}\epsilon_F^{e_1}\epsilon_P^{e_2} \ , \tag{9.34}$$

and

$$N_{FF} = N_{h_1h_2}\epsilon_F^{h_1}\epsilon_F^{h_2} + N_{h_1e_2}\epsilon_F^{h_1}\epsilon_F^{e_2} + N_{e_1h_2}\epsilon_F^{e_1}\epsilon_F^{h_2} + N_{e_1e_2}\epsilon_F^{e_1}\epsilon_F^{e_2} \ , \tag{9.35}$$

where $\epsilon_F^{h_1}, \epsilon_F^{h_2}, \epsilon_F^{e_1}, \epsilon_F^{e_2}$ stand for the the efficiencies of h_1, h_2, e_1, e_2 being selected into the `Fail` region respectively.

Using the above equations, one can derive

$$\begin{aligned} &N_{PP} - N_{e_1e_2}\epsilon_P^{e_1}\epsilon_P^{e_2} = \\ &\frac{\epsilon_P^{h_2}}{\epsilon_F^{h_2}}(N_{PF} - N_{e_1e_2}\epsilon_P^{e_1}\epsilon_F^{e_2}) + \frac{\epsilon_P^{h_1}}{\epsilon_F^{h_1}}(N_{FP} - N_{e_1e_2}\epsilon_F^{e_1}\epsilon_P^{e_2}) \\ &- \frac{\epsilon_P^{h_1}\epsilon_P^{h_2}}{\epsilon_F^{h_1}\epsilon_F^{h_2}}(N_{FF} - N_{e_1e_2}\epsilon_F^{e_1}\epsilon_F^{e_2}) \ . \end{aligned} \tag{9.36}$$

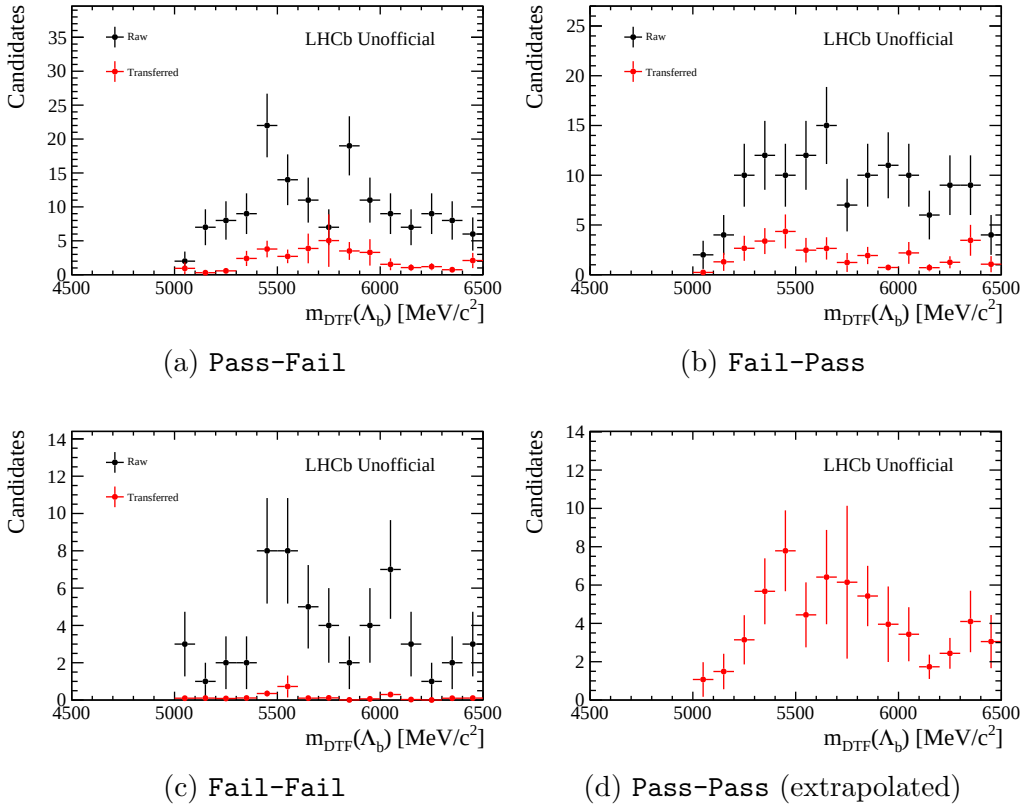
The factors such as $\epsilon_P^{h_1}/\epsilon_F^{h_1}$ are referred to as the transfer factors in the following discussion. Strictly speaking, the efficiency is kinematic dependent. In this case, the above argument still holds in individual kinematic bins. In practice, the transfer factors are extracted in kinematic bins (p, η) based on the PIDCalib tool [102, 103] and applied as weights to the lepton candidates. Furthermore, since kaons and pions have different PID responses, the kaon(pion) transfer factors are assigned according to the Artificial-Neural-Network(ANN)-based [17, 102] PID requirement `ProbNN`$k > 0.1$ (≤ 0.1).

The left-hand-side of Equation 9.36 is the total number of decay candidates minus the genuine dielectron component in the `Pass-Pass` region, which gives access to the misidentified background. The right-hand-side of Equation 9.36 consists of, other than the efficiency terms, N_{PF}, N_{FP}, N_{FF} and $N_{e_1e_2}$. The data in the `Pass-Fail`, `Fail-Pass` and `Fail-Fail` regions provide access to the N_{PF} N_{FP} and N_{FF}. The terms containing $N_{e_1e_2}$ correspond to the genuine dielectron components in the background-enhanced regions. These include the genuine $\Lambda_b^0 \to \Lambda^0 e^+e^-$ appearing in the `Pass-Fail`, `Fail-Pass` and `Fail-Fail` regions. To account for these $\Lambda_b^0 \to \Lambda^0 e^+e^-$ contributions in the background-enhanced regions, scaled $\Lambda_b^0 \to \Lambda^0 e^+e^-$ MC samples are used to subtract the background-enhanced data samples. The scaling factor is determined using the expected number of signal as described in Table 9.5. By taking this approach, the LFU assumption is implied. This is tolerated because the effect generated in this way by a hypothetical deviation from LFU is suppressed after applying the inverting PID selection and the transfer weights.

Using Equation 9.36, the extracted background includes not only the potential background from the decays like $\Lambda_b^0 \to \Lambda^0 h_1 h_2$ but also certain combinatorial components. The method does not differentiate between the hadrons

of the combinatorial origin and the hadrons of the hadronic decay origin. In addition, a component of the genuine dielectrons of the combinatorial origin also remains without being able to be subtracted by the scaled $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ simulation sample. This complication is further discussed in Section 10.2.

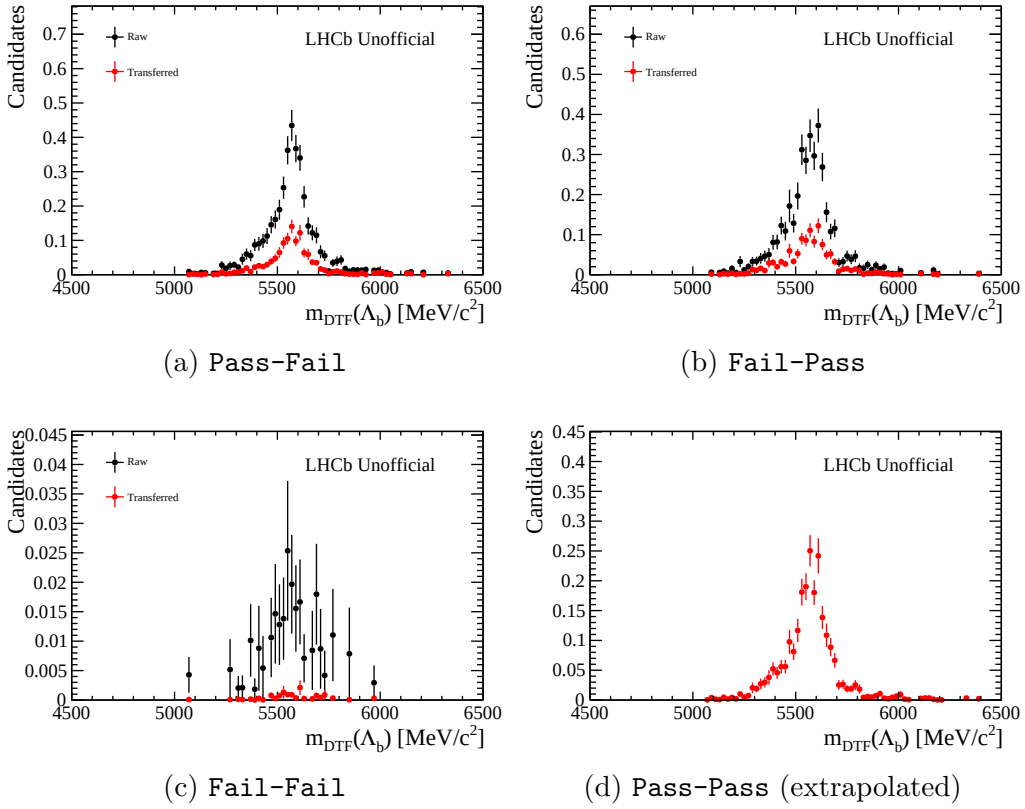
Figure 9.8 shows examples of invariant mass distribution in data before/after the transfer factor weighting in the `Pass-Fail`, `Fail-Pass` and `Fail-Fail` regions. Figure 9.9 shows the similar examples of the scaled $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ MC samples. The different regions of the data and signal MC samples are merged following Equation 9.36. To extract the shape, the different components are fitted with the kernel estimation method [124, 125] and merged in the fashion of Equation 9.36. The results are shown in Figure 9.10. The extracted shape and yields will be used for the rare q^2 region fits, which will be discussed in Section 9.4.2 and Section 9.4.3. As shown in the plot, the extracted shapes are largely affected by the statistical fluctuation. In Section 10.2, the systematic effect of this `Fail-To-Pass` method in the high q^2 region will be discussed.



(a) `Pass-Fail` (b) `Fail-Pass`

(c) `Fail-Fail` (d) `Pass-Pass` (extrapolated)

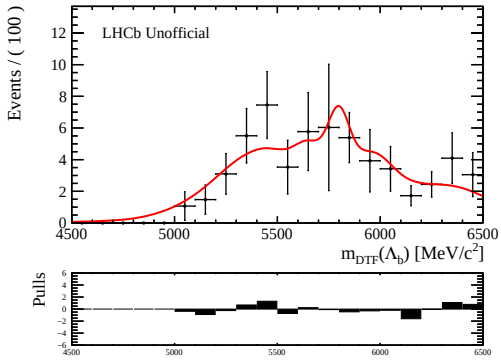
Figure 9.8: Examples (high q^2 region, DD category) of the raw and transferred (*i.e.* transfer weights applied) mass distribution in data for the `Pass-Fail`, `Fail-Pass` and `Fail-Fail` regions. Also shown is the extrapolated distribution in the `Pass-Pass` region, which is derived by merging the transferred distributions in the `Pass-Fail` and `Fail-Pass` regions and subtracting the transferred distribution in the `Fail-Fail` region.



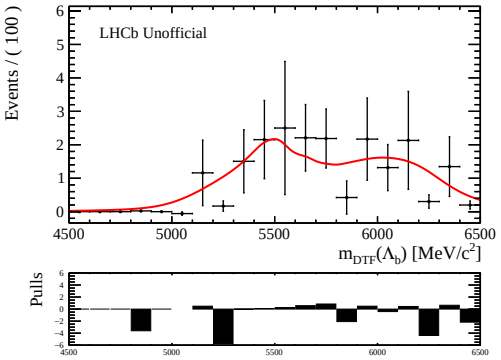
(a) `Pass-Fail` (b) `Fail-Pass`

(c) `Fail-Fail` (d) `Pass-Pass` (extrapolated)

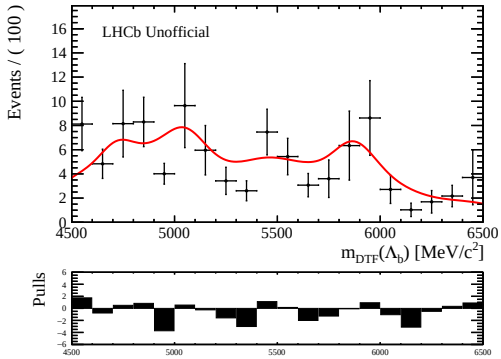
Figure 9.9: Examples (high q^2 region, DD category) of the raw and transferred (*i.e.* transfer weights applied) mass distribution in the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ MC (scaled) samples for the `Pass-Fail`, `Fail-Pass` and `Fail-Fail` region. Also shown is the extrapolated distribution in the `Pass-Pass` region, which is derived by merging the transferred distributions in the `Pass-Fail` and `Fail-Pass` regions and subtracting the transferred distribution in the `Fail-Fail` region.



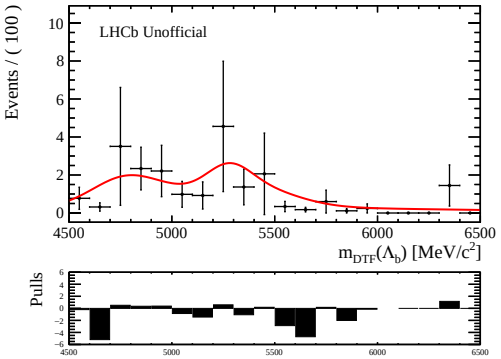
(a) High q^2, DD



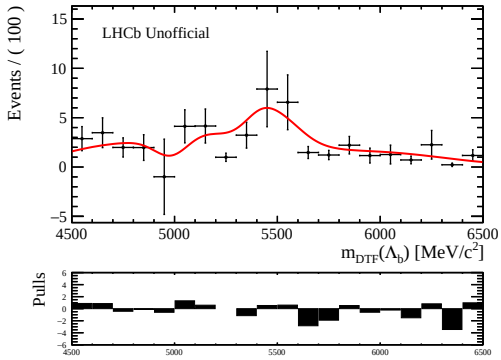
(b) High q^2, LL



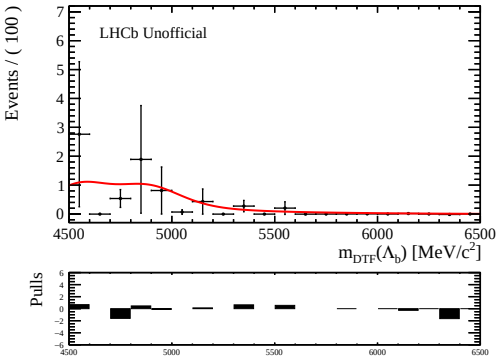
(c) Central q^2, DD



(d) Central q^2, LL



(e) Low q^2, DD



(f) Low q^2, LL

Figure 9.10: The merged distributions from the `Fail-To-Pass` method in different q^2 regions and $\Lambda^0 \to p\pi^-$ track types. The red line indicates the extracted shape based on the kernel estimation method [124].

9.4.2 Fits in the high q^2 region

Following a similar treatment as in the resonance fits discussed in Section 9.3, the rare electron signal yield in a certain data-taking period y and $\Lambda^0 \to p\pi^-$ track type category t can be parameterised using the observable R_Λ^{-1}, the efficiencies, the single ratio $r_{J/\psi}^{-1}$ and the muon channel signal yield:

$$N_{\rm sig}(ee)|_{y,t} = R_\Lambda^{-1} \times \left(\frac{\epsilon_{\rm sig}(ee)}{\epsilon_{\rm sig}(\mu\mu)} N_{\rm sig}(\mu\mu) r_{J/\psi}^{-1} \Big|_{y,t} \right) \ , \tag{9.37}$$

where $N_{\rm sig}(ee)$ stands for the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal yield, $N_{\rm sig}(\mu\mu)$ stands for the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yield and $\epsilon_{\rm sig}$ stands for the signal efficiency of the corresponding mode. Given the low statistics in the electron channel, the three data-taking periods (R1, R2p1 and R2p2) are merged. To preserve the benefit of using the double ratio, the signal yield of the electron channel in the merged dataset is parameterised as the sum of the yields of the three different data-taking periods with each of them parameterised as Equation 9.37. Mathematically, this is expressed as

$$N_{\rm sig}(ee)|_{t} = R_\Lambda^{-1} \times \left(\sum_{y={\rm R1,R2p1,R2p2}} \frac{\epsilon_{\rm sig}(ee)}{\epsilon_{\rm sig}(\mu\mu)} N_{\rm sig}(\mu\mu) r_{J/\psi}^{-1} \Big|_{y,t} \right) \ . \tag{9.38}$$

The efficiencies and the single ratios $r_{J/\psi}^{-1}$ are constrained based on their central values and uncertainties. The $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yields are allowed to be determined from the data fit. The observable R_Λ^{-1} can be determined by performing a simultaneous fit of the electron channel (with the two track type categories) and the muon channel (with the six combination of the track type and data-taking period categories).

In contrast to the fits in resonance q^2 regions, the dilepton mass cannot be constrained and thus the fit uses the mass variable $m_{\rm DTF}(\Lambda_b^0)$ (DTF reconstructed Λ_b^0 mass with PV and Λ^0 mass constrained) instead of $m_{\rm DTF}^{J/\psi}(\Lambda_b^0)$ or $m_{\rm DTF}^{\psi(2S)}(\Lambda_b^0)$ (DTF reconstructed Λ_b^0 mass with PV, Λ^0 mass and the dilepton mass constrained).

The shape of the signal component in the Λ_b^0 mass spectrum is fixed based on the simulation samples. It is likely the modelling of the signal shapes discussed here is not final because the discrepancy between data and simulation in the $m_{\rm DTF}(\Lambda_b^0)$ spectrum is not accounted for. In the resonance region fits discussed in Section 9.3, such a discrepancy is accommodated by allowing, in the data fit, a shift of the mean and a scaling of the width for the signal shape determined from simulation. This is not applied to the fits in

the rare q^2 regions, considering the limited data statistics. In similar analyses (*e.g.* Ref. [77]), the data/simulation discrepancy in the distribution of $m_{\rm DTF}(b\text{-hadron})$ is often treated by studying the resonance channels to determine a shift factor of the mean and a scaling factor of the width, which can then be applied to the rare signal channels. Since the $m_{\rm DTF}$ variable is reconstructed without the dilepton mass constraints, such shift and scaling factors are different from their counterparts in the cases of $m^{J/\psi}_{\rm DTF}$ and $m^{\psi(2S)}_{\rm DTF}$. The procedure to correct for the data/simulation discrepancy in the invariant mass distribution is sometimes referred as the mass calibration. For this analysis, the statistics is limited even in the case of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$. Without the dilepton mass constraint, a fit of the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ mode extracting precise mean shift and width scaling factors cannot be easily performed. An alternative for the mass calibration is to study the corresponding systematic uncertainties. However, it should be noted that this analysis is very limited in data statistics and this discrepancy is expected to be a small effect in comparison.

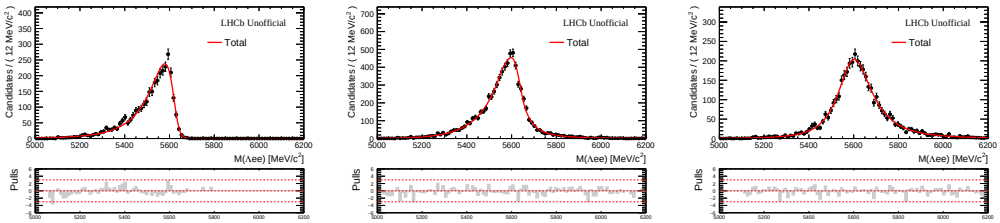
As discussed in Section 6.4, the bremsstrahlung recovery procedure is involved in the electron reconstruction. Without the dilepton mass constraints in the DTF algorithm, the electron channel signal peaks of the three different bremsstrahlung recovery categories (G0, G1, G2) have considerably different shapes. Thus, the signal peaks in different bremsstrahlung recovery categories are fitted separately and then merged according to the fraction of each category in the simulation [33].

Figure 9.11 shows the fits to the $\Lambda_b^0 \to \Lambda^0 e^+e^-$ signal MC samples in the different bremsstrahlung recovery and $\Lambda^0 \to p\pi^-$ track type categories. As shown in the plot, as more bremsstrahlung photons are recovered (from G0 to G2), the invariant mass $m_{\rm DTF}(\Lambda_b^0)$ distribution shifts towards the upper side. The G0 and G2 categories are modelled with the Johnson S_U function [122] while the G1 category is modelled with the AsymDSCB function [123] [34]. For the $\Lambda_b^0 \to \Lambda^0 \mu^+\mu^-$ signal MC fit, an example is shown in Figure 9.12, where the AsymDSCB function [123] is used.

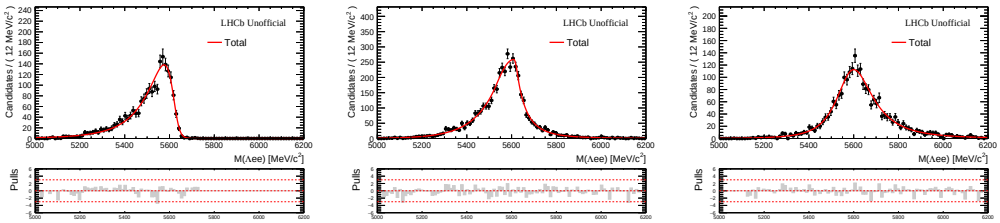
As discussed in Section 6.6, the combinatorial background in the high q^2 region is sculpted by the phase-space limit. To model the shape of the combinatorial background, the SS data samples are used. In order to retain more statistics in the SS data samples, the MVA selection, the very purpose of which is to suppress the combinatorial background, is not applied.

[33] It is likely that, by fixing the shape from MC and merging according to the MC fractions, the results might be similar as fitting merged categories. However, it is likely that the aforementioned data/simulation discrepancy in signal shapes needs to be studied in different bremsstrahlung recovery categories separately.

[34] Attempts were made to use the Johnson S_U function but the result was not ideal.



(a) G0 DD (b) G1 DD (c) G2 DD



(d) G0 LL (e) G1 LL (f) G2 LL

Figure 9.11: The fits of the high q^2 region $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal MC in the different bremsstrahlung recovery and track type categories.

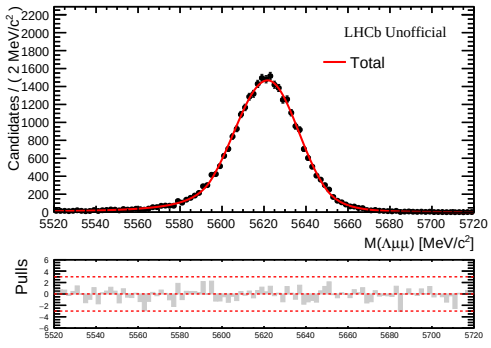


Figure 9.12: An example (R2p2-DD category) of the high q^2 region $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal MC fit.

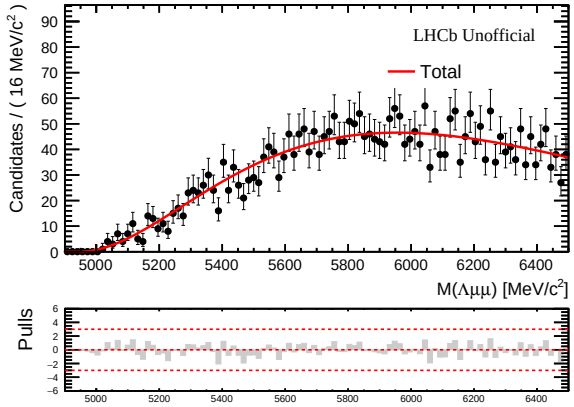


Figure 9.13: An example (muon R2p2 DD category) of the SS data fits in the high q^2 region.

Figure 9.13 shows an example of the SS data fits. The fit model used is the so-called `RooGamma` model [35] from RooFit [119], whose PDF has the following form [126]:

$$f_{\text{RooGamma}}(m|\mu,\gamma,\beta) = \frac{(m-\mu)^{\gamma-1}e^{-(m-\mu)/\beta}}{\Gamma(\gamma)\beta^{\gamma}} \ , \tag{9.39}$$

where $\Gamma(\gamma)$ is constant with respect to m and serves to normalise the PDF (Γ stands for the gamma function). The shape fixed from the SS data fit is used as the combinatorial background shape. The yield of the combinatorial background is allowed to be determined from the data fit.

Another component of the fit is the $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ decay reconstructed in the high q^2 region (leakage background). Its shape is extracted from the MC samples with the kernel estimation technique [124, 125]. Its yield is constrained using the expectations based on the resonance signal yields and the efficiencies (as shown in Table 9.5).

The results from the `Fail-To-Pass` method (as discussed in Section 9.4.1) are used to fix both the shapes and the yields of the misidentified hadronic component in the electron channel.

With the above components, the simultaneous fit to the high q^2 region data can be performed. Considering that the electron channel is blinded, the electron channel data are fitted without the signal region and R_Λ^{-1} is fixed to 1.0, which means the $\Lambda_b^0 \to \Lambda^0 e^+e^-$ signal yields are fixed to their expecta-

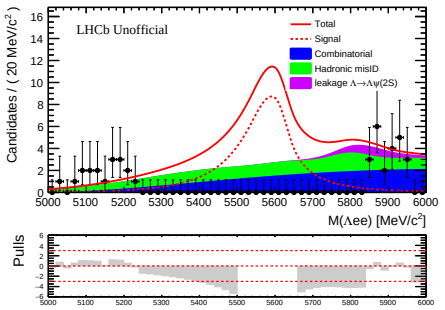
[35] This corresponds to the PDF of the Gamma distribution [126]. The motivation to use it to fit the SS data is empirically based.

tions under the LFU assumption. The simultaneous fit to the $m_{\mathrm{DTF}}(\Lambda_b^0)$ distribution in data is shown in Figure 9.14. As shown in Figure 9.14, comparing to the muon channel, the electron channel has very limited statistics, worse Λ_b^0 mass resolution and are affected by several backgrounds. The two main backgrounds in the electron channel are the combinatorial background and the misidentified hadronic background based on the `Fail-To-Pass` method (which can also contain candidates of the combinatorial origin). Between the electron and the muon channel, the electron channel is the main limiting factor for the sensitivity of measuring R_Λ^{-1}.

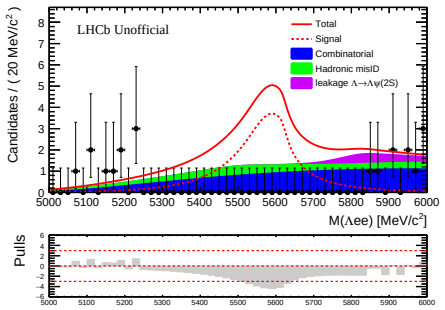
In the lower sidebands of the electron channel (LL and DD), several data points, while being largely dominated by statistical fluctuations, exhibit some slight upward trends. It is likely that the partially reconstructed background from $\Lambda_b^0 \to \Lambda^* e^+ e^-$ decays has a contribution. For the muon channel, due to the good mass resolution, a narrow mass window is used, which excludes these $\Lambda_b^0 \to \Lambda^* \mu^+ \mu^-$ backgrounds. Very little is known about the $\Lambda_b^0 \to \Lambda^* \ell^+ \ell^-$ decays in terms of their branching fractions (the $\Lambda_b^0 \to \Lambda(1520)\mu^+\mu^-$ branching fraction was measured [127]). Thus, there is no easy way to constrain the yields of these backgrounds in the fit. A systematic uncertainty study for the $\Lambda_b^0 \to \Lambda^* e^+ e^-$ background is discussed in Section 10.3.

An overview of the fit parameters determined during the data fit, with the exception of the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yields and efficiencies (constrained nuisance parameters), is shown in Appendix F Table F.1. Considering the ongoing dedicated $\mathcal{B}$ $(\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-)$ study within LHCb, the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yields and efficiencies are omitted to prevent indirectly giving an estimation of its branching fraction. The constant parameter of the data fit (fixed based on simulation, SS data *etc.*) are shown in Table F.2.

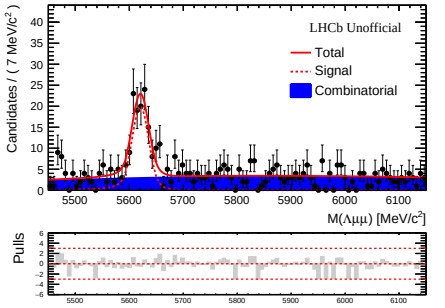
The statistical uncertainty (combining with the uncertainties induced by the nuisance parameter constraints) of the R_Λ^{-1} can be studied by generating toy data based on the aforementioned fit result and performing the refit with the observable R_Λ^{-1} allowed to be determined from the fit. The Gaussian constraints for the nuisance parameters are also sampled during the toy study to propagate their uncertainties. As shown in Figure 9.15, the toy R_Λ^{-1} sample shows a Gaussian-like distribution with the mean around 1.0 and a standard deviation around 0.158. This standard deviation incorporates the data statistical uncertainty (when assuming LFU) and the uncertainties induced by the nuisance parameters (efficiencies, single ratios $r_{J/\psi}^{-1}$ and $\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ leakage yields). The pull of the R_Λ^{-1} toy distribution shows a Gaussian-like distribution centered around 0.0 with a standard deviation around 1.0, which suggests good fit behaviour.



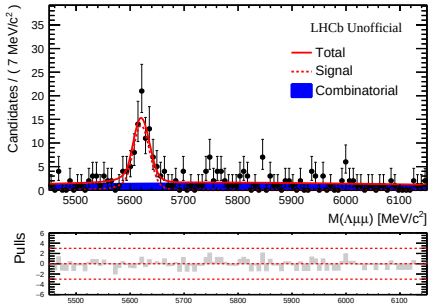
(a) Electron-DD



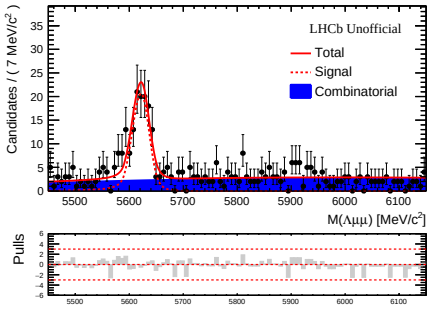
(b) Electron-LL



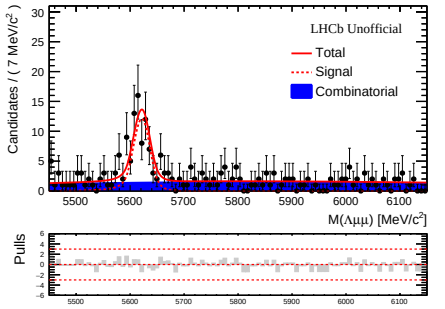
(c) Muon-R1-DD



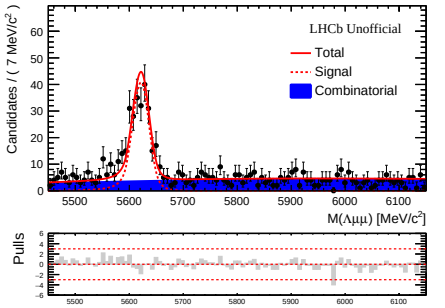
(d) Muon-R1-LL



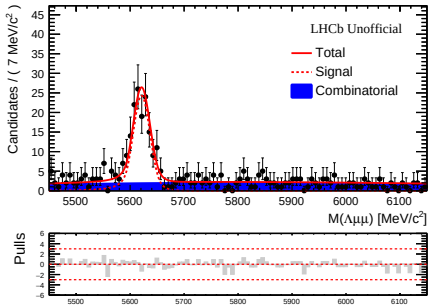
(e) Muon-R2p1-DD



(f) Muon-R2p1-LL



(g) Muon-R2p2-DD



(h) Muon-R2p2-LL

Figure 9.14: The data fit in the high q^2 region. The electron channel is blinded. The misidentified hadronic background is labelled as "Hadronic misID".

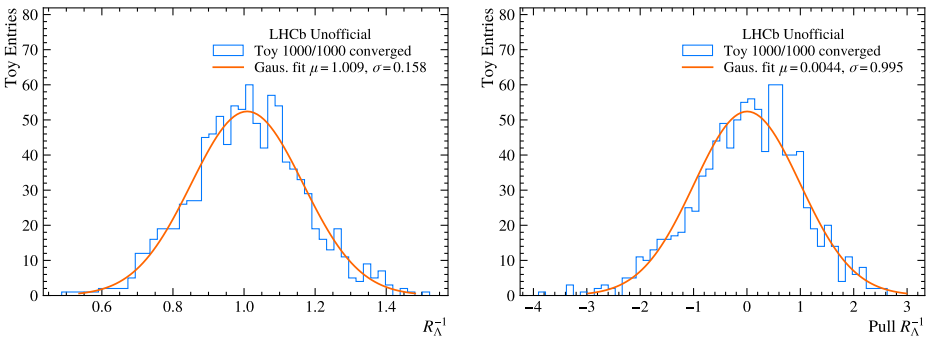


Figure 9.15: The R_Λ^{-1} (value and pull) distribution based on the toy data (1000 sets) fits.

9.4.3 Fits in the central and low q^2 regions

Given the expected signal yields assuming LFU (Table 9.5), no signal observation is expected in the low and central q^2 regions. The fits in these regions aim to estimate an upper limit of the branching fraction $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)$ relative to $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$.

Merging the three data-taking periods in a similar fashion as the fit in the high q^2 region, the electron channel yield in the track category $t = \text{LL/DD}$ can be parameterised as

$$N_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)|_t = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))} \times \left(\sum_{y=\text{R1,R2p1,R2p2}} \frac{\epsilon_{\text{sig}}(ee)}{\epsilon_{\text{sig}}(J/\psi, ee)} N_{\text{sig}}(J/\psi, ee)\Big|_{y,t} \right) , \tag{9.40}$$

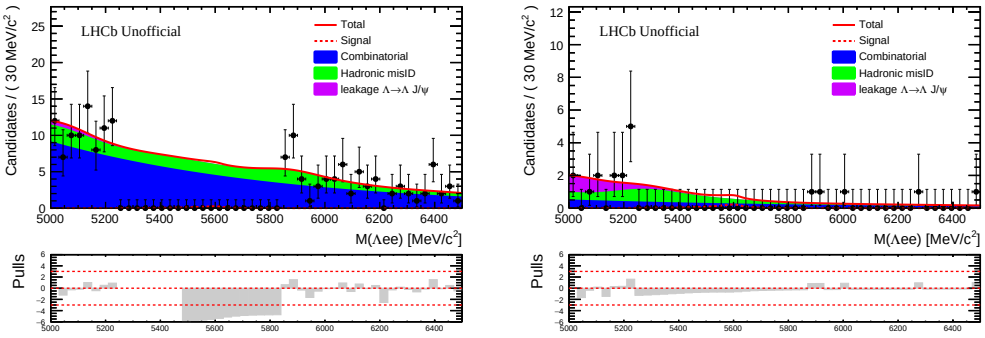
where $\epsilon_{\text{sig}}(ee)$ stands for the signal efficiency of $\Lambda_b^0 \to \Lambda^0 e^+ e^-$, $\epsilon_{\text{sig}}(J/\psi, ee)$ stands for the signal efficiency of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ and $N_{\text{sig}}(J/\psi, ee)$ stands for the signal yield of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$. These efficiencies and yields are evaluated following the discussion in Section 9.2 and Section 9.3. The following relative branching fraction is the parameter of interest:

$$r_{ee}^{\mathcal{B}} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))} . \tag{9.41}$$

The following factors (with $t = \text{LL/DD}$) are referred to as the normalisation factors:

$$\text{norm. factor}|_t = \sum_{y=\text{R1,R2p1,R2p2}} \frac{\epsilon_{\text{sig}}(ee)}{\epsilon_{\text{sig}}(J/\psi, ee)} N_{\text{sig}}(J/\psi, ee)\Big|_{y,t} . \tag{9.42}$$

The various fit components are treated in a similar fashion as for the fits in the high q^2 region with the exception of the combinatorial background. In the central and low q^2 regions, without the complication of the phase-space sculpting, an exponential shape is used for the combinatorial background with both the slope parameter and the yield allowed to be determined from the data fit. The signal shapes are, similar as in the high q^2 fit, fixed from the MC samples using Johnson S_U functions as fit models. The misidentified hadronic background is treated using the results from the `Fail-To-Pass` method. In the central q^2 region, the shape of the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ leakage is fixed from simulation and its yield is constrained using the expectations based on the resonance signal yields and the efficiencies (as shown in Table 9.5).



(a) DD (b) LL

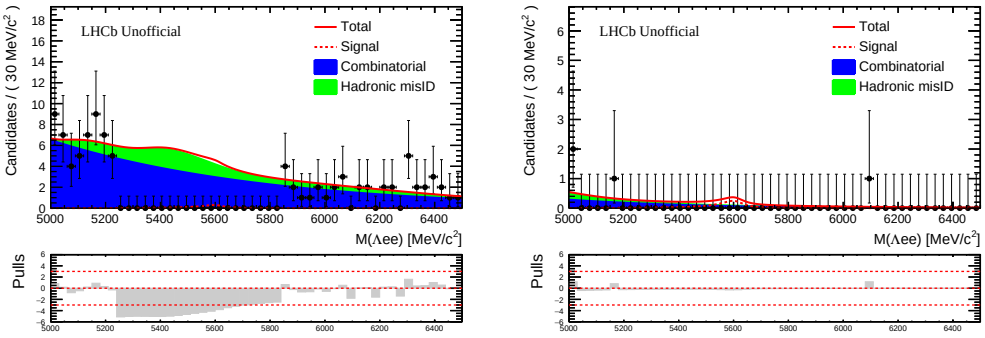
Figure 9.16: The blinded electron channel data fit in the central q^2 region. The misidentified hadronic background is labelled as "Hadronic misID".

Figure 9.16 and Figure 9.17 show the fits to the electron data (with the signal region blinded) in the central and low q^2 regions. The signal yields are fixed to their expected value under the LFU assumption. As shown in the plots, the two main backgrounds are the combinatorial background and the misidentified hadronic component based on the `Fail-To-Pass` method (which can contain candidates of the combinatorial origin). It is worth noting that, in both the central and the low q^2 regions, the DD category has considerably more statistics than the LL category. Although dedicated investigation to find the reason was not performed, such a large difference is not inconceivable. Considering the different track combination scenarios, *i.e.* in one case combining 2 long tracks together with 2 dowstream tracks and in the other case combining 4 long tracks, the combinatorial background behaviour is expected to be very different between the two categories.

The parameters of the data fits are shown in Appendix F.

As discussed in Section 9.1.3, in the case of no signal observation, the CL_s method can be used to estimate an upper limit on the relative branching fraction between $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$. Since the data is currently blinded, the expected CL_s curve under the background-only hypothesis can serve to provide an estimation of the upper limit.

The result of the aforementioned fit to the blinded data can be used to estimate the background composition in the mass window including the signal region. Using this as the background-only hypothesis, the expected median, $\pm\sigma$ and $\pm 2\sigma$ CL_s curves (as discussed in Section 9.1.3) can be determined. Using the asymptotic method described in Ref. [121] with its implementation



(a) DD (b) LL

Figure 9.17: The blinded electron channel data fit in the low q^2 region. The misidentified hadronic background is labelled as "Hadronic misID".

in `RooStats` [128, 129], the resulting CL_s curves are shown in Figure 9.18. The expected median upper limits corresponding to the 90 percent confidence level are

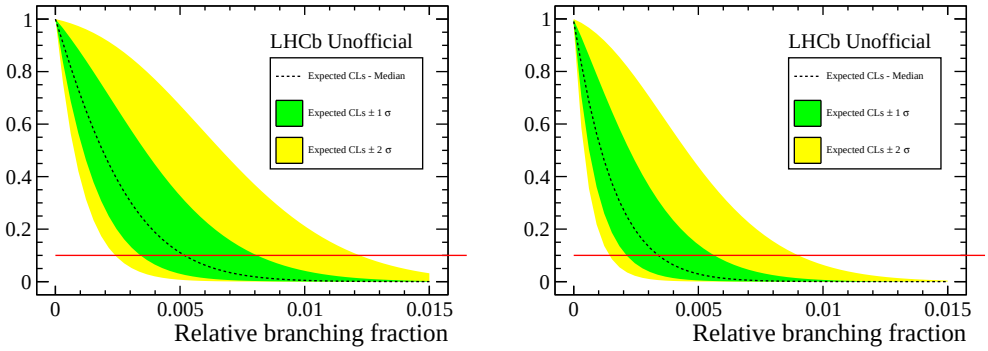
$$r_{ee}^{\mathcal{B}} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))} \lesssim \begin{cases} 5.2 \times 10^{-3} \text{ , central } q^2 \\ 3.4 \times 10^{-3} \text{ , low } q^2 \end{cases} . \tag{9.43}$$

In comparison, with the LFU assumption and the measurements in the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ channel in the relevant q^2 bins [84], the estimated relative branching fractions are [36]

$$r_{ee}^{\mathcal{B}} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))} \begin{cases} = (1.15^{+0.83}_{-0.74}) \times 10^{-3} \text{ , central } q^2 \\ \sim 1.37 \times 10^{-3} \text{ , low } q^2 \end{cases} . \tag{9.44}$$

Comparing the above two equations, an observation of the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal is indeed not expected under the LFU assumption.

[36] The uncertainty for the low q^2 region is not given, because the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ branching fraction for this q^2 region was not directly reported in Ref. [84] but inferred by combining several q^2 bins which are not all statistically independent.



(a) Central q^2 (b) Low q^2

Figure 9.18: The expected CL_s scan curves under the background-only hypothesis for the central and low q^2 regions. The curves are evaluated with the asymptotic method [121] implemented in `RooStats` [128, 129]. The red lines indicate the 90 percent confidence level.

10 Crosschecks and systematic uncertainties

In this chapter, several crosschecks and systematic uncertainties are discussed.

Following the discussion in Section 9.3, the determined $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ incorporate the data statistical uncertainty and the efficiency uncertainties induced by the MC sample statistics. In Section 10.1, several crosschecks for $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ are discussed.

For the observables in the rare q^2 regions, *i.e.* the R^{-1}_{Λ} ratio observable for the high q^2 region and the $r^{\mathcal{B}}_{ee}$ observable (relative branching fraction between $\Lambda^0_b \to \Lambda^0 e^+e^-$ and $\Lambda^0_b \to \Lambda^0 J/\psi(e^+e^-)$) for the central and low q^2 regions, the data statistical uncertainties and the uncertainties induced by the constrained nuisance parameters are already incorporated in the fitting procedure described in Section 9.4. In Section 10.2 and Section 10.3, two background-associated systematic uncertainties for R^{-1}_{Λ} in the high q^2 region are discussed. For the central and low q^2 regions, no studies for the systematic uncertainties, other than what has already been incorporated in the nuisance parameters during the fit discussed in Section 9.4.3, have been performed.

In Section 10.4, the already assigned uncertainties for R^{-1}_{Λ} (in the high q^2 region) are summarised. Plans for further investigations of the R^{-1}_{Λ} systematic uncertainties as well as the systematic uncertainty studies in the low and central q^2 regions are discussed.

10.1 Crosschecks with $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$

10.1.1 The simulation correction effects

As discussed in Section 7.3, several corrections are applied to the simulation samples, which are:

- The Λ_b^0 lifetime correction.
- The $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ angular correction.
- The Λ_b^0 kinematics and event multiplicity correction.
- The PID calibration correction.
- The electron tracking efficiency correction.
- The L0 trigger efficiency correction.

To illustrate the effects of these corrections, the efficiencies are calculated with various corrections switched on accumulatively one after the other. The fits in the resonance q^2 regions are then performed with these different versions of efficiencies as fixed parameters. Figure 10.1 illustrates how the $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ ratios evolves as the corrections get switched on one after the other. As shown in Figure 10.1, some of the simulation corrections can have considerable effects on the single ratio $r^{-1}_{J/\psi}$, especially the L0 trigger correction and the kinematic-multiplicity correction. The double ratio $R^{-1}_{\psi(2S)}$, on the other hand, is shown to be more robust against these corrections. This demonstrates the advantage of using the double ratio. In taking the double ratio, the effects from the efficiencies are better controlled. This crosscheck is illustrative, no corresponding systematic uncertainty is assigned.

In the fits of these crosschecks, the efficiencies are fixed to the new values without Gaussian constraints (which are applied for the nominal fits as discussed in Section 9.3). Thus, the uncertainties of these ratios shown in Figure 10.1 incorporate the data statistical effects but not the uncertainties associated with the efficiencies. The reasons for this treatment are: Firstly, the main purpose of this crosscheck is illustrative. Secondly, the evaluation of the uncertainties associated with the efficiencies using the toy MC method described in Section 9.2 costs computing resonances. The comparison of $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ between the scenario with Gaussian constrained efficiencies and the scenario with fixed efficiencies (both using the efficiencies calculated with the nominal MC corrections) are shown in Table 10.1 and Table 10.2. It is worth noting that, comparing the two scenarios, the central values of

Table 10.1: Comparison of $r^{-1}_{J/\psi}$ between the nominal fit scenario and the nominal fit scenario with fixed efficiency parameters.

	Nominal	Nominal (fixed eff.)
R1-LL	0.9538 ± 0.0746	0.951 ± 0.0686
R1-DD	0.9061 ± 0.06	0.9058 ± 0.0577
R2p1-LL	0.9818 ± 0.0532	0.9804 ± 0.0495
R2p1-DD	1.0278 ± 0.045	1.0261 ± 0.0423
R2p2-LL	0.9809 ± 0.0386	0.9798 ± 0.0361
R2p2-DD	1.0152 ± 0.0322	1.0144 ± 0.0304

Table 10.2: Comparison of $R^{-1}_{\psi(2S)}$ between the nominal fit scenario and the nominal fit scenario with fixed efficiency parameters.

	Nominal	Nominal (fixed eff.)
R1-LL	1.092 ± 0.247	1.095 ± 0.244
R1-DD	1.204 ± 0.234	1.205 ± 0.232
R2p1-LL	1.233 ± 0.199	1.235 ± 0.196
R2p1-DD	1.074 ± 0.15	1.075 ± 0.149
R2p2-LL	1.001 ± 0.126	1.002 ± 0.125
R2p2-DD	1.003 ± 0.105	1.003 ± 0.104

$r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ do not change much and the corresponding uncertainties are slightly smaller in the case of fixed efficiencies. This is expected considering that, when the efficiencies are fixed, their uncertainties are not propagated to $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$.

10.1.2 The Ξ_b background components

As discussed in Section 9.3, the fits in the resonance q^2 regions include the $\Xi_b^- \to \Xi^- J/\psi$ and the $\Xi_b^- \to \Xi^- \psi(2S)$ components. In some cases, given the available statistics in data, the fits are not able to provide meaningful estimations for the amplitudes of their contributions (in terms of the a_{partReco} parameters).

As a crosscheck, two alternative configurations are used to perform the fits. In one of the configurations, the $\Xi_b^- \to \Xi^- \psi(2S)$ components are removed. In the other configuration, both the $\Xi_b^- \to \Xi^- \psi(2S)$ components and the $\Xi_b^- \to \Xi^- J/\psi$ components are removed. For all cases, the nominal efficiency treatment is applied, *i.e.* treated as nuisance parameters with Gaussian constraints (as discussed in Section 9.3). The corresponding $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ ratios are shown in Figure 10.2. Their uncertainties incorporate the data statistical uncertainties and the uncertainties associated with the

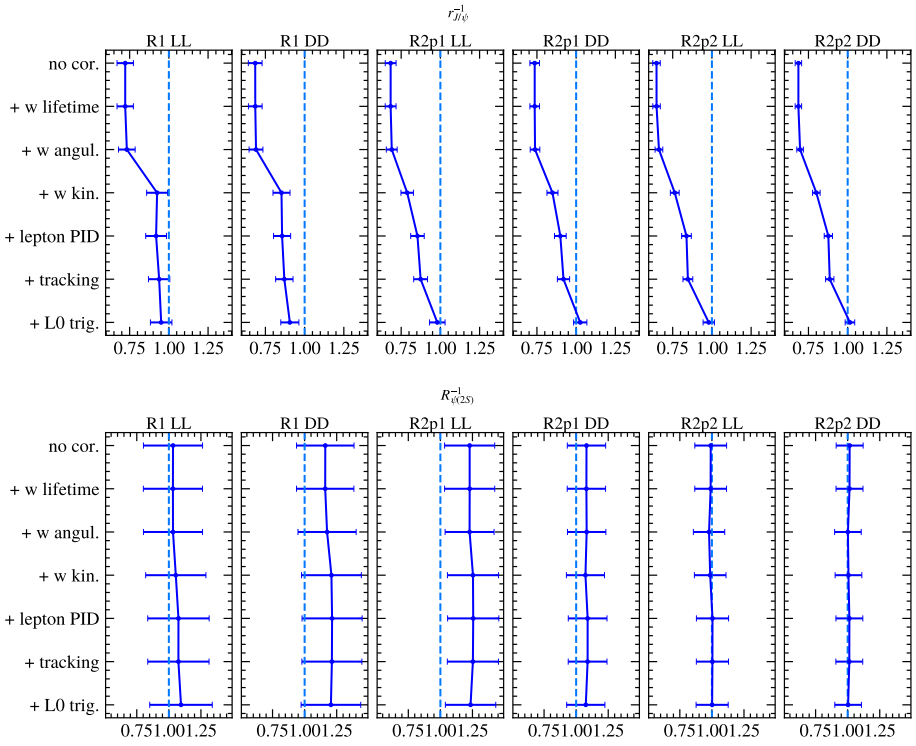


Figure 10.1: The $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ ratios with different efficiencies which incorporate different simulation corrections. The top plot corresponds to the $r^{-1}_{J/\psi}$ ratios. The bottom plot corresponds to the $R^{-1}_{\psi(2S)}$ ratios. In each plot, the different rows indicate the different weighting scenarios when calculating the efficiencies. The top row indicates the scenario without any MC correction weights. The following rows describe the cases with the weights switched on accumulatively in the sequence: the Λ_b^0 lifetime weights, the resonance mode angular weights, the kinematic(Λ_b^0)-multiplicity weights, the lepton PID calibration weights, the electron tracking efficiency weights and the L0 trigger calibration weights.

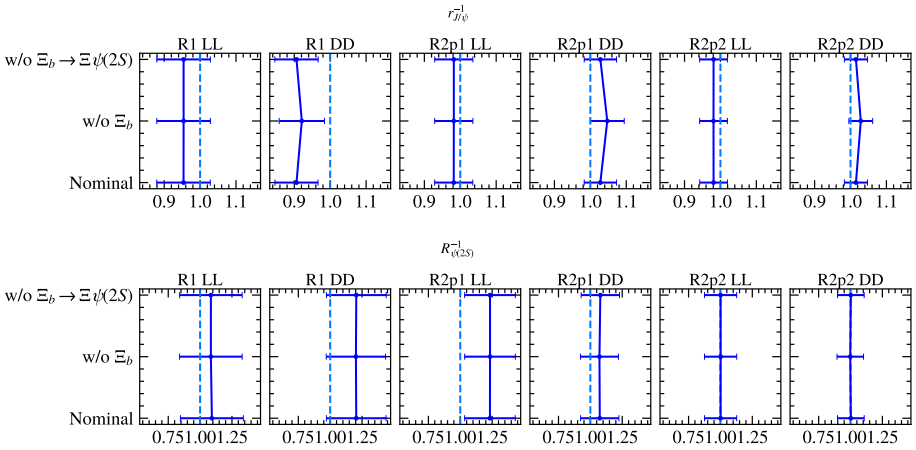


Figure 10.2: The comparison of the $r^{-1}_{J/\psi}$ (top plot) and $R^{-1}_{\psi(2S)}$ (bottom plot) ratios between the nominal scenario and two different treatments for the Ξ_b components. The label "w/o Ξ_b" indicates the treatment removing both the $\Xi_b^- \to \Xi^- J/\psi$ and $\Xi_b^- \to \Xi^-\psi(2S)$ components. The label "w/o $\Xi_b \to \Xi\psi(2S)$" indicates the treatment removing only the $\Xi_b^- \to \Xi^-\psi(2S)$ component.

efficiencies (induced by the MC statistics). As shown in the plots, for the two alternative treatments, the deviations of $r^{-1}_{J/\psi}$ from the nominal configuration becomes considerable to the scale of the associated uncertainties only in the DD category and only when removing both the $\Xi_b^- \to \Xi^- J/\psi$ component and the $\Xi_b^- \to \Xi^-\psi(2S)$ component. The deviations of $R^{-1}_{\psi(2S)}$ from the nominal configuration in all cases are small comparing to the associated uncertainties. As discussed in Section 9.3, in the DD category, the fits can provide meaningful estimations for a_{partReco} ($\Xi_b^- \to \Xi^- J/\psi$). Thus, the aforementioned considerable deviations happen when there are measurable background components being removed. This indicates that, in these cases, removing $\Xi_b^- \to \Xi^- J/\psi$ component is a mistreatment of the background. Thus it does not call for an assignment of systematic uncertainties.

The corresponding numerical results for $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$ are shown in Table 10.3 and Table 10.4. The aforementioned cases with considerable deviations in $r^{-1}_{J/\psi}$ are marked with red color. In other cases, the differences between the three scenarios for $r^{-1}_{J/\psi}$ are negligible. Thus, no corresponding systematic uncertainties are assigned for $r^{-1}_{J/\psi}$. For $R^{-1}_{\psi(2S)}$, the differences are small comparing with the associated uncertainties and they are not directly used for the evaluation of R^{-1}_{Λ}. Thus, no corresponding systematic

Table 10.3: Comparison between different resonance fit configurations for the Ξ_b background crosscheck. Parameter $r^{-1}_{J/\psi}$.

	Nominal	w/o Ξ_b	w/o $\Xi_b \to \Xi\psi(2S)$
R1-LL	0.9538 ± 0.0746	0.954 ± 0.0746	0.954 ± 0.0747
R1-DD	0.9061 ± 0.06	0.921 ± 0.0631	0.9062 ± 0.0603
R2p1-LL	0.9818 ± 0.0532	0.982 ± 0.0532	0.982 ± 0.0532
R2p1-DD	1.0278 ± 0.045	1.0471 ± 0.0472	1.0279 ± 0.045
R2p2-LL	0.9809 ± 0.0386	0.9807 ± 0.0386	0.9809 ± 0.0386
R2p2-DD	1.0152 ± 0.0322	1.0286 ± 0.0332	1.0152 ± 0.0322

Table 10.4: Comparison between different resonance fit configurations for the Ξ_b background crosscheck. Parameter $R^{-1}_{\psi(2S)}$.

	Nominal	w/o Ξ_b	w/o $\Xi_b \to \Xi\psi(2S)$
R1-LL	1.092 ± 0.247	1.084 ± 0.245	1.085 ± 0.245
R1-DD	1.204 ± 0.234	1.202 ± 0.233	1.205 ± 0.235
R2p1-LL	1.233 ± 0.199	1.234 ± 0.199	1.233 ± 0.199
R2p1-DD	1.074 ± 0.15	1.071 ± 0.149	1.078 ± 0.151
R2p2-LL	1.001 ± 0.126	1.001 ± 0.126	1.001 ± 0.126
R2p2-DD	1.003 ± 0.105	0.998 ± 0.104	1.002 ± 0.105

uncertainties are assigned to $R^{-1}_{\psi(2S)}$ either.

10.1.3 Alternative angular weights for $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$

As discussed in Section 7.3.2, the $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ simulation is weighted with the angular correction based on the $\Lambda_b^0 \to \Lambda^0 J/\psi$ measurement [107] due to the lack of the corresponding measurement in the $\psi(2S)$ channel. An alternative correction (referred to as the three-fold weights) based on the theory [108] can be applied to calculate the efficiency as a crosscheck. The resonance fits can be done with the efficiencies fixed to the nominal and the alternative values. The comparison of the resulting $r^{-1}_{J/\psi}$ and the $R^{-1}_{\psi(2S)}$ is shown in Figure 10.3. The difference between the two cases is negligible. No corresponding systematic uncertainty is assigned.

Due to the efficiencies being fixed during the fits, the uncertainties shown in Figure 10.3 incorporate the data statistical uncertainties but not the uncertainties associated with the efficiencies. The reason for this treatment is to avoid generating toy to estimate the alternative efficiency uncertainties, which cost computing resources.

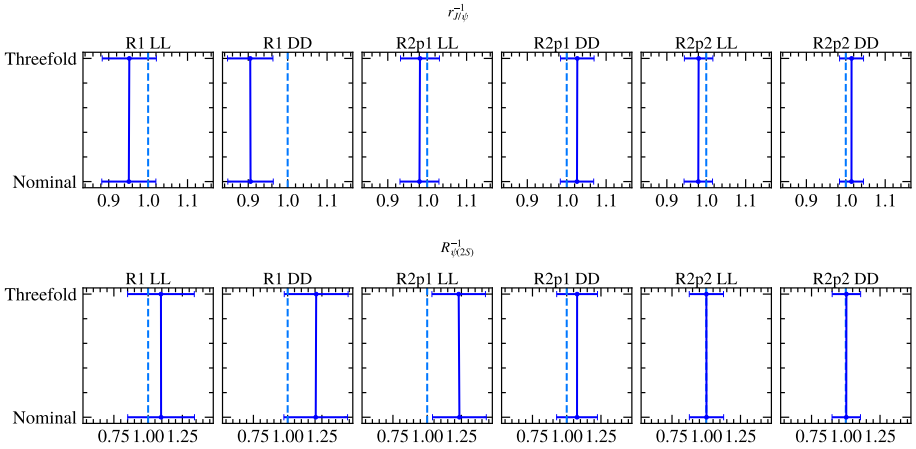


Figure 10.3: The comparison of the single ratio $r^{-1}_{J/\psi}$ (top plot) and the double ratio $R^{-1}_{\psi(2S)}$ (bottom plot) between the nominal scenario and the alternative scenario where the efficiencies are calculated with the three-fold angular weights for the $\Lambda_b^0 \to \Lambda^0\psi(2S)(\ell^+\ell^-)$ channels. The efficiencies are fixed during the fits.

10.1.4 $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ reconstructed in the J/ψ region

In the nominal resonance fit discussed in Section 9.3, the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ reconstructed in the $\psi(2S)$ region (leakage background) is modelled based on simulation and its yield is fitted simultaneously with the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ signal yield. Due to the much lower yield of $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$, its leakage in the J/ψ region is not included. As a crosscheck, such a leakage can be included in the fit with the shape fixed from simulation and the yield scaled from the signal $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ accounting for efficiencies. Figure 10.4 shows an example (R2p2-DD) of the resonance fit in the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ channel with this alternative configuration. The contribution of the $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ leakage is small. As shown in Figure 10.5, including such a component has almost negligible impact on $r^{-1}_{J/\psi}$ and $R^{-1}_{\psi(2S)}$. No corresponding systematic uncertainty is assigned.

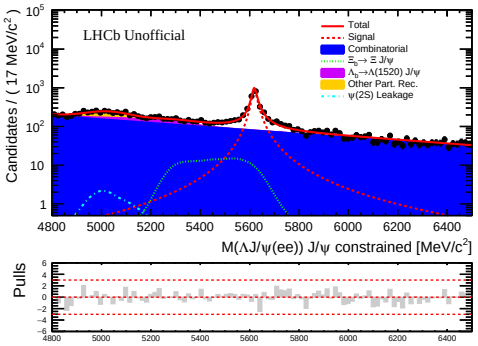


Figure 10.4: An example of resonance fit with the $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ leakage in the J/ψ region included. The plot shows the case of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ channel in the R2p2 DD category.

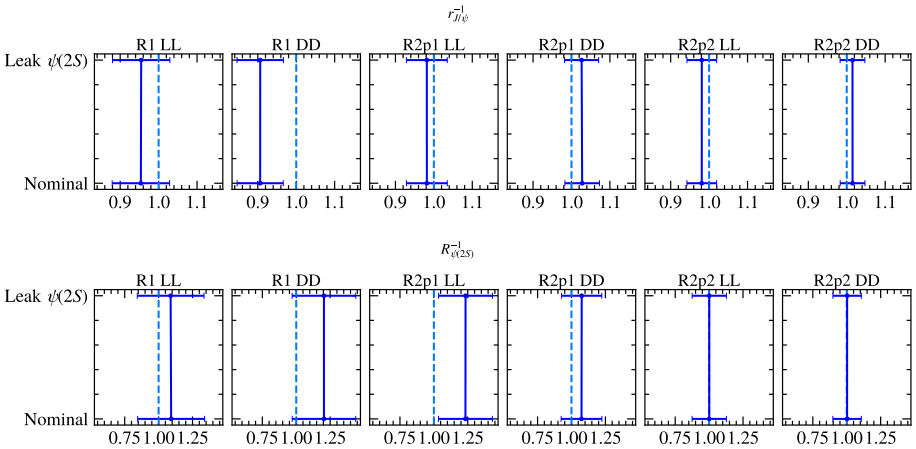


Figure 10.5: The comparison of $r^{-1}_{J/\psi}$ (top plot) and $R^{-1}_{\psi(2S)}$ (bottom plot) between the nominal case and the case including $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ leakage in the J/ψ region.

10.2 Misidentified hadronic background in the electron channel in high q^2 region

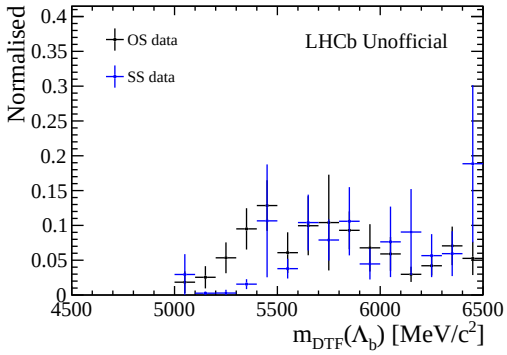
Given the limited statistics of the data sample in the electron channel, considerable uncertainties are associated with the shapes and yields extracted by the `Fail-To-Pass` method. The potential contribution from misidentified hadronic decays such as $\Lambda_b^0 \to \Lambda^0 h_1 h_2$ is modelled at the expense of introducing these uncertainties.

As discussed in Section 9.4.1, the extracted background component based on the `Fail-To-Pass` method also contains contributions of combinatorial origin. To check if the given statistics are able to resolve the existence of misidentified hadronic decays from the combinatorial contributions, a cross-check using the SS data can be performed (considering that the SS data are combinatorial).

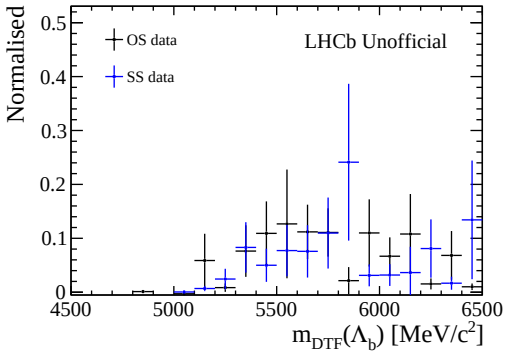
The `Fail-To-Pass` procedure can be applied to the SS data in a similar fashion as to the OS data with the modification that the subtraction of scaled signal MC is no longer needed. Figure 10.6 shows the comparisons of the resulting distributions between the OS and SS data samples in the high, central and low q^2 regions. These comparisons are shown in normalised histograms, considering that the absolute scales of the OS data and the SS data are different. Although considerable deviations can be seen in some bins, *e.g.* several bins on the lower side in the high q^2 DD category, the differences in most of the bins are small comparing to the scale of the fluctuations. This indicates that the majority of the contribution is of combinatorial origin.

Due to the above consideration, for the R_Λ^{-1} observable (in the high q^2 region), one way to preliminarily estimate the systematic uncertainty associated with this misidentified component from the `Fail-To-Pass` method is to perform toy studies using the alternative fit model without the component. In this alternative configuration, the misidentified component is expected to be, for a large part, absorbed in the usual combinatorial background, whose yield is allowed to be determined in the data fit. Following the same procedure described in Section 9.4.2, the alternative fit can be performed to the data of the muon and electron channels (with the electron channel blinded). Figure 10.7 shows the electron channel part of the fit. The free fit parameters determined during the data fit (excluding the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yields and efficiencies) are shown in Appendix F.

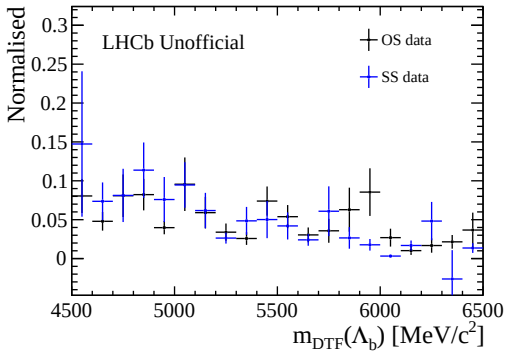
Table 10.5 shows, for the electron channel, the combinatorial background and the misidentified hadronic background yields in the nominal fit scenario (Section 9.4.2) comparing to the combinatorial background yields in the alternative fit (without the misidentified component). The combinatorial background yields are denoted as N_{comb}. The misidentified hadronic background



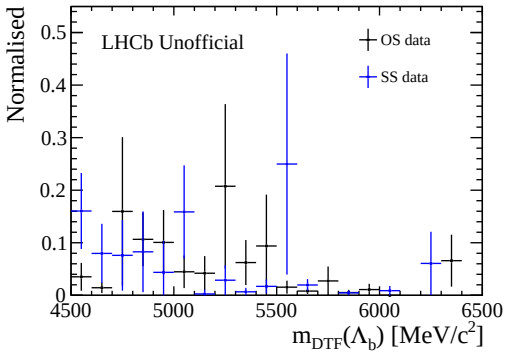
(a) High q^2 DD



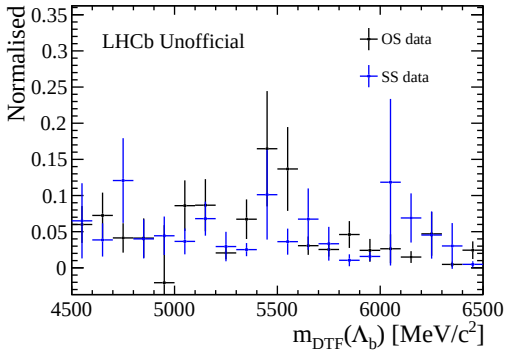
(b) High q^2 LL



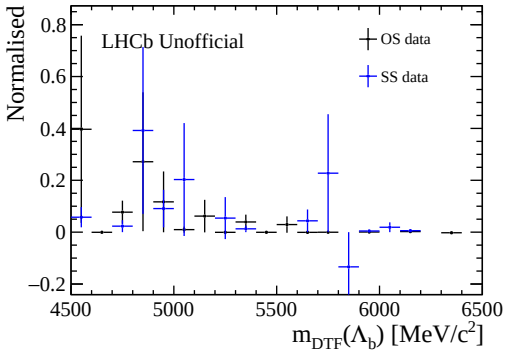
(c) Central q^2 DD



(d) Central q^2 LL

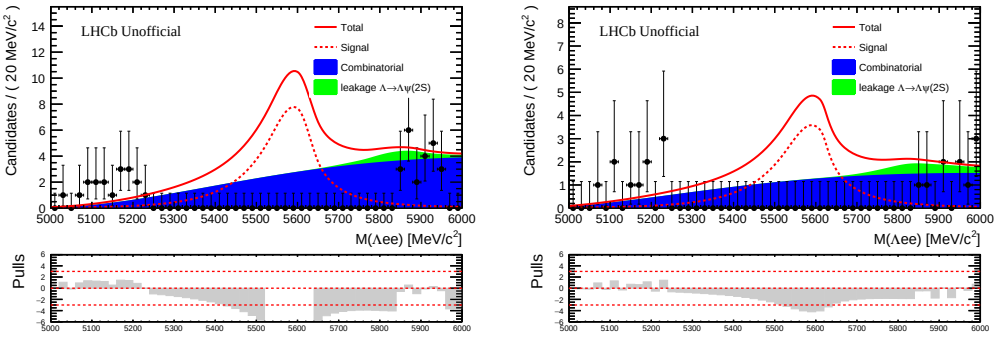


(e) Low q^2 DD



(f) Low q^2 LL

Figure 10.6: The comparison of the extracted distributions from the `Fail-To-Pass` method between the OS and SS data samples in the high, central and low q^2 regions.



(a) DD (b) LL

Figure 10.7: The $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ channel (blinded data) part of the alternative fit in the high q^2 region without the misidentified hadronic component.

yields are denoted as $N_{\text{Hadronic misID}}$. Figure 10.7 (comparing to Figure 9.14) and Table 10.5 indicates that the combinatorial background yields in the alternative fit increased to compensate for the removal of the misidentified hadronic background component. This reflects the aforementioned expected behaviour, *i.e.* the misidentified hadronic component would be, for a large part, absorbed in the combinatorial background in the alternative fit.

Table 10.5: The combinatorial background and misidentified hadronic background yields in the nominal fit scenario comparing to the combinatorial background yields in the alternative fit scenario (without misidentified hadronic component, denoted using "w/o $h \to e$") in the electron channel for the fits in the high q^2 region.

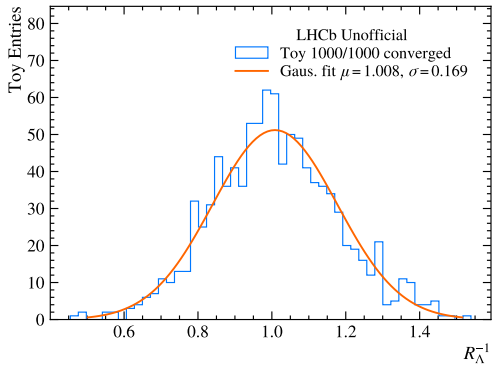
Nominal fit	
$N_{\text{Hadronic misID}}$ DD	43.2019
$N_{\text{Hadronic misID}}$ LL	13.6026
$N_{\text{comb}}(ee)$ DD	49.94 ± 18.56
$N_{\text{comb}}(ee)$ LL	36.82 ± 13.99
Alternative fit w/o $h \to e$	
$N_{\text{comb}}(ee)$ DD	102.53 ± 19.7
$N_{\text{comb}}(ee)$ LL	51.0 ± 14.02

Based on the alternative fit result, *i.e.* the fit configuration without the

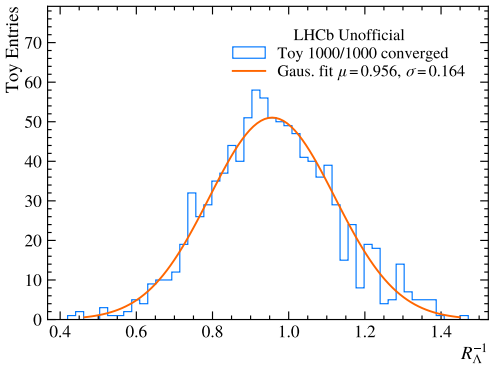
misidentified hadronic component, toy data can then be generated. Both the alternative fit model and the nominal fit model (including the misidentified hadronic component) are used to fit these generated toy data. [37] The resulting R_Λ^{-1} distributions (derived from 1000 sets of toy data) and the difference between them are shown in Figure 10.8. The distributions were fitted with Gaussian shapes. As shown in the plot, the estimated mean and standard deviation for the difference distribution are $\mu \sim 0.052$ and $\sigma \sim 0.008$. This gives $\sqrt{\mu^2 + \sigma^2} \sim 0.053$ [38], which can be considered as an estimation for the systematic uncertainty associated with the misidentified hadronic background in the high q^2 region. Comparing with the uncertainty estimated in Section 9.4.2 (0.158, including data statistical uncertainty and the uncertainties incorporated in the nuisance parameter constraints), this is still subdominant. It should also be noted that, this systematic uncertainty is not unrelated to the available statistics in the data. An increase of the statistics in the PID `Fail` region is expected to help improving the modelling of the misidentified hadronic background.

[37] This approach of generating toy data and refitting with alternative/nominal models follows the analysis in Ref. [77].

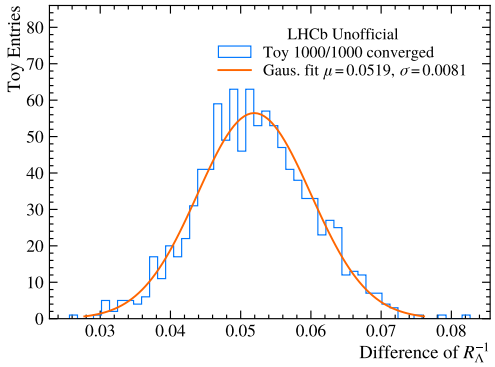
[38] This quantity reflects the uncertainty on the deviation. It was used in the analysis in Ref. [77]. When considering the shift itself as one Gaussian random variables (with standard deviation μ) and considering its fluctuation as another independent Gaussian random variable (with standard deviation σ), the variance of the sum of them is $\mu^2 + \sigma^2$.



(a) Fit without misidentified component



(b) Fit with misidentified component



(c) Difference between the two fits

Figure 10.8: The toy R_Λ^{-1} distributions (derived from 1000 sets of toy data) for the alternative fit model (without the misidentified component) and the nominal fit model. The distribution of the R_Λ^{-1} difference between the two configurations is also shown.

10.3 $\Lambda_b^0 \to \Lambda^* e^+ e^-$ background in the high q^2 region

In the nominal fits in the rare q^2 regions, the potential contribution from the partially reconstructed decay $\Lambda_b^0 \to \Lambda^* \ell^+ \ell^-$ is not included. Very little is known about these decays, especially for the electron channel (the $\Lambda_b^0 \to \Lambda(1520)\mu^+\mu^-$ branching fraction was measured [127]). For the muon channel, due to the good mass resolution (comparing to the electron channel), the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal peak is sufficiently separated from these partially reconstructed backgrounds. A narrow mass window can be applied for the muon channel fits excluding these backgrounds. Their effects on the muon mode are thus not further investigated. However, this is not the case for the electron mode, which has worse mass resolution.

Among the Λ^* particles, the heavier Λ^* modes are expected to lose more energy when decaying into Λ^0, and thus shift further to the lower side in the reconstructed Λ_b^0 mass spectrum, *i.e.* further away from the signal peak. In addition, for the high q^2 region, the heavier Λ^* modes are expected to have more limited phase-space (although the actual contributions also depend on the form factors). These two features are illustrated in Figure 10.9 using the MC samples produced with the fast simulation tool RapidSim [130]. The simulated decays are $\Lambda_b^0 \to \Lambda^* e^+ e^-$ with $\Lambda^* \to \Sigma^0(\to \Lambda^0\gamma)\pi^0$. They were simulated using a phase-space model for the decay. As shown in the left plot of Figure 10.9, from the Λ^0 (1405) mode to the Λ^0 (1690) mode (following the order of increasing mass), the combined invariant mass $m(\Lambda^0 e^+ e^-)$ distribution shifts further away from the known Λ_b^0 mass ($\sim 5620\,\mathrm{MeV}/c^2$ [64]). In the right plot of Figure 10.9, in the heavier Λ^* mode, the available phase space in the high q^2 region ($q^2 > 15\,\mathrm{GeV}^2/c^4$) is smaller.

Due to the above consideration, the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ decay is chosen to check the effects on the high-q^2 R_Λ^{-1} caused by these $\Lambda_b^0 \to \Lambda^* e^+ e^-$ decays. An alternative fit including the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ component can be performed in the same fashion as the nominal fit described in Section 9.4.2. The $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ shape is modelled using the MC sample (full simulation) with the kernel estimation technique [124, 125]. The yields for the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ contribution are allowed to be determined from the data fit. Figure 10.10 shows the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ channel (blinded data) part of the alternative fit. The free fit parameters determined during the data fit (excluding $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yields and efficiencies) are shown in Appendix F.

Toy data can then be generated using the alternative fit result and refit with both the alternative configuration and the nominal configuration. Figure 10.11 shows the resulting distributions for R_Λ^{-1} and the difference between the two configurations. The difference distribution exhibits an asymmetric behaviour. Using an similar method as described in Section 10.2,

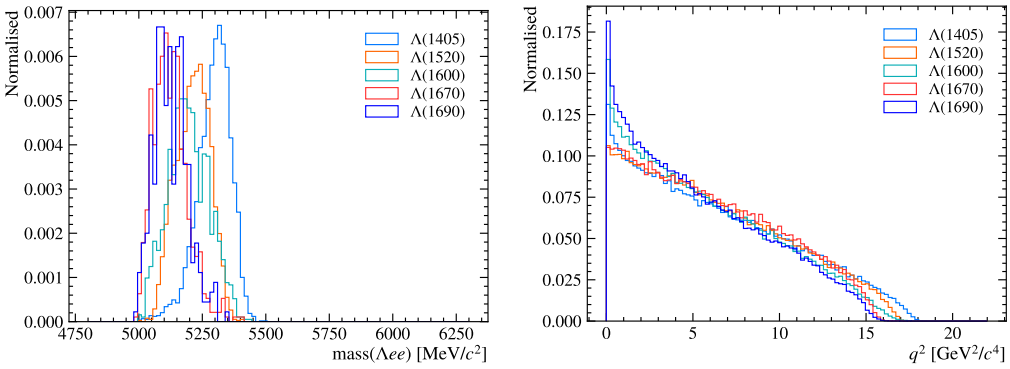
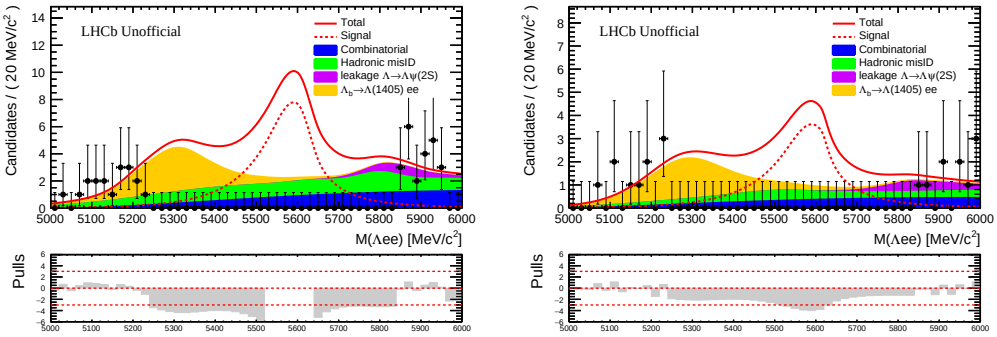
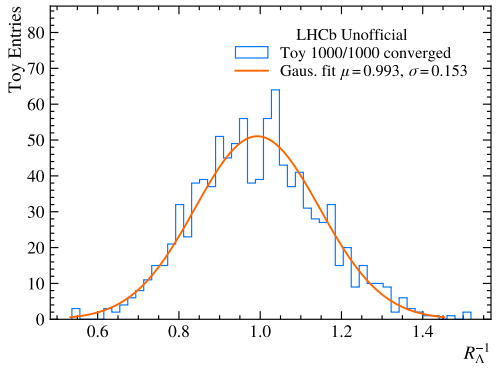


Figure 10.9: The left plot shows the distributions of the combined invariant mass $m(\Lambda^0 e^+ e^-)$ for several simulated $\Lambda_b^0 \to \Lambda^* e^+ e^-$ ($\Lambda^* \to \Sigma^0(\to \Lambda^0\gamma)\pi^0$) decays in the high q^2 region. The right plot shows the q^2 distributions for these samples. The simulation samples were produced with RapidSim [130]. They are phase-space simulations.

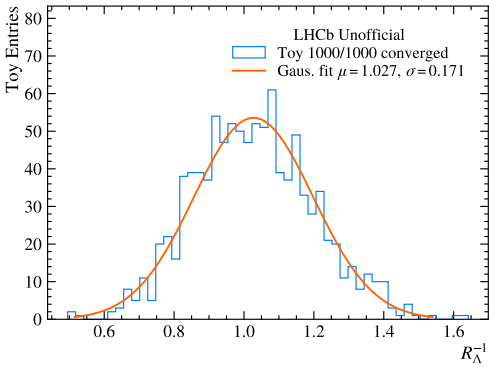


(a) DD (b) LL

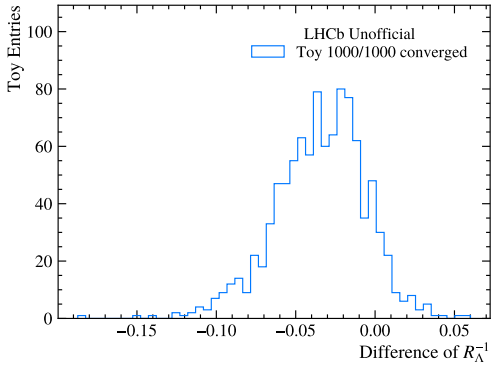
Figure 10.10: The $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ channel (blinded data) part of the alternative fit in the high q^2 region (including the $\Lambda_b^0 \to \Lambda(1405) e^+ e^-$ component).



(a) Fit with the $\Lambda(1405)$ component



(b) Fit without the $\Lambda(1405)$ component



(c) Difference between the two fits

Figure 10.11: The comparison of the toy R_Λ^{-1} distributions between the alternative (with the $\Lambda(1405)$ component) and the nominal fit configurations.

replacing the Gaussian-fitted μ with the sample mean and the σ with the sample standard deviation, the resulting uncertainty is about 0.045. This is small comparing to the expected statistical uncertainty (combined with uncertainties incorporated in nuisance parameter constraints) on R_Λ^{-1} (0.158, as discussed in Section 9.4.2). It should be noted that, due to the limited statistics in the electron mode data and the blinding, the estimation of the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ yields in the real data fit discussed above is not ideal. A considerable part of the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ component shape appears in the blinded region. An unblinding is likely to help the estimation of the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ yields.

10.4 Systematic uncertainty summary and plans

Following the procedure of this analysis and benefiting from the experiences of other similar studies (*e.g.* Refs. [74–77]), the already estimated systematic uncertainties and the plans for further investigations in the future are discussed in this section.

The already estimated uncertainties of R_{Λ}^{-1} (in the high q^2 region) is summarised in Table 10.6. These uncertainties are explained as follows.

- The "Data stat. + nuis. para." term refers to the data statistical uncertainties (assuming LFU) and the uncertainties incorporated in the fit as nuisance parameter constraints. These nuisance parameters include the signal efficiencies of $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ ($\ell = e, \mu$), $r_{J/\psi}^{-1}$ and the yields of the $\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ leakage background. In the case of efficiencies, the assigned uncertainties correspond to the limited statistics in the MC samples. In the case of $r_{J/\psi}^{-1}$ and the $\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ leakage yields, the assigned uncertainties correspond to the data statistics in the resonance region and the resonance mode efficiency uncertainties induced by the limited MC statistics.

- The "$h \to e$ background" term refers to the misidentified hadronic background in the electron channel. As the nominal approach, it is treated with the `Fail-To-Pass` method. The result of the method is limited by the data statistics in the inverted-PID selected regions. As discussed in Section 10.2, the systematic uncertainty estimation does not directly address the uncertainty of the result given by the `Fail-To-Pass` method. Instead, it employs a conservative approach, *i.e.* performing the fits with and without the background component in question.

- The "$\Lambda_b^0 \to \Lambda(1405)e^+e^-$ background" terms refers to the partially reconstructed background from the $\Lambda_b^0 \to \Lambda^* e^+ e^-$ decays. As discussed in Section 10.3, the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ decay is chosen to represent the $\Lambda_b^0 \to \Lambda^* e^+ e^-$ decays due to its proximity to the signal peak in the reconstructed Λ_b^0 mass spectrum and its larger (relative to the heavier Λ^* mode) available phase-space in high q^2 region. The systematic uncertainty is estimated by performing the fits with and without the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ component.

Combining these terms using uncorrelated uncertainty propagation method, the combined uncertainty is 0.173, which is dominated by the "Data stat. + nuis. para." term.

Table 10.6: An overview of the already estimated uncertainties for the R_{Λ}^{-1} observable in the high q^2 region.

Uncertainty origin	Assigned uncertainties
Data stat. + nuis. para.	0.158
$h \to e$ background	0.053
$\Lambda_b^0 \to \Lambda(1405)e^+e^-$ background	0.045
Combining the above terms	0.173

Besides the already estimated uncertainties, the fit models of the signal and background components can also constribute to the uncertainties. The plan for these uncertainties is to perform fits with alternative models. Their effects are expected to be small comparing to the already estimated uncertainty.

As discussed in Section 9.4.2, the discrepancy between simulation and data in the $\Lambda_b^0 \to \Lambda^0\ell^+\ell^-$ signal shapes in the Λ_b^0 mass spectrum is not yet accounted for, although its effect is expected to be small comparing to the already estimated uncertainty. The mass calibration procedure often used in other similar analyses are difficult to apply in this analysis due to the limited data statistics in the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ channel. Another possible way to approach this discrepancy is to assign the shape correction factors (mean-shift and width-scaling factors) to typical values based on the experiences of other similar analyses and perform toy studies to evaluate systematic uncertainties.

A similar discrepancy between data and simulation can also exist in the q^2 variable. This is expected to be a small effect comparing to the already estimated uncertainty. Similar strategy as the treatment for the Λ_b^0 mass variable can be employed.

For the uncertainties associated with the efficiencies, besides the already considered MC sample statistics contributions, the MC correction weights can also contribute. It should be noted that, considering the robustness of the double ratio against these corrections (as discussed in Section 10.1) as well as the experiences of the other similar studies (*e.g.* Ref. [77]), the uncertainties associated with the MC corrections are expected to be very small comparing to the already estimated uncertainty.

Nevertheless, for completeness, a plan for evaluating the systematic uncertainties associated with the MC corrections is explained as follows. These uncertainties can be estimated by calculating alternative efficiencies with alternative MC corrections. To propagate these efficiency uncertainties to the ratio observable R_{Λ}^{-1}, the plan is to add nuisance parameters to the fitting procedure of R_{Λ}^{-1} determination. In the nominal fitting procedure, the

$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal yield is parameterised with the R_Λ^{-1} observable, the efficiencies, the single ratio $r_{J/\psi}^{-1}$ and the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ signal yield (as shown in Equation 9.37). The single ratio $r_{J/\psi}^{-1}$ also depends on the efficiencies. In the following discussion, the alternative efficiency is denoted with $\overline{\epsilon_{\text{sig}}}$ and the alternative $r_{J/\psi}^{-1}$ is denoted with $\overline{r_{J/\psi}^{-1}}$. Assuming that the efficiency-dependence of $r_{J/\psi}^{-1}$ is captured by the efficiency ratio between $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$, the alternative version of Equation 9.37 can be written as:

$$\begin{aligned} N_{\text{sig}}(ee)|_{y,t} = R_\Lambda^{-1} \times \left(\frac{\overline{\epsilon_{\text{sig}}}(ee)}{\overline{\epsilon_{\text{sig}}}(\mu\mu)} N_{\text{sig}}(\mu\mu) \overline{r_{J/\psi}^{-1}} \Big|_{y,t} \right) = \\ R_\Lambda^{-1} \times \left(\frac{\overline{\epsilon_{\text{sig}}}(ee)}{\overline{\epsilon_{\text{sig}}}(\mu\mu)} \times \frac{\overline{\epsilon_{\text{sig}}}(J/\psi,\mu\mu)/\epsilon_{\text{sig}}(J/\psi,\mu\mu)}{\overline{\epsilon_{\text{sig}}}(J/\psi,ee)/\epsilon_{\text{sig}}(J/\psi,ee)} N_{\text{sig}}(\mu\mu) r_{J/\psi}^{-1} \Big|_{y,t} \right) , \end{aligned} \tag{10.1}$$

where y is used to denote the different data-taking periods, t is used to denote the $\Lambda^0 \to p\pi^-$ track types, (ee) is used to denote the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ mode, $(\mu\mu)$ is used to denote the $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ mode, $(J/\psi, ee)$ is used to denote the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ mode and $(J/\psi, \mu\mu)$ is used to denote the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ mode. A correction factor (C) can be defined using the alternative and the nominal efficiencies:

$$C = \frac{\overline{\epsilon_{\text{sig}}}(ee)/\epsilon_{\text{sig}}(ee)}{\overline{\epsilon_{\text{sig}}}(\mu\mu)/\epsilon_{\text{sig}}(\mu\mu)} \times \frac{\overline{\epsilon_{\text{sig}}}(J/\psi,\mu\mu)/\epsilon_{\text{sig}}(J/\psi,\mu\mu)}{\overline{\epsilon_{\text{sig}}}(J/\psi,ee)/\epsilon_{\text{sig}}(J/\psi,ee)} . \tag{10.2}$$

With the correction factor, Equation 10.1 can be rewritten as:

$$N_{\text{sig}}(ee)|_{y,t} = R_\Lambda^{-1} \times \left(C \frac{\epsilon_{\text{sig}}(ee)}{\epsilon_{\text{sig}}(\mu\mu)} N_{\text{sig}}(\mu\mu) r_{J/\psi}^{-1} \Big|_{y,t} \right) \tag{10.3}$$

Comparing to Equation 9.37, the nominal fit described in Section 9.4.2 can be considered as the special case with $C = 1$. Alternative fits can be performed with the correction factors C as additional nuisance parameters. When the alternative efficiencies are calculated with a single alternative configuration, the parameters C can be fixed to the corresponding values. When the alternative efficiencies are calculated with a bootstrapping method, the parameters C can be Gaussian constrained. The possible options for the various MC corrections include, for example, calculating alternative efficiencies without the correction in question and bootstrapping the correction maps.

The above discussion focuses on the R_Λ^{-1} observable in the high q^2 region. For the low and central q^2 regions, the observable is the following relative

branching fraction:

$$r_{ee}^{\mathcal{B}} = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+ e^-)}{\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))} \ . \tag{10.4}$$

Similar as the R_Λ^{-1} in the high q^2 region, the data statistical uncertainty and the uncertainties already assigned to the nuisance parameters are incorporated in the procedure described in Section 9.4.3. For the low q^2 region, the nuisance parameters are the normalisation factors (Equation 9.42). For the central q^2 region, the nuisance parameters are the normalisation factors and the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ leakage yields. The assigned uncertainties for the normalisation factors and the leakage yields correspond to the data statistical uncertainty in the resonance regions and the efficiency uncertainties caused by the limited MC sample statistics.

For the low and central q^2 regions, a technical difference with respect to the high q^2 region is the CL_s limit-setting procedure. Although it can incorporate the systematic uncertainties associated with constrained nuisance parameters, the procedure to treat the backgrounds as described in Section 10.2 and Section 10.3 cannot be directly integrated into the CL_s method (as far as the author is aware). A plan for the treatment of these backgrounds is to perform the CL_s limit setting with the different fit configurations and using the most conservative limit (the up limit with the highest value) as the result. Similar treatments can be applied to test alternative signal and background fit models. For the uncertainty treatments using nuisance parameters, the method used for R_Λ^{-1} can be adapted for $r_{ee}^{\mathcal{B}}$, which can then be incorporated into the CL_s method limit setting.

11 Summary

11.1 The time alignment of the SciFi Tracker

The SciFi Tracker was installed and commissioned as a part of the LHCb Upgrade I. The time alignment is indispensable for the detector to achieve good hit detection efficiency and ultimately the required tracking efficiency.

Following the discussion in Chapter 3 and Chapter 4, the method of the time alignment has been introduced and studies using the commissioning and early operation data have been presented. The method has demonstrated good sensitivity to detect time shift and the SciFi readout system has demonstrated good timing stability. Preliminary studies carried out by the collaborators indicated that the detector has achieved its design requirement in terms of the hit detection efficiency (as discussed in Section 4.5), which is the aim of the fine time alignment. This is a result of the collective effort of the commissioning and operation team, of which the time alignment is an important part.

While the overall hit detection performance has achieved this very positive result, further studies are needed to exactly understand the quantitative relationship between the time shift and the resulting hit efficiency.

11.2 The LFU study using the $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ decay

Tests of LFU in rare FCNC transitions such as $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ are sensitive probes for potential new physics beyond the Standard Model (SM).

The presented LFU study with $\Lambda_b^0 \to \Lambda^0 \ell^+ \ell^-$ utilises the data collected by LHCb during Run 1 and Run 2, corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$. The primary objective of the analysis is to measure R_Λ^{-1} in the high dilepton q^2 region. For the low and central q^2 regions, no signal observations are expected under the LFU assumption. In such a case, the secondary objective of the analysis is to evaluate the upper limit of the relative branching fraction between $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ and $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ (denoted as $r_{ee}^{\mathcal{B}}$) in the low and central q^2 region.

Two very important aspects for measuring R_Λ^{-1} (and estimating an upper limit on $r_{ee}^{\mathcal{B}}$) are the evaluation of the efficiency and the fit of the Λ_b^0 mass spectrum. To better control the systematic uncertainties associated with the efficiencies, the observable R_Λ^{-1} is constructed as a double ratio using the resonance modes $\Lambda_b^0 \to \Lambda^0 J/\psi(\ell^+\ell^-)$ for normalisation. For the same reason, in the low and central q^2 regions, $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ is used as the normalisation channel. The efficiencies are evaluated using the simulated samples (with corrections to improve the compatibility between data and simulation).

The fit of the Λ_b^0 mass spectrum is affected by the backgrounds. A dedicated MVA procedure was used to suppress the combinatorial background (Chapter 8) and two background-associated systematic uncertainties of R_Λ^{-1} are estimated (Section 10.2 and Section 10.3).

As a crosscheck of the analysis procedure, the resonance ratios $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ are evaluated. They are found to be compatible with the expected SM value (1.0).

In order to avoid biasing the analysis procedure, a blinding strategy is employed (as discussed in Section 6.3).

In the high q^2 region, the signal region of the electron channel data is blinded. For the preliminary sensitivity estimation of R_Λ^{-1} in the high q^2 region, the LFU assumption is used. A summary of the already estimated uncertainties (under LFU assumption) is discussed in Section 10.4, which is reiterated as follows [39] (x denotes the unknown value due to the blinding):

$$R_\Lambda^{-1}|_{q^2>15\,\mathrm{GeV}^2/c^4} = x \pm 0.158 \pm 0.053 \pm 0.045\ . \tag{11.1}$$

In the above formula, the first uncertainty term (0.158) is the dominating uncertainty, which corresponds to the data statistical uncertainty (under the LFU assumption) combined with the uncertainties assigned to the fit nuisance parameters. The second uncertainty term (0.053) corresponds to the misidentified hadronic background component in the electron channel. The third uncertainty term (0.045) corresponds to the $\Lambda_b^0 \to \Lambda^* \ell^+ \ell^-$ background in the electron channel. The electron channel data statistics contribute significantly to the first uncertainty term (0.158). Based on Table 9.5, combining three data-taking periods and two $\Lambda^0 \to p\pi^-$ track types, the total high q^2 region $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal yield is expected to be about 117. Taking the naive Poisson uncertainty ($\sqrt{117}$), the relative uncertainty of the electron channel signal yield would be ~ 0.09, which is already a significant fraction of 0.158.

In the low and central q^2 regions, the signal region of the electron channel data is blinded. The preliminary expected upper limits (90 percent confidence level) on $r_{ee}^{\mathcal{B}}$ under background-only hypothesis are discussed in Section 9.4.3. They are:

$$r_{ee}^{\mathcal{B}} \lesssim \begin{cases} 5.2 \times 10^{-3}\ , \ 1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4 \\ 3.4 \times 10^{-3}\ , \ 0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4 \end{cases}\ . \tag{11.2}$$

[39] Reminder: The high q^2 region definition used in this analysis is $15 - 22\,\mathrm{GeV}^2/c^4$. The upper bound is higher than the phase-space limit to accommodate the q^2 resolution effect. For the theoretical interpretation, this is effectively $q^2 > 15\,\mathrm{GeV}^2/c^4$.

Taking

$$\mathcal{B}(\Lambda_b^0 \to \Lambda^0 J/\psi) = (3.34 \pm 0.02 \pm 0.10 \pm 0.08 \pm 0.28) \times 10^{-4} \tag{11.3}$$

from Ref. [85] and $\mathcal{B}(J/\psi \to e^+e^-) = (5.971 \pm 0.032) \times 10^{-2}$ from Ref. [64], one can translate the upper limits on $r_{ee}^{\mathcal{B}}$ into the upper limits on $\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ (a naive combination without incorporating the uncertainties of these external inputs):

$$\mathcal{B}(\Lambda_b^0 \to \Lambda^0 e^+e^-) \lesssim \begin{cases} 1 \times 10^{-7} \ , \ 1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4 \\ 7 \times 10^{-8} \ , \ 0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4 \end{cases} . \tag{11.4}$$

It was not possible to completely cover all details of the analysis within this thesis. As discussed in Section 10.4, further systematic uncertainty studies are needed before the analysis can be finished.

The sensitivities on the observables are limited by data statistics, especially by the $\Lambda_b^0 \to \Lambda^0 e^+e^-$ statistics. Although copious Λ_b^0 baryons are produced by the LHC, the $\Lambda_b^0 \to \Lambda^0 \ell^+\ell^-$ process is extremely rare, which is one of the reasons why it is interesting (suppressed in the SM). In addition, the statistics suffer from several limiting factors in the efficiencies, *e.g.* the reconstruction of the long-lived Λ^0, the reconstruction of the electrons (affected by the bremsstrahlung emission) and the L0 trigger (especially for the low and central q^2 region electrons).

Since the beginning of Run 3 (2022), the LHCb Upgrade I detector has collected data corresponding to more than $22\,\mathrm{fb}^{-1}$ [131]. By the end of Run 4, the LHCb experiment is expected to have accumulated data corresponding to a total integrated luminosity of about $50\,\mathrm{fb}^{-1}$ [13, 132]. The increased luminosity together with the upgrades such as the removal of the L0 trigger is expected to greatly benefit the precision tests in rare processes such as $\Lambda_b^0 \to \Lambda^0 \ell^+\ell^-$ (among many other measurements currently limited by statistics).

Appendix

A Further details on the time alignment

Additional examples of the delay-occupancy curves in separate y-regions are shown in Figure A.1 - Figure A.6.

Additional illsutrations of the asymmetry t_0^{MC} and the optimised working point in the overlay with the delay-occupancy curves are shown in Figure A.7 (T1L0M3) and Figure A.8 (T1L0M4).

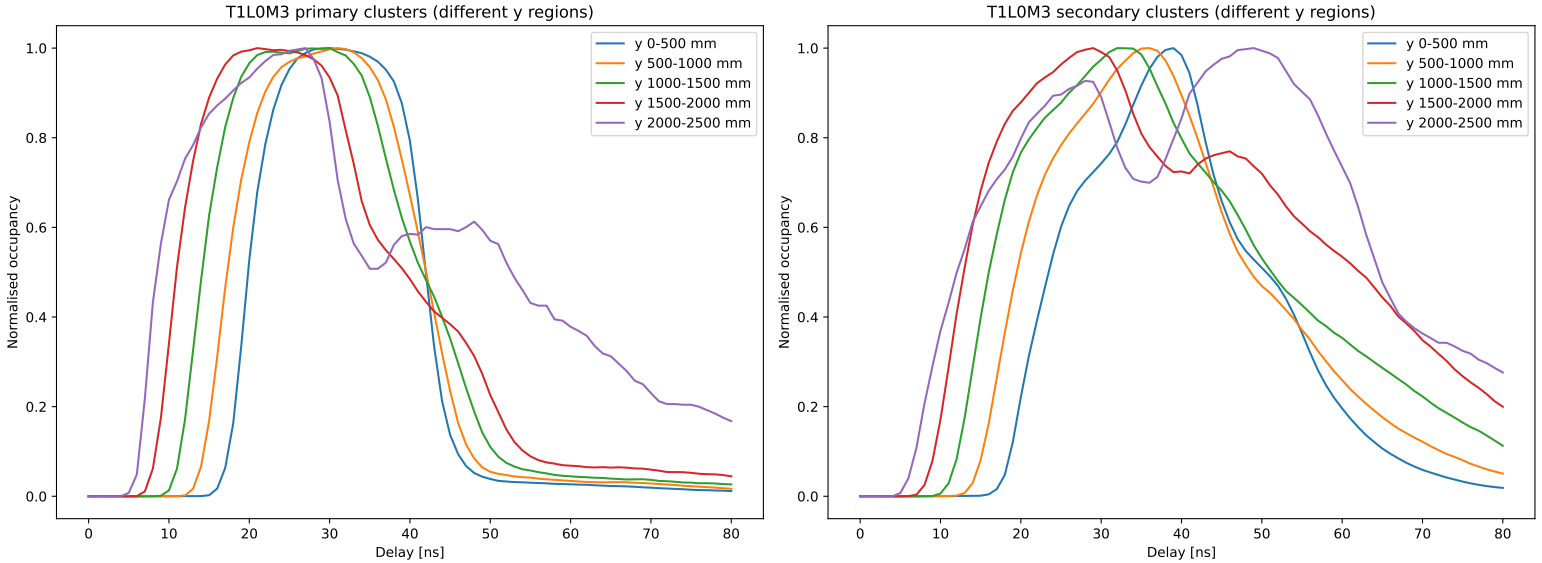


Figure A.1: An example (T1L0M3) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y such that their peak y-value are normalised to 1. The correspond half-modules in the 4 quarters are merged.

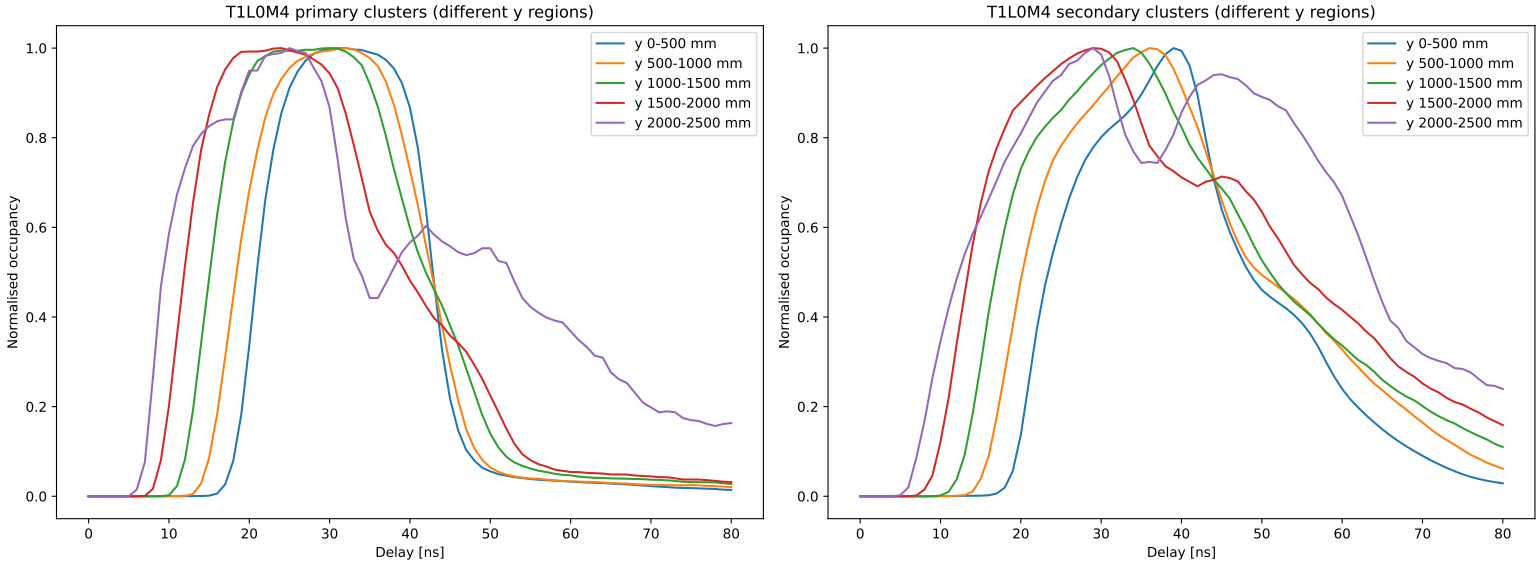


Figure A.2: An example (T1L0M4) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y such that their peak y-value are normalised to 1. The correspond half-modules in the 4 quarters are merged.

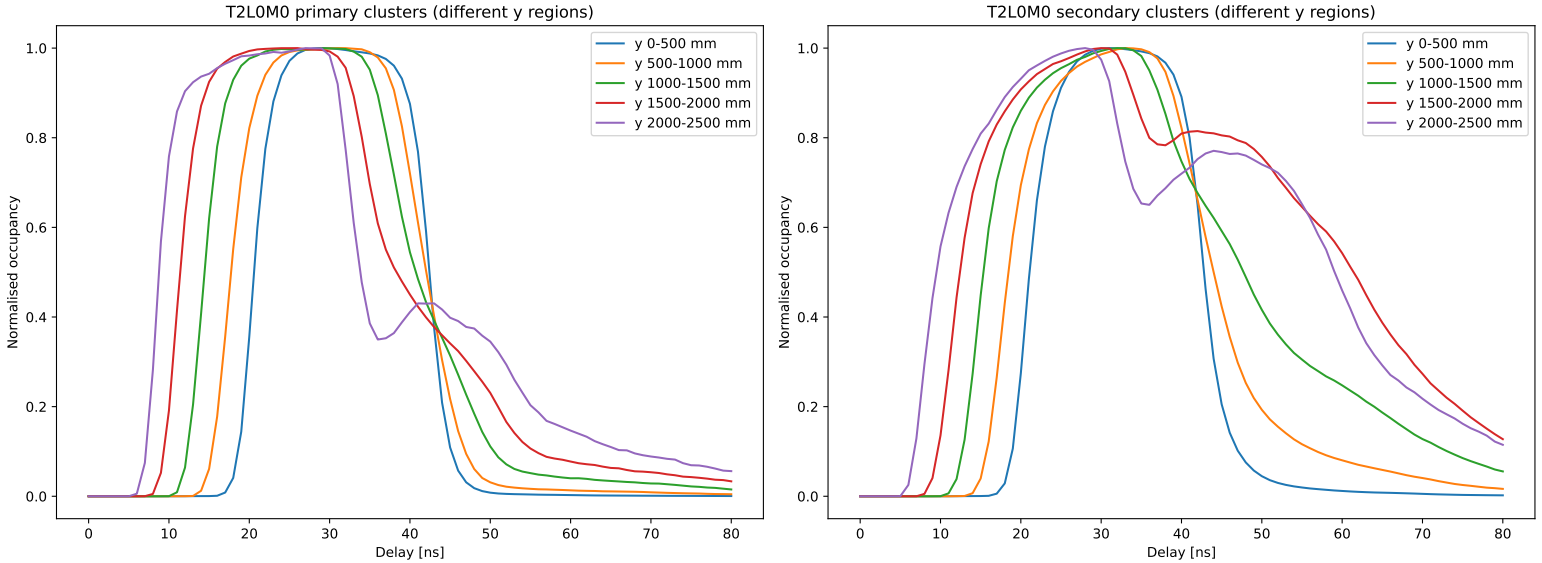


Figure A.3: An example (T2L0M0) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y such that their peak y-value are normalised to 1. The correspond half-modules in the 4 quarters are merged.

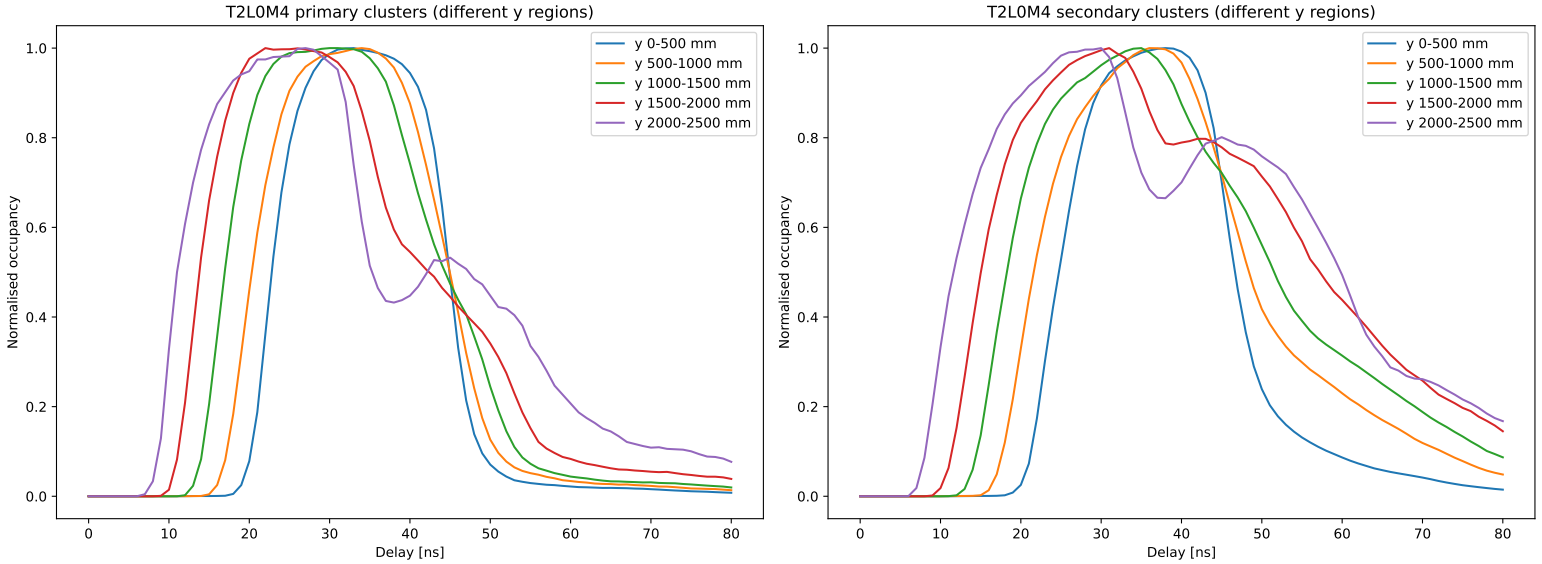


Figure A.4: An example (T2L0M4) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y such that their peak y-value are normalised to 1. The correspond half-modules in the 4 quarters are merged.

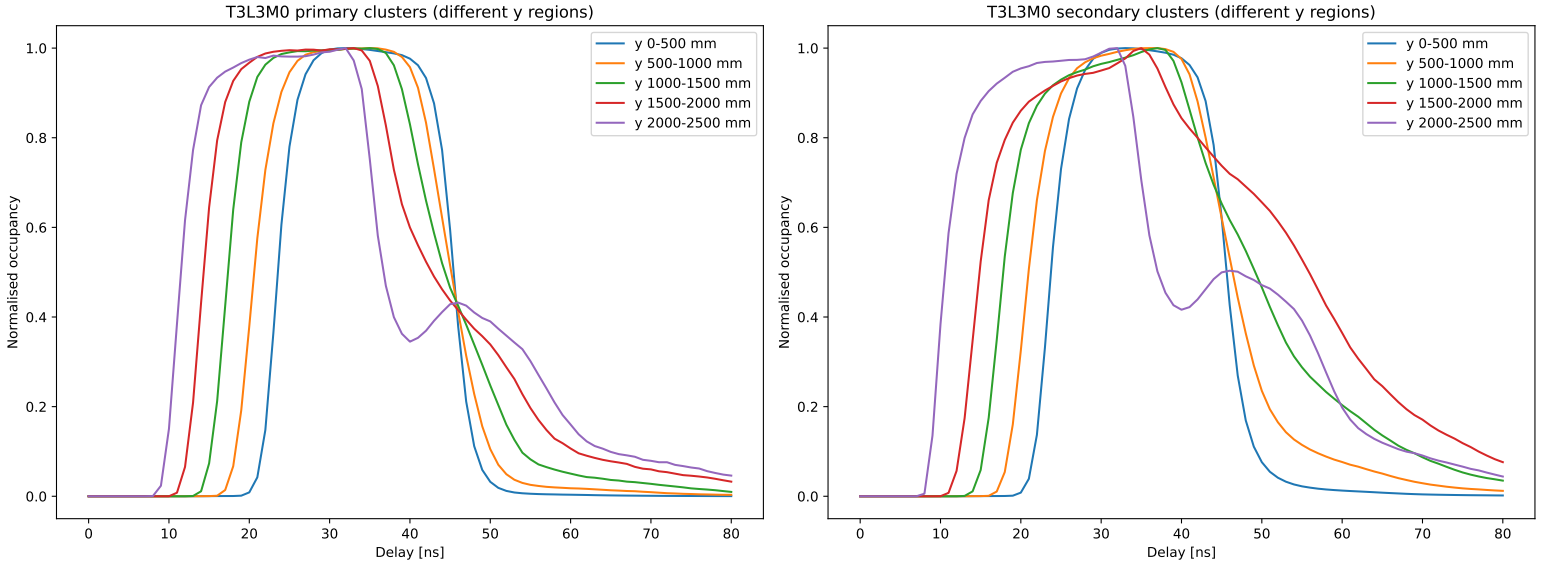


Figure A.5: An example (T3L3M0) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y such that their peak y-value are normalised to 1. The correspond half-modules in the 4 quarters are merged.

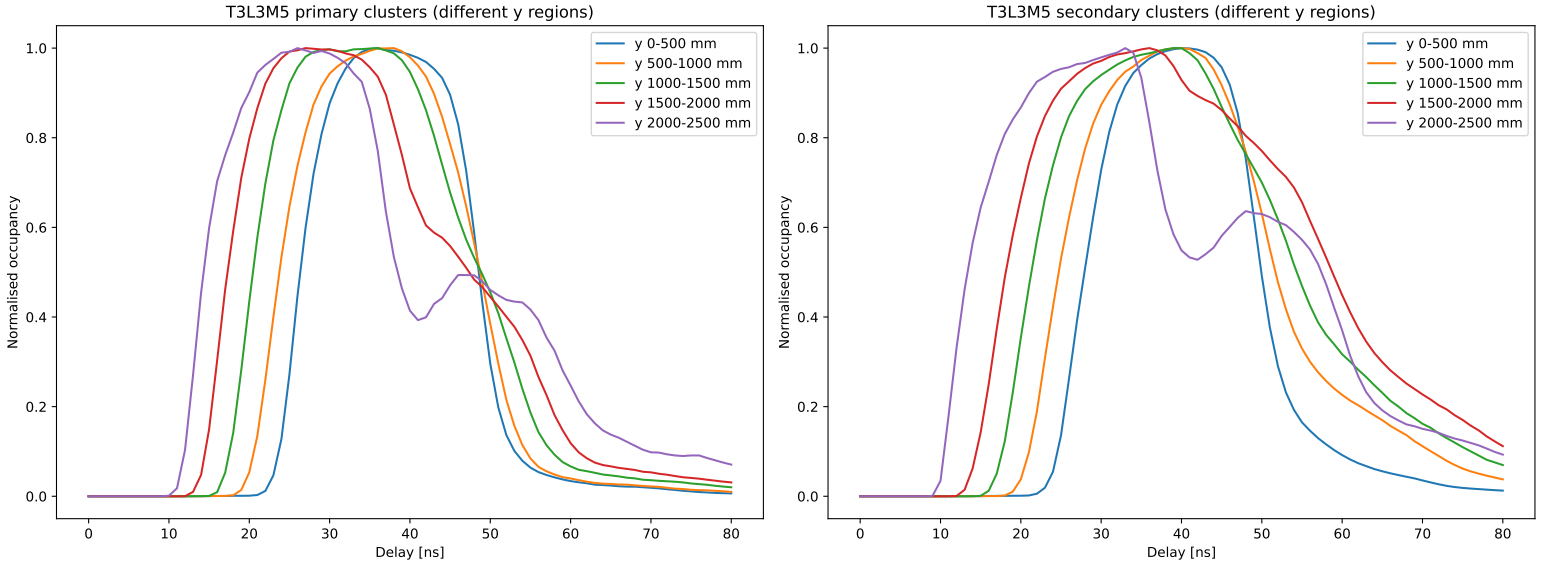


Figure A.6: An example (T3L3M5) of the simulated delay-occupancy curves of the primary (left) and secondary (right) clusters divided in different y-regions. The curves are scaled in y such that their peak y-value are normalised to 1. The correspond half-modules in the 4 quarters are merged.

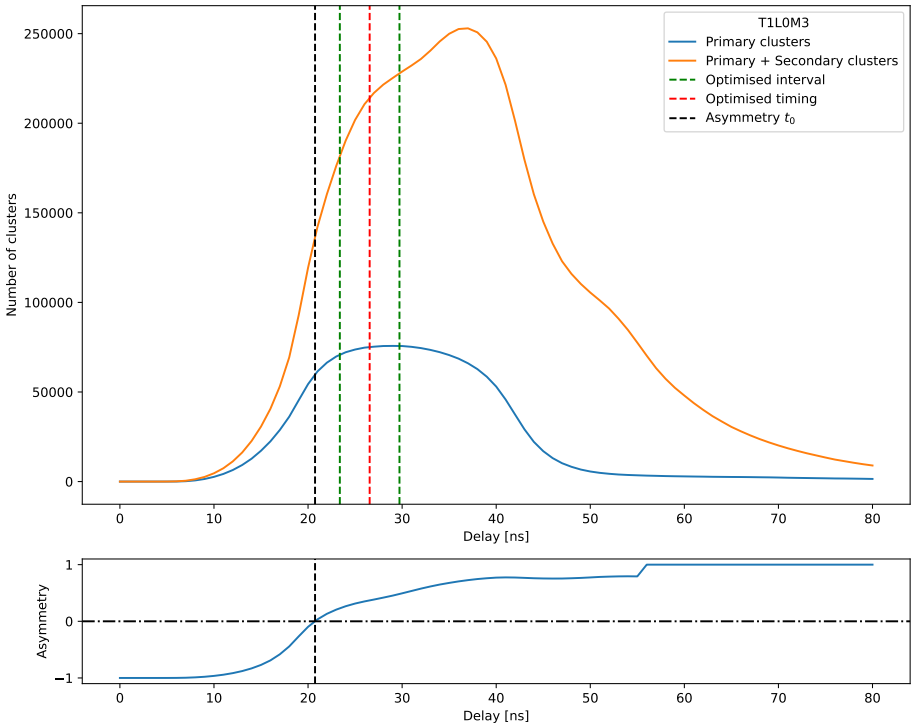


Figure A.7: An additional illsutration (T1L0M3) of the asymmetry t_0^{MC} and the optimised working point in the overlay with the delay-occupancy curves based on the MC samples. The optimised interval corresponds to the interval defined by the 0.96 line as discussed in Section 3.4.2. In this plot, the optimised working point was derived with the $1:1:1:1:1$ weight configuration and the corresponding half-modules in the 4 quarters are merged.

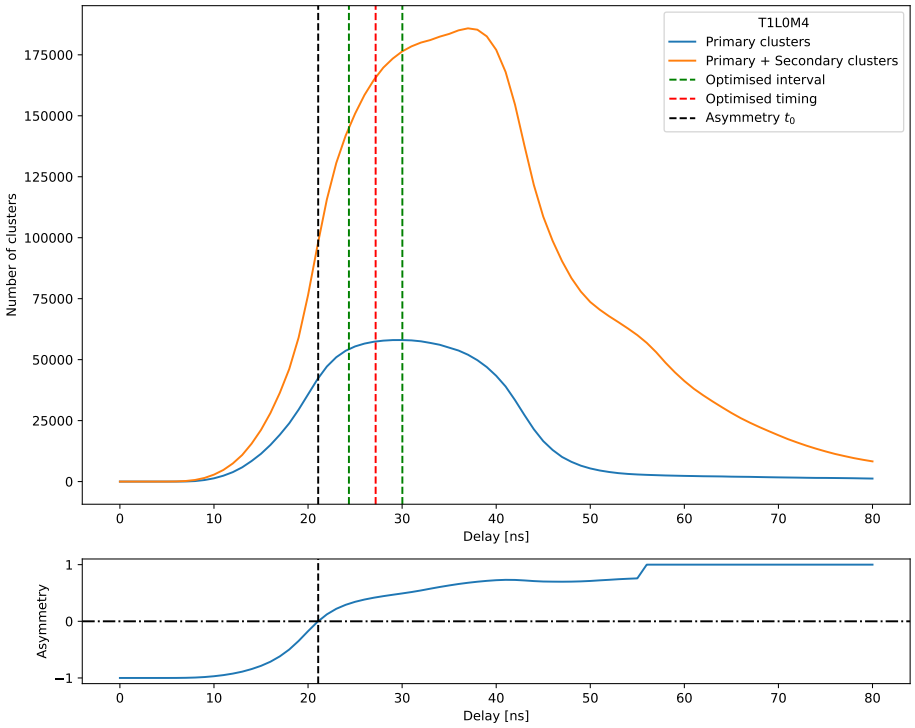


Figure A.8: An additional illsutration (T1L0M4) of the asymmetry t_0^{MC} and the optimised working point in the overlay with the delay-occupancy curves based on the MC samples. The optimised interval corresponds to the interval defined by the 0.96 line as discussed in Section 3.4.2. In this plot, the optimised working point was derived with the $1:1:1:1:1$ weight configuration and the corresponding half-modules in the 4 quarters are merged.

B Further details on the selections

As an example (version 34r0p1 for 2018), the details of the Stripping selection are shown in Table B.1 based on Refs [98–101],

An overview of the customised offline selection (which is adjusted based on the previous work of collaborators [36]) are shown in Table B.2 and Table B.3. The details of the HLT1 trigger lines are shown in Table B.4.

Some variables and cuts used in the tables are explained as follows.

- nSPDHits: the number of SPD hits.
- χ^2/n.d.f: the χ^2 per degree of freedom.
- χ^2_{IP}: the Impact Parameter (IP) χ^2 of a track relative to a vertex. It is calculated as the difference of the vertex fit χ^2 with or without the track.
- χ^2_{FD}: the Flight Distance (FD) significance (FD in the unit of χ^2).
- DIRA: the cosine of the angle between the momentum vector and the displacement vector.
- DOCA: Distance of closest approach of two tracks.
- `HasMuon`: the track is registered by the muon stations.
- `HasRich`: the track is registered by the RICH.
- `HasCalo`: the track is registered by the calorimeter.
- z_{endvtx}: the z (the axis along the beam-pipe) coordinate of the particle end vertex.
- `IsMuon`: as discussed in Section 2.2.1.
- `CombDLL`e: as discussed in Section 7.1.
- `ProbNN`$e(\mu)$: as discussed in Section 7.1.
- GhostProb: a MVA-based measure of a track being a ghost track.
- TT sensor veto: discussed in Ref. [36].
- HOP mass. As discussed in Ref. [133], M_{HOP} is the invariant mass of Λ_b^0 with some corrections for the bremsstrahlung.

Table B.1: The details of the Stripping selections of the version 34r0p1 [98–101], which is used for the year 2018. The exact cuts can vary for other versions or data taking years. For example, the Run 1 selection has nSPDHits < 600 instead of < 450.

Particle or variable	`Bu2LLK_ee(SS)Line2`	`Bu2LLK_mm(SS)Line`
nSPDHits	< 450	
Λ_b^0	$\lvert m - m_{\mathrm{PDG}}(B^+)\rvert < 1500\,\mathrm{MeV}/c^2$ End vertex χ^2/n.d.f < 9 χ^2_{IP} from PV < 25 DIRA relative to PV > 0.9995 χ^2_{FD} from PV > 100	
Λ^0 LL	$p_T > 400\,\mathrm{MeV}/c$ $m < 2600\,\mathrm{MeV}/c^2$ $\lvert m - m_{\mathrm{PDG}}(\Lambda^0)\rvert < 35\,\mathrm{MeV}/c^2$ DOCA χ^2 of daughter tracks < 30 End vertex $\chi^2 < 30$ χ^2_{FD} from PV > 4	
Λ^0 DD	$p_T > 400\,\mathrm{MeV}/c$ $m < 2600\,\mathrm{MeV}/c^2$ $\lvert m - m_{\mathrm{PDG}}(\Lambda^0)\rvert < 64\,\mathrm{MeV}/c^2$ DOCA χ^2 of daughter tracks < 25 End vertex $\chi^2 < 25$	
p, π	χ^2_{IP} from PV > 9 $p > 2\,\mathrm{GeV}/c$	
$\ell_1\ell_2$	$p_T > 0\,\mathrm{MeV}/c$ measured $m < 5500\,\mathrm{MeV}/c^2$ End vertex χ^2/n.d.f < 9 χ^2_{FD} from PV > 16 χ^2_{IP} from PV > 0	
ℓ	χ^2_{IP} from PV > 9 $p_T > 350(500)\,\mathrm{MeV}/c$ `CombDLL`$e > 0$	`HasMuon` & `IsMuon`

Table B.2: The customised offline selection (part 1).

Particles	Cuts	Comments
Λ^0	$\lvert m - m_{PDG}\rvert < 10\,\mathrm{MeV}/c^2$	
	$p_T(\Lambda^0) > 400\,\mathrm{MeV}/c$	
	$\chi^2_{\mathrm{FD}}(\mathrm{PV}) > 0$	
	$0 < z_{\mathrm{endvtx}} < 2250\,\mathrm{mm}$	
	$0.0005 < \tau < 2\,\mathrm{ns}$	
	$\frac{\tau}{\Delta\tau} > 6$	
	DIRA(PV)> 0	
	Only LL $z_{endvtx} < 640\,\mathrm{mm}$	
p, π and ℓ	$\chi^2_{\mathrm{IP}}(\mathrm{PV}) > 9$	
LL p, π and ℓ	Track χ^2/n.d.f < 4, GhostProb< 0.4	
DD p, π	Track χ^2/n.d.f < 4	
ℓ	$p_T(\ell) > 350\,\mathrm{MeV}/c$	
Different cuts for muon and electron		
μ	`HasMuon`	Detection
	`IsMuon`	
	$p_T(\ell_1\ell_2) > 0\,\mathrm{MeV}/c$	
	$p_T(\ell) > 800\,\mathrm{MeV}/c$	PID Fiducial
	`HasRich`	PID Fiducial
	$3 < p(\ell) < 300\,\mathrm{GeV}/c$	PID Fiducial
e	`HasCalo` & ECAL region$>= 0$	Detection
	$p_T(\ell_1\ell_2) > 200\,\mathrm{MeV}/c$	
	ECAL $\lvert x\rvert > 363.6\,\mathrm{mm}$ or $\lvert y\rvert > 282.6\,\mathrm{mm}$	ECAL cell veto
	$p_T(\ell) > 500\,\mathrm{MeV}/c$	PID Fiducial
	`HasRich`	PID Fiducial
	$3 < p(\ell) < 200\,\mathrm{GeV}/c$	PID Fiducial

Table B.3: The customised offline selection (part 2).

Cuts	Comments
nSPDHits< 600(450)	Run 1 (Run 2)
Track opening angles $> 0.5\,\mu$rad	Clone removal
$\sigma(z_{\text{endvtx}}(\Lambda_b^0)) < 30$	Photon conversion rejection
`ProbNN`$e(\mu) > 0.2$	Customised PID cut for $e(\mu)$
$\frac{-13.51}{235} < \frac{p_x}{p_z} < \frac{-3.87}{235}$&$0 < \frac{p_y}{p_z} < \frac{9.44}{235}$	TT sensor veto (only R2p2 DD)
ECAL distance$(e_1, e_2) > 100\,$mm	Reduce overlapping
$M_{\text{HOP}} > 2080 + 248 \cdot \log(\chi^2_{\text{FD}}(\Lambda_b^0))$	HOP mass cut (only ee)
$M(\pi(p \to \pi)) < 475$ or $> 525\,\text{MeV}/c^2$	Veto K^0_{S}
DTF status converged	
$M(\Lambda^0\ell) > 2300\,\text{MeV}/c^2$	Only low and central q^2
$M(\Lambda^0(\ell \to \pi)) > 2400\,\text{MeV}/c^2$	Only low and central q^2

Table B.4: The selection details of the used HLT1 trigger lines given as examples. The exact details can vary over the different time periods of the data-taking. The information of `TrackAllL0` was extracted from Ref. [33]. The information of `TrackMVA` was extracted from LHCb code records and the related LHCb software.

HLT1 line (Run 1)	`TrackAllL0`
number of VELO hits > 9	
number of missed VELO hits < 3	
IP$> 0.1\,$mm	
number of OT+IT$\times 2$ hits > 16	
$p_T > 1.7\,\text{GeV}/c$	
$p > 10\,\text{GeV}/c$	
track χ^2/n.d.f < 2.5	
$\chi^2_{\text{IP}} > 16$	
HLT1 line (Run 2)	`TrackMVA`
track χ^2/n.d.f < 2.5	
Ghost probability < 0.2	
Either one of the following (*1.* or *2.*)	
1. $p_T > 25\,\text{GeV}/c$ & $\chi^2_{\text{IP}} > 7.4$	
2. $1 < p_T < 25\,\text{GeV}/c$ & $\ln(\chi^2_{\text{IP}}) > \frac{1.0}{(\frac{p_T}{\text{GeV}/c}-1)^2} + 1.1 - \frac{1.1\times p_T}{2.5\,\text{GeV}/c} + \ln(7.4)$	

C Further details on the MVA training

The distributions for the MVA input variables are shown in Figure C.1 (the electron-DD category), Figure C.2 (the electron-LL category), Figure C.3 (the muon-DD category) and Figure C.4 (the muon-LL category).

The distributions of the MVA output on the signal and background proxy samples and the ROC curves of the classifiers are shown in Figure C.5 (electron-DD, Run 1), Figure C.6 (electron-LL, Run 1), Figure C.7 (electron-DD, Run 2), Figure C.8 (electron-LL, Run 2), Figure C.9 (muon-DD, Run 1), Figure C.10 (muon-LL, Run 1), Figure C.11 (muon-DD, Run 2) and Figure C.12 (muon-LL, Run 2).

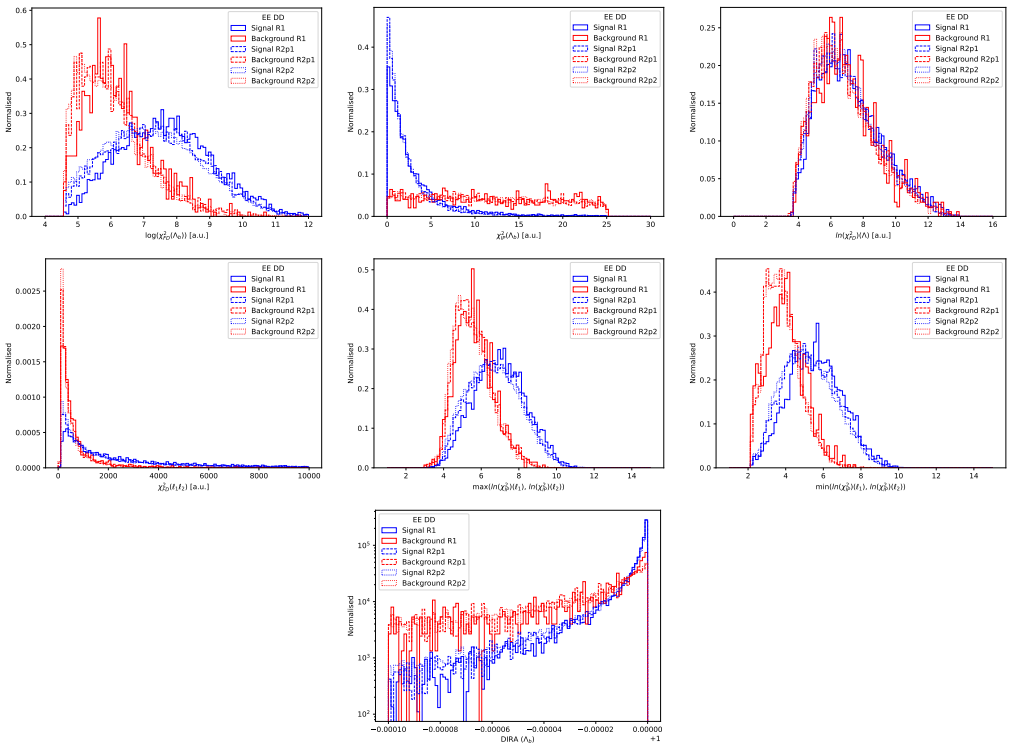


Figure C.1: The MVA input variables for the electron DD category.

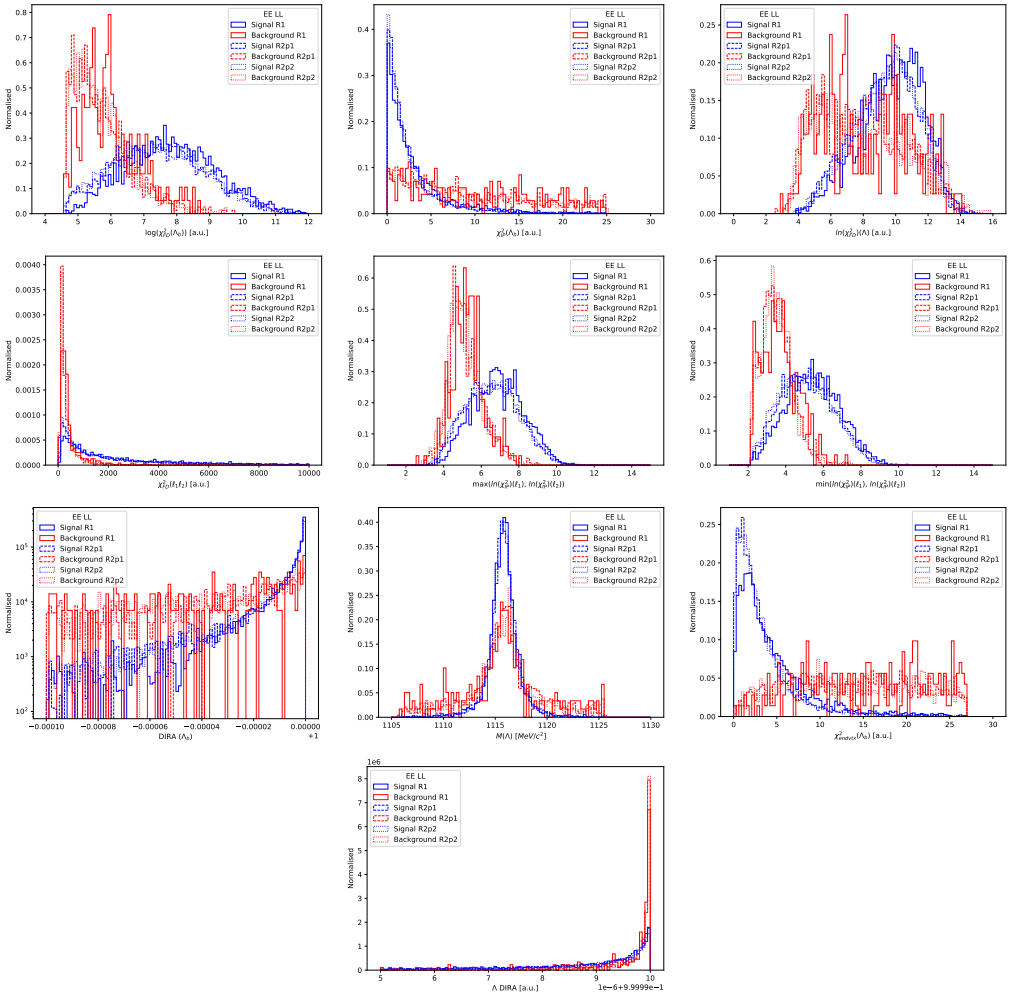


Figure C.2: The MVA input variables for the electron LL category.

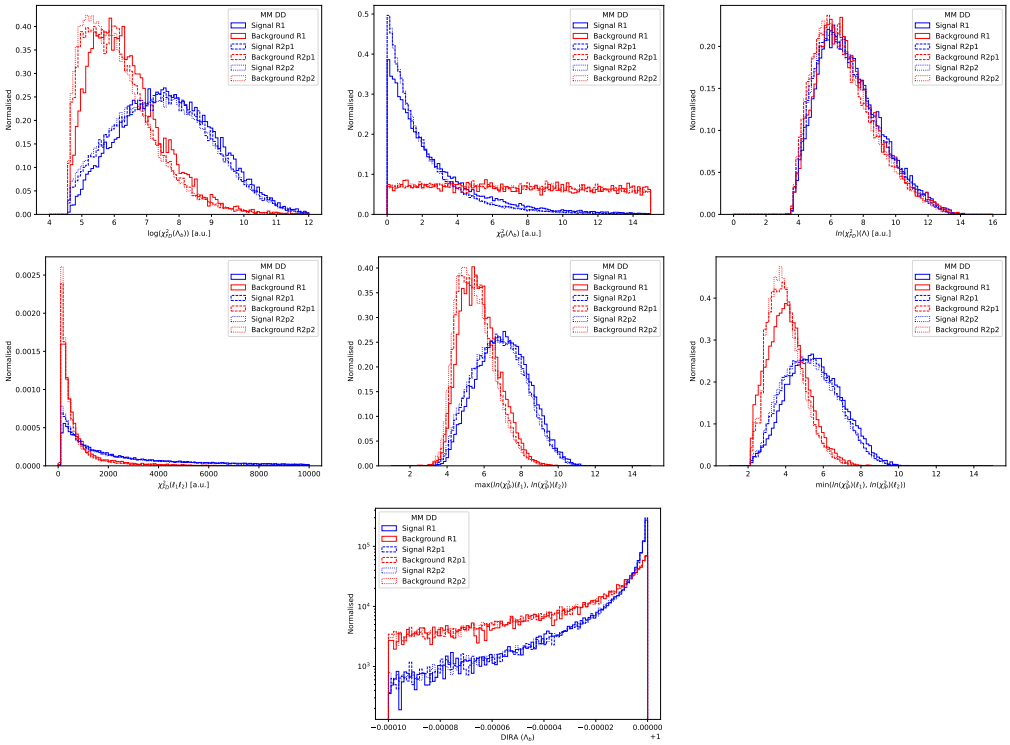


Figure C.3: The MVA input variables for the muon DD category.

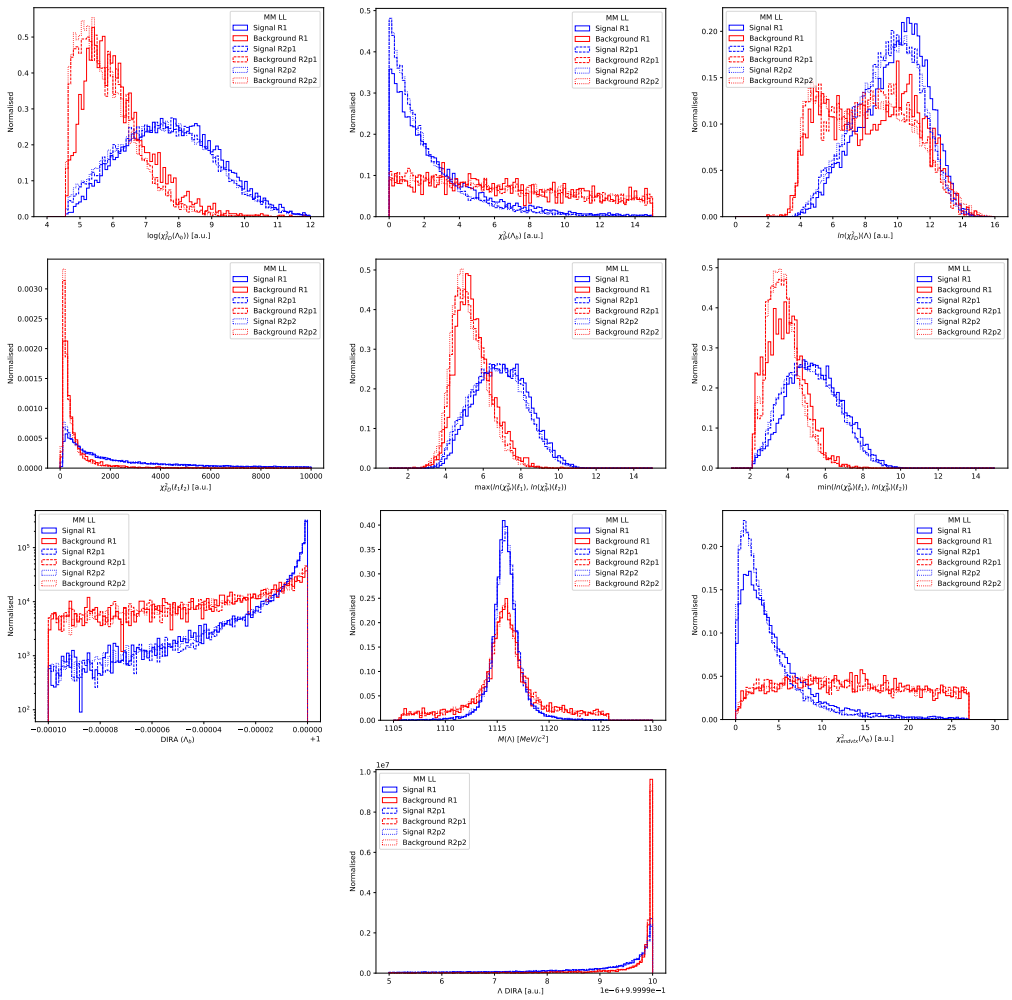


Figure C.4: The MVA input variables for the muon LL category.

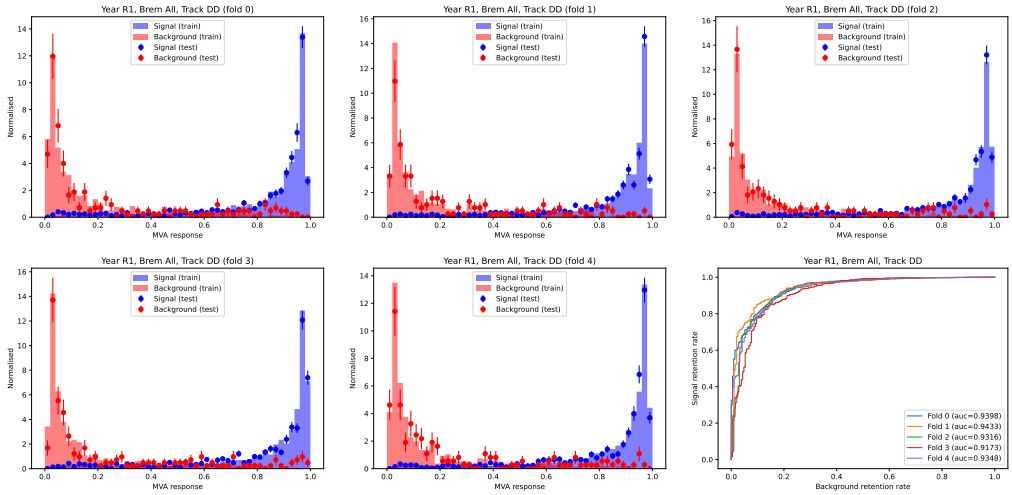


Figure C.5: The 5 folds MVA outputs and their ROC curves for the electron Run 1 DD category.

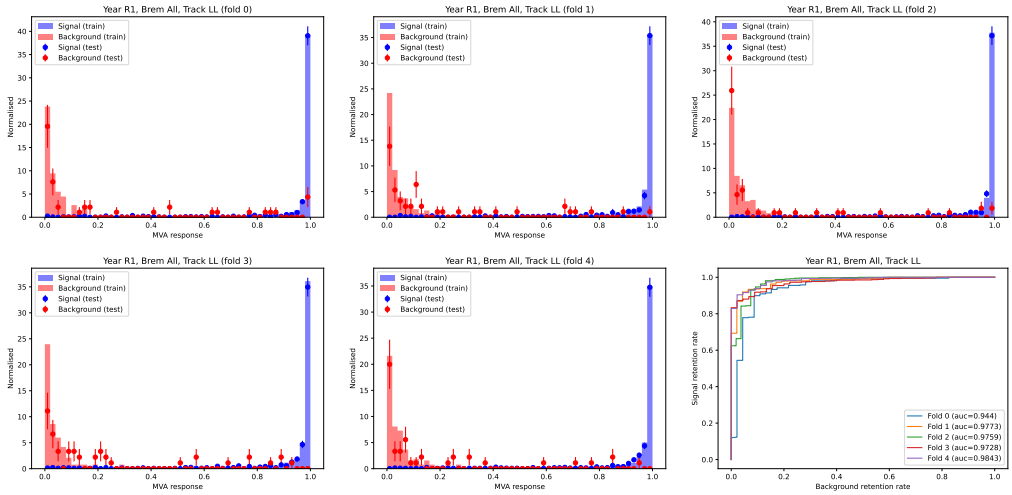


Figure C.6: The 5 folds MVA outputs and their ROC curves for the electron Run 1 LL category.

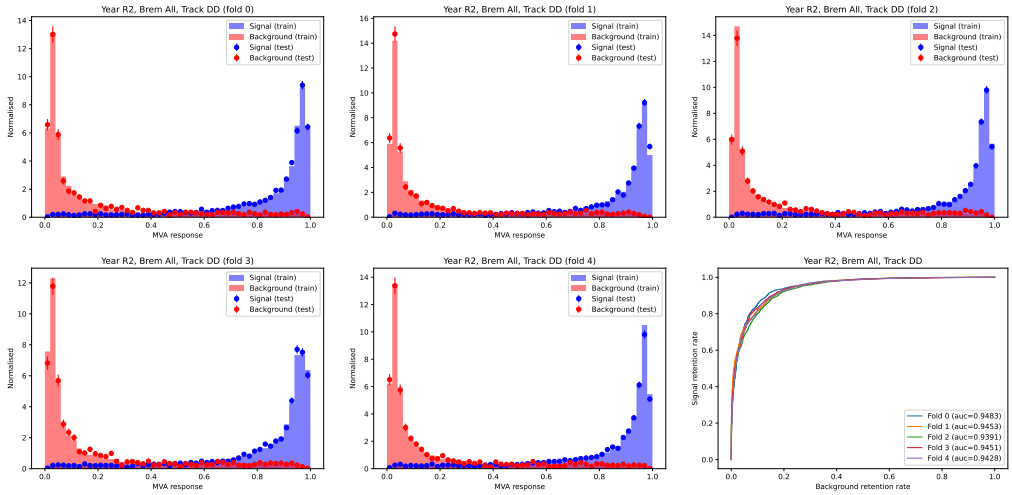


Figure C.7: The 5 folds MVA outputs and their ROC curves for the electron Run 2 DD category.

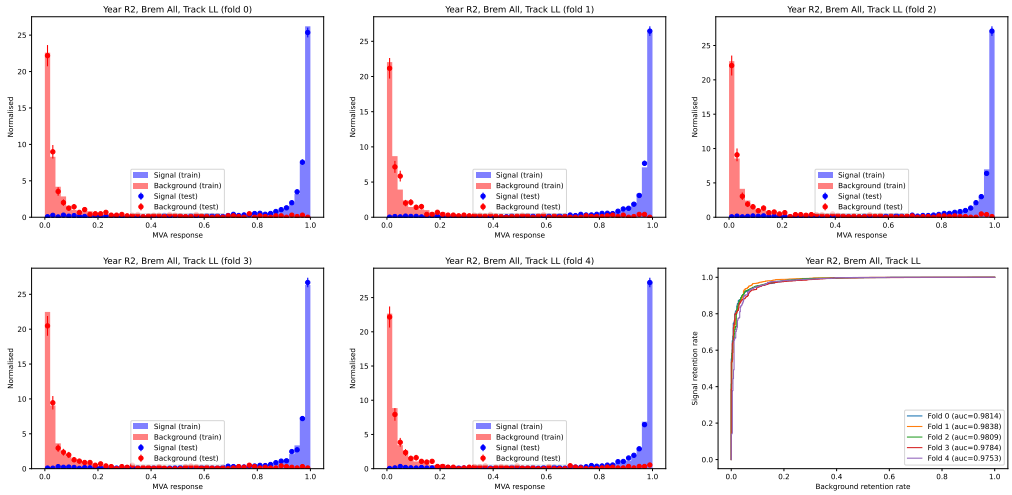


Figure C.8: The 5 folds MVA outputs and their ROC curves for the electron Run 2 LL category.

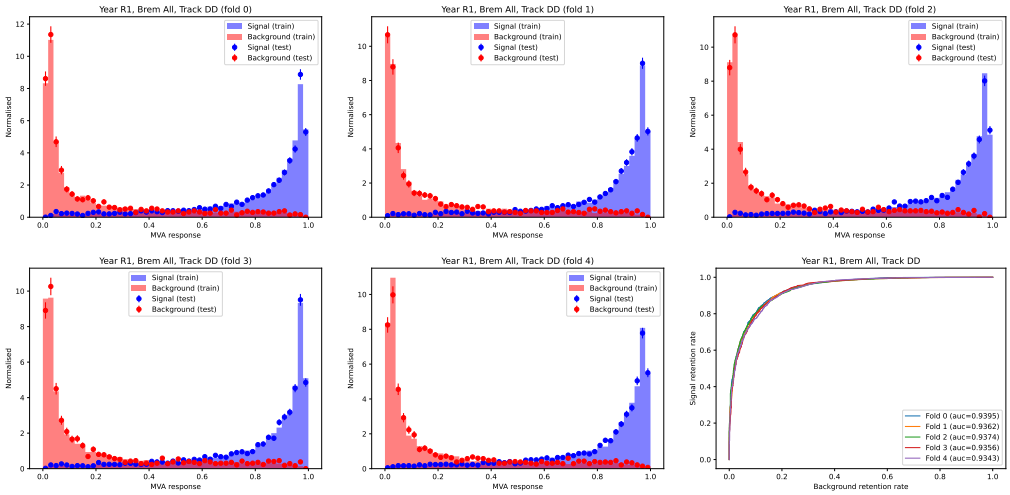


Figure C.9: The 5 folds MVA outputs and their ROC curves for the muon Run 1 DD category.

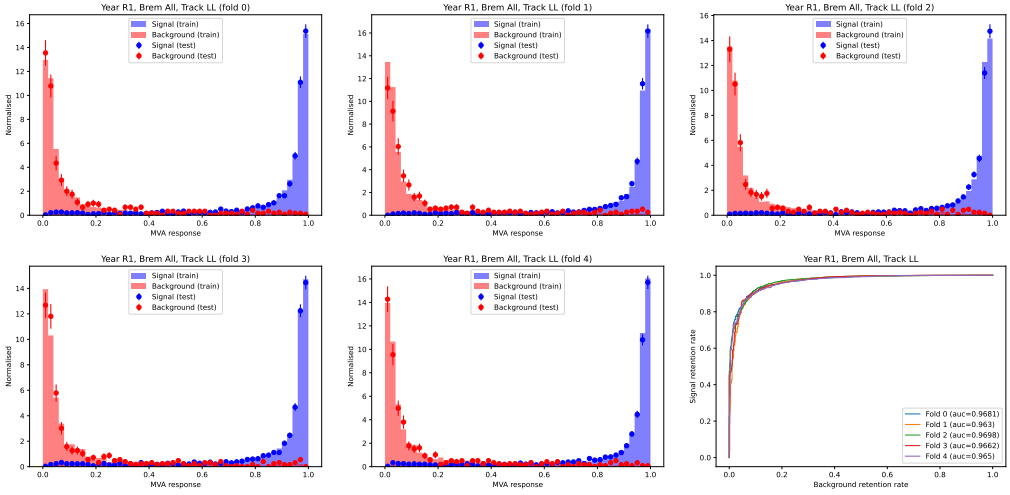


Figure C.10: The 5 folds MVA outputs and their ROC curves for the muon Run 1 LL category.

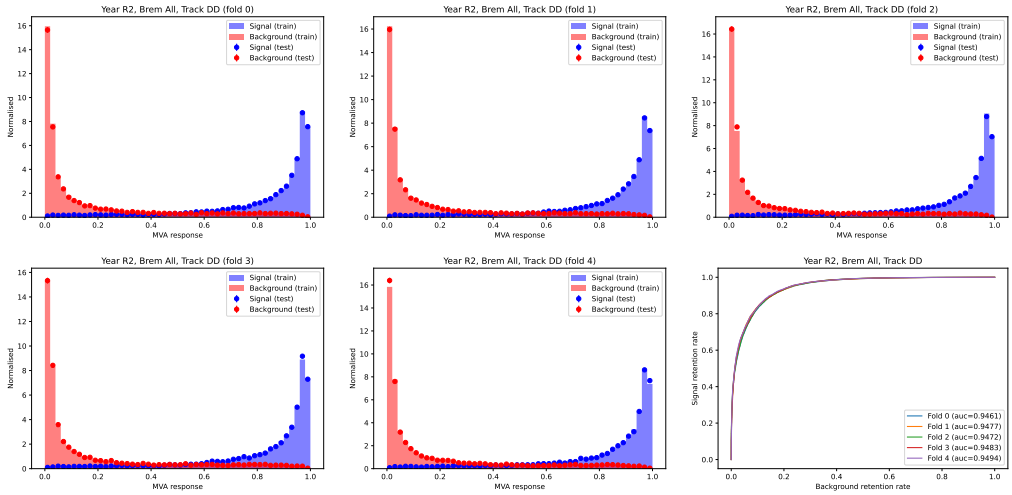


Figure C.11: The 5 folds MVA outputs and their ROC curves for the muon Run 2 DD category.

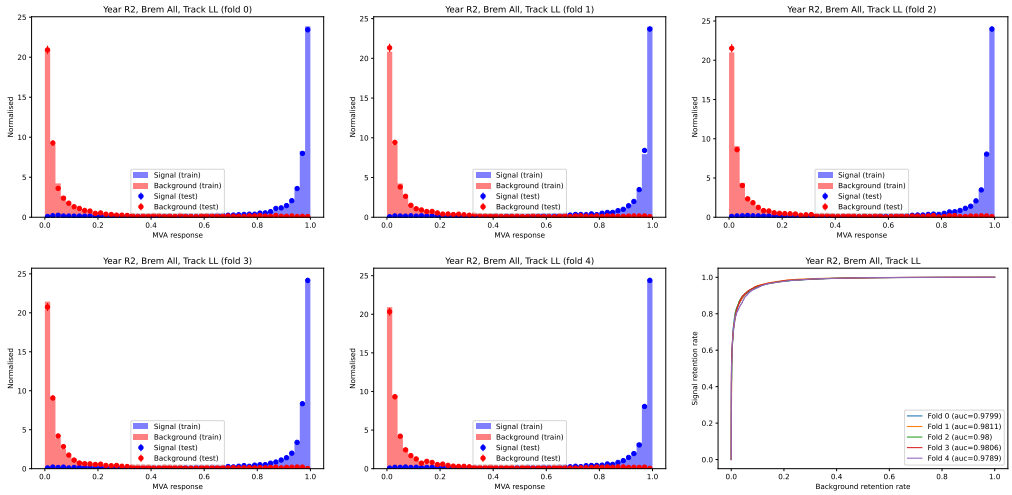


Figure C.12: The 5 folds MVA outputs and their ROC curves for the muon Run 2 LL category.

D Further details on the efficiencies

Table D.1 shows the signal efficiencies of the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ decay and its leakage efficiencies in the central and $\psi(2S)$ q^2 regions. Table D.2 shows the signal efficiencies of the $\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ decay and its leakage efficiencies in the J/ψ and the $\psi(2S)$ q^2 regions. Table D.3 shows the signal efficiencies of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ decay. Table D.4 shows the signal efficiencies of the $\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ decay. Table D.6 and Table D.5 show the efficiencies of several partially reconstructed backgrounds. The mass window requirements corresponding to the nominal fits are included in the calculation of these efficiencies.

Table D.1: Efficiencies of $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	Signal	Leakage in Central q^2	Leakage in $\psi(2S)$ q^2
R1 LL	0.833 ± 0.023	0.006 ± 0.002	0.022 ± 0.003
R1 DD	1.406 ± 0.026	0.001 ± 0.001	0.027 ± 0.004
R2p1 LL	1.861 ± 0.032	0.005 ± 0.002	0.043 ± 0.005
R2p1 DD	3.22 ± 0.042	0.002 ± 0.001	0.061 ± 0.006
R2p2 LL	1.849 ± 0.022	0.003 ± 0.001	0.045 ± 0.004
R2p2 DD	3.223 ± 0.03	0.003 ± 0.001	0.063 ± 0.004

Table D.2: Efficiencies of $\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	Signal	Leakage in J/ψ q^2	Leakage in High q^2
R1 LL	1.013 ± 0.024	0.15 ± 0.012	0.024 ± 0.004
R1 DD	1.954 ± 0.034	0.189 ± 0.011	0.032 ± 0.004
R2p1 LL	2.12 ± 0.033	0.34 ± 0.013	0.043 ± 0.005
R2p1 DD	3.855 ± 0.043	0.493 ± 0.016	0.071 ± 0.006
R2p2 LL	2.175 ± 0.025	0.325 ± 0.01	0.053 ± 0.004
R2p2 DD	3.731 ± 0.033	0.505 ± 0.011	0.067 ± 0.004

Table D.3: Efficiencies of $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	Signal
R1 LL	6.514 ± 0.069
R1 DD	11.449 ± 0.092
R2p1 LL	9.386 ± 0.09
R2p1 DD	15.896 ± 0.113
R2p2 LL	10.328 ± 0.07
R2p2 DD	17.385 ± 0.091

Table D.4: Efficiencies of $\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	Signal
R1 LL	10.105 ± 0.096
R1 DD	17.314 ± 0.114
R2p1 LL	12.785 ± 0.105
R2p1 DD	20.614 ± 0.134
R2p2 LL	14.045 ± 0.084
R2p2 DD	22.083 ± 0.104

Table D.5: Efficiencies in several muon background channels. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	R1 LL	R1 DD	R2p1 LL	R2p1 DD	R2p2 LL	R2p2 DD
$\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-)$, J/ψ q^2	0.251 ± 0.011	4.703 ± 0.057	0.417 ± 0.018	7.023 ± 0.094	0.47 ± 0.016	7.359 ± 0.08
$\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-)$, $\psi(2S)$ q^2	0.341 ± 0.026	5.432 ± 0.103	0.467 ± 0.027	7.077 ± 0.106	0.505 ± 0.02	7.397 ± 0.077
$\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-)$, J/ψ q^2	4.765 ± 0.043	8.172 ± 0.057	7.175 ± 0.064	11.582 ± 0.084	7.705 ± 0.076	11.933 ± 0.096

Table D.6: Efficiencies in several electron background channels. Nominal Λ_b^0 mass window requirements included. Numbers shown in $[10^{-4}]$.

	R1 LL	R1 DD	R2p1 LL	R2p1 DD	R2p2 LL	R2p2 DD
$\Xi_b^- \to \Xi^- J/\psi(e^+e^-)$, J/ψ q^2	0.031 ± 0.005	0.559 ± 0.018	0.079 ± 0.005	1.291 ± 0.024	0.072 ± 0.004	1.305 ± 0.016
$\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-)$, $\psi(2S)$ q^2	0.022 ± 0.004	0.613 ± 0.019	0.065 ± 0.005	1.203 ± 0.023	0.057 ± 0.003	1.097 ± 0.016
$\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-)$, J/ψ q^2	0.485 ± 0.026	0.992 ± 0.039	1.157 ± 0.036	2.144 ± 0.049	1.134 ± 0.025	2.149 ± 0.035

E Further details on the fits in the resonance q^2 regions

The data fit in the resonance q^2 regions for the different data-taking periods (R1,R2p1,R2p2) and $\Lambda^0 \to p\pi^-$ track types (LL,DD) are shown in Figure E.1-Figure E.6.

The corresponding free and constant parameters of the data fits are shown in Table E.1-Table E.12. The constant parameters are fixed based on the simulation. They are the shape parameters of the Johnson S_U functions (for $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$) and the AsymDSCB functions (for $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ and $\Lambda_b^0 \to \Lambda^0\psi(2S)(\mu^+\mu^-)$). For the free parameters of the data fit, the meaning of the different variables are explained as follows.

- $N_{\text{comb}}(J/\psi, ee)$, $N_{\text{comb}}(\psi(2S), ee)$, $N_{\text{comb}}(J/\psi, \mu\mu)$, $N_{\text{comb}}(\psi(2S), \mu\mu)$: the yields of the combinatorial background in different channels.

- $N_{\text{other part. rec.}}(J/\psi, ee)$, $N_{\text{other part. rec.}}(J/\psi, \mu\mu)$: the yields of the other partially reconstructed background in the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ and the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ channels.

- $N_{\text{sig}}(J/\psi, \mu\mu)$, $N_{\text{sig}}(\psi(2S), \mu\mu)$: the signal yield of the $\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ and the $\Lambda_b^0 \to \Lambda^0\psi(2S)(\mu^+\mu^-)$ channels.

- $C_{\text{comb}}(J/\psi, ee)$, $C_{\text{comb}}(\psi(2S), ee)$, $C_{\text{comb}}(J/\psi, \mu\mu)$, $C_{\text{comb}}(\psi(2S), \mu\mu)$: the slope parameters of the exponential shapes of the combinatorial backgrounds in different channels.

- $\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$: the efficiency for the decay $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ being reconstructed in the $\psi(2S)$ region (leakage). It is one of the constrained nuisance parameters.

- $\epsilon(\Lambda_b^0 \to \Lambda(1520)J/\psi(e^+e^-))$, $\epsilon(\Lambda_b^0 \to \Lambda(1520)J/\psi(\mu^+\mu^-))$, $\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$, $\epsilon(\Xi_b^- \to \Xi^-\psi(2S)(e^+e^-))$, $\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$, $\epsilon(\Xi_b^- \to \Xi^-\psi(2S)(\mu^+\mu^-))$: the background efficiencies, which are constrained as nuisance parameters.

- $\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$, $\epsilon(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$, $\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$, $\epsilon(\Lambda_b^0 \to \Lambda^0\psi(2S)(\mu^+\mu^-))$: the signal efficiencies, which are constrained as nuisance parameters.

- $a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520)J/\psi)$, $a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$, $a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$: as discussed in Section 9.3.
- $r_{J/\psi}^{-1}, R_{\psi(2S)}^{-1}$: as discussed in Chapter 6.
- mean shift (ee), mean shift $(\mu\mu)$: the mean-shifting parameters, which are used to accommodate the data/simulation discrepancy (as discussed in Section 9.3).
- other part. rec. mean $(J/\psi, ee)$, other part. rec. mean $(J/\psi, \mu\mu)$: the mean parameters of the Gaussian shapes, which are used to model the other partially reconstructed backgrounds.
- other part. rec. width $(J/\psi, ee)$, other part. rec. width $(J/\psi, \mu\mu)$: the width parameters of the Gaussian shapes, which are used to model the other partially reconstructed backgrounds.
- width scale (ee), width scale $(\mu\mu)$: the width-scaling parameters, which are used to accommodate the data/simulation discrepancy (as discussed in Section 9.3).

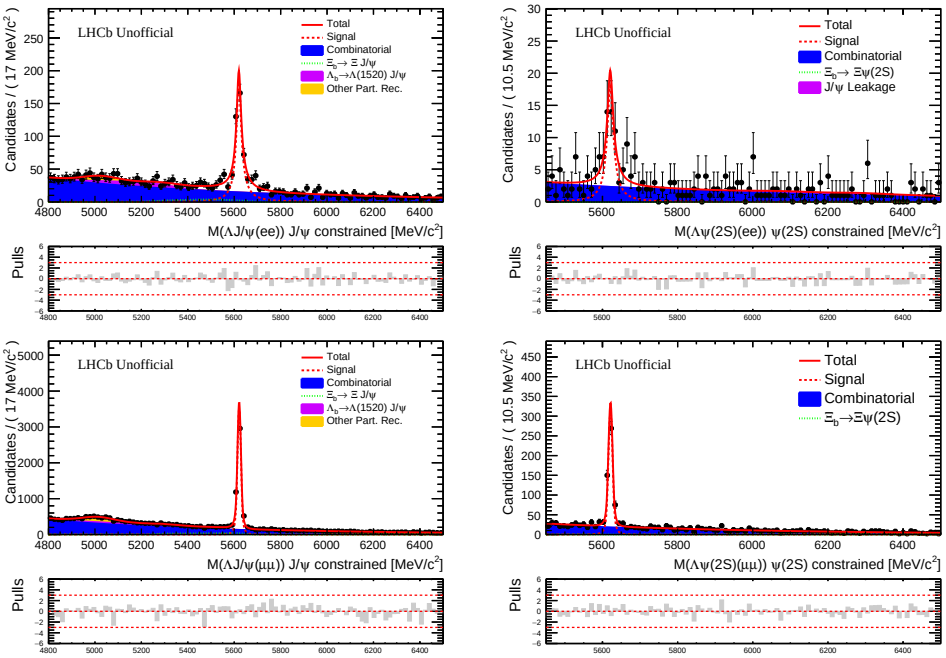


Figure E.1: The resonance mode data fit of the R1-DD category.

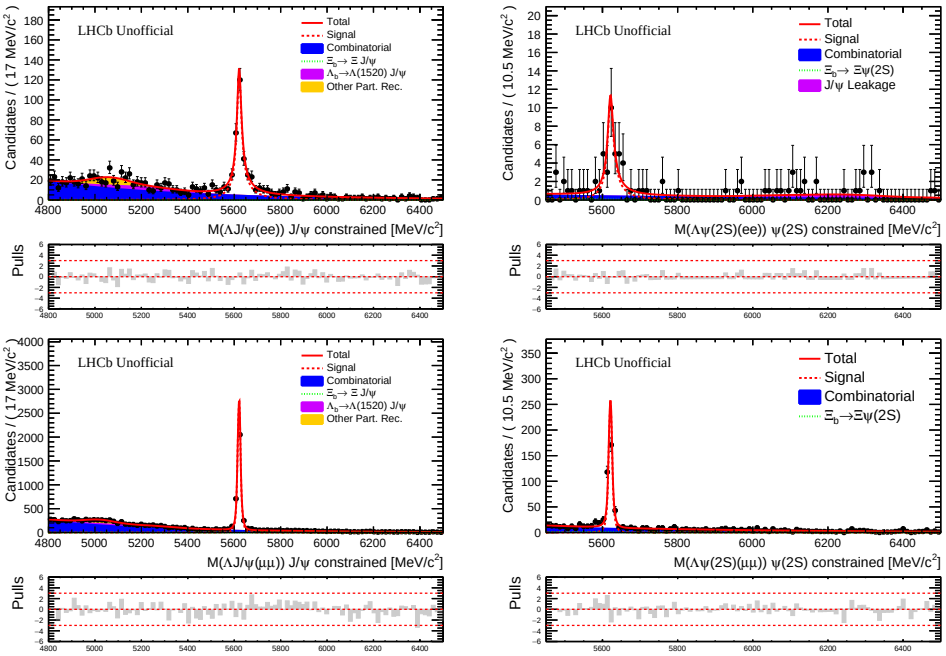


Figure E.2: The resonance mode data fit of the R1-LL category.

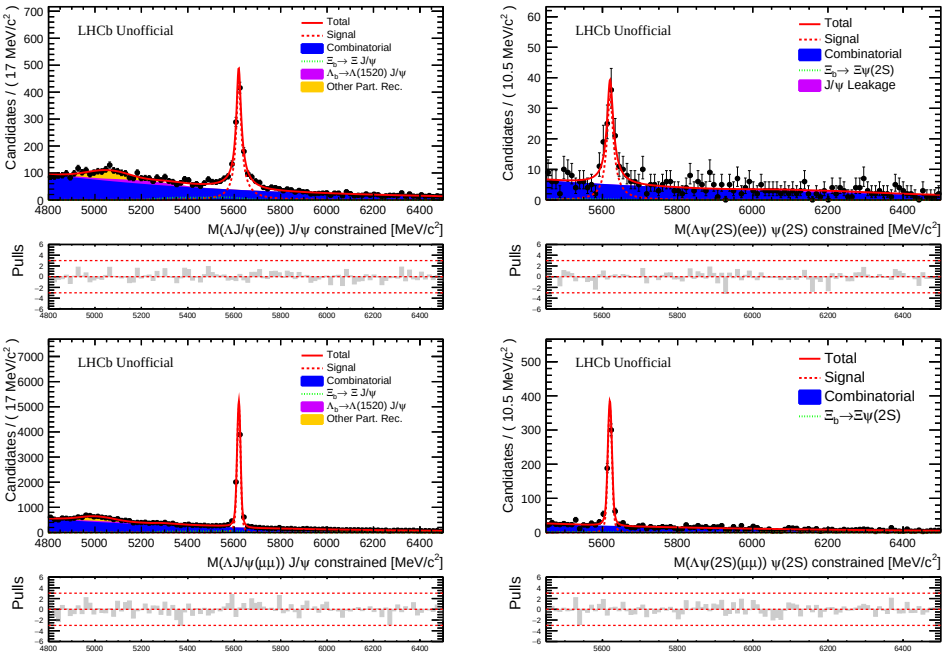


Figure E.3: The resonance mode data fit of the R2p1-DD category.

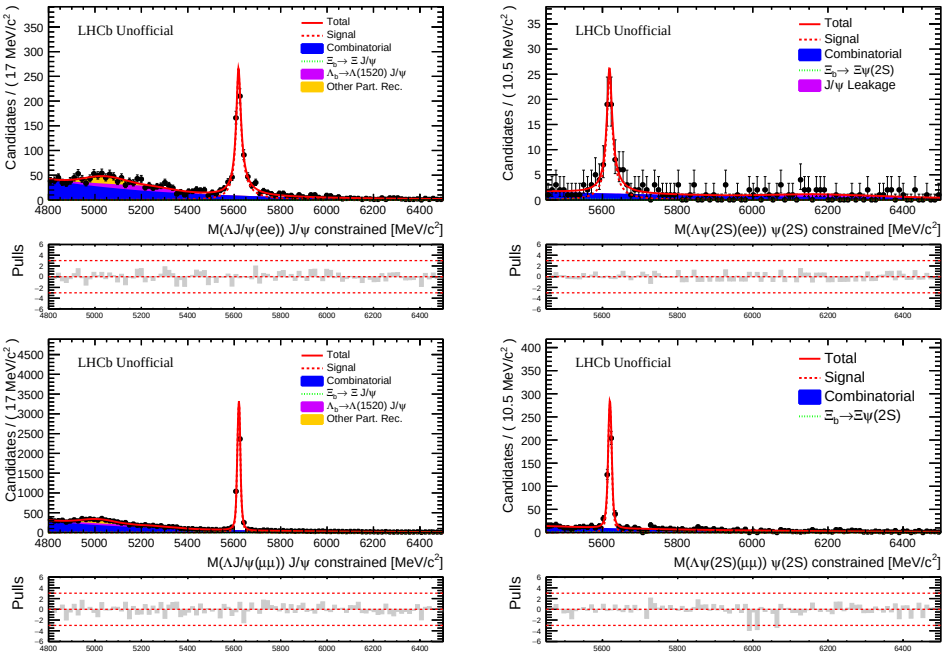


Figure E.4: The resonance mode data fit of the R2p1-LL category.

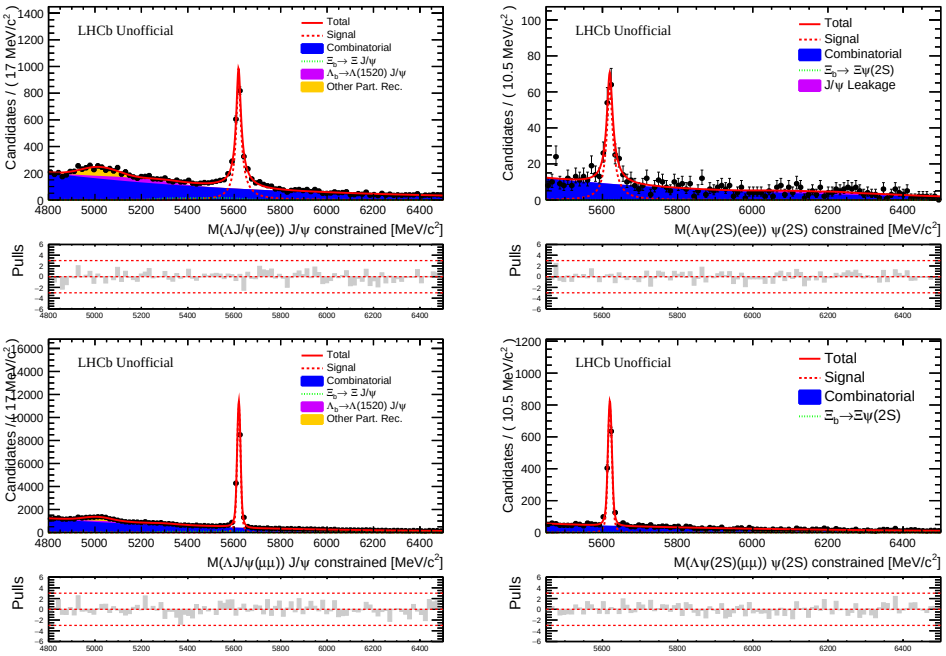


Figure E.5: The resonance mode data fit of the R2p2-DD category.

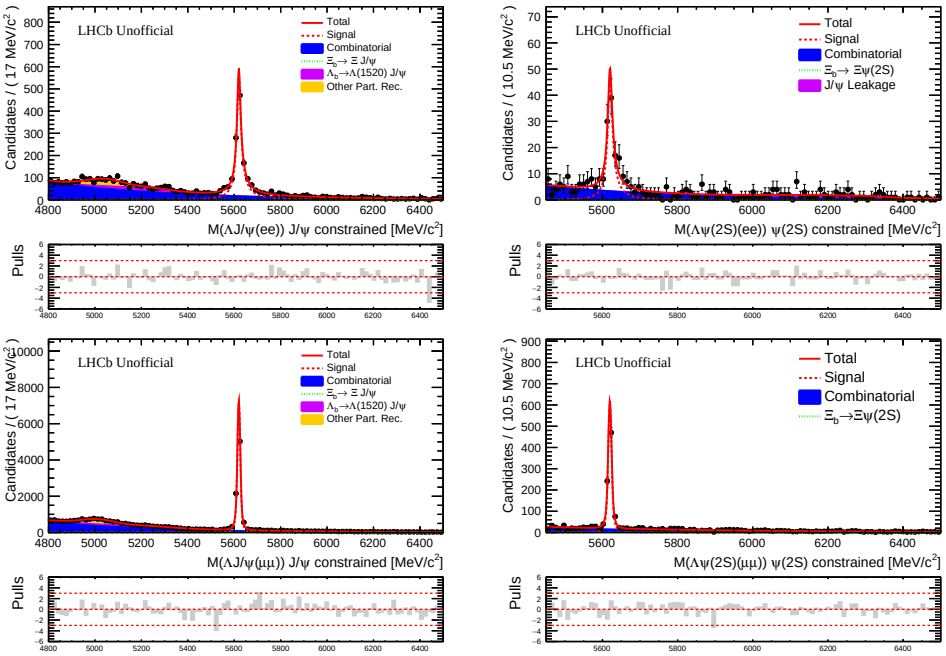


Figure E.6: The resonance mode data fit of the R2p2-LL category.

Table E.1: An overview of the data fit free parameters for the fits in the resonance regions (R1-DD).

	Fit result
$N_{\text{comb}}(J/\psi, ee)$	1740.92 ± 73.63
$N_{\text{comb}}(\psi(2S), ee)$	174.91 ± 15.36
$N_{\text{comb}}(J/\psi, \mu\mu)$	$(1.7832 \pm 0.0311) \times 10^4$
$N_{\text{comb}}(\psi(2S), \mu\mu)$	1262.71 ± 37.79
$N_{\text{other part. rec.}}(J/\psi, ee)$	67.47 ± 38.77
$N_{\text{other part. rec.}}(J/\psi, \mu\mu)$	1249.13 ± 175.7
$N_{\text{sig}}(J/\psi, \mu\mu)$	4305.98 ± 72.6
$N_{\text{sig}}(\psi(2S), \mu\mu)$	458.13 ± 23.61
$C_{\text{comb}}(J/\psi, ee)$	$(-1.013 \pm 0.0695) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), ee)$	$(-1.158 \pm 0.291) \times 10^{-3}$
$C_{\text{comb}}(J/\psi, \mu\mu)$	$(-1.2531 \pm 0.0263) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), \mu\mu)$	$(-1.775 \pm 0.105) \times 10^{-3}$
$\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$	$(2.689 \pm 0.388) \times 10^{-6}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-))$	$(9.895 \pm 0.386) \times 10^{-5}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-))$	$(8.1725 \pm 0.0567) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$	$(5.596 \pm 0.182) \times 10^{-5}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-))$	$(6.126 \pm 0.194) \times 10^{-5}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$	$(4.702 \pm 0.0572) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-))$	$(5.432 \pm 0.103) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$	$(1.4064 \pm 0.0264) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$	$(1.954 \pm 0.034) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$	$(1.14486 \pm 0.00916) \times 10^{-3}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))$	$(1.7314 \pm 0.0114) \times 10^{-3}$
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520) J/\psi)$	0.4382 ± 0.0648
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.3728 ± 0.0564
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	0.0 ± 1.15
$R^{-1}_{\psi(2S)}$	1.204 ± 0.234
$r^{-1}_{J/\psi}$	0.9061 ± 0.06
mean shift (ee)	1.381 ± 0.932
mean shift $(\mu\mu)$	1.559 ± 0.137
other part. rec. mean $(J/\psi, ee)$	5018.66 ± 51.34
other part. rec. width $(J/\psi, ee)$	69.3 ± 24.5
other part. rec. mean $(J/\psi, \mu\mu)$	5013.91 ± 6.71
other part. rec. width $(J/\psi, \mu\mu)$	68.49 ± 8.27
width scale (ee)	1.174 ± 0.115
width scale $(\mu\mu)$	1.1345 ± 0.0198

Table E.2: An overview of the data fit constant parameters for the fits in the resonance regions (R1-DD).

	Value
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U δ	0.54789
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U γ	-0.08117
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U λ	8.33473
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U μ	5618.31921
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U δ	0.58593
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U γ	-0.03182
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U λ	6.63878
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U μ	5619.21806
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_L	1.46573
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_R	1.71409
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB m_0	5619.75222
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_L	4.33218
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_R	3.3022
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_L	6.59372
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_R	6.9394
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_L	1.64967
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_R	1.58762
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB m_0	5619.85285
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_L	3.22594
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_R	3.58283
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_L	5.03291
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_R	4.91524

Table E.3: An overview of the data fit free parameters for the fits in the resonance regions (R1-LL).

	Fit result
$N_{\text{comb}}(J/\psi, ee)$	658.63 ± 48.6
$N_{\text{comb}}(\psi(2S), ee)$	34.49 ± 8.32
$N_{\text{comb}}(J/\psi, \mu\mu)$	7768.64 ± 187.06
$N_{\text{comb}}(\psi(2S), \mu\mu)$	557.82 ± 25.12
$N_{\text{other part. rec.}}(J/\psi, ee)$	88.25 ± 33.56
$N_{\text{other part. rec.}}(J/\psi, \mu\mu)$	682.95 ± 108.09
$N_{\text{sig}}(J/\psi, \mu\mu)$	2971.27 ± 57.24
$N_{\text{sig}}(\psi(2S), \mu\mu)$	325.57 ± 19.36
$C_{\text{comb}}(J/\psi, ee)$	$(-1.587 \pm 0.12) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), ee)$	$(-1.035 \pm 0.804) \times 10^{-3}$
$C_{\text{comb}}(J/\psi, \mu\mu)$	$(-1.8972 \pm 0.0401) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), \mu\mu)$	$(-1.92 \pm 0.16) \times 10^{-3}$
$\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$	$(2.35 \pm 0.346) \times 10^{-6}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-))$	$(4.864 \pm 0.26) \times 10^{-5}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-))$	$(4.764 \pm 0.043) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$	$(3.128 \pm 0.461) \times 10^{-6}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-))$	$(2.206 \pm 0.366) \times 10^{-6}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$	$(2.511 \pm 0.113) \times 10^{-5}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-))$	$(3.431 \pm 0.257) \times 10^{-5}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$	$(8.312 \pm 0.235) \times 10^{-5}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$	$(1.0129 \pm 0.0245) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$	$(6.5148 \pm 0.0691) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))$	$(1.01039 \pm 0.00963) \times 10^{-3}$
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520) J/\psi)$	0.4193 ± 0.0669
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.0 ± 0.337
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	1.5 ± 1.0
$R^{-1}_{\psi(2S)}$	1.092 ± 0.247
$r^{-1}_{J/\psi}$	0.9538 ± 0.0746
mean shift (ee)	1.9 ± 1.02
mean shift $(\mu\mu)$	1.243 ± 0.142
other part. rec. mean $(J/\psi, ee)$	5060.0 ± 13.79
other part. rec. width $(J/\psi, ee)$	82.42 ± 23.47
other part. rec. mean $(J/\psi, \mu\mu)$	5018.42 ± 8.58
other part. rec. width $(J/\psi, \mu\mu)$	65.39 ± 8.76
width scale (ee)	1.316 ± 0.152
width scale $(\mu\mu)$	1.1367 ± 0.0236

Table E.4: An overview of the data fit constant parameters for the fits in the resonance regions (R1-LL).

	Value
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U δ	0.47897
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U γ	-0.13043
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U λ	7.09005
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U μ	5618.93908
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U δ	0.55632
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U γ	-0.1724
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U λ	5.7173
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U μ	5618.94325
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_L	1.49879
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_R	1.07838
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB m_0	5620.30766
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_L	4.04873
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_R	7.09201
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_L	6.25563
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_R	5.11429
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_L	1.37197
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_R	1.20275
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB m_0	5620.05841
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_L	4.04541
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_R	5.33246
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_L	4.37169
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_R	4.03112

Table E.5: An overview of the data fit free parameters for the fits in the resonance regions (R2p1-DD).

	Fit result
$N_{\text{comb}}(J/\psi, ee)$	4174.12 ± 144.02
$N_{\text{comb}}(\psi(2S), ee)$	345.76 ± 22.4
$N_{\text{comb}}(J/\psi, \mu\mu)$	$(2.264 \pm 0.0382) \times 10^4$
$N_{\text{comb}}(\psi(2S), \mu\mu)$	1279.32 ± 51.41
$N_{\text{other part. rec.}}(J/\psi, ee)$	419.49 ± 104.17
$N_{\text{other part. rec.}}(J/\psi, \mu\mu)$	2026.1 ± 229.72
$N_{\text{sig}}(J/\psi, \mu\mu)$	6089.04 ± 86.0
$N_{\text{sig}}(\psi(2S), \mu\mu)$	531.13 ± 25.35
$C_{\text{comb}}(J/\psi, ee)$	$(-1.1449 \pm 0.0515) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), ee)$	$(-1.327 \pm 0.22) \times 10^{-3}$
$C_{\text{comb}}(J/\psi, \mu\mu)$	$(-1.2264 \pm 0.0258) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), \mu\mu)$	$(-1.507 \pm 0.142) \times 10^{-3}$
$\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$	$(6.014 \pm 0.566) \times 10^{-6}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-))$	$(2.1471 \pm 0.0491) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-))$	$(1.15807 \pm 0.00843) \times 10^{-3}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$	$(1.2933 \pm 0.0237) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-))$	$(1.2027 \pm 0.0228) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$	$(7.0165 \pm 0.0941) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-))$	$(7.077 \pm 0.106) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$	$(3.2169 \pm 0.0417) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$	$(3.8555 \pm 0.0426) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$	$(1.5902 \pm 0.0113) \times 10^{-3}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))$	$(2.0614 \pm 0.0134) \times 10^{-3}$
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520) J/\psi)$	0.3294 ± 0.0541
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.3864 ± 0.0438
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	0.124 ± 0.207
$R^{-1}_{\psi(2S)}$	1.074 ± 0.15
$r^{-1}_{J/\psi}$	1.0278 ± 0.045
mean shift (ee)	0.598 ± 0.616
mean shift $(\mu\mu)$	-0.146 ± 0.115
other part. rec. mean $(J/\psi, ee)$	5055.38 ± 12.83
other part. rec. width $(J/\psi, ee)$	90.33 ± 17.44
other part. rec. mean $(J/\psi, \mu\mu)$	5005.88 ± 5.58
other part. rec. width $(J/\psi, \mu\mu)$	75.28 ± 6.83
width scale (ee)	1.1835 ± 0.0783
width scale $(\mu\mu)$	1.1555 ± 0.017

Table E.6: An overview of the data fit constant parameters for the fits in the resonance regions (R2p1-DD).

	Value
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U δ	0.50642
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U γ	-0.10281
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U λ	8.26455
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U μ	5618.33614
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U δ	0.60463
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U γ	-0.09266
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U λ	6.90791
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U μ	5619.34738
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_L	1.64354
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_R	1.60061
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB m_0	5620.28946
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_L	3.57333
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_R	3.64373
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_L	6.94348
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_R	6.36381
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_L	1.61569
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_R	1.72719
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB m_0	5619.79185
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_L	3.56848
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_R	3.11927
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_L	4.86677
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_R	4.96922

Table E.7: An overview of the data fit free parameters for the fits in the resonance regions (R2p1-LL).

	Fit result
$N_{\text{comb}}(J/\psi, ee)$	1211.98 ± 82.95
$N_{\text{comb}}(\psi(2S), ee)$	85.13 ± 12.72
$N_{\text{comb}}(J/\psi, \mu\mu)$	8374.22 ± 203.98
$N_{\text{comb}}(\psi(2S), \mu\mu)$	559.91 ± 35.03
$N_{\text{other part. rec.}}(J/\psi, ee)$	188.34 ± 70.26
$N_{\text{other part. rec.}}(J/\psi, \mu\mu)$	1082.66 ± 121.21
$N_{\text{sig}}(J/\psi, \mu\mu)$	3666.91 ± 63.47
$N_{\text{sig}}(\psi(2S), \mu\mu)$	365.32 ± 20.85
$C_{\text{comb}}(J/\psi, ee)$	$(-1.903 \pm 0.108) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), ee)$	$(-1.381 \pm 0.513) \times 10^{-3}$
$C_{\text{comb}}(J/\psi, \mu\mu)$	$(-1.9182 \pm 0.0399) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), \mu\mu)$	$(-2.024 \pm 0.217) \times 10^{-3}$
$\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$	$(4.264 \pm 0.509) \times 10^{-6}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-))$	$(1.1608 \pm 0.0358) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-))$	$(7.1728 \pm 0.0635) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$	$(7.9 \pm 0.549) \times 10^{-6}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-))$	$(6.532 \pm 0.534) \times 10^{-6}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$	$(4.165 \pm 0.183) \times 10^{-5}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-))$	$(4.668 \pm 0.266) \times 10^{-5}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$	$(1.8591 \pm 0.0319) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$	$(2.12 \pm 0.0327) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$	$(9.389 \pm 0.0901) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))$	$(1.2785 \pm 0.0105) \times 10^{-3}$
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520) J/\psi)$	0.5302 ± 0.0599
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.0 ± 0.268
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	0.886 ± 0.972
$R^{-1}_{\psi(2S)}$	1.233 ± 0.199
$r^{-1}_{J/\psi}$	0.9818 ± 0.0532
mean shift (ee)	-0.826 ± 0.724
mean shift $(\mu\mu)$	-0.044 ± 0.131
other part. rec. mean $(J/\psi, ee)$	5044.19 ± 18.03
other part. rec. width $(J/\psi, ee)$	85.78 ± 26.05
other part. rec. mean $(J/\psi, \mu\mu)$	5015.09 ± 6.28
other part. rec. width $(J/\psi, \mu\mu)$	68.21 ± 6.31
width scale (ee)	1.1498 ± 0.09
width scale $(\mu\mu)$	1.1458 ± 0.0211

Table E.8: An overview of the data fit constant parameters for the fits in the resonance regions (R2p1-LL).

	Value
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U δ	0.51611
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U γ	-0.18131
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U λ	8.38646
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U μ	5618.58494
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U δ	0.55148
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U γ	-0.12424
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U λ	5.96305
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U μ	5619.08519
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_L	1.64749
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_R	1.17581
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB m_0	5620.57704
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_L	2.49651
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_R	5.36856
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_L	6.39127
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_R	5.26362
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_L	1.25264
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_R	1.46058
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB m_0	5619.76504
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_L	4.30093
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_R	3.64772
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_L	4.12006
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_R	4.3907

Table E.9: An overview of the data fit free parameters for the fits in the resonance regions (R2p2-DD).

	Fit result
$N_{\text{comb}}(J/\psi, ee)$	9262.0 ± 197.45
$N_{\text{comb}}(\psi(2S), ee)$	547.15 ± 28.42
$N_{\text{comb}}(J/\psi, \mu\mu)$	$(4.851 \pm 0.0502) \times 10^4$
$N_{\text{comb}}(\psi(2S), \mu\mu)$	2757.85 ± 73.04
$N_{\text{other part. rec.}}(J/\psi, ee)$	787.69 ± 124.25
$N_{\text{other part. rec.}}(J/\psi, \mu\mu)$	3102.53 ± 244.93
$N_{\text{sig}}(J/\psi, \mu\mu)$	$(1.323 \pm 0.0126) \times 10^4$
$N_{\text{sig}}(\psi(2S), \mu\mu)$	1117.19 ± 37.22
$C_{\text{comb}}(J/\psi, ee)$	$(-1.0755 \pm 0.0336) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), ee)$	$(-1.791 \pm 0.184) \times 10^{-3}$
$C_{\text{comb}}(J/\psi, \mu\mu)$	$(-1.2767 \pm 0.0158) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), \mu\mu)$	$(-1.5725 \pm 0.0934) \times 10^{-3}$
$\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$	$(6.286 \pm 0.433) \times 10^{-6}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-))$	$(2.1511 \pm 0.0346) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-))$	$(1.19295 \pm 0.00955) \times 10^{-3}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$	$(1.3056 \pm 0.0162) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-))$	$(1.0974 \pm 0.0155) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$	$(7.3558 \pm 0.0796) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-))$	$(7.3966 \pm 0.0771) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$	$(3.2205 \pm 0.0301) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$	$(3.7307 \pm 0.033) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$	$(1.73891 \pm 0.00907) \times 10^{-3}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))$	$(2.2083 \pm 0.0104) \times 10^{-3}$
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520) J/\psi)$	0.4206 ± 0.0355
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.3104 ± 0.0284
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	0.0 ± 0.231
$R^{-1}_{\psi(2S)}$	1.003 ± 0.105
$r^{-1}_{J/\psi}$	1.0152 ± 0.0322
mean shift (ee)	-0.143 ± 0.441
mean shift $(\mu\mu)$	$(-5.81 \pm 7.74) \times 10^{-2}$
other part. rec. mean $(J/\psi, ee)$	5018.76 ± 9.09
other part. rec. width $(J/\psi, ee)$	79.91 ± 9.77
other part. rec. mean $(J/\psi, \mu\mu)$	5010.55 ± 3.71
other part. rec. width $(J/\psi, \mu\mu)$	60.51 ± 4.15
width scale (ee)	1.1846 ± 0.056
width scale $(\mu\mu)$	1.1364 ± 0.0113

Table E.10: An overview of the data fit constant parameters for the fits in the resonance regions (R2p2-DD).

	Value
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U δ	0.51285
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U γ	-0.12009
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U λ	8.1891
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U μ	5618.39562
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U δ	0.59122
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U γ	-0.0862
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U λ	6.56388
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U μ	5619.23024
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_L	1.62869
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_R	1.60688
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB m_0	5620.03944
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_L	3.34199
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_R	3.74551
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_L	6.74869
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_R	6.61595
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_L	1.54743
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_R	1.63558
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB m_0	5619.66129
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_L	3.64847
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_R	3.62777
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_L	4.72446
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_R	4.98644

Table E.11: An overview of the data fit free parameters for the fits in the resonance regions (R2p2-LL).

	Fit result
$N_{\text{comb}}(J/\psi, ee)$	2723.76 ± 104.07
$N_{\text{comb}}(\psi(2S), ee)$	197.62 ± 17.56
$N_{\text{comb}}(J/\psi, \mu\mu)$	$(1.8947 \pm 0.0299) \times 10^4$
$N_{\text{comb}}(\psi(2S), \mu\mu)$	1128.83 ± 35.01
$N_{\text{other part. rec.}}(J/\psi, ee)$	328.63 ± 72.24
$N_{\text{other part. rec.}}(J/\psi, \mu\mu)$	2374.74 ± 185.34
$N_{\text{sig}}(J/\psi, \mu\mu)$	7728.79 ± 92.4
$N_{\text{sig}}(\psi(2S), \mu\mu)$	774.21 ± 29.52
$C_{\text{comb}}(J/\psi, ee)$	$(-1.693 \pm 0.0644) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), ee)$	$(-2.53 \pm 0.347) \times 10^{-3}$
$C_{\text{comb}}(J/\psi, \mu\mu)$	$(-1.895 \pm 0.0262) \times 10^{-3}$
$C_{\text{comb}}(\psi(2S), \mu\mu)$	$(-2.015 \pm 0.111) \times 10^{-3}$
$\epsilon_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ into $\psi(2S))$	$(4.585 \pm 0.349) \times 10^{-6}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(e^+e^-))$	$(1.1368 \pm 0.0251) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda(1520) J/\psi(\mu^+\mu^-))$	$(7.702 \pm 0.0758) \times 10^{-4}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(e^+e^-))$	$(7.233 \pm 0.369) \times 10^{-6}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(e^+e^-))$	$(5.715 \pm 0.334) \times 10^{-6}$
$\epsilon(\Xi_b^- \to \Xi^- J/\psi(\mu^+\mu^-))$	$(4.698 \pm 0.157) \times 10^{-5}$
$\epsilon(\Xi_b^- \to \Xi^- \psi(2S)(\mu^+\mu^-))$	$(5.053 \pm 0.202) \times 10^{-5}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$	$(1.8477 \pm 0.0222) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$	$(2.1755 \pm 0.025) \times 10^{-4}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-))$	$(1.03299 \pm 0.00704) \times 10^{-3}$
$\epsilon(\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-))$	$(1.40451 \pm 0.00838) \times 10^{-3}$
$a_{\text{partRec}}(\Lambda_b^0 \to \Lambda(1520) J/\psi)$	0.4444 ± 0.0424
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- J/\psi)$	0.0 ± 0.0755
$a_{\text{partRec}}(\Xi_b^- \to \Xi^- \psi(2S))$	0.0 ± 1.0
$R^{-1}_{\psi(2S)}$	1.001 ± 0.126
$r^{-1}_{J/\psi}$	0.9809 ± 0.0386
mean shift (ee)	0.352 ± 0.46
mean shift $(\mu\mu)$	-0.2192 ± 0.0877
other part. rec. mean $(J/\psi, ee)$	5037.69 ± 12.87
other part. rec. width $(J/\psi, ee)$	77.89 ± 13.84
other part. rec. mean $(J/\psi, \mu\mu)$	5012.63 ± 4.11
other part. rec. width $(J/\psi, \mu\mu)$	67.39 ± 4.56
width scale (ee)	0.9802 ± 0.0583
width scale $(\mu\mu)$	1.1207 ± 0.0146

Table E.12: An overview of the data fit constant parameters for the fits in the resonance regions (R2p2-LL).

	Value
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U δ	0.46804
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U γ	-0.15787
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U λ	7.68717
$\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ Johnson S_U μ	5619.15216
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U δ	0.56048
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U γ	-0.1497
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U λ	6.06674
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-)$ Johnson S_U μ	5619.37834
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_L	1.32434
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB α_R	1.50863
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB m_0	5619.81562
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_L	3.45398
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB n_R	3.55975
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_L	5.43386
$\Lambda_b^0 \to \Lambda^0 J/\psi(\mu^+\mu^-)$ AsymDSCB σ_R	6.10664
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_L	1.34345
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB α_R	1.40704
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB m_0	5619.83279
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_L	3.74746
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB n_R	4.02002
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_L	4.11883
$\Lambda_b^0 \to \Lambda^0 \psi(2S)(\mu^+\mu^-)$ AsymDSCB σ_R	4.37475

F Further details on the fits in the rare q^2 regions

The free and constant fit parameters of the nominal data fit in the high q^2 regions are shown in Table F.1 and Table F.2.

The free fit parameters of the nominal data fit in the central and low q^2 regions are shown in Table F.3 (central q^2) and Table F.4 (low q^2).

The constant fit parameters of the nominal data fit in the central and low q^2 regions are shown in Table F.5 (central q^2) and Table F.6 (low q^2).

Table F.7 shows the free fit parameters determined during the alternative fit in the high q^2 region without the misidentified hadronic component. Table F.8 shows the free fit parameters determined during the alternative fit in the high q^2 region with the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ component.

The signal yields and efficiencies of $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ in the high q^2 region have been omitted to prevent indirectly estimating its branching fraction.

The meaning of the free parameters for these data fits are explained as follows.

- N_{comb}: the yield of the combinatorial background.
- C_{comb}: the slope parameter of the exponential shape of the combinatorial background in the central or low q^2 region.
- $N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$, $N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$: the yields of the $\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-)$ leakage background in the high q^2 region or the yields of the $\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-)$ leakage background in the central q^2 region. These leakage background yields are shown in the three data-taking periods separately because of the technical implementation of the data fit. The leakage yields for the merged data-taking periods are parameterised as the sum of the yields in separate data-taking periods, which are constrained nuisance parameters.
- norm. factor: the normalisation factor as described in Equation 9.42 for the central or the low q^2 region.
- $N(\Lambda_b^0 \to \Lambda(1405)e^+e^-)$: the yields of the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ component.

The meaning of the constant parameters in Table F.2, Table F.5 and Table F.6 are explained as follows.

- $N_{\text{Hadronic misID}}$: the yield parameters for the misidentified hadronic background in the electron channel based on the `Fail-To-Pass` method.

- Comb. ($ee/\mu\mu$, DD/LL/$\cdots$) `RooGamma` $\beta/\gamma/\mu$: shape parameters of the `RooGamma` functions determined using the SS data (in order to model the combinatorial background in the high q^2 region).

- G0/G1 fraction : the fraction of G0/G1 bremsstrahlung recovery category for the $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ signal. They are determined based on simulation.

- R_Λ^{-1}: the observable in the high q^2 region, which is fixed in the data fit with the blinded electron channel data.

- mean shift : the mean shift parameters for the signal shapes, which are fixed to 0 at the moment. It can potentially be used in the future to accommodate possible data/simulation discrepancy depending on the mass calibration study.

- width scale : the width scale parameters for the signal shapes, which are fixed to 1 at the moment. It can potentially be used in the future to accommodate possible data/simulation discrepancy depending on the mass calibration study.

- $r_{ee}^{\mathcal{B}}$: the observable in the central and low q^2 regions, which is fixed in the data fit with the blinded electron channel data.

- In addition, there are shapes parameters of the Johnson S_U functions and AsymDSCB functions determined based on simulation to model the signal shapes of $\Lambda_b^0 \to \Lambda^0 e^+ e^-$ or $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ in different categories.

Table F.1: An overview of the data fit free parameters for the fit in the high q^2 region. The $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ yields and efficiencies are omitted.

	Fit result
$N_{\text{comb}}(ee)$ DD	49.94 ± 18.56
$N_{\text{comb}}(ee)$ LL	36.82 ± 13.99
$N_{\text{comb}}(\mu\mu)$ R1-DD	319.49 ± 19.18
$N_{\text{comb}}(\mu\mu)$ R1-LL	149.01 ± 13.2
$N_{\text{comb}}(\mu\mu)$ R2p1-DD	263.03 ± 17.5
$N_{\text{comb}}(\mu\mu)$ R2p1-LL	153.69 ± 13.46
$N_{\text{comb}}(\mu\mu)$ R2p2-DD	424.19 ± 22.34
$N_{\text{comb}}(\mu\mu)$ R2p2-LL	210.32 ± 15.82
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R1-DD	0.911 ± 0.2
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R1-LL	0.805 ± 0.214
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p1-DD	2.026 ± 0.311
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p1-LL	1.48 ± 0.271
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p2-DD	3.448 ± 0.403
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p2-LL	2.814 ± 0.395
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R1-DD	$(2.1533 \pm 0.0395) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R1-LL	$(1.2164 \pm 0.0322) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p1-DD	$(3.3908 \pm 0.0492) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p1-LL	$(1.9819 \pm 0.0404) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p2-DD	$(3.1158 \pm 0.0322) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p2-LL	$(2.231 \pm 0.0278) \times 10^{-4}$
$r_{J/\psi}^{-1}$ R1-DD	0.9071 ± 0.06
$r_{J/\psi}^{-1}$ R1-LL	0.9544 ± 0.0746
$r_{J/\psi}^{-1}$ R2p1-DD	1.0285 ± 0.0449
$r_{J/\psi}^{-1}$ R2p1-LL	0.9822 ± 0.0531
$r_{J/\psi}^{-1}$ R2p2-DD	1.0158 ± 0.0322
$r_{J/\psi}^{-1}$ R2p2-LL	0.9813 ± 0.0386

Table F.2: An overview of the data fit constant parameters for the fit in the high q^2 region.

	Value
$N_{\text{Hadronic misID}}$ DD	43.20189
$N_{\text{Hadronic misID}}$ LL	13.60258
Comb. (ee, DD) `RooGamma` β	446.27469
Comb. (ee, DD) `RooGamma` γ	3.65174
Comb. (ee, DD) `RooGamma` μ	4922.13558
Comb. (ee, LL) `RooGamma` β	480.84486
Comb. (ee, LL) `RooGamma` γ	3.26995
Comb. (ee, LL) `RooGamma` μ	4881.59365
Comb. ($\mu\mu$, R1-DD) `RooGamma` β	377.97978
Comb. ($\mu\mu$, R1-DD) `RooGamma` γ	3.3271
Comb. ($\mu\mu$, R1-DD) `RooGamma` μ	4966.3637
Comb. ($\mu\mu$, R1-LL) `RooGamma` β	256.85907
Comb. ($\mu\mu$, R1-LL) `RooGamma` γ	4.16059
Comb. ($\mu\mu$, R1-LL) `RooGamma` μ	4903.75273
Comb. ($\mu\mu$, R2p1-DD) `RooGamma` β	550.81067
Comb. ($\mu\mu$, R2p1-DD) `RooGamma` γ	2.80772
Comb. ($\mu\mu$, R2p1-DD) `RooGamma` μ	4993.11565
Comb. ($\mu\mu$, R2p1-LL) `RooGamma` β	384.72702
Comb. ($\mu\mu$, R2p1-LL) `RooGamma` γ	3.27041
Comb. ($\mu\mu$, R2p1-LL) `RooGamma` μ	4920.42539
Comb. ($\mu\mu$, R2p2-DD) `RooGamma` β	468.68783
Comb. ($\mu\mu$, R2p2-DD) `RooGamma` γ	3.16272
Comb. ($\mu\mu$, R2p2-DD) `RooGamma` μ	4941.34748
Comb. ($\mu\mu$, R2p2-LL) `RooGamma` β	350.96834
Comb. ($\mu\mu$, R2p2-LL) `RooGamma` γ	3.34016
Comb. ($\mu\mu$, R2p2-LL) `RooGamma` μ	4940.15135
G0 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) DD	0.22481
G0 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) LL	0.23261
G1 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) DD	0.50167
G1 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) LL	0.49837
R_Λ^{-1}	1.0
mean shift (ee, DD)	0.0
mean shift (ee, LL)	0.0
mean shift ($\mu\mu$, R1-DD)	0.0
mean shift ($\mu\mu$, R1-LL)	0.0

Continue to next page

	Value
mean shift ($\mu\mu$, R2p1-DD)	0.0
mean shift ($\mu\mu$, R2p1-LL)	0.0
mean shift ($\mu\mu$, R2p2-DD)	0.0
mean shift ($\mu\mu$, R2p2-LL)	0.0
width scale (ee, DD)	1.0
width scale (ee, LL)	1.0
width scale ($\mu\mu$, R1-DD)	1.0
width scale ($\mu\mu$, R1-LL)	1.0
width scale ($\mu\mu$, R2p1-DD)	1.0
width scale ($\mu\mu$, R2p1-LL)	1.0
width scale ($\mu\mu$, R2p2-DD)	1.0
width scale ($\mu\mu$, R2p2-LL)	1.0
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U δ	1.31893
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U γ	2.15068
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U λ	40.64589
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U μ	5637.04886
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB α_L	0.78999
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB α_R	1.0984
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB m_0	5599.94722
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB n_L	135.03244
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB n_R	3.08048
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB σ_L	74.31661
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) AsymDSCB σ_R	41.50019
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U δ	0.97633
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U γ	-0.25316
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U λ	89.26822
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U μ	5591.42371
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U δ	1.30109
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U γ	2.17721
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U λ	41.97647
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U μ	5637.87173
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB α_L	0.88217
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB α_R	0.61072
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB m_0	5609.59177
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB n_L	139.20104
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB n_R	3.58732
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB σ_L	88.47881
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) AsymDSCB σ_R	25.74829
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U δ	0.99603

Continue to next page

	Value
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U γ	-0.33603
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U λ	96.95365
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U μ	5584.00948
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB α_L	1.75911
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB α_R	1.77646
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB m_0	5620.66772
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB n_L	1.82165
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB n_R	8.57098
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB σ_L	15.92713
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-DD) AsymDSCB σ_R	15.42255
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB α_L	1.57358
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB α_R	1.86109
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB m_0	5620.1378
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB n_L	2.2256
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB n_R	9.99724
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB σ_L	14.83745
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R1-LL) AsymDSCB σ_R	15.90762
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB α_L	1.91525
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB α_R	1.80216
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB m_0	5621.87649
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB n_L	1.43802
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB n_R	9.99969
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB σ_L	16.63642
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-DD) AsymDSCB σ_R	14.8894
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB α_L	1.66331
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB α_R	1.73191
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB m_0	5621.77332
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB n_L	2.21082
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB n_R	9.99995
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB σ_L	16.03975
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p1-LL) AsymDSCB σ_R	14.78057
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB α_L	1.6774
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB α_R	1.87313
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB m_0	5621.71112
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB n_L	2.41151
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB n_R	7.00498
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB σ_L	16.06781
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-DD) AsymDSCB σ_R	14.94363
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB α_L	1.66329

Continue to next page

	Value
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB α_R	1.92943
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB m_0	5621.08623
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB n_L	2.11395
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB n_R	9.99921
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB σ_L	15.57162
$\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ (R2p2-LL) AsymDSCB σ_R	15.75558
End of the table	

Table F.3: An overview of the data fit free parameters for the fit in the central q^2 region.

	Fit result
N_{comb} DD	217.16 ± 23.36
N_{comb} LL	13.62 ± 8.2
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$ R1-DD	0.463 ± 0.296
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$ R1-LL	2.401 ± 0.714
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$ R2p1-DD	0.604 ± 0.51
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$ R2p1-LL	1.999 ± 0.762
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$ R2p2-DD	2.559 ± 0.724
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 J/\psi(e^+e^-))$ R2p2-LL	1.941 ± 0.575
C_{comb} DD	$(-1.164 \pm 0.203) \times 10^{-3}$
C_{comb} LL	$(-1.02 \pm 1.03) \times 10^{-3}$
norm. factor DD	1165.88 ± 33.86
norm. factor LL	1099.12 ± 37.29

Table F.4: An overview of the data fit free parameters for the fit in the low q^2 region.

	Fit result
N_{comb} DD	147.34 ± 17.86
N_{comb} LL	5.05 ± 3.59
C_{comb} DD	$(-1.282 \pm 0.238) \times 10^{-3}$
C_{comb} LL	$(-1.94 \pm 1.53) \times 10^{-3}$
norm. factor DD	1120.24 ± 43.7
norm. factor LL	928.31 ± 40.42

Table F.5: An overview of the data fit constant parameters for the fit in the central q^2 region.

	Value
$N_{\text{Hadronic misID}}$ DD	70.22628
$N_{\text{Hadronic misID}}$ LL	14.07243
$r_{ee}^{\mathcal{B}}$	0.00115
G0 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) DD	0.32572
G0 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) LL	0.29872
G1 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) DD	0.48574
G1 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) LL	0.49525
mean shift (ee, DD)	0.0
mean shift (ee, LL)	0.0
width scale (ee, DD)	1.0
width scale (ee, LL)	1.0
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U δ	1.05224
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U γ	2.4452
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U λ	18.93652
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U μ	5634.18843
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U δ	0.79376
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U γ	0.5957
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U λ	46.82125
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U μ	5607.12731
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U δ	0.87466
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U γ	0.12521
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U λ	59.5018
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U μ	5606.23252
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U δ	0.90043
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U γ	1.95808
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U λ	23.60875
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U μ	5626.79635
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U δ	0.70101
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U γ	0.69578
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U λ	42.5522
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U μ	5612.36827
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U δ	0.69748
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U γ	0.34877
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U λ	45.72026
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U μ	5623.2093

Table F.6: An overview of the data fit constant parameters for the fit in the low q^2 region.

	Value
$N_{\text{Hadronic misID}}$ DD	39.04029
$N_{\text{Hadronic misID}}$ LL	1.6212
$r_{ee}^{\mathcal{B}}$	0.00137
G0 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) DD	0.33705
G0 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) LL	0.3349
G1 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) DD	0.50072
G1 fraction ($\Lambda_b^0 \to \Lambda^0 e^+ e^-$) LL	0.5159
mean shift (ee, DD)	0.0
mean shift (ee, LL)	0.0
width scale (ee, DD)	1.0
width scale (ee, LL)	1.0
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U δ	0.91282
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U γ	1.7507
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U λ	26.37136
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G0) Johnson S_U μ	5621.92144
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U δ	0.77414
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U γ	0.47648
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U λ	45.21882
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G1) Johnson S_U μ	5606.6947
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U δ	0.79188
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U γ	$-3e-05$
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U λ	46.40662
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (DD-G2) Johnson S_U μ	5609.37927
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U δ	0.76629
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U γ	1.28274
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U λ	28.98592
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G0) Johnson S_U μ	5608.96361
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U δ	0.78246
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U γ	0.69054
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U λ	39.9757
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G1) Johnson S_U μ	5622.70596
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U δ	0.85833
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U γ	0.05711
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U λ	66.67065
$\Lambda_b^0 \to \Lambda^0 e^+ e^-$ (LL-G2) Johnson S_U μ	5613.2633

Table F.7: An overview of the data fit free parameters for an alternative fit (without the misidentified hadronic component in the electron channel) in the high q^2 region. The $\Lambda_b^0 \to \Lambda^0 \mu^+ \mu^-$ yields and efficiencies are omitted.

	Fit result
$N_{\text{comb}}(ee)$ DD	102.53 ± 19.7
$N_{\text{comb}}(ee)$ LL	51.0 ± 14.02
$N_{\text{comb}}(\mu\mu)$ R1-DD	319.38 ± 19.17
$N_{\text{comb}}(\mu\mu)$ R1-LL	149.01 ± 13.2
$N_{\text{comb}}(\mu\mu)$ R2p1-DD	262.91 ± 17.49
$N_{\text{comb}}(\mu\mu)$ R2p1-LL	153.68 ± 13.46
$N_{\text{comb}}(\mu\mu)$ R2p2-DD	424.01 ± 22.33
$N_{\text{comb}}(\mu\mu)$ R2p2-LL	210.31 ± 15.82
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R1-DD	0.907 ± 0.2
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R1-LL	0.804 ± 0.214
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p1-DD	2.016 ± 0.311
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p1-LL	1.478 ± 0.271
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p2-DD	3.431 ± 0.403
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0 \psi(2S)(e^+e^-))$ R2p2-LL	2.81 ± 0.395
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R1-DD	$(2.1535 \pm 0.0395) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R1-LL	$(1.2165 \pm 0.0322) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p1-DD	$(3.391 \pm 0.0492) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p1-LL	$(1.9819 \pm 0.0404) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p2-DD	$(3.1161 \pm 0.0322) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p2-LL	$(2.231 \pm 0.0278) \times 10^{-4}$
$r_{J/\psi}^{-1}$ R1-DD	0.908 ± 0.06
$r_{J/\psi}^{-1}$ R1-LL	0.9545 ± 0.0746
$r_{J/\psi}^{-1}$ R2p1-DD	1.0292 ± 0.0449
$r_{J/\psi}^{-1}$ R2p1-LL	0.9822 ± 0.0531
$r_{J/\psi}^{-1}$ R2p2-DD	1.0165 ± 0.0322
$r_{J/\psi}^{-1}$ R2p2-LL	0.9814 ± 0.0386

Table F.8: An overview of the data fit free parameters for an alternative fit (with the $\Lambda_b^0 \to \Lambda(1405)e^+e^-$ component in the electron channel) in the high q^2 region. The $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ yields and efficiencies are omitted.

	Fit result
$N_{\text{comb}}(ee)$ DD	34.65 ± 17.14
$N_{\text{comb}}(ee)$ LL	15.92 ± 12.92
$N_{\text{comb}}(\mu\mu)$ R1-DD	319.56 ± 19.18
$N_{\text{comb}}(\mu\mu)$ R1-LL	149.05 ± 13.2
$N_{\text{comb}}(\mu\mu)$ R2p1-DD	263.13 ± 17.5
$N_{\text{comb}}(\mu\mu)$ R2p1-LL	153.75 ± 13.46
$N_{\text{comb}}(\mu\mu)$ R2p2-DD	424.35 ± 22.35
$N_{\text{comb}}(\mu\mu)$ R2p2-LL	210.42 ± 15.83
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$ R1-DD	0.915 ± 0.2
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$ R1-LL	0.813 ± 0.214
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$ R2p1-DD	2.035 ± 0.311
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$ R2p1-LL	1.492 ± 0.271
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$ R2p2-DD	3.462 ± 0.403
$N_{\text{leak}}(\Lambda_b^0 \to \Lambda^0\psi(2S)(e^+e^-))$ R2p2-LL	2.84 ± 0.395
$N(\Lambda_b^0 \to \Lambda(1405)e^+e^-)$ DD	38.4 ± 18.07
$N(\Lambda_b^0 \to \Lambda(1405)e^+e^-)$ LL	25.13 ± 12.29
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R1-DD	$(2.1532 \pm 0.0395) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R1-LL	$(1.2163 \pm 0.0322) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p1-DD	$(3.3906 \pm 0.0492) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p1-LL	$(1.9818 \pm 0.0404) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p2-DD	$(3.1157 \pm 0.0322) \times 10^{-4}$
$\epsilon_{\text{sig}}(\Lambda_b^0 \to \Lambda^0 e^+e^-)$ R2p2-LL	$(2.2309 \pm 0.0278) \times 10^{-4}$
$r_{J/\psi}^{-1}$ R1-DD	0.9064 ± 0.06
$r_{J/\psi}^{-1}$ R1-LL	0.9537 ± 0.0746
$r_{J/\psi}^{-1}$ R2p1-DD	1.028 ± 0.0449
$r_{J/\psi}^{-1}$ R2p1-LL	0.9817 ± 0.0531
$r_{J/\psi}^{-1}$ R2p2-DD	1.0153 ± 0.0322
$r_{J/\psi}^{-1}$ R2p2-LL	0.9808 ± 0.0386

References

[1] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[2] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[3] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[4] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[5] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[6] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[7] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398–410, 1983.

[8] P. Bagnaia et al. Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider. *Phys. Lett. B*, 129:130–140, 1983.

[9] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[10] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[11] A. A. Alves Jr. et al. The LHCb detector at the LHC. *JINST*, 3(LHCb-DP-2008-001):S08005, 2008.

[12] LHC Machine. *JINST*, 3:S08001, 2008.

[13] Roel Aaij et al. The LHCb Upgrade I. *JINST*, 19:P05065, 2024.

[14] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[15] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[16] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[17] R. Aaij et al. LHCb detector performance. *Int. J. Mod. Phys.*, A30:1530022, 2015.

[18] Jorg Wenninger. Operation and Configuration of the LHC in Run 2. 2019. https://cds.cern.ch/record/2668326.

[19] G. Papotti, X. Buffat, W. Herr, R. Giachino, and T. Pieloni. Observations of beam-beam effects at the LHC. In *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, pages 1–5, 2014.

[20] O. Bruning, H. Burkhardt, and S. Myers. The Large Hadron Collider. *Prog. Part. Nucl. Phys.*, 67:705–734, 2012.

[21] Roel Aaij et al. Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. *JINST*, 14(LHCb-DP-2019-001):P04013, 2019.

[22] S. Amato et al. LHCb technical proposal: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays. 2 1998.

[23] LHCb reoptimized detector design and performance: Technical Design Report, 2003.

[24] LHCb magnet: Technical Design Report, 2000.

[25] LHCb VELO (VErtex LOcator): Technical Design Report, 2001.

[26] R. Aaij et al. Performance of the LHCb Vertex Locator. *JINST*, 9:P09007, 2014.

[27] LHCb inner tracker: Technical Design Report, 2002.

[28] M. Adinolfi et al. Performance of the LHCb RICH detector at the LHC. *Eur. Phys. J.*, C73:2431, 2013.

[29] LHCb calorimeters: Technical Design Report, 2000.

[30] C. Abellan Beteta et al. Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC. 2020. submitted to JINST.

[31] G. Lanfranchi, X. Cid Vidal, S. Furcas, M. Gandelman, J. A. Hernando, J. H. Lopez, E. Polycarpo, and A. Sarti. The Muon Identification Procedure of the LHCb Experiment for the First Data. 8 2009.

[32] F. Archilli et al. Performance of the muon identification at LHCb. *JINST*, 8:P10020, 2013.

[33] R. Aaij et al. The LHCb trigger and its performance in 2011. *JINST*, 8:P04022, 2013.

[34] R. Aaij et al. Tesla: an application for real-time data analysis in high energy physics. *Comput. Phys. Commun.*, 208:35–42, 2016.

[35] Roel Aaij et al. A comprehensive real-time analysis model at the LHCb experiment. *JINST*, 14:P04006, 2019.

[36] Lex Marinus Greeven. Decoding Beauty: rare baryonic decays & SciFi detector commissioning. 2024. PhD thesis.

[37] Letter of Intent for the LHCb Upgrade. 3 2011.

[38] Conor Fitzpatrick and Vladimir V. Gligorov. The anatomy of an LHCb event in the upgrade era, and implications for the LHCb trigger. 2014.

[39] LHCb Trigger and Online Upgrade Technical Design Report, 2014.

[40] LHCb VELO Upgrade Technical Design Report, 2013.

[41] Christian Joram, Ulrich Uwer, Blake Dean Leverington, Thomas Kirn, Sebastian Bachmann, Robert Jan Ekelhof, and Janine Mueller. LHCb Scintillating Fibre Tracker Engineering Design Review Report: Fibres, Mat and Modules, 2015. https://cds.cern.ch/record/2004811.

[42] Daniel Berninghoff. Commissioning of the Front-End Electronics of the LHCb Scintillating Fibre Tracker. 2022. Dissertation.

[43] LHCb Tracker Upgrade Technical Design Report, 2014.

[44] A. Comerma and J. Mazorra. PACIFICr5 - 64 channels SiPM readout ASIC for the SciFi Tracker detector, 2020. https://edms.cern.ch/document/1841761/1.

[45] Lukas Witola. Calibration and performance studies of the readout ASIC for the LHCb SciFi Tracker. 2019. Master Thesis.

[46] Lukas Witola. Commissioning, Calibration & Early Performance of the LHCb Scintillating Fibre Tracker. 2024. Dissertation.

[47] A. Gabrielli, S. Bonacini, K. Kloukinas, A. Marchioro, P. Moreira, A. Ranieri, and D. De Robertis. The GBT-SCA, a radiation tolerant ASIC for detector control applications in SLHC experiments. In *Topical Workshop on Electronics for Particle Physics*. CERN, 2009.

[48] P. Moreira et al. The GBT Project. In *Topical Workshop on Electronics for Particle Physics*. CERN, 2009.

[49] L. Amaral et al. The versatile link, a common project for super-LHC. *JINST*, 4:P12003, 2009.

[50] J. P. Cachemiche, P. Y. Duval, F. Hachon, R. Le Gac, and F. Réthoré. The PCIe-based readout system for the LHCb experiment. *JINST*, 11(02):P02013, 2016.

[51] Federico Alessio, Paolo Durante, and Guillaume Vouters. The readout supervisor firmware for controlling the upgraded LHCb detector and readout system. 6 2018.

[52] F. Alessio and R. Jacobsson. Readout Control Specifications for the Front-End and Back-End of the LHCb Upgrade. 3 2013.

[53] F. Alessio, Z. Guzik, and R. Jacobsson. An LHCb general-purpose acquisition board for beam and background monitoring at the LHC. *JINST*, 6:C01001, 2011.

[54] F. Alessio, Z. Guzik, and R. Jacobsson. LHCb Global Timing and Monitoring of the LHC Filling Scheme. 1 2011.

[55] W. Vink and H. Verkooijen. Master Board, technical manual, 2024.

[56] M. Adinolfi, F. Archilli, W. Baldini, A. Baranov, D. Derkach, A. Panin, A. Pearce, and A. Ustyuzhanin. LHCb data quality monitoring. *J. Phys. Conf. Ser.*, 898(9):092027, 2017.

[57] Jessy Daniel, Xiaoxue Han, Blake Leverington, Pascal Perret, Chishuai Wang, and Zehua Xu. Time alignment of the front end electronics of the LHCb SciFi, 2024. https://cds.cern.ch/record/2917298.

[58] C. Abellan Beteta et al. Time alignment of the front end electronics of the LHCb calorimeters. *JINST*, 7:P08020, 2012.

[59] Y. S. Amhis. Time alignment of the electromagnetic and hadronic calorimeters, reconstruction of the $B \to D^- \rho(770)^+, B_s \to D_s^- \rho(770)^+$ and $B_s \to D_s^- K^{*+}(892)$ decay channels with the LHCb detector. 2009.

[60] Jessy Daniel. Measurement of the CKM angle γ in $B^\pm \to [K_S^0 \pi^+ \pi^- \pi^0]_D h^\pm$ decays at LHCb and commissioning of the SciFi tracker. 2024.

[61] The LHCb collaboration. SciFi Performance plots July 2024. 2024. https://cds.cern.ch/record/2904659.

[62] Navas S. et al. Review of particle physics. *to be published in Phys. Rev*, D110(8):030001, 2024.

[63] Antonio Pich. The Standard Model of Electroweak Interactions. pages 1–50, 1 2012.

[64] Navas S. et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024. and 2025 update.

[65] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[66] Makoto Kobayashi and Toshihide Maskawa. CP-violation in the renormalizable theory of weak interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[67] B. Pontecorvo. Inverse Beta Processes and Nonconservation of Lepton Charge. *Sov. Phys. JETP*, 7:172–173, 1958.

[68] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962.

[69] Ben Gripaios, Marco Nardecchia, and S. A. Renner. Composite leptoquarks and anomalies in B-meson decays. *JHEP*, 05:006, 2015.

[70] Alejandro Celis, Javier Fuentes-Martin, Martín Jung, and Hugo Serodio. Family nonuniversal Z' models with protected flavor-changing interactions. *Phys. Rev. D*, 92(1):015007, 2015.

[71] Gudrun Hiller and Martin Schmaltz. R_K and future $b \to s\ell\ell$ physics beyond the standard model opportunities. *Phys. Rev. D*, 90:054014, 2014.

[72] Christoph Bobeth, Gudrun Hiller, and Giorgi Piranishvili. Angular distributions of $\bar{B} \to \bar{K}\ell^+\ell^-$ decays. *JHEP*, 12:040, 2007.

[73] Marzia Bordone, Gino Isidori, and Andrea Pattori. On the Standard Model predictions for R_K and R_{K^*}. *Eur. Phys. J. C*, 76(8):440, 2016.

[74] R. Aaij et al. Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays. *Phys. Rev.*, D108:032002, 2023.

[75] R. Aaij et al. Test of lepton universality in $b \to s\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 131:051803, 2023.

[76] Roel Aaij et al. Measurement of the branching fraction ratio R_K at large dilepton invariant mass. *JHEP*, 07:198, 2025.

[77] Roel Aaij et al. Test of Lepton Flavor Universality with Bs0$\to\phi\ell+\ell$-Decays. *Phys. Rev. Lett.*, 134(12):121803, 2025.

[78] R. Aaij et al. Test of lepton universality using $\Lambda_b^0 \to pK^-\ell^+\ell^-$ decays. *JHEP*, 05:040, 2020.

[79] William Detmold and Stefan Meinel. $\Lambda_b \to \Lambda\ell^+\ell^-$ form factors, differential branching fraction, and angular observables from lattice QCD with relativistic b quarks. *Phys. Rev. D*, 93(7):074501, 2016.

[80] R. Aaij et al. Branching fraction measurements of the rare $B_s^0 \to \phi\mu^+\mu^-$ and $B_s^0 \to f_2'(1525)\mu^+\mu^-$ decays. *Phys. Rev. Lett.*, 127:151801, 2021.

[81] R. Aaij et al. Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction. *JHEP*, 11:047, 2016.

[82] R. Aaij et al. Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays. *JHEP*, 06:133, 2014.

[83] R. Aaij et al. Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay. *Phys. Rev. Lett.*, 125:011802, 2020.

[84] R. Aaij et al. Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ decays. *JHEP*, 06:115, 2015.

[85] Roel Aaij et al. Measurement of the branching fraction of the $\Lambda_b^0 \to J/\psi\Lambda$ decay and isospin asymmetry of $B \to J/\psi K$ decays. 9 2025.

[86] Wouter D. Hulsbergen. Decay chain fitting with a Kalman filter. *Nucl. Instrum. Meth.*, A552:566–575, 2005.

[87] R. Aaij et al. Measurement of the track reconstruction efficiency at LHCb. *JINST*, 10:P02007, 2015.

[88] R. Aaij et al. Measurement of b-hadron fractions in 13 TeV $p\,p$ collisions. *Phys. Rev.*, D100:031102(R), 2019.

[89] R. Aaij et al. Isospin amplitudes in $\Lambda_b^0 \to J/\psi\Lambda^0(\Sigma^0)$ and $\Xi_b^0 \to J/\psi\Xi^0(\Lambda^0)$ decays. *Phys. Rev. Lett.*, 124:111802, 2020.

[90] Muriel Pivk and Francois R. Le Diberder. sPlot: A statistical tool to unfold data distributions. *Nucl. Instrum. Meth.*, A555:356–369, 2005.

[91] The LHCb collaboration. The stripping project. online at https://lhcbdoc.web.cern.ch/lhcbdoc/stripping/.

[92] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. PYTHIA 6.4 physics and manual. *JHEP*, 05:026, 2006.

[93] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[94] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

[95] S. Agostinelli et al. Geant4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250, 2003.

[96] John Allison, K. Amako, J. Apostolakis, H. Araujo, P.A. Dubois, et al. Geant4 developments and applications. *IEEE Trans.Nucl.Sci.*, 53:270, 2006.

[97] R. Aaij et al. Tests of lepton universality using $B^0 \to K^0_S\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 128:191802, 2022.

[98] The LHCb collaboration. Stripping project `Bu2LLK_eeLine2`. online at Line `Bu2LLK_eeLine2` version 34r0p1.

[99] The LHCb collaboration. Stripping project `Bu2LLK_mmLine`. online at Line `Bu2LLK_mmLine` version 34r0p1.

[100] The LHCb collaboration. Stripping project `Bu2LLK_eeSSLine2`. online at Line `Bu2LLK_eeSSLine2` version 34r0p1.

[101] The LHCb collaboration. Stripping project Bu2LLK_mmSSLine. online at Line Bu2LLK_mmSSLine version 34r0p1.

[102] R. Aaij et al. Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2. *Eur. Phys. J. Tech. Instr.*, 6:1, 2019.

[103] Lucio Anderlini, Andrea Contu, Christopher Rob Jones, Sneha Sirirshkumar Malde, Dominik Muller, Stephen Ogilvy, Juan Martin Otalora Goicochea, Alex Pearce, Ivan Polyakov, Wenbin Qian, Barbara Sciascia, Ricardo Vazquez Gomez, and Yanxi Zhang. The PIDCalib package, 2016.

[104] V. V. Gligorov and M. Williams. Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. *JINST*, 8:P02013, 2013.

[105] T. Likhomanenko et al. LHCb topological trigger reoptimization. *J. Phys. Conf. Ser.*, 664:082025, Oct 2015.

[106] Y. Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021. *Phys. Rev.*, D107:052008, 2023. updated results and plots available at https://hflav.web.cern.ch.

[107] R. Aaij et al. Measurement of the $\Lambda_b^0 \to J/\psi\Lambda^0$ angular distribution and the Λ^0 polarisation in pp collisions. *JHEP*, 06:110, 2020.

[108] Thomas Gutsche, Mikhail A. Ivanov, Jürgen G. Körner, Valery E. Lyubovitskij, and Pietro Santorelli. Polarization effects in the cascade decay $\Lambda_b \to \Lambda(\to p\pi^-) + J/\psi(\to \ell^+\ell^-)$ in the covariant confined quark model. *Phys. Rev. D*, 88(11):114018, 2013.

[109] Roel Aaij et al. Measurement of the electron reconstruction efficiency at LHCb. *JINST*, 14:P11023, 2019.

[110] S Tolk, J Albrecht, F Dettori, and A Pellegrino. Data driven trigger efficiency determination at LHCb, 2014.

[111] A A Alves Jr. et al. Performance of the LHCb muon system. *JINST*, 8:P02022, 2013.

[112] A. Rogozhnikov. Reweighting with boosted decision trees. *J. Phys. Conf. Ser.*, 762(1):012036, 2016. https://github.com/arogozhnikov/hep_ml.

[113] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and regression trees*. Wadsworth international group, Belmont, California, USA, 1984.

[114] Fabian Pedregosa et al. Scikit-learn: Machine learning in Python. *J. Machine Learning Res.*, 12:2825, 2011. and online at http://scikit-learn.org/stable/.

[115] The scikit-learn developers. scikit-learn User Guide, Supervised learning. online at https://scikit-learn.org/stable/supervised_learning.html.

[116] Andreas Hoecker et al. TMVA 4 — Toolkit for Multivariate Data Analysis with ROOT. Users Guide. 2009.

[117] The scikit-learn developers. scikit-learn User Guide, Cross-validation: evaluating estimator performance. online at https://scikit-learn.org/stable/modules/cross_validation.html.

[118] Giovanni Punzi. Sensitivity of searches for new signals and its optimization. *eConf*, C030908:MODT002, 2003.

[119] The RooFit developers. Roofit manual. online at https://root.cern/manual/roofit/.

[120] A L Read. Presentation of search results: The CL_s technique. *J. Phys.*, G28:2693, 2002.

[121] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[122] N. L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1/2):149–176, 1945.

[123] Tomasz Skwarnicki. *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*. PhD thesis, Institute of Nuclear Physics, Krakow, 1986. DESY-F31-86-02.

[124] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput. Phys. Commun.*, 136:198–207, 2001.

[125] The RooFit developers. `RooKeysPdf` class reference. online at https://root.cern.ch/doc/master/classRooKeysPdf.html.

[126] The RooFit developers. `RooGamma` class reference. online at https://root.cern.ch/doc/master/classRooGamma.html.

[127] R. Aaij et al. Measurement of the $\Lambda_b^0 \to \Lambda(1520)\mu^+\mu^-$ differential branching fraction. *Phys. Rev. Lett.*, 131:151801, 2023.

[128] Lorenzo Moneta, Kevin Belasco, Kyle Cranmer, Sven Kreiss, Alfio Lazzaro, Danilo Piparo, Gregory Schott, Wouter Verkerke, and Matthias Wolf. The RooStats Project, 2011.

[129] The RooStats developers. `AsymptoticCalculator` class reference. online at https://root.cern.ch/doc/master/classRooStats_1_1AsymptoticCalculator.html.

[130] G. A. Cowan, D. C. Craik, and M. D. Needham. RapidSim: an application for the fast simulation of heavy-quark hadron decays. *Comput. Phys. Commun.*, 214:239–246, 2017.

[131] The LHCb collaboration. End of the 2025 proton-proton collision run. online at https://lhcb-outreach.web.cern.ch/2025/11/11/end-of-the-2025-proton-proton-collision-run/.

[132] Framework TDR for the LHCb Upgrade: Technical Design Report, 2012.

[133] Marie-Helene Schune, Francesco Polci, and Martino Borsato. Hop an additional tool for decays involving electrons. Technical report, CERN, Geneva, 2015.