Searching for Charged Lepton Flavour Violation at LHCb and Long-Lived Particles with CODEX-b

By

Paul N. Swallow

A thesis submitted to the University of Birmingham
for the degree of *Doctor Of Philosophy*

UNIVERSITY OF BIRMINGHAM

Particle Physics Research Group
School of Physics and Astronomy
College of Engineering and Physical Sciences
University of Birmingham
August 2023

Abstract

The CODEX-b detector is an RPC-based experiment proposed to search for long-lived particles that may be produced in pp collisions at the LHC. The first comprehensive investigation on the impact of misalignments of CODEX-b are presented alongside reports of hardware quality assurance tests of the readout electronics for this detector.

The mechanical precision ($< 5\,\text{mm}$) when constructing the RPC modules will be more than sufficient to ensure track reconstruction is unaffected by misalignments. The first major quality control process for the evaluation of CODEX-b's RPC readout boards and its results are also presented, with more than 95% of the 440 tested boards validated for installation.

The first experimental upper limits of the branching fraction of the rare decay, $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$, are presented based on analysis of data taken by the LHCb experiment between 2011 and 2018 (Run 1 & 2), corresponding to an integrated luminosity of $9\,\text{fb}^{-1}$.

Although the dataset remains blinded, estimated upper limits of the branching fraction of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ assuming no signal is present were obtained using a background-only proxy dataset derived from the data sidebands. This is evaluated for two scenarios, one with higher efficiency and one with lower background contributions (including a Λ_c^+ decay veto), leading to results of

$$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)_{\text{std}} \lesssim 2.09 \times 10^{-7} \text{ @ } 95\% \text{ CL},$$
$$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)_{\text{veto}} \lesssim 6.45 \times 10^{-8} \text{ @ } 95\% \text{ CL},$$

respectively. These limits are comparable to that from a theoretical analysis of a proposed model-independent limit.

Acknowledgements

Throughout the past few years that have culminated in the writing of this thesis, I have been blessed with the advice and support of a great many people, whose efforts are reflected both in this work and will be in my future contributions. I cannot thank all of you that have made me the person and scientist I am today, but will highlight a few.

First, thanks to my supervisors Nigel, Phil and Niladri. You have each taught me a great deal at different stages of my career and I would like to thank you all for being there whenever I needed you. In particular, I must mention Nigel: your patience, knowledge and the way you happily give your time to others to help however you can are an inspiration to me.

I am exceedingly grateful for the wonderful Lb2L analysis team that I have been privileged to be part of. They are Mick, Lex, Chishuai, Jan, Niladri, Niels, Flavio, and Nigel. You all have made the past few years a joy despite the challenges of working on three analyses in parallel. I especially have to thank Mick, whose deep knowledge and patient support whenever I had questions were incredibly helpful and a template of the PostDoc that I aspire to be.

I want to thank my companions in the Birmingham LHCb office, it has been a pleasure to learn from and teach you. Alongside my friends within particle physics that I have come to know at Birmingham and at CERN, especially the LTA crew, you have made the past few years entertaining, if a bit less sane.

I want to express my thanks to my friends and family who have helped me immensely through not just the past few years, but the steps to this point. I'm very grateful for my parents for their love and support, and mum in particular for correcting my worms when I've misspelt something. For my brother, Joel, I don't think I can show enough gratitude as he has literally led the way for me, or, as he might say, I've copied him even down to the university attended and subsequent career. Thank you Joel for your encouragement throughout the years. I also have to thank my cat, Molly, who would like to be credited as a second author of this thesis after sleeping on the keyboard many times during its writing. I feel it's best to honour Molly via some of her own words: "cfxcvmasaaaa,,,,,,,,,,,,mc".

Finally, to my lovely wife Rebecca, you have been a wonderful part of this journey. You have provided support, an ear to listen to my complaints when my code isn't working, and patience when I had to work late. But you also made sure that I did not burn myself out and enjoyed life outside of the PhD.

A heartfelt thank you to you all.

Declaration of Author's contribution

The work detailed in Chapter 5 was performed entirely by myself, with some support on acquiring and setting up the testing set-up from Section 5.1 from CODEX-b colleagues. For the analysis detailed in Part II I was one of the main authors alongside a small team of LHCb colleagues at the University of Birmingham and Nikhef. I was the main contributor or actively collaborated to all aspects of the analysis; with the exception of the production of samples used, which was predominantly performed by Mick Mulder; the optimisation of the PID selection and its efficiency determination which was performed by Marco Santimaria and Lex Greeven respectively; the development of the trigger selection and adjustment of the TCK selections which was performed by Lex; the re-weighting of the simulated data samples, which was performed by Mick and Lex; and the derivation of the systematic efficiencies, which was performed predominantly by Mick.

Contents

List of Figures

List of Tables

List of Acronyms

ALICE A Large Ion Collider Experiment

ALP Axion-like Particle

ANUBIS AN Underground Belayed In-Shaft experiment

ASIC Application Specific Integrated Circuits

ATLAS A Toroidal Large Hadron Collider ApparatuS

AUC Area Under Curve

BDT Boosted Decision Tree

BSM Beyond the Standard Model

CB Crystal Ball

CERN Conseil européen pour la recherche nucléaire

CKM Cabibbo-Kobayashi-Maskawa

cLFV Charged Lepton Flavour Violation

CMS Compact Muon Solenoid

CODEX-b COmpact Detector for EXotics at LHCb

CODEX-β COmpact Detector for EXotics - Beta

CP Charge-Parity

CPV Charge-Parity Violation

DD Down-Down

DIRA DIRection Angle

DOCA Distance Of Closest Approach

DSCB Double-Sided Crystal Ball

DTF Decay Tree Fitter

ECAL Electromagnetic CALorimeter

EFT Effective Field Theory

EM Electromagnetism

EW ElectroWeak

FASER ForwArd Search ExpeRiment

FCNC Flavour-Changing-Neutral-Current

GBDT Gradient Boosted Decision Tree

GBR Gradient Boosted Re-weighter

GDML Geometry Description Markup Language

GEM Gas Electron Multiplier

GIM Glashow-Iliopoulous-Maiani

HCAL Hadronic CALorimeter

HeRSCheL High Rapidity Shower Counters for LHCb

HLT High Level Trigger

HNL Heavy Neutral Lepton

HPD Hybrid Photodetector

IT Inner Tracker

IP Impact Parameter

IP Interaction Point

L0 Level-0

LEP Large Electron-Positron collider

LFU Lepton Flavour Universality

LFV Lepton Flavour Violation

LHC Large Hadron Collider

LHCb Large Hadron Collider beauty

LINAC2 Linear Accelerator 2

LINAC4 Linear Accelerator 4

LL Long-Long

LLP Long-Lived Particle

LVDS Low-Voltage Differential Signalling

MC Monte Carlo

MD Mag Down

MU Mag Up

MWPC Multi-wire Proportional Chamber

OPE Operator-Product Expansion

OS Opposite Sign

OT Outer Tracker

PCB Printed Circuit Board

PDF Probability Density Function

PID Particle Identification

PMNS Pontecorvo–Maki–Nakagawa–Sakata

PMT Photomultipler Tube

PS PreShower

PS Proton Synchrotron

PSB Proton Synchrotron Booster

PV Primary Vertex

QCD Quantum ChromoDynamics

QED Quantum ElectroDynamics

QFT Quantum Field Theory

RF Radio-frequency

RICH Ring Imaging CHerenkov

RM Rotational Misalignment

ROC Receiver Operator Characteristic

RPC Resistive Plate Chamber

SM Standard Model

SPD Scintillating Pad Detector

SPS Super Proton Synchrotron

SS Same Sign

TCK Trigger Configuration Key

TM Translational Misalignment

TT Tracker Turicensis

VELO VErtex LOcator

WC Wilson coefficient

WIMP Weakly-Interacting Massive Particle

Chapter 1

Introduction

Throughout history, people have attempted to explain the nature of the universe in which they live. The study of physics has taken the world from the early astronomy of ancient Greece to the invention of modern-day computers. The goal of physics has not changed: to understand our universe. The Standard Model (SM) of particle physics provides the best theoretical understanding of the fundamental building blocks of the universe and their interactions. Although the SM provides some of the most precise and experimentally verified predictions in science, such as (g-2) [1–6], it remains incomplete and there are many mysteries left to understand.

After its initial development in the 1960s and 1970s, the SM has been thoroughly tested, with the goal of precisely measuring and exploring its predictions. Several decades later, the fundamental structure of the SM is now rigorously defined, and experiments are now attempting to find inconsistencies in its description of data that may provide insight into the currently unknown complete model of particle interactions. These searches are looking for new physics *i.e.* particles or interactions not defined within the SM, and are informed by expanded theoretical models of the SM. For example, though gravity is a well-understood property within the macroscopic world, the SM does not include any gravitational interactions between the particles. Additionally, observations of the structure of galaxies have revealed so-called 'Dark Matter' [7, 8] that does not interact with matter described by the SM, as well as an energy that appears to be leading to the acceleration of the universe's expansion known as 'Dark Energy' [9, 10]. Neither of these are described by the SM. Furthermore, the SM predicts that matter and antimatter should be produced in an almost identical ratio but the universe itself is dominated by matter. These are a few of the known problems of the SM that are being studied, and imply the possible existence of new structures Beyond the Standard Model (BSM).

The Large Hadron Collider beauty (LHCb) experiment at CERN is performing both indirect and direct searches for BSM physics. Indirect searches use precise measurements of SM predictions, with any significant deviation being potential evidence for BSM particles or interactions. In contrast, direct searches attempt to observe decays from BSM particles that produce detectable particles within the LHCb detector but cannot be explained by a known decay in the SM. However, LHCb is one experiment of many within the field of particle physics attempting to test the SM to its breaking point. Direct searches for new physics can also be performed by searching for Long-Lived Particles (LLPs) with experiments such as the proposed COmpact Detector for EXotics at LHCb (CODEX-b).

This thesis describes an overview of the theoretical basis for the SM and the LHCb detector; the potential of the CODEX-b detector to search for BSM physics in Part I; and the use of LHCb data to test the SM in Part II. Specifically, Chapter 2 describes the theoretical basis of the SM and relevant extensions that can be explored via the study of rare decays or direct searches for BSM particles. Then, Chapters 3 and 4 outline the design of the LHCb detector and the CODEX-b proposal respectively. Chapter 5 investigates the sensitivity of the CODEX-b tracking performance when subject to unknown mechanical misalignments of the detector elements, and hardware quality assurance testing of the readout boards. Chapters 6 to 9 describe a search for the charged lepton flavour violating mode of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$, utilising LHCb data from 2011–2012 and 2015–2018. Finally, Chapter 10 summarises the results of the work described.

Chapter 2

Theoretical Foundations

2.1 The Standard Model of Particle Physics

The SM is a self-consistent, renormalisable, non-abelian, gauge Quantum Field Theory (QFT) to describe the phenomena and properties of particles and their interactions. It has a symmetry group of $SU(3)_c \times SU(2)_L \times U(1)_Y$, which leads to 21 fundamental fields. Of these, 18 correspond to particles, with 12 spin-$\frac{1}{2}$ fermions, five spin-1 bosons and the spin-0 Higgs boson. The remaining three fields correspond to the strong, weak and electromagnetic interactions. The fundamental fermions can appear as particles or antiparticles, where the latter are identical to the former travelling backwards in time with the opposite charge and parity.

The non-abelian gauge group of $SU(3)_c$ in the SM symmetry group corresponds to the strong interaction and Quantum ChromoDynamics (QCD). Quantum ElectroDynamics (QED) is an abelian gauge theory related to Electromagnetism (EM) with the group $U(1)_Y$ [11], and when combined with the group $SU(2)_L$, which describes the weak interaction [12–14], they form a combined local gauge symmetry ($SU(2)_L \times U(1)_Y$) that is spontaneously broken called the ElectroWeak (EW) interaction [15–17]. Each of these interactions can be associated with one or more gauge bosons, which are particles that mediate the underlying interaction. The Standard Model has four gauge bosons: the photon (γ) that mediates electromagnetism; two weak bosons ($W^{\pm}$, Z^0) that mediate the weak interaction; and gluons that mediate the strong force with eight different associated colors. The Higgs field gives rise to the spontaneous symmetry breaking in EW interactions, and this also leads to the associated weak bosons acquiring mass, with the Higgs boson serving as a pseudo-Nambu-Goldstone boson of the Higgs field alongside these weak bosons [18–20]. The other massive particles acquire their mass via Yukawa

couplings with the Higgs field. Overall, the SM has 19 parameters[1]: three coupling parameters related to each of the three fundamental interactions (α_{EM}, G_F, and g_s); two parameters related to the Higgs field, its vacuum expectation energy, v, and the mass of the Higgs boson, m_H; nine fermion masses; four mixing parameters from the Cabibbo-Kobayashi-Maskawa (CKM) matrix; and the QCD phase responsible for Charge-Parity Violation (CPV), θ_{QCD} [21]. These are summarised in Table 2.1, with Figure 2.1 showing a diagram of the fundamental particles and their interactions.

Table 2.1: The 19 free SM parameters in the form of physical observables, with the parameter values from Ref. [22].

Parameter	**Value**
Fine Structure constant, $\alpha_{EM} \equiv \frac{e^2}{4\pi}$	$(137.035999084(21))^{-1}$
Strong coupling, g_s (at the m_Z scale)	0.1179(9)
Fermi Constant, G_F (GeV^{-2})	$1.1663788(6) \times 10^{-5}$
Higgs Mass, m_H (GeV/c^2)	125.25(17)
Vacuum Expectation value, v (GeV/c^2)	~ 246.22
CKM mixing angles: $\sin\theta_{12}$, $\sin\theta_{23}$, $\sin\theta_{13}$	0.22500 ± 0.00067, $0.04182^{+0.00085}_{-0.00074}$, 0.00369 ± 0.00011
CKM matrix CPV phase, δ	1.144(27)
Quark masses: m_u, m_c, m_t (MeV/c^2)	$2.16^{+0.49}_{-0.26}$, $(1.27 \pm 0.020) \times 10^3$, $(172.69 \pm 0.30) \times 10^3$
Quark masses: m_d, m_s, m_b (MeV/c^2)	$4.67^{+0.48}_{-0.17}$, $93.4^{+8.6}_{-3.4}$, $4.18^{+0.03}_{-0.02} \times 10^3$
Lepton masses: m_e, m_μ, m_τ (MeV/c^2)	0.51099895000(15), 105.6583755(23), 1776.86(12)
QCD phase angle, θ_{QCD} (rad)	$\lesssim 2 \times 10^{-10}$

The right-handed fermion fields form right-handed singlets. However, the left-handed fermion fields form $SU(2)_L$ doublets in three distinct generations with matching properties except for their masses, with each subsequent generation more massive than the previous. The fermions are classified as leptons and quarks, where both can interact via the EW interaction but only the quarks interact via the strong force.

For the leptons, the left-handed doublets contain a massive particle with a single EM charge *e.g.* an electron, e, and a massless, chargeless particle *e.g.* the electron neutrino, ν_e. Each generation of the lepton doublet has an associated 'flavour' that must be conserved in its interactions, named for the charged lepton in that generation.

For the quarks, the left-handed doublets contain an up-type quark (u, c, t) with an EM charge[2] of $+\frac{2}{3}|e|$, and a down-type quark (d, s, b) with an EM charge of $-\frac{1}{3}|e|$. Each quark also has an associated QCD charge known as color, labelled as red, green or blue. These quarks cannot exist in a free state and must always be bound within

[1] In minimal extensions to the SM that include massive neutrinos, there are 7 additional potential parameters: the three neutrinos masses; the three mixing angles and phase of the PMNS matrix.

[2] This has units of the magnitude of the charge of an electron, $|e|$.

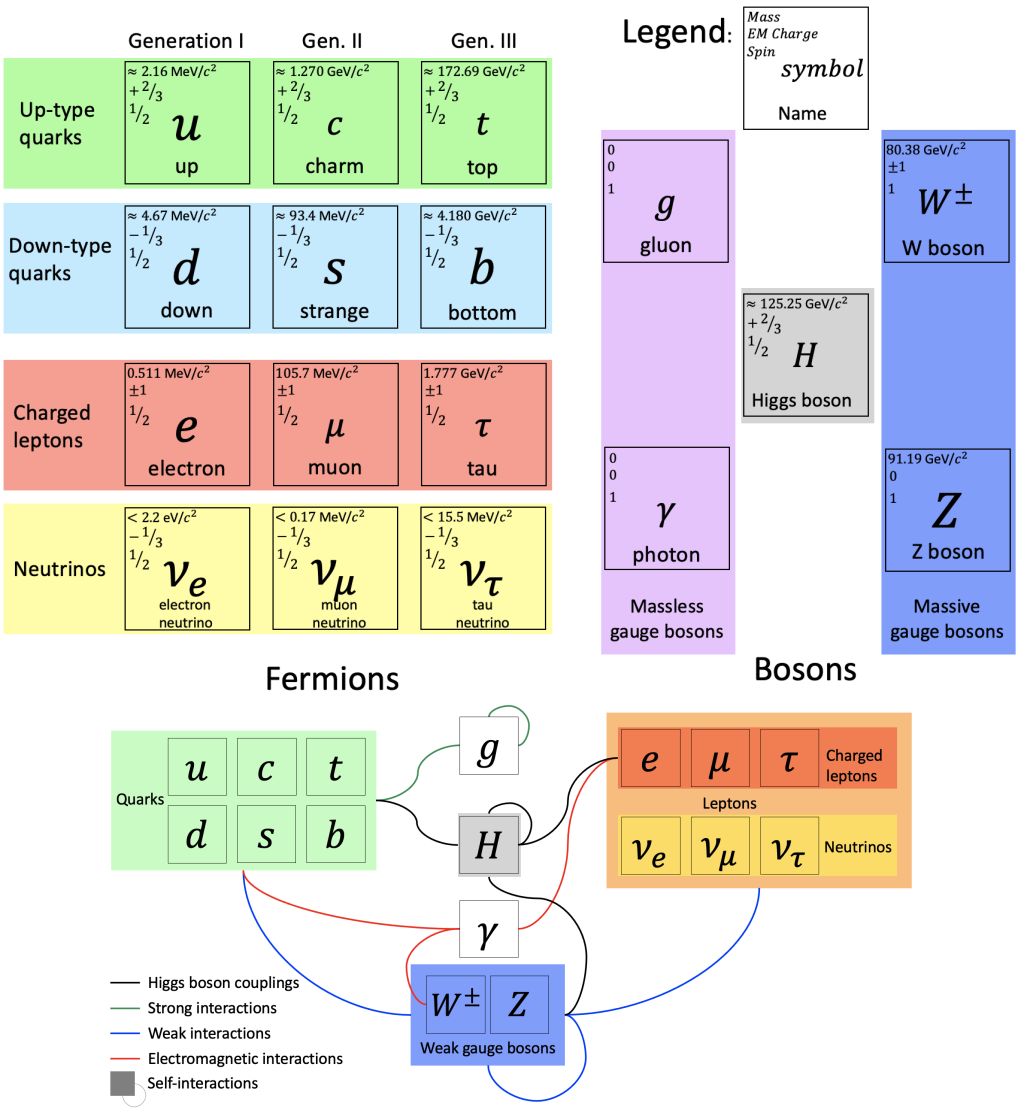


Figure 2.1: (Top) The fundamental particles of the SM, divided into leptons, quarks and bosons. (Bottom) The SM interactions between the particles. Picture adapted from Ref. [23].

multi-quark particles known as hadrons or as virtual particles within the hadrons in a so-called 'virtual sea'. These hadrons are most commonly divided into mesons, which contain a quark and an anti-quark; and baryons which contain a combination of three quarks or three antiquarks. This leads to different properties as the mesons will have an integer spin, whereas baryons can have a non-integer spin. There are other complex bound hadron states such as tetraquarks [24–26] and pentaquarks [27] that have four and five quarks respectively.

The process by which hadrons are created is known as hadronisation, and follows directly from the lack of free quarks due to color confinement, where the gluon self-interaction leads to a potential that is proportional to the quark separation distance.

Therefore, as a $q\overline{q}$ pair is separated the energy stored in the color field lines (often described as a color string or flux tube) increases until it is sufficiently high that new $q\overline{q}$ pairs are produced and the process repeats until the initial pair has fragmented into many, lower-energy pairs that can combine to form colorless hadrons. This is often referred to as the string model [21, 28, 29]. Another hadronisation model, known as the cluster model, uses the tendency of partons to be arranged in localised, color-singlet clusters (pre-confinement), which can then 'decay' or split to produce hadrons [30].

In the SM only the left-handed states can interact with the $W^{\pm}$ bosons to undergo a Vector-Axial Vector (V-A) charged current interaction. For quarks, these left-handed states do not coincide in their mass (strong) eigenstate and their flavour (weak) eigenstate. This was originally formulated by Cabibbo in 1963 [31], by introducing mixing between the mass and flavour eigenstates for down-type quarks,

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}, \tag{2.1.1}$$

where q' denotes the flavour eigenstate that undergoes the charged current interaction, q is the mass eigenstate, and $\theta_c \approx 13^{\circ}$ is the Cabibbo mixing angle [22]. This was later generalised to the CKM matrix to allow for mixing in all three generations of quarks, given by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.1.2}$$

where the CKM matrix, V_{CKM}, is parameterised by three mixing angles, $s_{ij} = \sin\theta_{ij}$ and $c_{ij} = \cos\theta_{ij}$, and a Charge-Parity (CP) violating phase, δ [22]. The CKM matrix is a 3×3 unitary matrix by construction, where each component gives the relative scaling of an interaction between an up- and a down-type quark. The relative strength of interactions varies dramatically, with preferred decays predominantly being interactions involving diagonal CKM matrix elements. By contrast, interactions with off-diagonal CKM matrix elements are suppressed. This can be highlighted using the Wolfenstein parameterisation to $\mathcal{O}(\lambda^4)$,

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & V_{ts} & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{2.1.3}$$

where $s_{12} = \lambda = \frac{|V_{us}|}{\sqrt{|V_{us}|^2+|V_{ud}|^2}}$, $s_{23} = A\lambda^2 = \lambda\left|\frac{V_{cb}}{V_{us}}\right|$, and $s_{13}e^{i\delta} = A\lambda^3(\rho + i\eta) = V_{ub}^*$. The values for the individual components, given in Equation 2.1.4, have been determined using a variety of methods with different decays summarised in Ref. [22]:

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{pmatrix}. \tag{2.1.4}$$

This has the overall effect of introducing a non-universal coupling between quarks and the weak boson, the universal weak coupling being modified by the CKM matrix element. This leads to a difference in decay rate, which can be seen through the form of the matrix element, $\mathcal{M}_{fi}$. The matrix element is related to the fundamental interactions that transform the initial particles into the final state particles and can be determined through the application of the 'Feynman rules' to a particular Feynman diagram [32]. For weak charged current interactions of quarks, $q_i \to q_j$, the coupling factor for each $q_i W q_j$ vertex is $g_W V_{q_i q_j}$, where g_W is the universal weak coupling strength. Hence, $\mathcal{M}_{fi} \propto g_W V_{q_i q_j}$.

The differential decay rate of an n-body decay in the centre-of-mass frame is given by,

$$d\Gamma(A \to B + ... + n) = \frac{1}{2M_A} \|\mathcal{M}_{fi}\|^2 \left(\prod_f^n \frac{d^3\vec{p_f}}{(2\pi)^3(2E_f)} \right) (2\pi)^4 \delta^{(4)} \left(p_A - \sum_f^n (p_f) \right), \tag{2.1.5}$$

where the product accounts for the available phase space and the delta function ensures four-momenta conservation [33]. Therefore, as the rate is proportional to the CKM matrix elements, certain decays that include off-diagonal CKM matrix elements will have a suppressed rate, known as CKM suppression.

Alongside CKM suppression, the rates of SM Flavour-Changing-Neutral-Current (FCNC) processes are further reduced, because they must occur through loop diagrams. The total matrix element, $\mathcal{M}_{fi}$, used within Equation 2.1.5 for a single-loop diagram in a $b \to s$ transition, has the approximate form $\mathcal{M}_{fi} \propto \sum_{i,j=u,c,t} V_{ib}^* V_{is} V_{jb} V_{js}^*$. The CKM matrix is a unitary matrix which implies that $\sum_i V_{ib}^* V_{is} = \delta_{bs} = 0$, therefore for a non-zero decay rate the factors contributed by different generations of quark within the loop do not cancel out. This is due to the fermion masses in the loops which give $\mathcal{O}(\frac{m_q^2}{m_W^2})$ corrections to the propagators. Therefore, lower mass quarks are more strongly suppressed in the loop than the heavier quarks, leading to top quark loops typically dominating. The suppression of FCNC transitions due to the mass of the quark in a quark loop and the unitarity of the CKM matrix is known as the Glashow-Iliopoulous-Maiani (GIM) mechanism [14, 34].

The lepton sector has an equivalent mixing matrix between the mass and flavour eigenstates of leptons, known as the PMNS matrix [35, 36]. In this case, the neutrinos mix analogously to the down-type quarks in the CKM matrix formalism. The PMNS matrix does not show the same hierarchy as the CKM matrix, with some off-diagonal terms being at a similar level as the on-diagonal ones. Another property of neutrinos that is not present in the quark sector is the ability to oscillate. This is because a propagating neutrino is a superposition of the possible flavour eigenstates, and so can convert from its initial flavour into either of the other two flavours. However, this is only possible if there is a mass difference between neutrino mass eigenstates, which then mix to form the flavour eigenstates, and so at least two neutrino mass eigenstates must be massive. This reveals one of the limitations of the SM, which in its original formulation assumes that neutrinos are massless, which is inconsistent with experimental evidence of neutrino oscillations, first reported in Ref. [37]. There have been models that seek to incorporate massive neutrinos within the SM, such as adding Dirac or Majorana neutrino mass terms to the SM lagrangian [38, 39]. This non-conservation of lepton flavour when neutrinos propagate over large distances suggests that lepton flavour may also not be conserved in their interactions. This is known as Lepton Flavour Violation (LFV).

However, within the SM, the weak couplings of charged leptons to a given gauge boson are identical regardless of their flavour. This is an accidental symmetry of the SM known as Lepton Flavour Universality (LFU) and it has been measured through many precise experiments [40–44] that do not show any BSM behaviour. This has recently been explored at LHCb through LFU ratios that examine the difference in branching fractions between two final states that only differ by the flavour of the lepton(s) [40, 41], such as

$$R_X = \frac{\mathcal{B}(H \to X\mu^+\mu^-)}{\mathcal{B}(H \to Xe^+e^-)} \cdot \frac{\mathcal{B}(H \to XJ/\psi(\to e^+e^-))}{\mathcal{B}(H \to XJ/\psi(\to \mu^+\mu^-))}, \tag{2.1.6}$$

where H denotes a b-hadron decay to a dilepton pair and another particle, X [40, 41]. This ratio also includes the resonant modes as these have already been precisely measured to be $\sim$1 and enables many systematic effects to be reduced, a similar procedure is described in Chapter 6. These R ratios typically have an approximate value of 1 within the SM. Despite some potential deviations from the SM, reported at the $\mathcal{O}(2-3)\sigma$ level, the most precise LFU measurement performed to date by LHCb provides values that are compatible with the SM [40, 41]. This demonstrates that charged leptons cannot oscillate between generations, and therefore the flavour of charged leptons is conserved within decays *i.e.* for massless neutrinos $L_\ell = N_\ell + N_{\nu_\ell} - (N_{\bar{\ell}} + N_{\bar{\nu}_\ell})$ where $\ell = e, \mu, \tau$, and $N_{\ell/\bar{\ell}}$ or $N_{\nu_\ell/\bar{\nu}_\ell}$ are the number of leptons or neutrinos present in an interaction. Two other conserved quantities are the overall number of leptons, L, and baryons, B, given as $L = N_\ell - N_{\bar{\ell}}$ and $B = \frac{1}{3}(N_q - N_{\bar{q}}) = N_{\text{baryons}} - N_{\text{anti-baryons}}$ respectively, where $N_{q/\bar{q}}$ are the number of

quarks/anti-quarks present. These arise from symmetries within the SM, in particular, the simultaneous rotation of the phase of the fermion fields.

Recent studies of a variety of different methods at LHCb have shown consistency with SM predictions, typically at the level of $\mathcal{O}(1-2)\sigma$. These highlight potential deviations from the SM arising in different measurements of b-hadron decays such as: differential branching fractions [45, 46]; and angular measurements such as the P_5' variable [47, 48]. These are collectively referred to as the 'B anomalies' within LHCb, and highlight a potential route through which new physics may be observed but no conclusive evidence has yet been obtained. Additionally, before the recent LFU measurements [40, 41] showed the R_K and R_{K^*} values to be consistent with the SM, LFU measurements were another component of the 'B anomalies'. Further data will ultimately clarify the consistency with the SM of these other measurements that comprise the 'B anomalies'.

2.1.1 Known Limitations of the Standard Model

To date, there are several known limitations of the SM, which include both observed phenomena that are not predicted in the SM, and theoretical issues with the SM's framework. A brief, non-exhaustive list of these limitations are:

- **Gravity:** gravitational interactions are not incorporated within the SM.
- **Dark Matter:** cosmological observations have identified a large portion of matter in the universe that does not interact with the matter predicted by the SM [7, 8].
- **Dark Energy:** observations of the universe's expansion imply an increasing expansion rate rather than a steady-state or contracting universe, and suggests a form of abundant energy that, like dark matter, does not interact with the SM [9, 10].
- **Neutrino Masses:** as described in the previous section, neutrinos are massless in the SM, but experimental evidence of neutrino oscillations require at least two of the three neutrino mass eigenstates to be massive [34, 37].
- **Naturalness/The Gauge Hierarchy Problem:** attempts to unify the strong and the EW interactions predict new physics at high energy scales ($\mathcal{O}(10^{11}\,\text{TeV})$), however EW symmetry breaking leads to a Higgs boson of mass $\sim$125 GeV/c^2. There is a significant difference between the two scales and it requires intensive fine-tuning of the Higgs field in order to satisfy both [34, 49, 50].
- **The Flavour Problem:** there is a pattern of fermion masses but the exact origin of these masses and the structure of the mixing matrices is not known [34].

- **The Strong CP Problem:** the combined symmetry of charge conjugation and parity is known to be broken in the weak interaction. However, despite not being forbidden, no CPV has been observed with the strong interaction. Therefore, the CPV phase in QCD, θ_{QCD}, is considered to be fine-tuned to be small but non-zero [34, 51].

These limitations inform the development of additional theoretical models that attempt to extend the SM and account for them. Additionally, the SM is rigorously measured to ensure that it can fully describe particle phenomena.

2.2 Rare Decays and Charged Lepton Flavour Violation

One way of precisely testing the SM is through FCNCs. These are interactions in which a quark changes flavour via a neutral current mediator. As described in Section 2.1, FCNCs such as $b \to s\ell^-\ell^+$ are heavily suppressed in the SM due to GIM suppression and, in this case, CKM suppression as well. The only interaction that can change the flavour of particles within the SM is the weak interaction with charged W bosons, and so no SM tree-level $b \to s\ell^-\ell^+$ decays are possible. These must occur at loop level with at least two charged current interactions, leading to the high CKM suppression. For example, in $b \to s\ell^-\ell^+$ decays such as $\Lambda_b^0 \to \Lambda\ell^+\ell^-$ there are two single-loop Feynman diagrams, as seen in Figure 2.2: a box diagram and a penguin diagram. Overall, these factors lead to branching fractions of $\mathcal{O}(10^{-6})$, such as $\mathcal{B}(\Lambda_b^0 \to \Lambda\mu^-\mu^+) = (1.08 \pm 0.28) \times 10^{-6}$ [22], which is small but not negligible. This branching fraction can then be measured and any significant deviation from this predicted value would be direct evidence for BSM physics. This, typically, is predicted to be due to BSM models that provide additional interactions or tree-level diagrams that enhance or suppress the branching fraction. One example is the introduction of a new particle, such as a Z' [52] or a leptoquark [53, 54], that can directly couple quarks and leptons to provide a tree-level diagram rather than a suppressed loop level decay, thus enhancing the $b \to s\ell^-\ell^+$ branching fraction.

This search for new physics can be enhanced by looking at forbidden decays. Due to LFU there is no Charged Lepton Flavour Violation (cLFV), and to date no evidence of cLFV exists. The only practical way to obtain a SM-like cLFV decay such as $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ is via a neutrino oscillation, where that neutrino then produces another different flavour charged lepton as shown in Figure 2.3. This requires a minimal extension to the SM that includes massive neutrinos in order to describe the neutrino oscillation. This neutrino oscillation would introduce an additional factor of approximately $\frac{m_\nu^4}{m_W^4} \approx 10^{-44}$

for $m_\nu = 1\,\mathrm{eV}/c^2$. Therefore, to estimate the branching fraction $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$, the following is used: $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm) \approx \mathcal{B}(\Lambda_b^0 \to \Lambda \mu^- \mu^+) \cdot \frac{m_\nu^4}{m_W^4} \approx 10^{-50}$. This is such a low branching fraction that it is effectively forbidden within the minimal SM extension and entirely forbidden in the SM where neutrino oscillations cannot occur. However, if this channel was observed this would be clear and direct evidence of BSM physics, such as new tree-level decays or BSM interactions, that enhanced the value of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$.

One way of characterising the potential BSM contributions to measurements of the properties of rare decays, such as $\Lambda_b^0 \to \Lambda \mu^- \mu^+$ and $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$, is through the use of an Effective Field Theory (EFT).

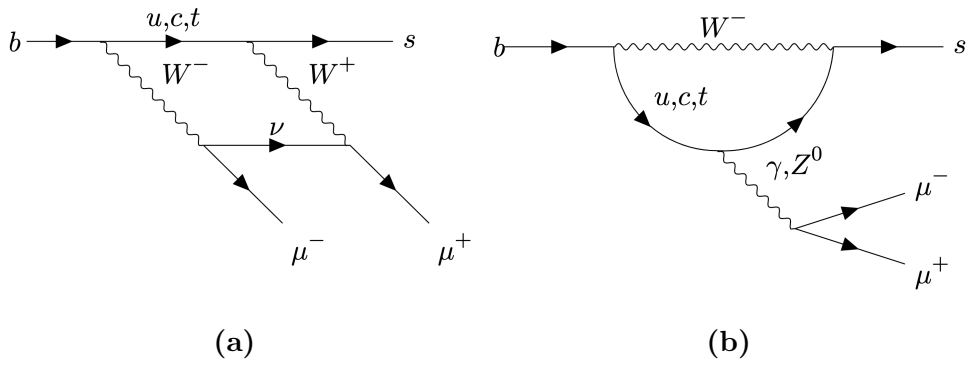


Figure 2.2: Two possible Feynman diagrams of the $b \to s\ell^-\ell^+$ decay.

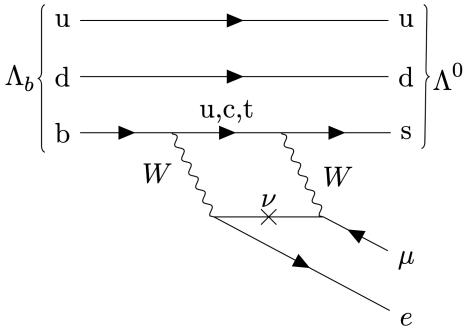


Figure 2.3: A possible SM-like Feynman diagram of the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ decay.

2.2.1 Effective Field Theories and Wilson Coefficients

For meson and baryon decays there is a hierarchy of scales involved, $m_W \gg m_b > \Lambda_{QCD}$, which allows the interaction to be described using an EFT approach. The general principle of an EFT is to consider a QFT that approximates the behaviour at given energy scale. The most well-known example of an EFT is the Fermi model of the weak decay, which approximates the W boson propagator as a four-particle point-like interaction that can be

used to derive an effective model of nuclear beta decay. Generally, if a QFT is considered at a high-energy scale, M, with interest in the low-energy particles at E, where $E \ll M$, a cutoff energy, Λ_{cut}, can be defined. This is such that the low and high-frequency modes of the field, ϕ, can be decoupled into $\phi = \phi_L + \phi_H$ where $\phi_{L,H}$ describes the low or high energy physics, and the behaviour above Λ_{cut} is not known. The high-frequency fields can be integrated out following a path integral approach and using the 'Wilsonian Effective Action', $S_{\Lambda_{cut}}(\phi_L)$. This effective action is non-local but can be expanded in terms of local operators in a process called Operator-Product Expansion (OPE). This is equivalent to integrating $S_{\Lambda_{cut}}(\phi_L)$ over an effective Lagrangian where OPE is performed,

$$\mathcal{L}^{eff}_{\Lambda_{cut}(x)} = \sum_i g_i \mathcal{O}_i(\phi_L(x)) = \sum_i M^{-\gamma_i} \mathcal{C}_i \mathcal{O}_i(\phi_L(x)), \tag{2.2.1}$$

where $g_i = M^{-\gamma_i}\mathcal{C}_i$ is a coupling constant with mass units of $M^{-\gamma_i}$, and γ_i is an integer that describes the mass dimension of g_i [55]. The terms $\mathcal{O}_i$ are local operators that describe the 'long distance'/low energy contribution of an interaction. Finally, $\mathcal{C}_i$ are dimensionless coupling constants referred to as a Wilson coefficient (WC), which encodes the 'short distance'/high energy contributions. Within the SM, the WCs should typically be close to unity due to "naturalness", where a single fundamental scale, M, is assumed.

Using this formalism an effective Hamiltonian for $\Lambda_b^0 \to \Lambda \ell_1^- \ell_2^+$ transitions can be described by [56],

$$\mathcal{H}_{eff} = \frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_i \left(\mathcal{C}_i^{\ell_1 \ell_2} \mathcal{O}_i^{\ell_1 \ell_2} + \mathcal{C}_i^{\prime \ell_1 \ell_2} \mathcal{O}_i^{\prime \ell_1 \ell_2} \right), \tag{2.2.2}$$

where the top quark is assumed to dominate the quark loop. The set of gauge-invariant operators that are allowed from symmetry and quantum number conservation have a mass dimension of 6 and are given by [57]:

$$\mathcal{O}_1^{(p)} = (\bar{s}_L \gamma_\mu T^\alpha p_L)(\bar{p}_L \gamma^\mu T^\alpha b_L), \tag{2.2.3}$$

$$\mathcal{O}_2^{(p)} = (\bar{s}_L \gamma_\mu T^\alpha p_L)(\bar{p}_L \gamma^\mu b_L), \tag{2.2.4}$$

$$\mathcal{O}_3 = (\bar{s}_L \gamma_\mu p_L) \sum_q (\bar{q}_L \gamma^\mu q_L), \tag{2.2.5}$$

$$\mathcal{O}_4 = (\bar{s}_L \gamma_\mu T^\alpha p_L) \sum_q (\bar{q}_L \gamma^\mu T^\alpha q_L), \tag{2.2.6}$$

$$\mathcal{O}_5 = (\bar{s}_L \gamma_\mu \gamma_\nu \gamma_\rho p_L) \sum_q (\bar{q}_L \gamma^\mu \gamma^\nu \gamma^\rho q_L), \tag{2.2.7}$$

$$\mathcal{O}_6 = (\bar{s}_L \gamma_\mu \gamma_\nu \gamma_\rho T^\alpha p_L) \sum_q (\bar{q}_L \gamma^\mu \gamma^\nu \gamma^\rho T^\alpha q_L), \tag{2.2.8}$$

$$\mathcal{O}_7 = \frac{e}{(4\pi)^2} m_b \left(\overline{s}_R \sigma^{\mu\nu} b_L\right) F_{\mu\nu}, \tag{2.2.9}$$

$$\mathcal{O}_8 = \frac{g_s}{(4\pi)^2} m_b \left(\overline{s}_R \sigma^{\mu\nu} T^\alpha b_L\right) G_{\mu\nu}, \tag{2.2.10}$$

$$\mathcal{O}_9 = \frac{e^2}{(4\pi)^2} m_b \left(\overline{s}_L \gamma_\mu b_L\right) \left(\bar{l}\gamma_\mu l\right), \tag{2.2.11}$$

$$\mathcal{O}_{10} = \frac{e^2}{(4\pi)^2} m_b \left(\overline{s}_L \gamma_\mu b_L\right) \left(\bar{l}\gamma_\mu \gamma_5 l\right), \tag{2.2.12}$$

$$\mathcal{O}_S = \frac{e^2}{(4\pi)^2} \left(\overline{s}_L b_R\right) \left(\bar{l}l\right), \tag{2.2.13}$$

$$\mathcal{O}_P = \frac{e^2}{(4\pi)^2} \left(\overline{s}_L b_R\right) \left(\bar{l}\gamma_S l\right), \tag{2.2.14}$$

where $p = (u, c)$, and T^α are the $SU(3)$ generators (Gell-mann matrices). These operators can be summarised as: $\mathcal{O}_{1,2}$ are scalar, charged-current operators; $\mathcal{O}_{3-6}$ are penguin operators; $\mathcal{O}_7$ is the photon dipole operator; $\mathcal{O}_8$ is the gluon dipole operator [58]; $\mathcal{O}_9$ is a vector current semileptonic operator; and $\mathcal{O}_{10}$ is an axial-vector current semileptonic operator. The associated SM WCs have theoretical values of $\mathcal{C}_{1c,2c}(m_b) \sim \mathcal{O}(1)$, $\mathcal{C}_9(m_b) \approx -\mathcal{C}_{10} \sim -4.2$, $\mathcal{C}_7(m_b) \sim 0.3$, and all others are $\sim\mathcal{O}(10^{-2})$ [57].

These are all SM operators, except for $\mathcal{O}_{S,P}$ which are flavour-changing scalar and pseudoscalar operators not present in the SM. Another operator that is not present in the SM is $\mathcal{O}'_7 = \frac{e}{(4\pi)^2} m_b \left(\overline{s}_L \sigma^{\mu\nu} b_R\right) F_{\mu\nu}$. This is because it is related to right-handed weak couplings which are absent from the SM. The equivalent 'primed'/right-handed operators can be found for each of the SM operators, so a non-zero WC for these operators would then be evidence of BSM physics.

A source of new physics may arise from new particles in particle loops, which would provide an additional contribution to certain WCs, therefore a modified WC is measured of the form: $\mathcal{C}_9 = \mathcal{C}_9^{SM} + \mathcal{C}_9^{BSM}$ which is model agnostic. This allows for the construction of theoretical BSM models that predict the branching fraction of some decays in that model. By measuring the branching fraction of decays and comparing it to the SM it may also be possible to narrow down the particular BSM model based on the modification to the WCs.

2.2.2 A Scalar Leptoquark Model of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$

One example of a BSM model of the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ decay includes the introduction of a scalar leptoquark. There are two scalar leptoquarks that potentially accommodate the 'B anomalies', which have a hypercharge of $Y = \frac{1}{3}$: one is a weak singlet, S_1, and the other is a weak triplet, S_3. The main principle of this model is the application of the Froggatt-

Nielsen mechanism [59], which also attempts to explain the quark mass hierarchy and hence the flavour problem. This model is discussed more fully in Ref. [60]. The resultant $S_1 + S_3$ model can then be introduced in terms of WCs to observe the overall effect on the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ hamiltonian, which is given by Equation 2.2.2 with $\ell_{1,2} = e, \mu$ and the contributing operators are $\mathcal{O}_9$, $\mathcal{O}_{10}$, $\mathcal{O}_S$ and $\mathcal{O}_P$. From this it can be observed that S_1 does not contribute to $b \to s\ell^-\ell^+$ decays, and the associated value for the BSM Wilson coefficients for S_3 are $\mathcal{C}_9^{e^-\mu^+} = -\mathcal{C}_{10}^{e^-\mu^+} \propto 10^{-3}$ and a lower value for $\mathcal{C}_9^{e^+\mu^-} = -\mathcal{C}_{10}^{e^+\mu^-}$ implying a possible charge dependence [56].

A model independent limit of $\mathcal{B}\left(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}\right)$ has also been performed in Ref. [56], that does not require a specific value for the BSM contributions to $\mathcal{C}_9$, $\mathcal{C}_{10}$, $\mathcal{C}_S$ or $\mathcal{C}_P$. Instead, the limit is evaluated with $\mathcal{C}_{9,10} \neq 0$ and $\mathcal{C}_{9,10} = 0$ giving $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}) < 1.1 \times 10^{-8}$; $\mathcal{C}_{9,10} = 0$ and $\mathcal{C}_{9,10} \neq 0$ giving $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}) < 0.06 \times 10^{-8}$; and $\mathcal{C}_9 = -\mathcal{C}_{10}$ and $\mathcal{C}_S = -\mathcal{C}_P$ giving $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}) < 1.1 \times 10^{-8}$. This is within the sensitivity of LHCb to measure.

Part II of this thesis shows the efforts to perform a search for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ decay. This particular channel was selected as any observation of this forbidden channel would be direct evidence of BSM physics and the first observation of cLFV. Additionally, the value of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ with potential BSM contributions has been determined to be within LHCb sensitivity and therefore there are definite prospects for either observation or additional constraints upon these contributions. This analysis of the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ decay would also be the first cLFV search performed within LHCb that uses a baryonic decay instead of a mesonic one [61–67]. This is important to note, due to the different helicity structure of the decay as the Λ_b^0 has a non-integer spin, and the sensitivity to different theoretical form factors. Thus, this provides an independent search of cLFV in LHCb while being complementary to those prior searches.

2.3 Long-Lived Particles

Another way of testing the SM is through direct searches for BSM particles instead of precisely measuring standard processes with SM particles. This is a natural way to attempt to explain limitations to the SM where theoretical extensions introduce BSM particles that have not been observed. The most natural example is that of dark matter, which has been observed through astronomical measurements, such as the flattening of galactic rotation curves implying the presence of additional unseen mass [8]. Also, the energy densities due to dark matter and baryonic matter have been surveyed by collaborations such as Planck, concluding that dark matter dominates over regular matter, $\Omega_{DM}/\Omega_B = 5.357 \pm 0.001$ [68].

Therefore, some form of matter that is not described by the SM and which appears to be 'sterile' to SM interactions, only interacting via gravity, appears to exist. As such it can be natural to imagine a new 'dark sector' of particles that behaves like the existing SM but that does not couple strongly to it. One potential example of a dark matter candidate, is the Weakly-Interacting Massive Particle (WIMP). This annihilates with itself to produce SM particles and thus it has a very weak coupling to the SM and a long lifetime [69].

Similar theoretical models, known as 'Hidden sector' models which analogously have a set of particles that do not couple strongly to the SM, arise generically in many BSM theories [70–75]. These hidden sector models often have one or more 'portal particles', which are particles that couple both to the hidden sector and the SM, much like the WIMP does via annihilation. The weak coupling of this portal particle leads to it having a long lifetime, and these can often be referred to as Long-lived particles. LLPs exist within the SM such as the neutron or K_{L}^{0}, but typically an LLP refers to exotic particles not present within the SM. These are also not limited to arising from hidden sector models, but also cases such as heavy BSM particles that are too massive to be produced directly by current particle experiments. In this example, the couplings of these new BSM particles may lead to one being an LLP candidate.

These LLPs then can form the basis of direct searches for BSM physics as they can have distinctive experimental signatures due to their extended lifetimes and lack of interaction with SM particles. For collider experiments this may be represented by a displaced vertex with no prior SM tracks to which it can be linked. In this case the LLP is produced in the pp collision and leaves no observable trace in the detector until it later decays to SM particles that can be detected. However, the ubiquity of these LLPs and the large variety of models from which they can arise poses an experimental challenge. In each individual BSM model the LLP decay width, and hence lifetime, can have strong power-law dependencies on unknown ratios of various scales. Thus LLPs may span a very large range of phenomenological parameter space. To perform searches, experiments are necessarily restricted to specific intervals of these parameter spaces due to their finite size. For example, the Large Hadron Collider (LHC) may have sensitivity to an LLP mass of $\mathcal{O}(1)\mathrm{MeV}/c^2 - \mathcal{O}(1)\mathrm{TeV}/c^2$ and lifetimes of $\lesssim 0.1\,\mathrm{s}$.

Generally, detector sensitivity is dependent on: (1) the available centre of mass energy; (2) the fiducial volume of the detector; (3) the distance from the interaction point; (4) limitations of triggering and reconstruction systems; and (5) irreducible backgrounds [76]. The first determines the LLP masses that can be probed. Both (2) and (3) determine the lifetimes that can be observed: a detector further from the Interaction Point (IP) would require an LLP to travel further, and therefore have a longer lifetime, but short enough that the decay is contained in the fiducial volume. This constrains the lower

and upper limits of the lifetimes examined. The last two, (4) and (5), are experimental difficulties which are hard to overcome at experiments that were not specifically designed for LLP searches. For example, the Compact Muon Solenoid (CMS) and A Toroidal Large Hadron Collider ApparatuS (ATLAS) experiments can perform LLP searches but are limited by the high QCD background and trigger systems that are not optimised specifically for the detection of LLPs. An additional complication to consider is that the topology of an LLP signal is unknown a priori, though the simplest scenario would be a 2-body decay to charged products. More complex scenarios such as: decays to photons present in axion models; or high multiplicity events, present in dark-QCD models, are possible and would have different identifiable signals and backgrounds. Thus, no single experiment is able to provide full coverage of the parameter space, instead, many experiments with overlapping coverage must be used [76–82].

Part I of this thesis describes a dedicated LLP experiment, CODEX-b, and the work performed to test the potential tracking ability of a demonstrator unit and quality assurance tests of that demonstrator's readout hardware.

Chapter 3

The LHCb Detector

3.1 The Large Hadron Collider

The LHC is a 27 km circumference hadron collider located on the border of France and Switzerland at the CERN laboratory. The LHC collides hadrons at a high centre of mass energy, $\sqrt{s}$, with a set of detectors placed at different points of the accelerator complex with a wide variety of physics goals. The LHC collides proton or ion beams, with this thesis analysing pp collision data.

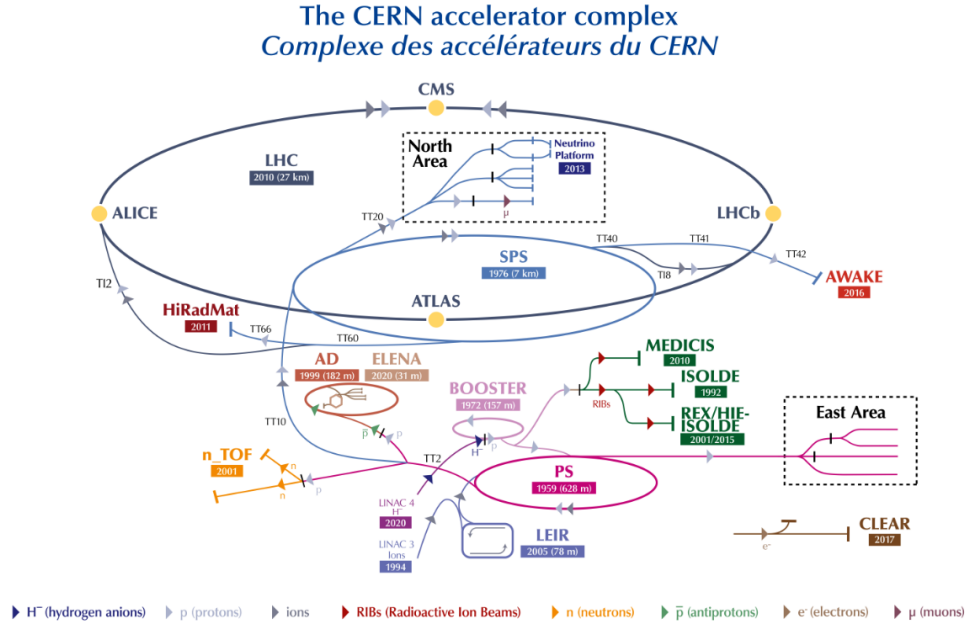


Figure 3.1: The CERN accelerator complex, showing the process by which the different particles are accelerated and transported around the CERN site, in particular, showing the injection of protons to the LHC that hosts experiments including LHCb [83].

Figure 3.1 shows the full CERN accelerator complex. Firstly an electric field strips electrons from hydrogen gas to obtain 'bunches' of 1×10^{11} protons. These bunches are accelerated by[1] LINAC2 up to an energy of 50 MeV into the Proton Synchrotron Booster (PSB), a small synchrotron with four rings, that accelerates the protons up to 1.4 GeV before reaching the Proton Synchrotron (PS). The PSB fills the PS, a 628 m circumference synchrotron, with a total of six bunches in a two-step process. In each step, a bunch from three of the PSB rings is transferred to the PS. Within the PS, Radio-frequency (RF) systems split these six original bunches into 72 bunches, each of width ∼4 ns with a spacing between bunches of ∼25 ns. During this fill the protons have been boosted further to 25 GeV before moving on to the Super Proton Synchrotron (SPS) (7 km circumference). This is the last step before injection into the main LHC ring and by this point the protons have an energy of 450 GeV. The 72 proton bunches define a batch and there is a gap equivalent to 20 bunches between adjacent batches, corresponding to ∼555 ns. The SPS can hold a maximum of four such batches when it is filled by the PS. These proton batches are then sent to the main LHC to be accelerated to the nominal collision energy. They are sent in a 'bunch train' with a set of three or four filled proton batches from the SPS in a defined pattern that covers the full LHC ring: 334-334-334-333. Between each batch in each group separated by a hyphen in this pattern (*e.g.* between 3 and 4 in 334) there is a spacing of 38 'empty' bunches (∼1077 ns), each hyphen represents a gap of 39 empty bunches (∼1106 ns), and at the end of the train is a gap of 119 empty bunches (∼3426 ns). The full 'bunch train' is shown in Figure 3.2, with a total of 2808 proton bunches circulating the LHC ring at any time [84]. This is the nominal procedure outlined in Ref. [84] and shows the general methodology, although the exact scheme may have since changed.

The operation of the LHC is divided into periods known as 'Runs', where Run 1 took place between 2010 and 2012 and Run 2 took place between 2015 and 2018. For 2011-2012 the proton bunches had spacings of 50 ns, though 2011 also had some instances of 75 ns, and throughout Run 2 the bunch spacing was 25 ns [85, 86]. The 2010 data were used for commissioning and will not be discussed further in this thesis. The data and simulated data samples considered for the remainder of this thesis that are labelled as Run 1 explicitly refer to the 2011-2012 period from this point forward.

The proton bunches are injected into the LHC as two separate, counter-rotating beams, each in its own beam pipe. These beams each have an energy of 4 TeV in Run 1[2] and 6.5 TeV in Run 2 giving a centre of mass energy, $\sqrt{s}$, of 8 TeV and 13 TeV respectively [85, 86]. There are a set of points available around the LHC ring where these beams can be made to collide at IPs [84]. These are instrumented by large-scale experiments:

[1] As of 2020, Linear Accelerator 4 (LINAC4) replaced Linear Accelerator 2 (LINAC2).
[2] In 2011 the beam energy was 3.5 TeV, giving $\sqrt{s} = 7$ TeV. This was increased in 2012.

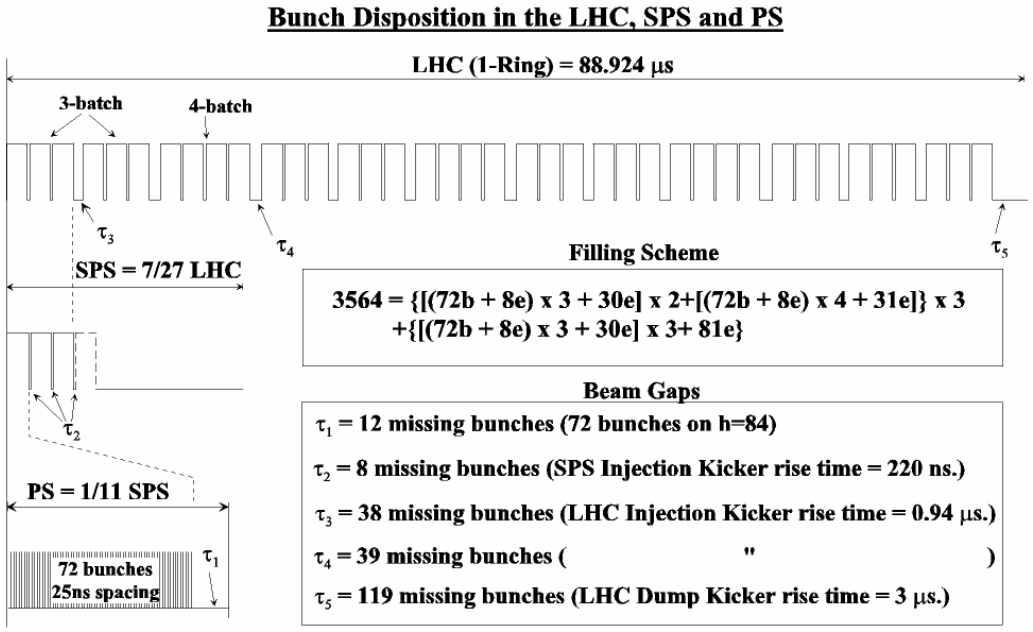


Figure 3.2: The nominal pattern for proton bunches from the PS, SPS and LHC rings during the running of the LHC machine, including the gaps between beams that allow for delays from the LHC machine procedure. Note that the SPS is filled with 3 or 4 PS batches depending on the point in the LHC bunch train [84].

A Large Ion Collider Experiment (ALICE), ATLAS, CMS and LHCb. These detectors have different physics goals which can overlap with one another. For example, ATLAS and CMS are both general purpose detectors that have a similar detector structure and one of their main goals was to observe the Higgs boson, achieved in 2012 [87, 88].

Figure 3.3 shows the LHCb detector [89, 90], which is a single-arm forward spectrometer that covers the pseudorapidity[3] range $2 < \eta < 5$, rather than the full angular coverage of the other three main LHC experiments. This detector design originates from the physics motivation of LHCb, namely to search for evidence of BSM physics in CPV and rare decays of beauty and charm hadrons [89]. The LHC produces a large amount of $b\bar{b}$ pairs due to their large production cross-section in pp collisions and the high luminosity achieved. Gluon-gluon fusion where the momenta of constituent partons are strongly asymmetric in the laboratory frame[4] is the dominant production mechanism, leading to the resultant $b\bar{b}$ pair being highly boosted along the direction of the higher momentum gluon. Hence, the subsequent two b-hadrons are both produced in either the forward or backward direction [92]. Figure 3.4 shows the $b\bar{b}$ production angle for $\sqrt{s} = 14\,\text{TeV}$, with dominant peaks in the forward and backwards regions. Thus, LHCb adopts a narrow η range $(2 < \eta < 5)$ that corresponds to this forward region and does not instrument the

[3]Pseudorapidity is defined as $\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right]$, where θ is the polar angle.
[4]It is statistically less probable for the gluons to have similar and opposite momenta [91].

backward direction. This corresponds to $\sim 24\%$ of the $b\bar{b}$ pairs [93] produced at its IP, IP8. However, the number of recorded events in this η interval is more than sufficient for LHCb's physics programme, with the cost to develop, install and run the detector reduced by the 'single-arm' spectrometer design.

Figure 3.3 also shows the coordinate system used at LHCb, where z is along the direction of the beam pipe the positive direction going from the collision point into the LHCb detector; x is the horizontal axis transverse to the beam pipe in a right-handed system where the positive direction points to the centre of the LHC ring, and y is the vertical axis. The associated polar coordinates use the same definition for z, with the polar angle, θ, being relative to the z axis and the azimuthal angle, ϕ, relative to the x axis in a right-handed system.

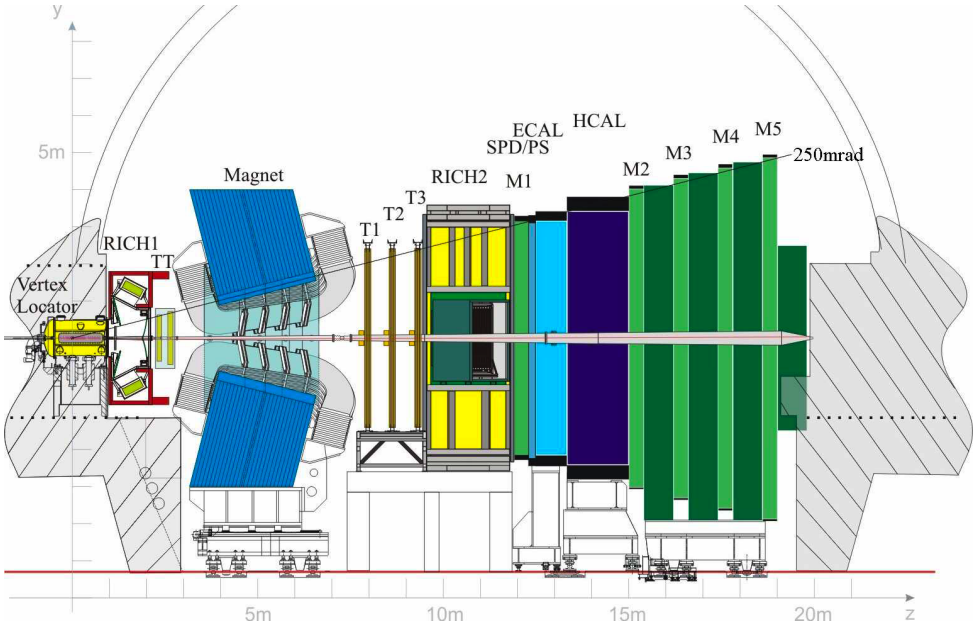


Figure 3.3: The LHCb detector and its sub-detectors [94].

The LHCb detector (Figure 3.3) includes a high-precision tracking system consisting of a vertex detector close to the IP, instrumented with a silicon-strip detector, and several large area detectors downstream from the IP, which are described in Section 3.2. There is a dipole magnet with a bending power of about 4 T m that allows the momentum of charged particles to be measured using the tracking systems. To identify different types of particles LHCb has two Ring Imaging CHerenkov (RICH) detectors, RICH1 and RICH2, (see Section 3.3) as well as an Electromagnetic CALorimeter (ECAL) and a Hadronic CALorimeter (HCAL) (see Section 3.4). The calorimeters also allow for the energies of the detected particles to be determined. Finally, there are several stations with alternating layers of iron and Multi-wire Proportional Chambers (MWPCs) that are used to identify muons (see Section 3.2), as these particles interact minimally with the preceding aspects

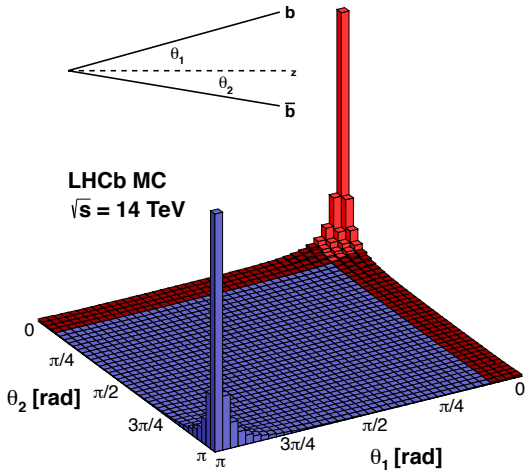


Figure 3.4: Simulated production angles for a $b\bar{b}$ pair, showing the strong preference for the forward and backwards directions. The LHCb detector acceptance is overlaid in red.

of the detector. Alongside the hardware of LHCb, there is also a variety of software that is utilised in the running of LHCb, triggering (see Section 3.6), and offline data analyses including data processing applications (see Section 3.7). The software trigger works alongside the hardware trigger to reduce the rate of obtained data, ∼40 MHz, to a level that can be processed and eventually stored, ∼5 kHz in Run 1 and ∼12.5 kHz in Run 2.

The intensity of the colliding beams is characterised by the instantaneous luminosity, $\mathcal{L}$. For a collider like the LHC that collides two beams with roughly Gaussian beam profiles this can be described by [95],

$$\mathcal{L} = \frac{N_1 N_2 f_{rev} n_b}{2\pi\sqrt{\sigma_{1x}^2 + \sigma_{2x}^2}\sqrt{\sigma_{1y}^2 + \sigma_{2y}^2}} \cdot S \cdot W \cdot e^{\frac{B^2}{A}}, \tag{3.1.1}$$

where,

$$S = \frac{1}{\sqrt{1 + \left(\frac{\sigma_x}{\sigma_s}\tan\frac{\phi}{2}\right)^2}} \frac{1}{\sqrt{1 + \left(\frac{\sigma_s}{\sigma_x}\tan\frac{\phi}{2}\right)^2}}, \tag{3.1.2}$$

$$A = \frac{\sin^2\frac{\phi}{2}}{\sigma_x^2} + \frac{\cos^2\frac{\phi}{2}}{\sigma_s^2}, \tag{3.1.3}$$

$$W = e^{-\frac{(d_2-d_1)^2}{4\sigma_x^2}}, \tag{3.1.4}$$

$$B = \frac{(d_2 - d_1)\sin\frac{\phi}{2}}{2\sigma_x^2}, \tag{3.1.5}$$

and where N_i is the number of particles per bunch for beam $i = [1, 2]$; f_{rev} is the bunch revolution frequency; $\sigma_{i(x,y)}$ is the beam width of beam i in the x/y direction; σ_s is the bunch length; S is the luminosity reduction factor that accounts for the beams colliding at

a crossing angle given by ϕ; W is a correction factor for if there is a transverse offset given by d_i such that the beams do not collide head on, and $e^{\frac{B^2}{A}}$ is a correction for if there is both a non-zero crossing angle and a transverse offset. Additional considerations regarding the luminosity at colliders from a variety of factors are described greater detail in Ref. [96]. The LHC was designed to operate at a nominal value of $\mathcal{L} = 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ [84]. This leads to multiple pp collisions in the same bunch crossing (referred to as 'pile-up', μ_{vis}) which would make the decay vertices of particular b-hadrons at LHCb hard to distinguish from those of other decays. Therefore, LHCb initially focused the beam and reduced the $\mathcal{L}$ by a factor $\sim$100 down to $\mathcal{L} = \mathcal{O}(10^{32})\,\text{cm}^{-2}\,\text{s}^{-1}$, reducing the pile-up to a single interaction per bunch crossing in most instances [92]. For 2011, data was taken at an instantaneous luminosity of $3.5 \times 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$. This was increased to $4 \times 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$ in 2012 [90], and was maintained at that level throughout Run 2.

Integrated luminosity, $\mathcal{L}_{int}$, is defined as the instantaneous luminosity accumulated over a period of time. From this the number of events observed, N, for a process of a given production cross-section, σ_{prod}, and branching fraction, $\mathcal{B}$, can be derived,

$$N = \mathcal{L}_{int} \cdot \sigma_{prod} \cdot \mathcal{B}. \tag{3.1.6}$$

This enables the calculation of yields from a measured branching fraction, a concept utilised in Part II, particularly in Chapters 6 and 8.

For LHCb the recorded integrated luminosities over the years of Run 1 and Run 2 for the pp collisions used for the data in Part II are given in Table 3.1.

Table 3.1: The pp run conditions at LHCb for each year of Run 1 and Run 2 used for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis [97].

Run Condition	**Year**					
	2011	2012	2015	2016	2017	2018
$\sqrt{s}$ (TeV)	7	8	13			
Median $\mathcal{L}$ ($10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$)	3.1	3.9	2.3	3.2	3.3	4.3
$\mathcal{L}_{int}$ (fb^{-1})	0.99	1.99	0.28	1.62	1.46	2.00
Median $N_{bunches}$	1296	1262	1278	1767	1749	2332
Pile-up, μ_{vis}	1.4	1.7	1.1			
Average $f_{collision}$ (MHz)	8.5	10.9	9.1	12.9	12.9	16.2
$f_{rev,LHC}$ (kHz)	11.2					

3.2 Tracking Detectors

The tracking system at LHCb consists of the VErtex LOcator (VELO) located around the IP, four planar tracking stations known as Tracker Turicensis (TT) and T1-T3 where the TT is located upstream of the magnet and the other stations downstream. These form the majority of the tracking detectors at LHCb, but there are also a set of five large area planar tracking stations much further downstream from the IP. These comprise the muon detectors, M1-M5, that mostly lie after all other detector components except for M1 which lies between RICH2 and the ECAL.

The tracking system provides a measurement of the momentum of charged particles by observing the deflection of the particle track in a magnetic field, generated by the dipole magnet. This deflection can be directly related to the momentum (in $\mathrm{GeV}\,c^{-1}$) of the particle transverse to the magnetic field by $p = 0.3 \cdot qBR$, where q is the particle's charge, B the magnetic field strength and R the radius of curvature. The LHCb tracking system provides a high quality measurement of the momenta with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at $200\,\mathrm{GeV}/c$. Additionally, the minimum distance of a track to a primary pp collision vertex (PV), the Impact Parameter (IP), can also be measured with a resolution of $(15 + 29/p_{\mathrm{T}})\,\mu\mathrm{m}$, where p_{T} is the momentum transverse to the LHC beam-line, *i.e.* perpendicular to z [89].

The LHCb magnet is a warm dipole magnet with two identical conical saddle-shaped coils placed horizontally opposite each other within the magnet yoke, seen in Figure 3.5, and provides a bending power of $4\,\mathrm{T\,m}$. The primary component of the magnetic field is oriented in the y direction, and thus the charged particles are bent in the $\pm x$ direction in the $x - z$ plane [89]. The polarity of the magnet is occasionally switched during data-taking, these two polarities are referred to as Mag Up (MU) and Mag Down (MD), with the data for each considered separately to avoid bias from detector asymmetries. The asymmetries arise from minor differences in detector response for particles deflected in the positive or negative x directions downstream of the magnet. For the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis in Part II the two polarities are combined within the datasets considered.

3.2.1 The Vertex Locator

The VELO is the detector element that lies closest to the IP, and its role is to measure precisely charged particle tracks and assist in reconstructing the decay topology. Of particular interest are secondary vertices which are a distinctive feature of c- and b-hadron decays. The long mean lifetimes for these hadrons ($\tau_c \approx \mathcal{O}(0.1 - 1)\,\mathrm{ps}$ [22, 98] & $\tau_b \approx 1.5\,\mathrm{ps}$ [99]) can lead to measurable flight distances of $\mathcal{O}(\mathrm{cm})$ at $\sqrt{s} = 13\,\mathrm{TeV}$ before

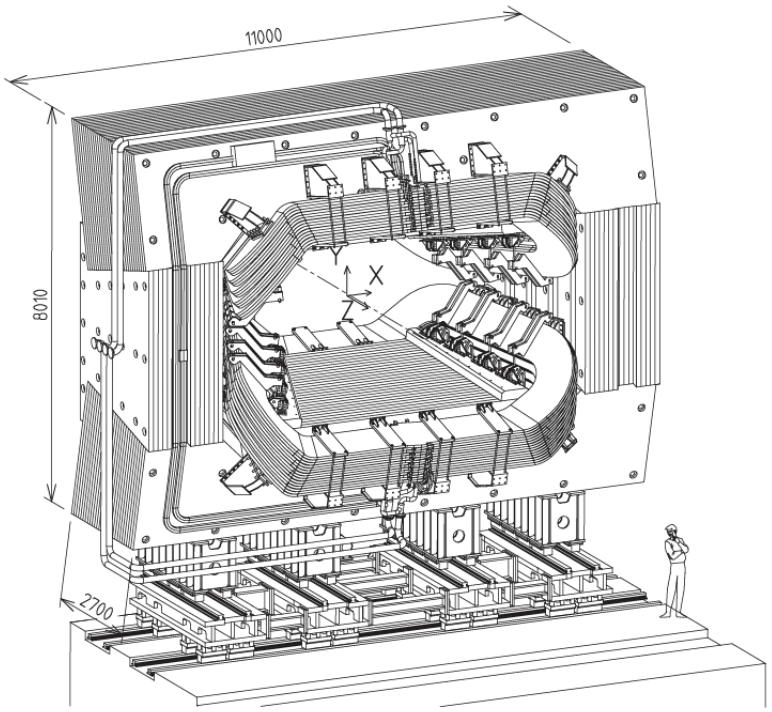


Figure 3.5: A view of the LHCb dipole magnet, with its two 'saddle' shape coils housed in a magnet yoke. The interaction point is situated behind the magnet [89].

decaying, in contrast to decays of light-quark hadrons. The VELO consists of a set of 42 silicon modules arranged along the beam and close to the pp collision point, as seen in Figure 3.6. Each module has an almost hemispherical geometry[5], such that two modules act as a pair, and when overlapped form a circular cross-section. These modules are placed radially closer to the beam than the LHC's required injection aperture, and thus these paired overlapping modules retract during injection and move into position during data-taking. This also requires the VELO detector modules to be contained in a vessel that maintains a vacuum different to the LHC's vacuum in the beam pipe. This vessel is known as the 'RF-foil' and there is one for each side of the VELO. The shape of this RF-foil is also precisely machined and installed, such that its form lies very close to the modules. This is done to ensure as little material as possible is present in front of the detectors. Alongside the 42 silicon modules there are four additional identical modules that are used as the pile-up veto, and these are located before any of the other VELO modules [89].

Each VELO module uses two silicon sensors on top of one another, one has silicon strips oriented radially that measure the azimuthal angle (ϕ-sensor) and the other has strips that are concentric semi-circles that measure the radial distance from the beam,

[5]This was true in Run 1 and Run 2, but from 2022 onwards (Run 3) the modules have an L-shaped geometry and are instrumented with silicon pixel detectors [100].

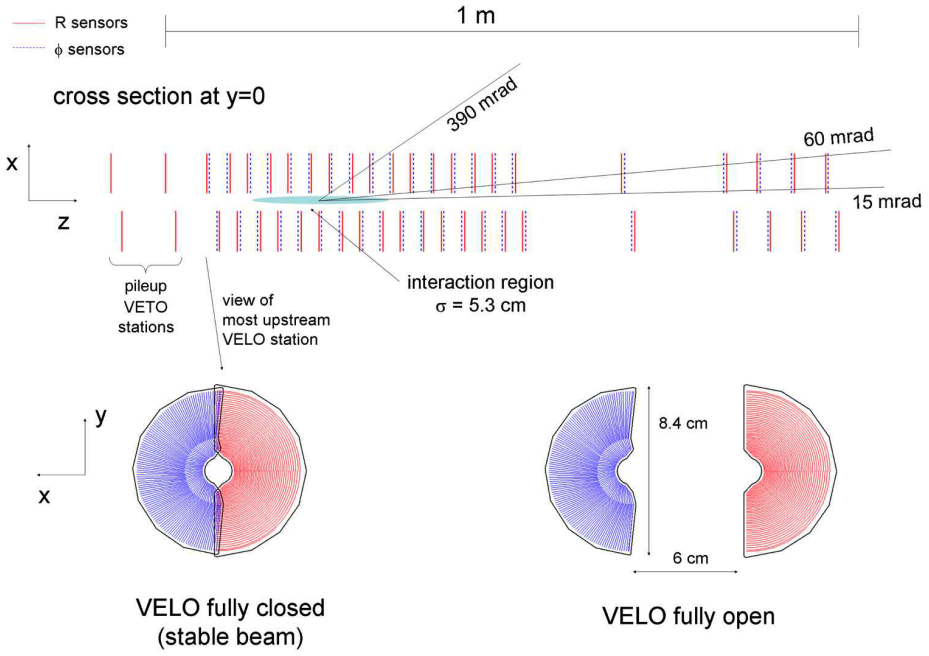
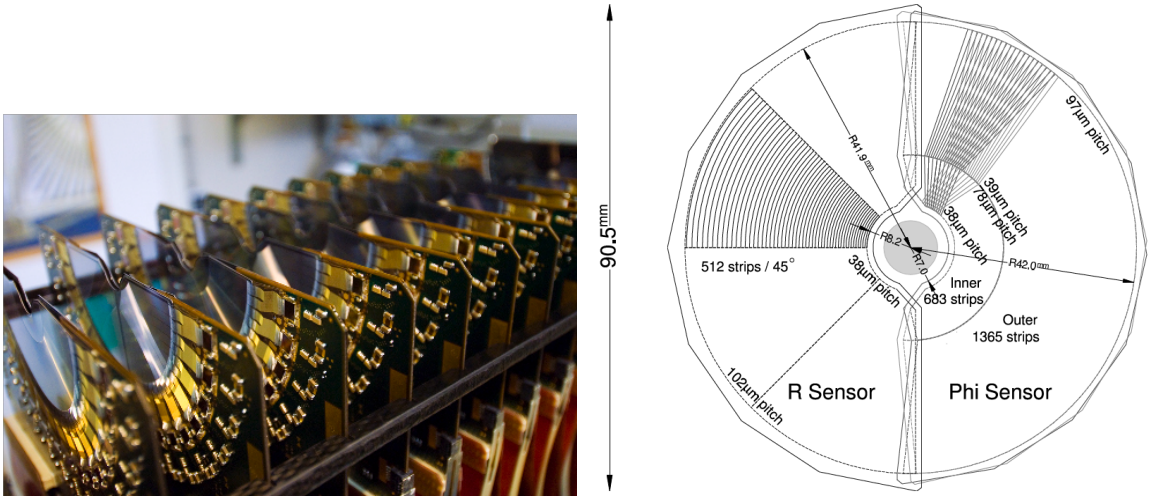


Figure 3.6: (Top) The 42 VELO modules separated into their two halves along the beam-line, with four modules at the far left acting as the pile-up veto stations. (Bottom) A pair of VELO modules when they are closed or open [89].

(r-sensor). This allows the VELO to output in cylindrical polar coordinates r–ϕ–z, where the z coordinate is related to which of the 21 pairs of modules records the hit. This system requires precise alignment of the VELO modules to ensure good tracking, specifically, the two halves need to be aligned relatively to better than 100 µm [89].

For the r-sensors each strip is subdivided into four regions of 45° to reduce the occupancy in individual strips, as seen in Figure 3.7b. The semi-circles also have finer minimal strip separation (pitch) closer to the beam axis to allow for better resolution there, with the pitch being 28 µm, increasing linearly to 101.6 µm at the outer edge of the sensor. Likewise the ϕ-sensor also has different regions, with an inner region up to 17.25 mm radially from the centre in which the pitch is roughly half the size of that in the outer region. The strips are not exactly radial in the ϕ-sensor, instead there is some skew of the strips that reduces the overall occupancy of the strips and also ensures there is not a large pitch at the outer edge of the module. The skew of the inner(outer) region is ∼20°(∼10°) to the radius and the skew is in the opposite direction for the inner and outer regions. Then adjacent ϕ-sensors are such that they also have the opposite skew [89].

The VELO has shown good performance during its lifetime as shown in Ref. [101], with the impact parameter being measured to a resolution of < 35 µm for particles with $p_{\mathrm{T}} > 1\,\mathrm{GeV}/c$ and is typically able to obtain a decay time resolution of ∼50 fs [101], with



(a) **(b)**

Figure 3.7: (a) The installed VELO modules in-situ [101]. (b) The geometry of the $r-$ and ϕ-sensors used in a VELO module, highlighting the strips and their pitches in different regions of the detector. It should be noted that the strips for the ϕ-sensor are for two adjacent modules to highlight the different skews and their overlap [89].

all tracks inside LHCb acceptance passing through at least three modules [89]. Additionally, the VELO has a hit resolution of $5-25\,\mu$m with a tracking efficiency typically better than 98%, and resolutions for a primary vertex with 25 tracks are $13\,\mu$m in the $x-y$ plane and $71\,\mu$m in the z direction [101].

3.2.2 The TT and Downstream Tracking Stations

The TT and downstream tracking stations (T1–T3) are large area planar detectors, each covering an overall rectangular area of $150\times130\text{cm}^2$, but using different detector elements.

The TT uses silicon micro-strip sensors in four different layers with pitches of $\sim200\,\mu$m. The outer layers are oriented such that the strips run in the y direction to measure x and the two inner layers have strips running at angles of $\pm5°$ with respect to y forming an $x-u-v-x$ arrangement. This allows for accurate 2D reconstruction. Figure 3.8 shows the layout of the third detector layer, split into two half modules of $7(8)\times14$ sensors for the inner(outer) layers that are read out in three sectors [89]. In Run 1 the TT had a hit efficiency of $>99.7\%$ with resolution of 52.6–53.4 µm with performance matching the experiment's requirements, as further reported in Ref. [90].

The downstream trackers consist of an Inner Tracker (IT) and an Outer Tracker (OT). The inner tracker is instrumented using similar silicon micro-strip sensors to the TT, but covers a cross-shaped region close to the beam-line; and the outer tracker is a

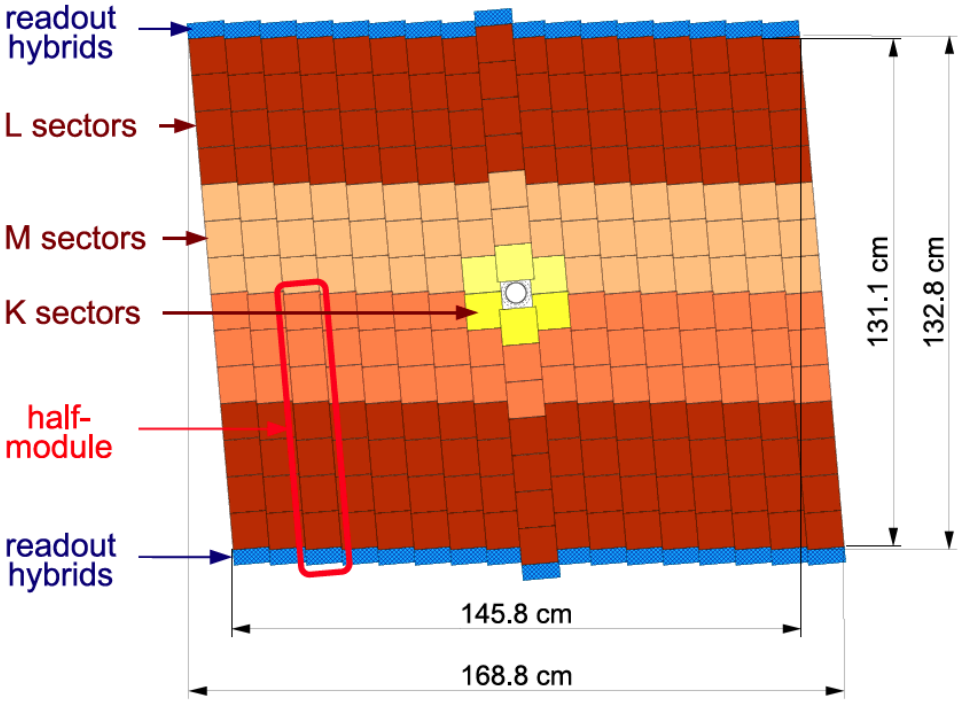
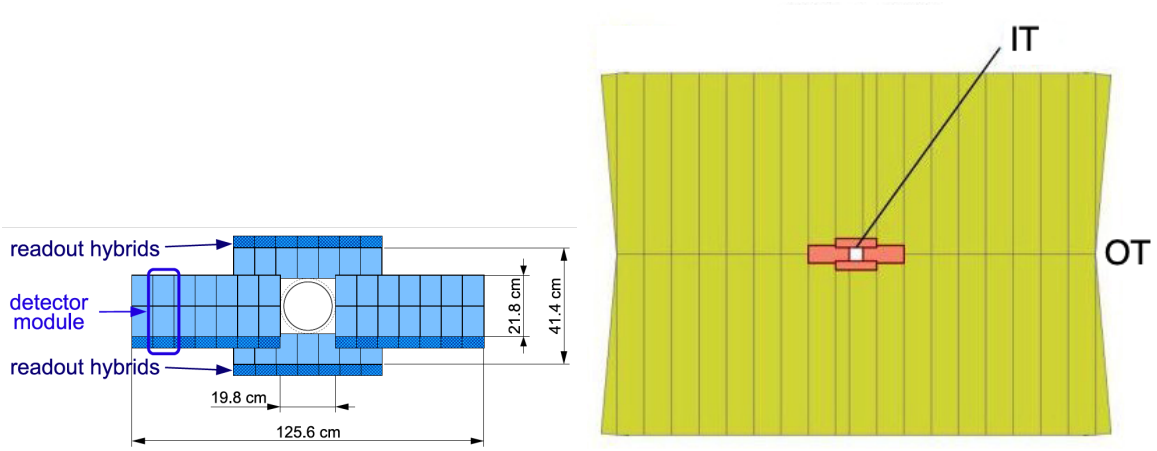


Figure 3.8: Layout of the third detector layer of the TT subdetector, showing the three different sectors that have differing resolutions [89].

large area drift chamber covering the remaining area. Both the IT and OT also have four layers in the $x{-}u{-}v{-}x$ arrangement used in the TT [89].

Figure 3.9a shows the shape of the inner tracker. It lies close to the beam-line, and hence has a very high density of particles traversing it. Similar silicon micro-strip sensors to the TT are used to cope with this environment. The IT is composed of 14 full detector modules with two sensors in each that lie horizontally on either side of the beam pipe, and 14 half detector modules with just one sensor. Seven of these half detector modules lie above the beam pipe and seven below, centred on the beam-line but overlapping slightly with the horizontal modules [89]. Overall, these provide an active area of $4\,\mathrm{m}^2$ and the performance required for LHCb as seen in Ref. [90]. In Run 1 the IT had a hit efficiency of $> 99.8\%$ with a resolution of 50.3–54.9 µm while keeping occupancy $< 2\%$ close to the beam-line, dropping to $\sim$0.2% further away.

The OT uses a different detector technology from the TT and IT. It is a drift detector with modules that contain two staggered layers of 5 mm straw drift-tubes to cover the rest of the $150 \times 130\mathrm{cm}^2$ area not instrumented by the IT. These drift-tubes contain a mixture of Ar (70%) and CO_2 (30%), chosen to ensure a < 50 ns drift time and a resolution of 200 µm [89]. The OT covers an active area of $\sim 29\,\mathrm{m}^2$, with each tracking station having four module layers arranged as $x{-}u{-}v{-}x$, as in the TT and IT. Overall, the OT showed an average single cell efficiency in Run 1 of 99.2%, and single hit resolution



(a) (b)

Figure 3.9: The layout for (a) The Inner Tracker [89] and (b) The Outer Tracker with inner tracker overlaid. Figure was modified from Ref. [102].

of 205 µm which is close to the design resolution quoted earlier [90].

A single IT and OT detector together form one of LHCb's 'T' stations. Unlike the TT, which is placed before the magnet, all three of the T stations lie after LHCb's magnet with a spacing of ∼1 m between them.

At LHCb the majority of charged particle tracks are reconstructed based on the information obtained from the VELO, TT and T stations, with the muon stations acting primarily to identify muons. Tracks are then categorised according to the subdetector information that contributes to their reconstruction, which can affect the track properties *e.g.* resolution. In the instance where a given particle is reconstructed with multiple track type hypotheses, the single most appropriate for analysis is used [103]. There are five main track types, shown in Figure 3.10:

1. A **VELO track**, which is only reconstructed in the VELO.
2. A **T-track**, which is only reconstructed in the T stations.
3. An **upstream track**, which is reconstructed in the VELO and TT.
4. A **downstream track**, which is reconstructed in the TT and T stations, often referred to simply as a 'down' track.
5. A **long track**, which is reconstructed in the VELO, T stations and possibly the TT too.

The analysis presented in Part II is searching for the decay $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$, where the Λ_b^0 itself and possibly the Λ will decay within the VELO, and therefore the focus is

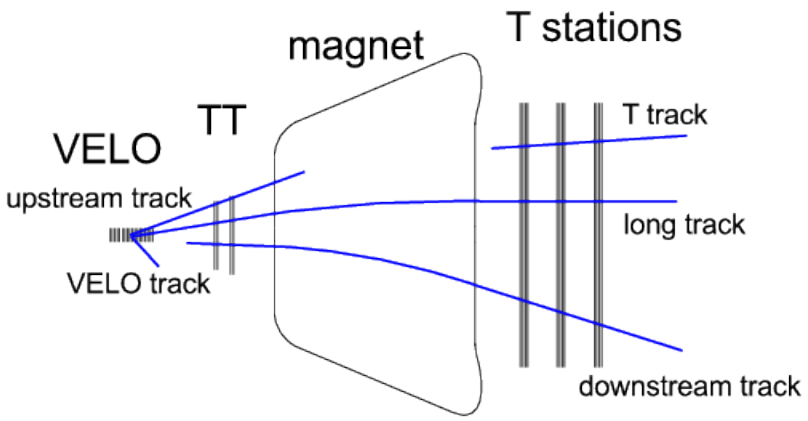


Figure 3.10: The different reconstructed track types at LHCb overlaid on a diagram of the tracking detectors [103].

on downstream and long tracks to ensure good momentum determination of the Λ decay products. Of the track types, long tracks have the best spatial and momentum resolution due to the inclusion of the VELO, however downstream tracks can give information on long-lived decay products at a cost of lower overall resolution.

Another aspect to note is that the design of the LHCb tracking system does have a known complication that has to be accounted for when considering electrons, due to the magnet being upstream of the calorimeters. A charged particle undergoing bremsstrahlung upstream of the magnet produces a photon which is not deflected by the magnetic field. This causes the energy deposition of the charged particle and the photon to be detected in different cells of the calorimeter, leading to a lower-energy particle being reconstructed and biasing analyses. In contrast, a particle that undergoes bremsstrahlung after the magnet would not show this decreased energy as the photon and the charged particle would be detected in the same cell of the calorimeter, giving the correct energy reading. These situations are demonstrated schematically in Figure 3.11. The charged particle undergoing bremsstrahlung is almost always an electron due to its light mass.

To correct for the bremsstrahlung losses the direction of the track before the magnet, as measured by the VELO and TT, is extrapolated to the calorimeter as if no magnetic field were present. If there is an energy deposit in the cells close to this extrapolated direction, with an energy of $> 75\,\text{MeV}$ that is not associated with another charged particle track, the two energies can be recombined to obtain the true energy of the charged particle. If two electrons are considered in the final state and the same bremsstrahlung photon is associated with both electrons, the energy is randomly assigned to one to avoid bias from double-counting energies [104]. This bremsstrahlung-recovery procedure works well to reduce the overall impact on analyses involving electrons. However, when bremsstrahlung

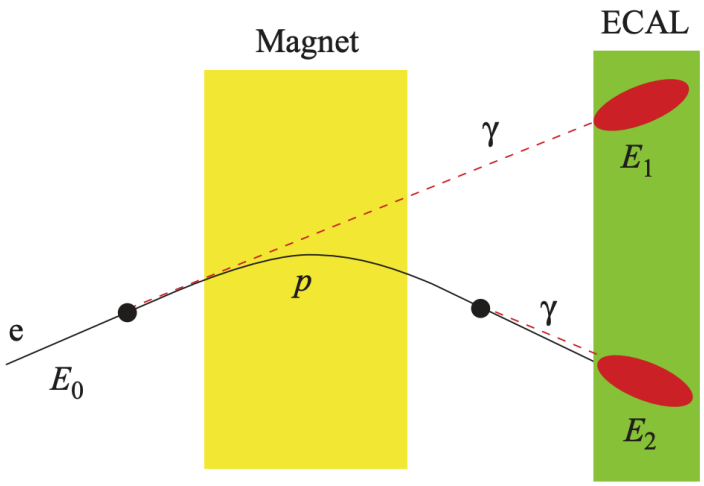


Figure 3.11: The effect of bremsstrahlung before and after the LHCb magnet, showing the former requires some correction to associate clusters $E1$ and $E2$ to the original particle [89].

photons are not detected there is a degradation of the mass resolution of the reconstructed parent particle, which is often seen as a long tail below the mass peak.

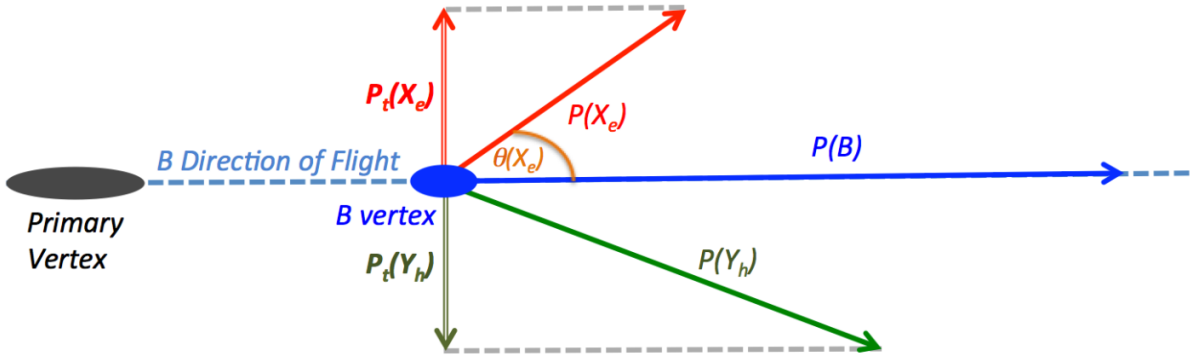


Figure 3.12: The kinematics in a $B \to Y_h X_e$ decay highlighting relevant quantities for the 'HOP' variables [104].

A variable has been developed from Ref. [104] to exploit the kinematics of the parent b-hadron in decays with electrons in the final state. This is shown in Figure 3.12 for a $B \to Y_h X_e$ decay where Y_h represents all particles except the electrons, and X_e all the electrons. Two important points follow from this: the sum of momenta transverse to the b-hadron flight direction, P_t, must be zero; and as bremsstrahlung is emitted in the same direction as the electrons' flight distance the electrons' direction should be unaffected. Using these quantities, the ratio, α_{HOP} can be defined,

$$\alpha_{HOP} = \frac{P_t(Y_h)}{P_t(X_e)}. \tag{3.2.1}$$

As the bremsstrahlung recovery procedure is not 100% efficient this ratio differs from unity, which would be the ideal case. Instead, this ratio is used to correct the momenta of the electron system, as the difference from unity represents the energy lost by bremsstrahlung recovery, giving $\vec{P}_{corr}(X_e) = \alpha_{HOP} \cdot \vec{P}(X_e)$, which in turn can be used to determine the invariant mass of the parent b-hadron. This corrected invariant mass is often referred to as the 'HOP' mass.

Despite its effectiveness, this correction is negatively affected by reconstruction effects, such as the angular resolution of the flight direction of the b-hadron which is used as the reference for P_t, and the resolution of the angle between the electron system and the b-hadron flight distance. From the former, it is expected that the HOP mass is correlated with the χ^2_{FD} of the b-hadron. This is the significance of its flight distance with respect to the primary vertex, which gives a combined representation of the quality of reconstructed vertices and the flight distance that both affect the resolution of the flight direction [104]. Overall, this correction provides good discrimination of partially reconstructed backgrounds and is utilised in many analyses at LHCb involving electrons in the final state.

3.2.3 Muon Stations

The muon system consists of five stations, M1-M5, that are placed far downstream of the IP, with M1 placed just before the calorimeters, as seen in Figure 3.13, to improve the trigger's p_{T} measurement and M2-M5 placed as the furthest downstream detector elements of LHCb, each separated by $\sim$1 m with a $\sim$0.8 m thick iron absorber between them. This iron absorber is used to select highly penetrating muons, as a minimum momentum of 6 GeV/c is required to pass through all five muon stations. The sizes of the muon stations scale with their distance from the IP to ensure uniform angular coverage for all the muon stations, with M1-M5 having dimensions of $7.7 \times 6.4\,\mathrm{m}^2$, $9.6 \times 8.0\,\mathrm{m}^2$, $10.4 \times 8.6\,\mathrm{m}^2$, $11.1 \times 9.3\,\mathrm{m}^2$, and $11.9 \times 9.9\,\mathrm{m}^2$ respectively [105]. Overall the muon system covers an area of $435\,\mathrm{m}^2$.

Each muon station is divided into four regions which are shown in Figure 3.14. Each of these regions have different levels of granularity depending on the distance from the beam-line with segmentation sizes that roughly double for each subsequent region. This has the effect of distributing the particle flux and occupancy uniformly across the four regions, though the spatial resolution is reduced at distances farther from the beam axis and IP because of this. These regions are instrumented with MWPCs for all the muon stations except for region 1 of M1 where the particle rate is too high during the expected lifetime of the detector, in this case triple-Gas Electron Multiplier (GEM) detectors are

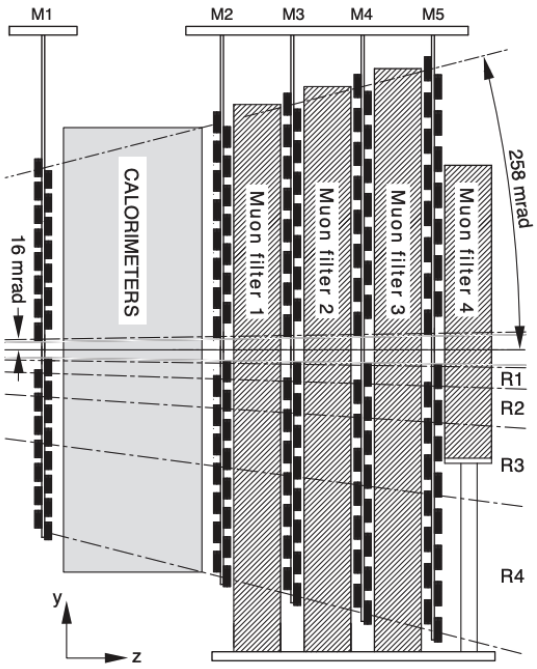


Figure 3.13: The LHCb muon system including all five muon stations relative to the calorimeters and the iron absorbers between M2-M5. The lines R1-R4 show the projections of the different regions of the muon stations [89].

used. For M2-M5 the chambers have four gas gaps in two sensitive layers, whereas M1 has two gas gaps, which reduces the material before the ECAL, and two overlaid GEM detectors in the central region.

There are a total of 1368 MWPCs in the muon system, using a gas mixture of Ar, CO_2 and CF_4 in the ratio 40:55:5. This, combined with linking two adjacent gas gaps via an OR logic gate, leads to a predicted efficiency $> 95\%$ in a 20 ns window for a gain of 10^5 [89]. The general principle of an MWPC is that there are two parallel conductive planes both with a set of wires at high voltage running parallel to each other and a gas mixture between them. A charged particle passing through the system then ionises the gas mixture, and the electric field from the wires collects the avalanche electrons in the closest wire to the initial charged particle which can be read-out and converted to a position.

There are a total of 24 triple-GEM detectors in the 12 chambers of region 1 of M1. These triple-GEM detectors are composed of three GEM foils that sit stacked between an anode and a cathode plane with gaps between each of these as shown in Figure 3.15. The same gas mixture as the MWPCs is used in the triple-GEMs but in the ratio 45:15:40 instead [89]. Similarly to the MWPCs, a charged particle passing through the detector initiates an electron avalanche in the drift gap, with an electric field set up between the anode and cathode planes causing these electrons to drift towards the GEM foils. Each GEM foil then multiplies the number of electrons until a measurable signal is detected

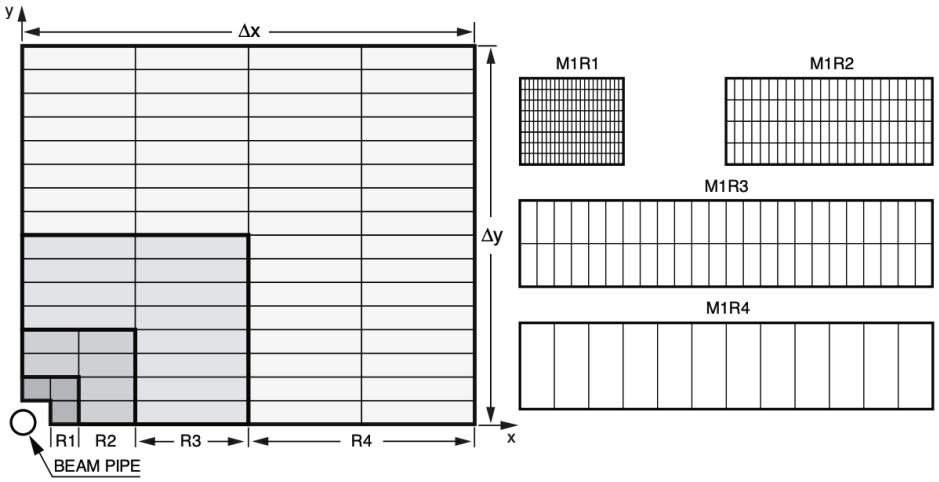


Figure 3.14: (Left) A quarter of an LHCb muon station with the four detector regions highlighted with the chambers represented as rectangles. (Right) The division of these chambers in each region into pads showing the granularity of the detectors in that region [89].

on the anode, which is divided into pads to allow for spatial resolution. Two triple-GEM detectors are then overlaid to form two layers with the outputs linked with an OR logic as in the MWPCs, to form the chambers in region 1 of M1 [89].

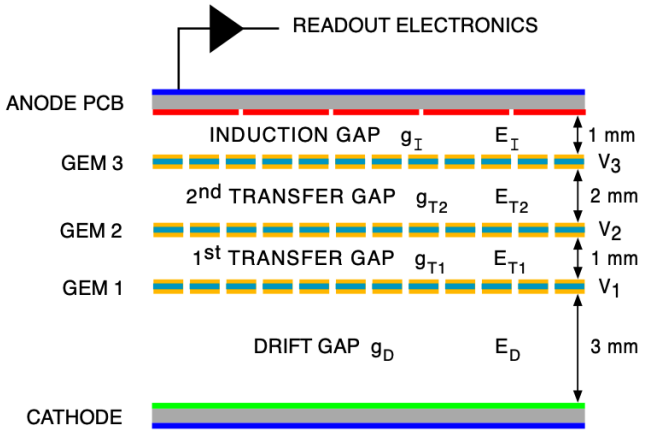


Figure 3.15: Cross-section of a triple-GEM detector used in the inner region of the first muon tracking station [89].

The muon stations are used to identify muons with high purity signals of penetrating muons. This is possible for high p_{T} muons if hits are observed in all five muon stations, however lower momentum muons may only provide hits in the two or three closest stations after the calorimeter. The need to identify muons at LHCb for a wide array of analyses means that a stringent limit of at least 99% detection efficiency is required for each station, a requirement that the muon tracking system easily achieves [90]. If there are any tracks in the muon stations then the event is flagged with the algorithm `HasMuon`.

These hits from the muon stations are then identified as muons or not using the algorithm `isMuon`, which extrapolates the trajectory in the muon system and attempts to associate them with previous reconstructed tracks from upstream tracking stations. This provides a muon identification efficiency in the trigger and offline reconstruction of ∼97%, with a ∼1% probability of misidentifying a pion as a muon for momenta $> 3\,\text{GeV}/c$ [106].

3.3 RICH Detectors

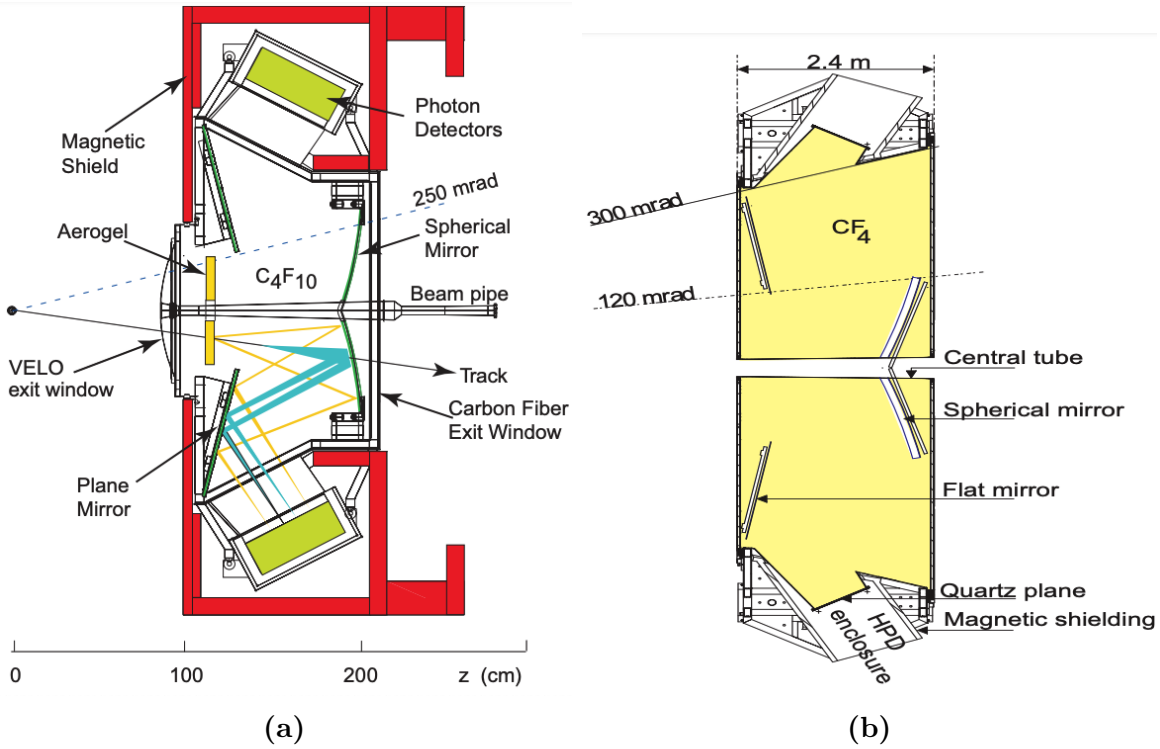
An important aspect of many particle physics experiments is the ability to identify the particles that are detected. At LHCb this is fundamental to allow for the separation of pions and kaons which often arise in b-hadron decays as backgrounds for one another. One part of the Particle Identification (PID) system at LHCb is the two RICH detectors, which use the emission of Cherenkov photons from charged particles that pass through a gas-filled region to determine their velocity. This can be combined with the momentum measurement provided by the tracking system to determine their masses and, therefore, identity. This is possible as the Cherenkov photons are emitted at a characteristic angle, the cosine of which is inversely proportional to the velocity. These photons are guided and focused by a mirror system in the RICH detectors which read the photons as a set of rings that can be thought of as a cone with an 'opening angle' of the characteristic Cherenkov angle.

A large momentum range needs to be covered at LHCb, and so the two RICH detectors are optimised to cover different momentum regimes. RICH1 lies upstream of the magnet, just after the VELO and before the TT, and covers the low momentum range $\sim[1, 60]\,\text{GeV}/c$ [89]. By contrast RICH2 lies downstream of the magnet between T3 and M1 and covers the higher momentum range $\sim[15, >100]\,\text{GeV}/c$. Each RICH detector employs a different implementation of the basic concept of a RICH detector, with different optical set-ups and materials used, as shown in Figure 3.16. Additionally, RICH1 has to ensure it has minimal material that could interact with the particles before the magnet and calorimeters.

The RICH1 detector contains a mixture of an aerogel plane and C_4F_{10} gas that acts as the radiator of Cherenkov photons when charged particles pass through it [89]. These photons are guided by primary spherical mirrors onto a secondary planar mirror, and then onto photon detection planes instrumented by Hybrid Photodetectors (HPDs). The HPDs are not designed to operate in high magnetic field environments (60 mT) and require the local magnetic field near RICH1 to be attenuated by a factor of ∼20 [89] to 3 mT. This is the reason for the planar mirrors as they allow for the HPDs to be situated

outside the LHCb acceptance in magnetic shielding boxes.

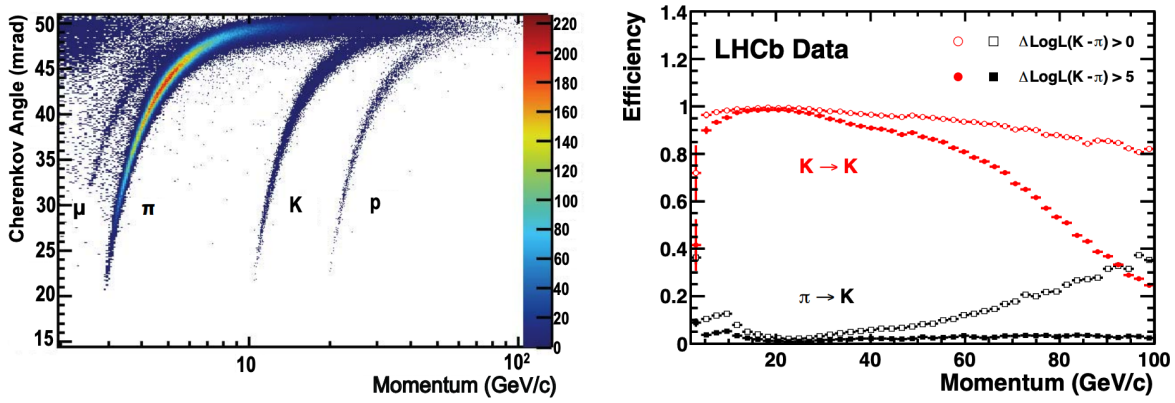
The RICH2 detector is similar in its optical arrangement, with a set of primary spherical and secondary planar mirrors, though the radius of curvature for the spherical mirror is much higher. It contains a different gas, CF_4, and no additional radiating material [89]. The planar mirrors here ensure that the length of the detector is kept as short as possible and limit the material before the calorimeters, and once more reflect the Cherenkov rings onto a plane of HPDs in a magnetically shielded location just outside the LHCb acceptance.



(a) **(b)**

Figure 3.16: (a) The Layout of RICH1 from a side-view showing the internal optical system. (b) The layout of RICH2 from an overhead view, showing the internal optical system [89].

Overall, the measurement of the Cherenkov angle at a range of momenta performed using the RICH detectors at LHCb provides a way to distinguish different particle identities. Figure 3.17a shows how the different particles form different 'branches' in Cherenkov angle as a function of momentum due to their differing masses as measured by RICH2. With this, pions and kaons can be distinguished for a large momentum range, though at higher momenta the identities of the particles are harder to discern. For the momentum range 2–100 GeV/c in Run 1 the kaon identification efficiency is 95% with the misidentification rate of a pion as a kaon at 3% [90]. Figure 3.17b shows these two quantities in data for Run 1.



(a) (b)

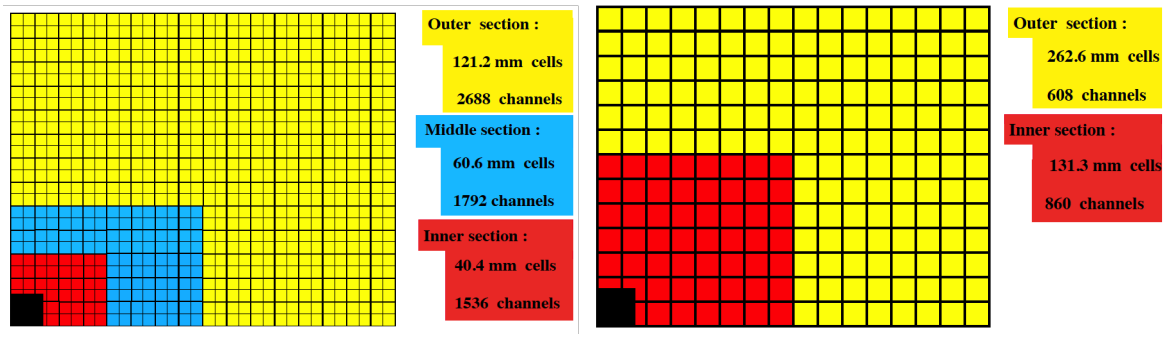
Figure 3.17: (a) The reconstructed Cherenkov angle as a function of the track momentum in RICH2 for isolated tracks. Showing clear 'branches' that can be identified as muons, pions, kaons and protons respectively. (b) The kaon identification efficiency and pion to kaon misidentification rate as a function of track momentum in data for two separate requirements on the likelihood for the kaon mass hypothesis to be larger than the pion hypothesis, $\Delta \log \mathcal{L}(K - \pi)$ [90].

3.4 Calorimeters

LHCb has two calorimeters that are optimised to measure the energy of particles from a b-hadron decay that primarily interact either electromagnetically or hadronically. The calorimeters are located downstream of RICH2, M1 and two sub-detectors that work in tandem with the calorimeters: the PreShower (PS) and the Scintillating Pad Detector (SPD).

The SPD is a scintillator pad detector with high granularity that allows for the distinction of charged particles compared to neutrals. In particular, this is used to distinguish between electrons and neutral particles (*i.e.* photons or π^0). This is possible as the charged particle will interact with the scintillator and deposit a small amount of energy, whereas neutral particles do not interact with the scintillator. The PreShower detector is then placed after the SPD with a 15 mm, or 2.5 X_0 radiation lengths, thick plane of lead between them [89]. This lead plane acts as a converter, initiating an electromagnetic shower for electrons and photons. The PS is also a scintillator pad detector and so can also distinguish charged particles from neutrals, but is predominantly used to initiate the shower for the ECAL whilst also recording the deposited energies which can be used to separate electrons and charged pions. Both the SPD, PS and the lead plane have dimensions of $7.6 \times 6.2\,\mathrm{m}^2$, with the planes of the SPD and PS divided into three regions with granularity varying with distance from the beam-line. In the inner, middle and outer sections, the pad is divided into cells of dimension $4 \times 4\,\mathrm{cm}^2$, $6 \times 6\,\mathrm{cm}^2$, and $12 \times 12\,\mathrm{cm}^2$

respectively [89]. These align with the segmentation of the ECAL shown in Figure 3.18a.



(a) **(b)**

Figure 3.18: The segmentation perpendicular to the beam-line for (a) the SPD, PS and ECAL and (b) the HCAL in a single quadrant of the detector. The cell dimensions in (a) are reported for the ECAL [89].

3.4.1 Electromagnetic Calorimeter

The ECAL is a sampling calorimeter that has 66 layers perpendicular to the beam-line. The layers alternate between 4 mm thick active scintillator material and 2 mm thick passive lead converters [89]. The passive converters help to initiate showers, whereas the active scintillators measure energy deposited by the shower through the emitted scintillation photons that are read out by plastic wavelength-shifting fibres connected to a Photomultipler Tube (PMT). The main purpose of the ECAL is to measure the energy for particles that interact electromagnetically, *i.e.* charged particles, photons and neutral pions. As in the SPD and PS, the ECAL is divided into three sections radially out from the beam-line that cover the same projected area relative to the IP, so the exact dimensions of the detectors are scaled up further away [89]. The inner sections have higher granularity to handle the increased density of events near to the beam-line. These sections and the cells dimensions for the ECAL are shown in Figure 3.18a. The ECAL has a design energy resolution of

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%, \tag{3.4.1}$$

where E is the measured energy in GeV [89].

3.4.2 Hadronic Calorimeter

The HCAL, instead of attempting to measure the energy of electromagnetically interacting particles measures the energy of hadrons, and also absorbs the hadronic showers entirely

before the outer muon chambers. This prevents a large background of charged hadrons that would be detected in the muon chambers. Like the ECAL, the HCAL is a sampling calorimeter but, instead of lead, the passive converter material is iron. However, unlike the ECAL, the six layers of the calorimeter are not perpendicular to the beam-line but parallel to it, as shown in Figure 3.19. Additionally, the HCAL is only divided into two regions of granularity in the transverse direction to the beam-line for the same reason as the ECAL, with cells of size 131.3 mm in the inner section and 262.6 mm in the outer [89], as shown in Figure 3.18b. The HCAL measures the energy of the shower in the same way as the ECAL, with a design energy resolution of

$$\frac{\sigma_E}{E} = \frac{(69 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%, \tag{3.4.2}$$

where E is the measured energy in GeV. The HCAL has a longitudinal dimension of $\sim 5.6\lambda_I$, where λ_I is the hadronic interaction length, which was limited by the available space but sufficiently contains the hadronic showers to a controlled level [89].

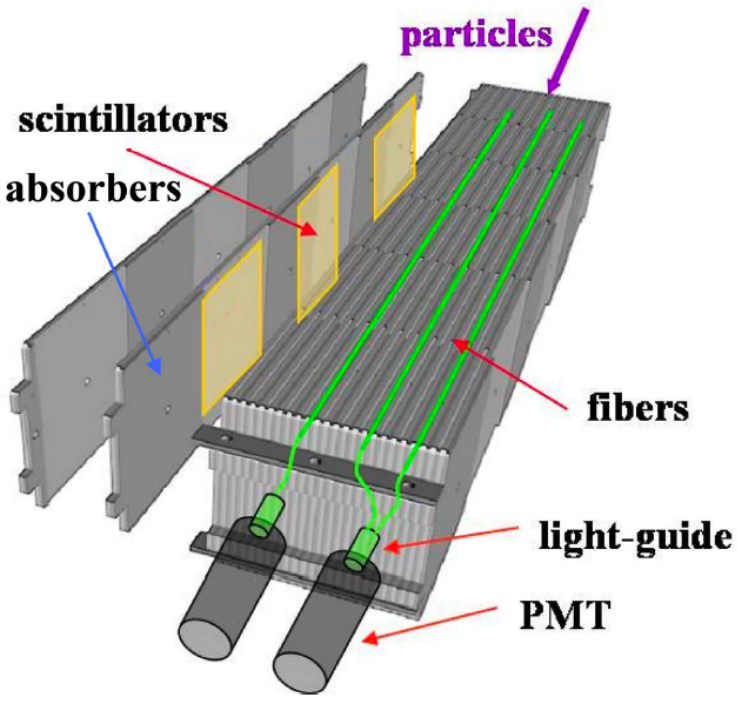


Figure 3.19: The internal structure of a cell within the HCAL, with two tiles that show the six layers of alternating scintillators and absorbers in the longitudinal direction [89].

3.5 Particle Identification

The measurements of particle clusters within both the ECAL and the HCAL can also be used for particle identification due to the calorimeters inherently interacting with different types of particles. This information can be utilised alongside the information obtained by the RICH detectors, the SPD and the PS to further improve the identification of particles within the LHCb detector. However, there is still the possibility of misidentification which must be considered when studying potential backgrounds in the measurement of a particular decay. A representation of how the information throughout the LHCb detector can be combined to determine particle identity is shown in Figure 3.20. In practice, the RICH detectors, calorimeters and muon stations provide a likelihood for each particle hypothesis considered. These likelihoods can then be used as PID variables to perform offline analyses, where Delta Log Likelihood (DLL_{X-Y}) variables are the sum from each of the sub-detectors of the log-likelihood differences between the hypothesis of the particle, X, and a reference particle, Y. There are also the ProbNN_X variables that use these likelihoods together with other information such that from the tracking system to train a neural network which gives a pseudo-probability that the particle matches a particular hypothesis, X.

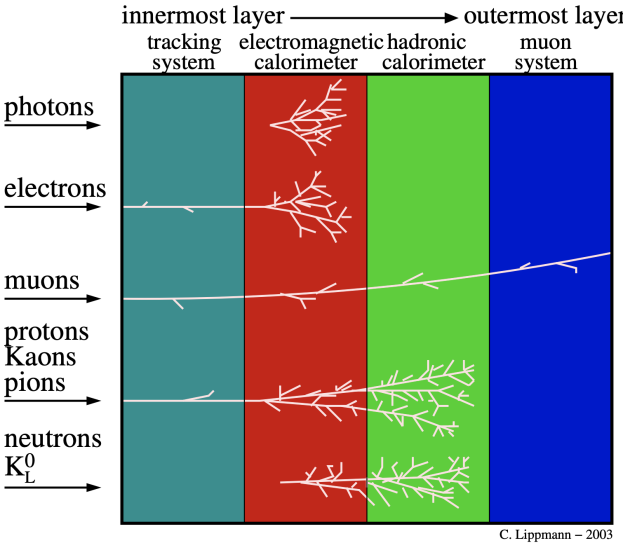


Figure 3.20: The detector signatures for different particle types in the detector elements of a standard particle physics experiment [107].

3.6 Trigger

The rate of pp collisions is around 10–20 MHz, which was too high for all the relevant detector information to be processed and stored during Run 1 and Run 2. Therefore,

LHCb used a trigger system that reduces this data frequency to a manageable 5 kHz in Run 1, and 12.5 kHz in Run 2. This allows for the data to be stored for offline analysis. If an event does not pass the trigger requirements it is lost and so the trigger configuration is designed to isolate the maximum number of 'interesting' physics events.

There were two types of trigger used in Run 1 and Run 2, a hardware trigger known as Level-0 (L0), and a software trigger known as the High Level Trigger (HLT). The hardware trigger uses the basic signals from detectors to distinguish 'useful' physics events using the pile-up system in the VELO, which identifies collisions with single or multiple interactions. The L0 calorimeter trigger identifies the presence of particles with high transverse energy such as photons, pions or hadrons, and the L0 muon trigger that identifies two muons with the highest p_{T} in each quadrant of the muon trackers [89]. This tends to reduce the data rate to ∼1 MHz. The HLT is divided into two parts, HLT1 and HLT2 both using LHCb's Moore application. HLT1 reconstructs particles in the VELO and T-stations that correspond to objects in the L0 trigger to confirm the L0 candidate, whereas HLT2 uses a combination of inclusive and exclusive trigger algorithms to partially or fully reconstruct b-hadron decays [89]. HLT1 reduces the data rate to ∼40 kHz or ∼110 kHz, and HLT2 reduces it further to ∼5 kHz or ∼12.5 kHz for Run 1 [108] and Run 2 [109] respectively. The overall effect of the trigger at different stages in Run 1 and Run 2 is shown in Figure 3.21. For Run 3, the HLT trigger has become sufficient to reduce the data rate to a manageable level and so the L0 trigger is not used. This was done to optimise the possible physics events that can be observed and allow for the detector to adapt its trigger more easily as it only relies on software and not hardware.

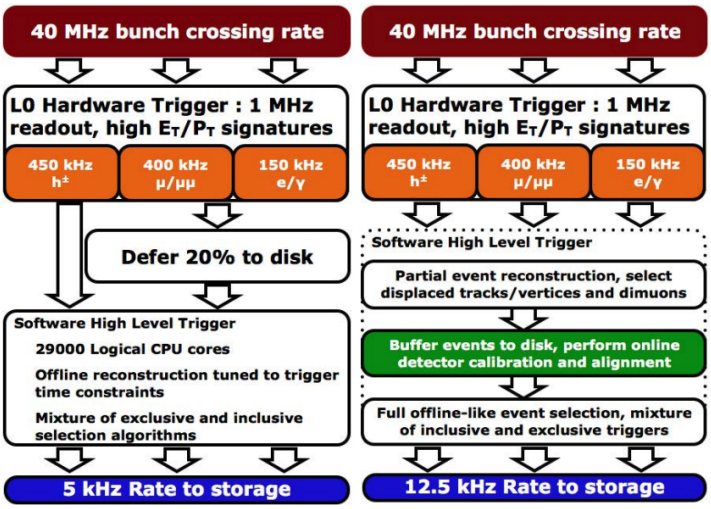


Figure 3.21: Overview of the LHCb trigger and its effect on the data rate at different trigger stages for Run 1 (left) and Run 2 (right).

Although the hardware trigger remained the same for both Run 1 and Run 2, the HLT was improved in Run 2. This was due to the HLT1 events being buffered with 10 PB of storage on the online computing farm, allowing data-taking for up to two weeks before execution of HLT2 and the data stored. This delay allows the alignment and calibration of the subdetectors that HLT2 uses to be performed, and the reconstruction and selection could then be both carried out entirely with the software trigger [109].

LHCb has pre-defined 'trigger-lines' that each have a sequence of reconstruction algorithms and selections optimised for particular physics cases. The trigger line combined with the L0 configuration create a unique trigger that is identified by a Trigger Configuration Key (TCK) which is stored alongside raw data. A topological trigger line combines a set of N "Topo-tracks" into an N-body candidate, that can be used by the trigger to identify particular decay topologies. This N-body candidate can also be determined using a Boosted Decision Tree (BDT), which is a multi-variate classifier that determines the optimal candidates. In this case a topological 'MVA' trigger line is used [108]. These Topo lines are used in the analysis described in Part II, and referenced in Chapter 7.

When an event has been triggered it can be classified as 'TOS' (Trigger on Signal) where the trigger was activated by the signal candidate, or 'TIS' (Trigger Independently of Signal) where the trigger was activated by the rest of the event. Additionally, an event can be classed as both TOS and TIS simultaneously. This leads to the 'TISTOS' method that allows for the trigger efficiency to be determined by,

$$\varepsilon_{trig} = \frac{N_{TOS}}{N_{TOS} + N_{TIS\,!TOS}}, \tag{3.6.1}$$

where ε_{trig} is the trigger efficiency, N_{TOS} is the yield of events classified as TOS, and $N_{TIS\,!TOS}$ is the yield of events classified as TIS that are not also TOS [110].

3.7 Data and Simulation

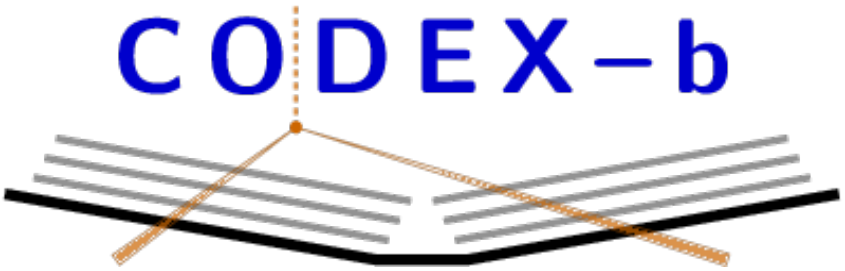
After an event passes the trigger configuration it is stored to tape and the Brunel application is used to perform reconstruction of the underlying physics event. This extracts the hits in detectors and converts them into track objects, allowing for the association of multiple tracks with decay vertices and thus different potential interactions that can be analysed. The reconstructed events can then be processed into different general decay 'archetypes' using low-level selections. This allows for easier portioning of data to the archetype most relevant to the particular case. These selections are known as 'Stripping Lines'. For example, the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis uses the `Bu2LLK_meLine` stripping line that selects decays of a b-hadron to a hadron and a pair of leptons. Stripping is handled by a

software application known as DaVinci that can also process the reconstructed variables into other useful variables that analysts may want to employ, such as combined invariant mass combinations of different final state particles which can be used to identify partially reconstructed backgrounds. DaVinci also processes the stripped dataset into N-tuples. This is a database format that provides all requested information from the data events over a specified time-frame, where this information can be obtained on a per-event basis and often stored as Root files. It is these N-tuples, or 'tuples' that analysts use to perform searches or measurements by manipulating and fitting the data contained within.

To determine the effect selections have on data and to examine the kinematic structure of the different decays examined for an analysis, simulated data are used extensively. Within LHCb these are generated using the Gauss framework. The underlying pp event without modelling of experimental effects is produced using Pythia [111, 112] with a specific configuration for LHCb [113] as the event generator. The subsequent decay of b-hadrons is simulated using the EvtGen application [114]. Final state radiation can also be generated where appropriate using Photos [115, 116]. Properties relating to these simulated events without experimental modelling are referred to as 'generator-level' variables. The interaction of these particles with the LHCb detector is modelled using the Geant4 toolkit [117, 118]. To ensure that the simulated data can be compared directly with LHCb data, they follow the same processing procedure: Brunel to DaVinci to N-tuples. Hereafter, the simulated data will be referred to as Monte Carlo (MC).

Part I

CODEX-b



Chapter 4

The CODEX-b Detector Proposal

The CODEX-b project introduces a detector proposed to perform a dedicated direct search for LLPs [76], which would complement investigations carried out at the LHC experiments. Searches may be either direct or indirect in their approach. Direct searches may be through the observation of potentially anomalous energy losses that can not be related to a SM particle could indicate a charged LLP that interacts with the SM via ionisation [119]. An alternative direct search method uses a 'disappearing track' signature, where a charged LLP forms a track that starts at the IP but is missing hits further away as it decays to BSM particles that do not interact with the detector [120]. Indirect LLP searches are also possible, for example at the LHC experiments using displaced jets of SM particles that cannot be associated with an observed SM track. These arise from an LLP that does not interact with the detector decaying to SM particles [121]. However, LLP searches are not the primary goals of the LHC experiments, therefore they have not been optimised for that purpose. For example, they have been designed to reconstruct the underlying pp collision as precisely as possible and therefore concentrate their tracking detectors close to the IP. This could mean that longer lifetime exotic LLPs may not leave a measurable signature over the volume covered by the detector. Additionally, there can be hardware/software limitations, such as triggering. Thus, the LHC experiments may not be able to handle directly the more complex topologies that require isolated tracks without associated hits, which could be identified as poor reconstruction rather than a potential LLP event. Dedicated trigger configurations for LLPs are developed, or existing ones manipulated for this purpose, but this can cause additional difficulties in searches.

Therefore, there have been proposals at CERN for multiple dedicated LLP experiments [78–82] specifically optimised for observation of a variety of potential LLP models. This is essential as LLPs can arise generically in many different models, as described in Chapter 2, with a parameter space that is too large to be covered by any one experiment.

Hence, multiple experiments work together to explore different aspects of these parameter spaces to constrain the existing models, and in turn inform future experiments what areas still need exploring to constrain further or find an LLP signal. These experiments also cover overlapping areas in parameter space to cross-check and complement the obtained results.

The advantages that the proposed CODEX-b detector would provide are: additional sensitivity to a range of BSM LLP models or complementing existing sensitivity; allowing searches in a minimal background environment such that any detected signal has a strong likelihood of being an LLP candidate; exploiting the nearby LHCb detector to provide independent identification of pp events that give rise to detected LLP candidates in CODEX-b; and its compact size leading to a relatively low-cost detector [122].

Before CODEX-b is constructed there will be a smaller scale detector, 125 times smaller in volume, known as the COmpact Detector for EXotics - Beta (CODEX-β), which will be set up in the nominal location and detector configuration of CODEX-b. This demonstrator unit will allow installation and operational experience to be developed and refined before the full-scale installation. Data can also be taken to observe any potential backgrounds or LLP signal, despite the much lower sensitivity CODEX-β would provide due to its smaller size.

4.1 Detector Location

The currently proposed location for the CODEX-b detector is in a counting barrack next to the LHCb cavern, as seen in Figure 4.1, roughly 25 m away from LHCb's interaction point, IP8. The detector itself has a cubic geometry and will occupy a fiducial volume of $10 \times 10 \times 10\,\mathrm{m}^3$, corresponding to a pseudorapidity range of $\eta \in [0.13, 0.54]$ [76]. This location itself provides several benefits such as the presence of the concrete UXA radiation wall, which is 3 m thick and will strongly reduce potential backgrounds from SM LLPs that could potentially traverse the 25 m distance to CODEX-b. This is already shown from existing background studies that are discussed in Section 4.4. These studies imply that the goal of a minimal background detector could be achieved by the inclusion of an additional lead shield and embedded scintillator located close to IP8.

Another benefit provided by the location is its infrastructure, as the counting barrack is accessible during LHC runs. Additionally, the counting barrack itself is well connected to the electrical and gas supply that would be required for the continual running of the detector and therefore there would be minimal additional infrastructure required to serve CODEX-b, reducing the overall cost of the detector.

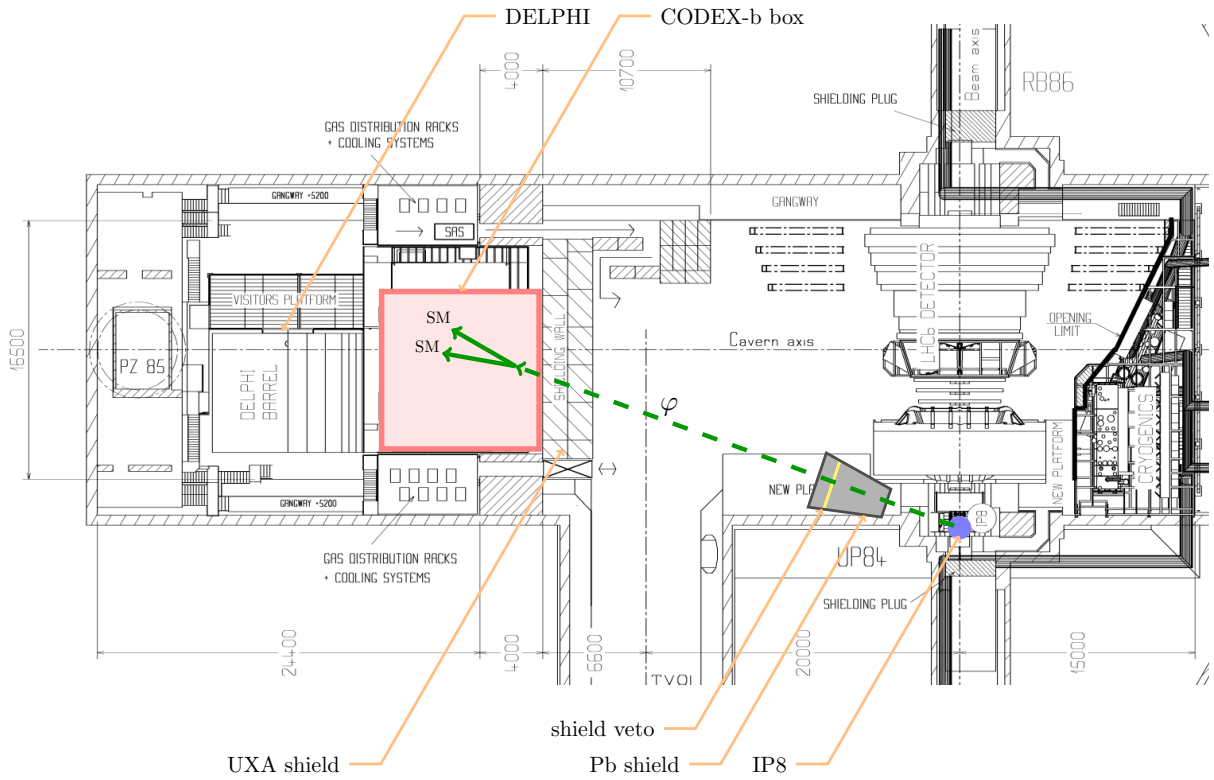


Figure 4.1: The layout of the LHCb experimental cavern UX85 at interaction point 8 of the LHC, with the CODEX-b detector and proposed prompt shield overlaid [123].

The location of CODEX-b will also determine the types of physics models to which it is sensitive. It is transverse to IP8 and so directly probes processes with a high parton centre-of-mass energy, $\sqrt{\hat{s}}$, such as Higgs or Z production for relatively light ($m_{LLP} \lesssim 10$–$100\,\text{GeV}$) LLPs with lifetimes of $1 \lesssim c\tau \lesssim 10^7\,\text{m}$. This can be compared to other LLP detectors where: ATLAS and CMS typically look for heavy LLPs ($m_{LLP} \gtrsim 10\,\text{GeV}$) with $c\tau \lesssim 10^7\,\text{m}$; LHCb looks for light LLPs ($0.1 \lesssim m_{LLP} \lesssim 10\,\text{GeV}$) with lifetimes $c\tau \lesssim 1\,\text{m}$ [76]; forward/beam dump detectors like the ForwArd Search ExpeRiment (FASER) [78] look for light LLPs ($m_{LLP} \lesssim \mathcal{O}(1)\,\text{GeV}$) with lifetimes $0.1 \lesssim c\tau \lesssim 10^7\,\text{m}$ and low $\sqrt{\hat{s}}$ [76]. The differences and overlaps between these different detectors are shown schematically in Figure 4.2.

The proximity to LHCb also could allow direct identification of the underlying event that produces an LLP detected in CODEX-b, because the future trigger-less readout of LHCb has the potential to integrate CODEX-b as a sub-detector. Of the existing and proposed LLP experiments at the LHC, none have had the opportunity for integration on this level [76]. This means that CODEX-b could have the potential to not only observe LLP candidates, but also to determine detailed topological information of the underlying physics event and so provide powerful discrimination between different models.

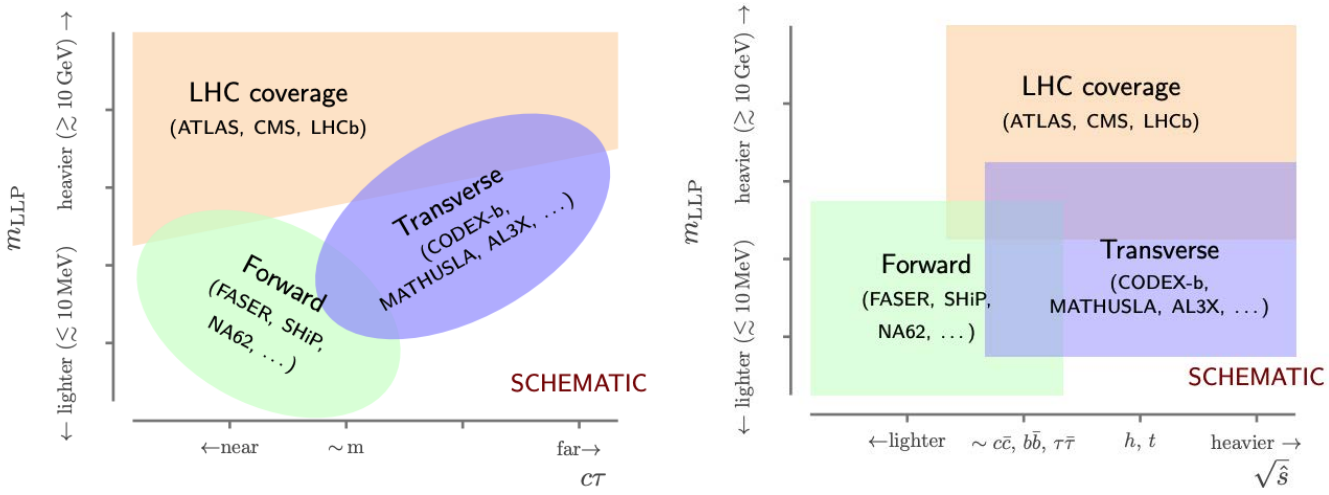


Figure 4.2: The experimental coverage of current, planned or proposed experiments in terms of the mass of the LLP and its lifetime (left) or the required parton centre-of-mass energy (right), $\sqrt{\hat{s}}$ [76].

4.2 Detector Technology: BIS78 RPCs

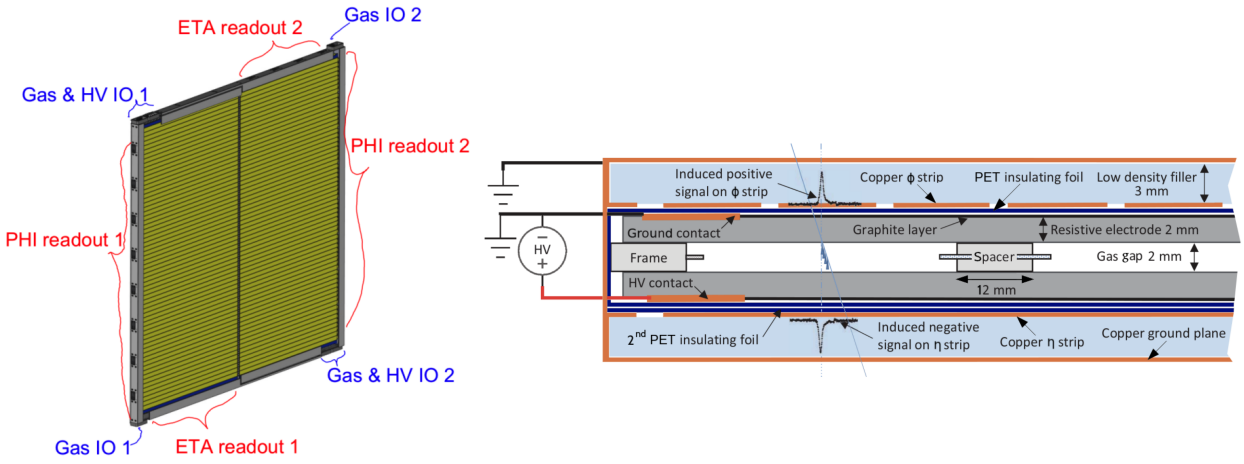
For both CODEX-β and CODEX-b the detector will be composed of a set of ATLAS BIS78 Resistive Plate Chambers (RPCs) designed for the ATLAS phase-II upgrade. Each RPC is ~0.7 cm thick and has an area of $2.0 \times 1.1\,\mathrm{m}^2$, where the long(short) edge is referred to as the 'Phi'('Eta') edge[1] as shown in Figure 4.3a, which has two RPC modules next to each other. These RPCs are composed of 13 layers of materials including a gas-gap with a mixture of gases[2]: $C_2H_2F_4$(94.7%), C_4H_{10}(5%), and SF_6(0.3%) [124]. A schematic diagram of the BIS78 RPC singlet is presented in Figure 4.3b. This shows two planes with orthogonal strips, having strip pitches of 20–25 mm, that allow for two orthogonal coordinates to be read out by front-end electronics mounted on one of the Phi and Eta edges [125]. These Phi and Eta readouts directly correspond to the strips on either side of the gas-gap, so all Phi signals are for one particular orthogonal coordinate, and Eta for the other.

These front-end electronics are read out using a Printed Circuit Board (PCB) developed by the BIS78 team. It has a total of eight channels of a pre-amplifier, coupled with two custom ASIC discriminators, which each have four channels. The boards also have integrated Low-Voltage Differential Signalling (LVDS) transmitters [125]. Figure 4.4 shows one of these readout PCBs, highlighting some of the key aspects. The avalanche electrons generated by a charged particle passing through the RPC will drift toward the

[1] This nomenclature follows from the ATLAS documentation of the RPCs where these edges would explicitly align with their definition of ϕ and η relative to the ATLAS detector. For CODEX-b, they are used in x, y, z coordinates in three alignments so this labelling is just for ease of reference.

[2] The $C_2H_2F_4$ (freon) gas will be replaced with a more environmentally friendly alternative.

Eta readout strips, and hence the associated Eta readout boards will only ever receive a negative signal and equivalently the Phi boards receive a positive signal from the same event. However, the ASIC discriminator for these boards requires a negative signal input and so there must be two versions of the readout boards: a Phi and an Eta version. A Phi board requires a signal inverter before the discriminator to change the positive input into a negative one, whereas an Eta board does not. This is controlled on the board itself by whether a capacitor is placed on the left set of pads (corresponding to an Eta board) or the right (Phi) for each input channel, as shown in the zoomed-in view of the channel in Figure 4.4. This shows the capacitor on the left in the Eta set-up, with the adjacent soldered pads showing where the capacitor would be located for a Phi PCB. There are a total of 12 front-end PCBs required per singlet, eight of which will be set up as Phi boards and four as Eta boards. This is because there are 64 Phi and 16 Eta readout strips within the BIS78 RPC singlet.



(a) **(b)**

Figure 4.3: (a) The layout of a CODEX-β face with two RPC modules including their mechanical support frames, with sides labelled 'Phi' and 'Eta' corresponding to the orthogonal readouts [125]. (b) A schematic drawing of the ATLAS BIS78 RPC singlet [124] showing the different layers of material and the gas gap.

Each RPC module is composed of a triplet of BIS78 RPCs that are stacked directly on top of one another, housed in an aluminium frame referred to as the "CX1 Frame". This mechanical framework encloses the RPCs with access available to readout the different RPC planes on both an Eta and Phi side. A set of 'skins', which are two thin panels the same width and length as the singlets, are put in place to cover the top and bottom of the triplet, alongside five equally spaced crossbars on both sides. A set of shims underneath those crossbars act to secure the RPCs in position within the module and to counteract the pressure from the gas gaps within each RPC in the triplet potentially warping them [125]. The assembled CX1 frame without the skins and shims is shown

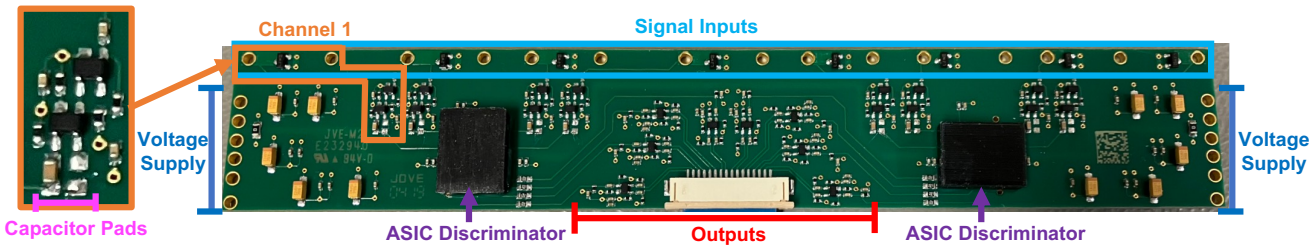


Figure 4.4: The readout PCB used for front-end electronics in a BIS78 RPC singlet in Eta mode, highlighting the signal inputs, outputs, voltage supply contacts and discriminators. On the left is a zoomed-in portion of channel one showing the capacitor layout for the Eta mode.

in Figure 4.5. Additionally, the internal dimensions of the CX1 frame match those of the RPC, except in thickness where the frame is thicker than the triplet such that the triplet is held tight with the addition of the shims and skins, ensuring that the alignments of the RPCs in the triplet relative to one another are maintained. These RPC modules then form the basic detector element that is tiled to create the CODEX-β and CODEX-b geometries.

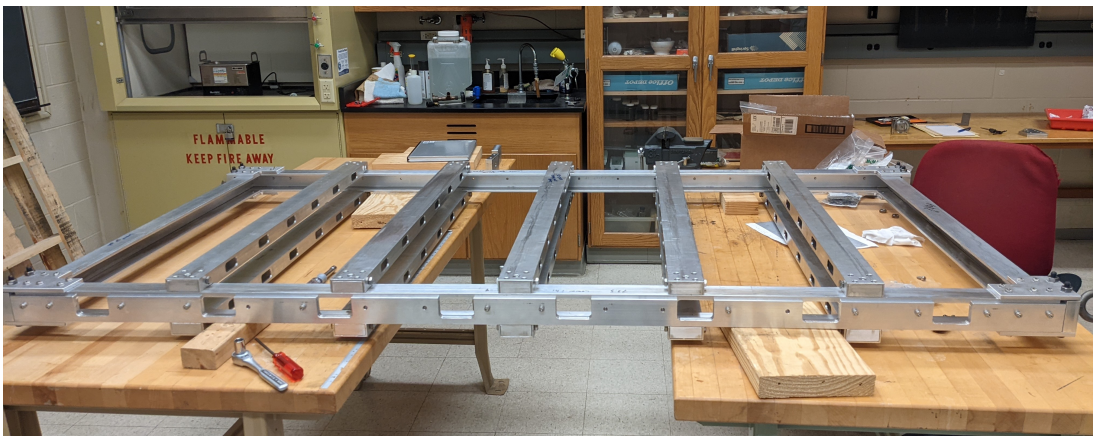


Figure 4.5: The CX1 mechanical support frame for the RPC module, without the skins, shims or singlets installed.

4.3 Detector Geometry

The baseline CODEX-b detector has a cubic geometry (see Figure 4.6) where each face has an area of $10 \times 10\,\mathrm{m}^2$, and each of these faces is instrumented by a sextet of RPCs composed of two layers of a 5×10 tiled arrangement of the RPC modules described in the previous section. These external faces of the cube create a fully hermetic detector with a fiducial volume of $10 \times 10 \times 10\,\mathrm{m}^3$ that should allow for the detection of any significant

LLP decay to charged products within its volume as the tracks from these decays could have a wide range of possible angles. There is also a set of four internal detector stations composed of a single layer of 5×10 tiled RPC modules and hence are effectively a triplet of RPCs. The internal stations are aligned along the x axis (in LHCb's beamline coordinates) and therefore are each parallel to the beamline. Their purpose is to provide good vertex resolution for LLP events, with the reconstruction of a decay vertex to good resolution requiring at least six hits per track as close as possible to the vertex. Some additional requirements have been proposed to improve track performance: a minimum threshold of 600 MeV, due to multiple rescattering of soft tracks; and hit separation of > 2 cm on any singlet due to the finite resolution of hits [122].

The exact geometry has not yet been finalised as there are some ongoing optimisation studies. These studies decomposed the baseline CODEX-b geometry into a set of $2 \times 2\,\mathrm{m}^2$ panels and compared the vertex reconstruction efficiency of a detector geometry with the optimal arrangement of a reduced number of these panels for a few different simulated LLP decays. It was shown that a Higgs to di-dark scalar model, $h \to A'A'$, with $A' \to e^-e^+$ decay, could have a minimal reconstruction efficiency of $\sim$80% for a geometry with half the area coverage. However, for a different LLP decay with an exotic b-hadron decaying to a dark scalar, $b \to sS$ with $S \to e^-e^+$, this reconstruction efficiency could be as low as $\sim$40% [126]. The overall effect of the changing number of panels for each of these cases can be seen in Figure 4.7. This implies that the number of detector elements required could be potentially reduced to save cost but would also introduce an element of bias towards particular LLP models. An additional benefit of this study is that it can perform this modelling in a variety of potential geometry arrangements and physical locations beyond that discussed in Section 4.1. This means that, for CODEX-b, both its geometry and its location have the potential to change from the baseline proposal before the complete installation, whereas the CODEX-β detector design and location are more definitely fixed.

The CODEX-β detector has the same basic cubic geometry as the baseline version of CODEX-b but is 125 times smaller by volume, with each of the external faces having a $2 \times 2\,\mathrm{m}^2$ area coverage, providing a fiducial volume of $2 \times 2 \times 2\,\mathrm{m}^3$ and are triplets rather than sextets. The smaller size also means that there is only a single RPC station within the cube, which is also a triplet. Each of these faces and the internal station are composed of a pair of adjacent RPC modules that are aligned along the Phi side as shown in Figure 4.3a. This arrangement allows for easy access to the Phi and Eta readouts for both RPC triplets in the module. The full CODEX-β geometry including the labelling of the different RPC modules is shown in Figure 4.8a. This shows that the orientation of the RPC modules in each station depends on which face of the cube they are on, and is likely to be identical for CODEX-b for the same reasons given.

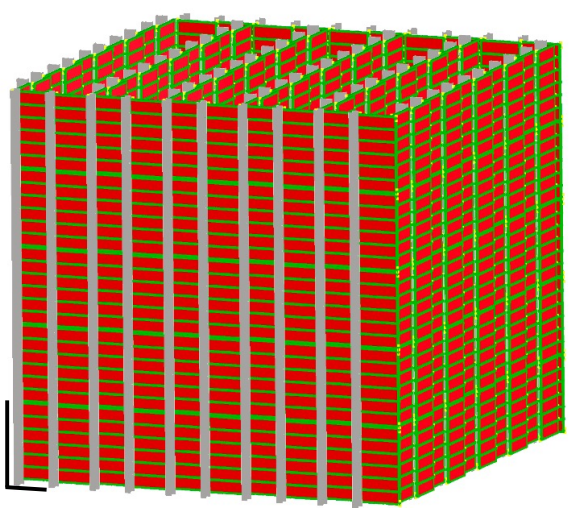
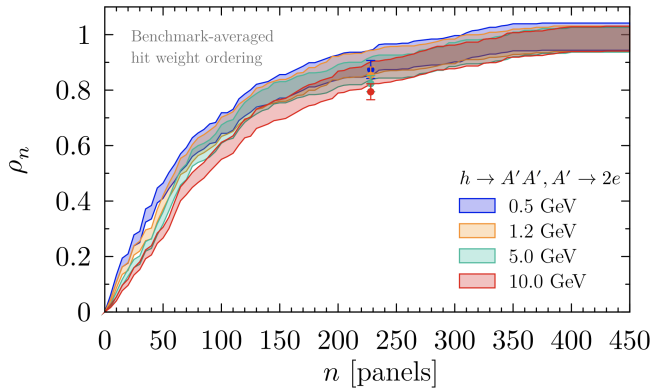
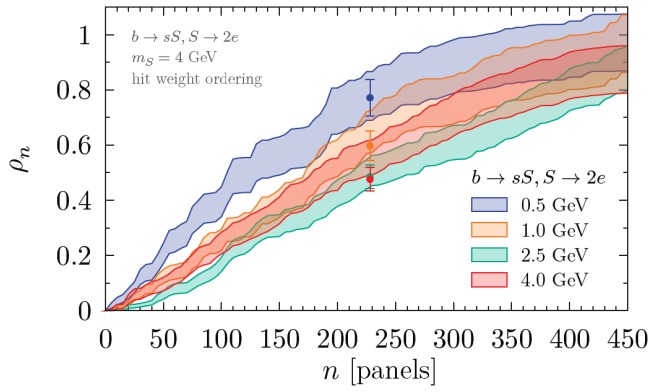


Figure 4.6: The CODEX-b baseline design with outer faces of a sextet of RPCs and four internal RPC triplets. The top face has been removed for clarity. The RPC triplet panels are shown in red, with support frames in green and structural steel supports in dark grey. A single RPC module is highlighted in black in the bottom-left corner [122].



(a)



(b)

Figure 4.7: The 1σ confidence level bands for the relative vertex reconstruction efficiencies for an optimal detector geometry with a different number of panels compared to the baseline geometry. With two potential LLP models: (a) $h \to A'A'$, $A' \to e^-e^+$ and (b) $b \to sS$, $S \to e^-e^+$ [126].

The Sky (S) and Ground (G) modules are oriented such that the Phi(Eta) sides are perpendicular(parallel) to the incoming particles. The Phi side has finer resolution compared to the Eta one, and so is placed such that LLP candidates originating from the LHCb IP will obtain the best tracking. The modules on the Front (F), Back (B), and Centre (C) faces are arranged so the Phi sides are parallel to the vertical axis and both Phi and Eta are perpendicular to the incoming particles. The modules on the Left (L) and Right (R) faces are oriented such that the Phi sides are parallel to the vertical axis and the Eta sides are parallel to the incoming particles.

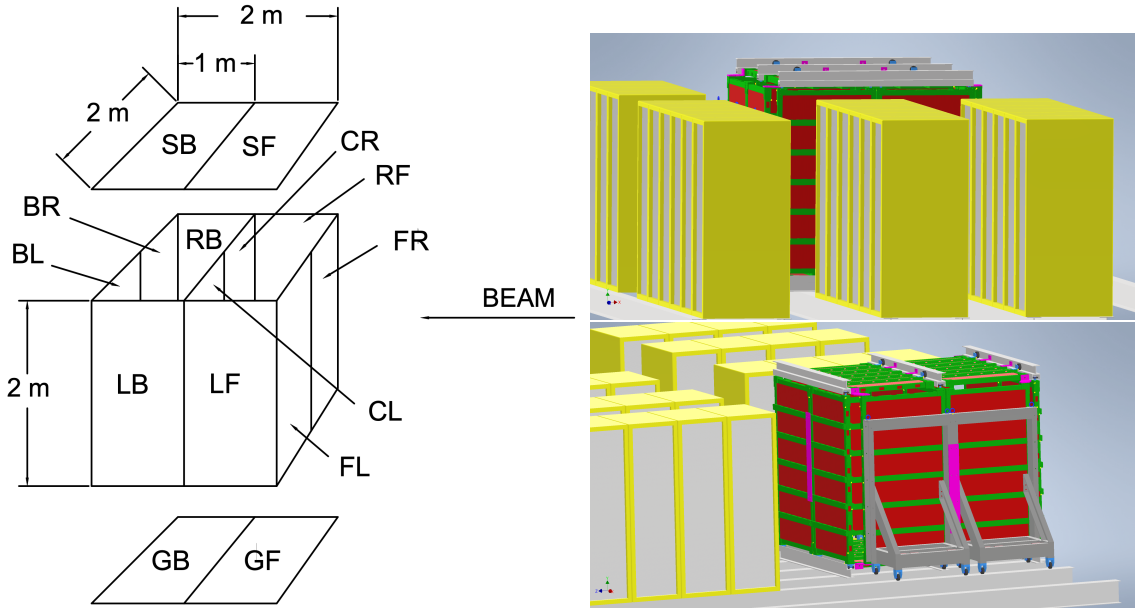


Figure 4.8: (a) Technical drawing of the CODEX-β baseline design with the outer faces and an internal station composed of a triplet of RPCs. Each RPC module is given a two letter code with the first letter being the face (Front, Centre, Back, Left, Right, Ground, or Sky) and the second being the relative position in that face (Left or Right, Front or Back) [125]. (b) A representation of CODEX-β within the LHCb counting barrack, its nominal location [125].

These different modules also require slightly different installation methods and mechanical support due to their locations and orientations. The modules on the F, B, and C faces form the main rigid support structure for CODEX-β within a set of rails in the installation location, and the L face modules are mounted on a trolley that can be moved for access to the C face modules. The G face modules are installed horizontally below the other modules within the set of rails, and the S face modules are also installed in the same way but lifted by a crane to the 2 m height above the other modules [125].

4.4 Background Studies at the Proposed Location

The CODEX-b detector is sensitive to charged particles that pass through its fiducial volume. Due to its displaced, transverse location compared to the LHCb IP and the presence of the UXA concrete wall, it is expected that these charged particles will originate predominantly from LLP candidate events produced at IP8. However, there are a number of possible backgrounds to be considered: vertically travelling charged cosmic rays; neutral SM particles with long lifetimes; and muons.

For the cosmic rays, these should be easily vetoed using the top and bottom horizontal stations (S and G) as a coincidence detector. However, neutral SM particles with long lifetimes, *e.g.* (anti)neutrons and K_L^0's have to be considered more carefully. These have the potential to travel the required distance to the CODEX-b detector and produce tracks in its volume. For the expected integrated luminosity of LHCb's Upgrade II, $300\,\text{fb}^{-1}$, simulations suggest that $\sim 10^{14}$ neutrons and K_L^0's are produced, requiring $\sim$32 nuclear interaction lengths, λ, for full attenuation. The existing UXA concrete wall is conservatively estimated to provide 7λ of shielding, and therefore part of CODEX-b's proposed design is the inclusion of a 25λ ($\sim$4.5 m) lead shield [76] to be placed close to IP8 as seen in Figures 4.1 and 4.9. This would effectively reduce the impact of neutral SM particles produced directly from the IP.

It is also important to consider the background due to muons. This is because they can act as a generator of neutral long-lived SM particles through material interactions with the detector shielding. If this occurs within the lead shield then the produced secondary SM LLPs will likely be stopped in the UXA concrete shielding, but if this occurs in the UXA wall itself then these secondaries can be a potentially significant background for CODEX-b. These are referred to as upstream and downstream 'stopped-parent secondaries' respectively, as shown in Figure 4.9. This background can be reduced by the inclusion of an active veto from a scintillator embedded in the lead shield. This would have to be placed such that it is deep enough to be shielded from the IP and effectively veto the downstream stopped-parent secondaries, while also providing sufficient shielding to reduce the upstream component at the same time. This led to the proposed '$(20{+}5)\lambda$' configuration with 20λ of lead before the shield veto and 5λ afterward which has been tested with simulations [76].

Simulations have been performed, using Pythia [111, 112] to generate the flux of primary particles from the IP and Geant4 [117, 118] to propagate those primary particles through the shielding arrangements, to determine the effect of the primary and secondary backgrounds. From this study using $300\,\text{fb}^{-1}$, given fully in Ref. [76], there are only four particle types that have a non-zero net yield and these are negligible: neutrons ($\lesssim 1$),

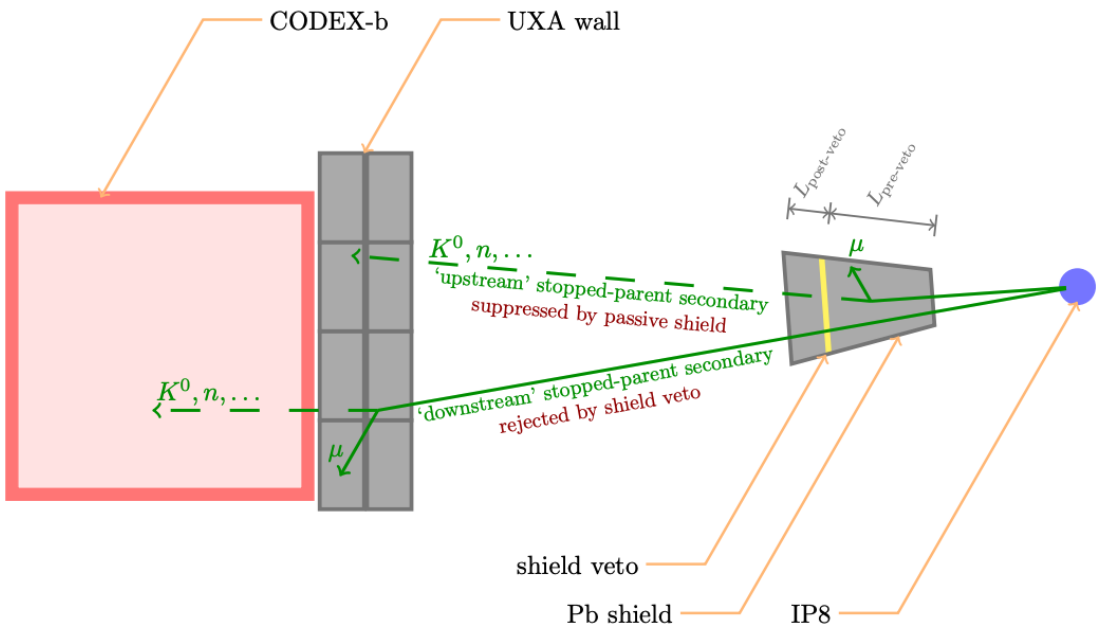


Figure 4.9: Proposed shielding arrangement for CODEX-b with the lead shield containing the active shield veto (yellow) and the concrete UXA wall. The typical topologies for backgrounds from the production of upstream and downstream stopped-parent secondaries are also shown in green [76].

antineutrons ($\ll 1$), K_{L}^0's ($\lesssim 0.1$), and K_{S}^0's ($\ll 1$). An additional in-situ background study has been performed using scintillators and PMTs from the High Rapidity Shower Counters for LHCb (HeRSCheL) experiment in different areas of the nominal CODEX-b location, with and without the LHC beam running and with only the UXA concrete shield. One location gave a hit rate of (167.10 ± 1.43)mHz, which can be directly compared to a value for the same place in simulation giving ∼5–10 Hz [76, 127]. However, the simulation does not yet account for potential interactions with the full LHCb cavern which may decrease the simulated rate. This implies that the expected backgrounds from simulation could be very conservative despite already being negligible, and hence the requirement for a minimal background environment for CODEX-b is likely to be achievable. However, the difference compared to the scintillator hit rate does suggest that further tests should be performed to ensure that the backgrounds are definitively understood. This can be provided by CODEX-β which is planned to be located in the same location as CODEX-b without the additional lead shielding and so during its data-taking run it can observe the effect of potential backgrounds on the same hardware as the full-sized detector.

4.5 Potential Sensitivity to LLP Models

A variety of LLP models have been studied to determine the potential sensitivity of CODEX-b in the nominal location near LHCb. In particular, a set of four main minimal models that are simple dimension 4/5 SM extensions with a single new particle. These can represent the more complex models which give rise to BSM physics. For these minimal models, the possible couplings of a new and neutral state to the SM are restricted by SM symmetries. These four main models are shown alongside potential operators for these models:

1. An abelian hidden sector: $F_{\mu\nu}F'^{\mu\nu}$, $hA'_{\mu}A'^{\mu}$.
2. A dark Higgs portal: $S^2H^{\dagger}H$, $SH^{\dagger}H$.
3. A Heavy Neutral Lepton (HNL) portal: $\tilde{H}\bar{L}N$.
4. An Axion-like Particle (ALP) portal: $\partial^{\mu}a\bar{\psi}\gamma_{\mu}\gamma_5\psi$, $aW_{\mu\nu}\tilde{W}^{\mu\nu}$, $aB_{\mu\nu}\tilde{B}^{\mu\nu}$, $aG_{\mu\nu}\tilde{G}^{\mu\nu}$.

Where $F_{\mu\nu}$ is the field strength operator to vector field field A'; h is the physical SM Higgs boson; H is the SM Higgs doublet; S is a scalar; N is a fermion state; L is the SM lepton doublets; a is a pseudoscalar; ψ is any SM fermion; and $B^{\mu\nu}$, $W^{\mu\nu}$, and $G^{\mu\nu}$, are the field strengths for the SM hypercharge, $SU(2)$, and strong forces respectively [122]. Each of these models have been simulated to estimate the parameter space to which CODEX-b would be sensitive, given a total luminosity of $300\,\mathrm{fb}^{-1}$.

The first model, an abelian hidden sector, introduces a new vector field, A', which is a $U(1)$ gauge boson and is often identified as a 'dark photon' in analogy to the SM photon. This dark photon can then kinetically mix with the SM photon and decay to visible products. The production via an associated dark higgs decay, $h \to A'A'$, and the coupling to the SM photon can be associated with the mass and lifetime of the A'. Therefore the sensitivity range for CODEX-b is given over a lifetime range and for two mass values, seen in Figure 4.10a. This shows that CODEX-b has the possibility to probe a large amount of unexplored parameter space.

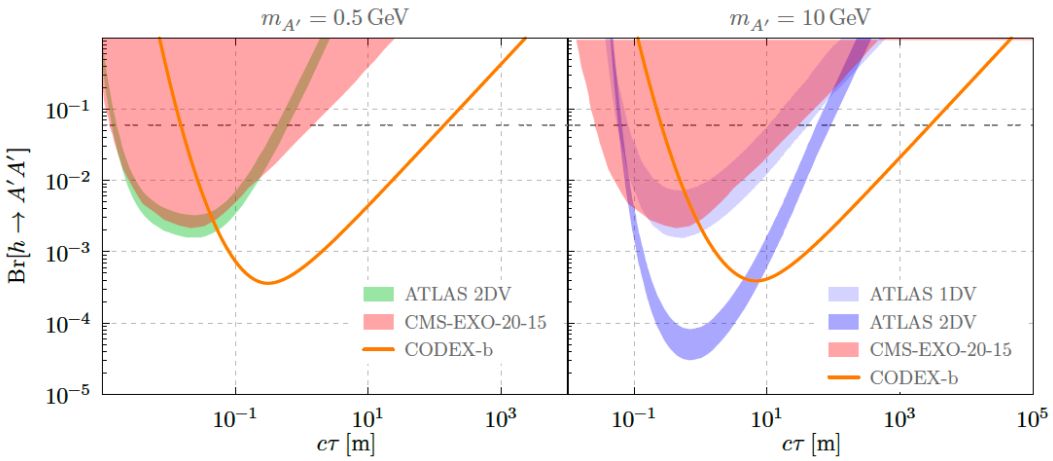
The dark Higgs portal introduces a scalar, S, that couples to a hidden sector and the SM Higgs boson. This leads to a model with three free parameters: the scalar's mass (m_S), its mixing angle with the SM Higgs boson ($\sin\theta$), and the mixed quartic coupling to the dark Higgs (λ_D). For the model considered for CODEX-b an exotic B decay primarily produces the scalar, though it can be pair produced from an exotic Higgs boson as well which is controlled via the mixed quartic coupling [122]. The sensitivity for CODEX-b is given in Figure 4.10b and 4.10c over a range of m_S and $\sin^2\theta$, for a case without any pair

production and a case where 1% of exotic Higgs pair produce. For both cases, CODEX-b would dramatically increase the area of parameter space explored even for the smaller luminosity of $10\,\text{fb}^{-1}$, with the model with pair-production having a much larger possible coverage.

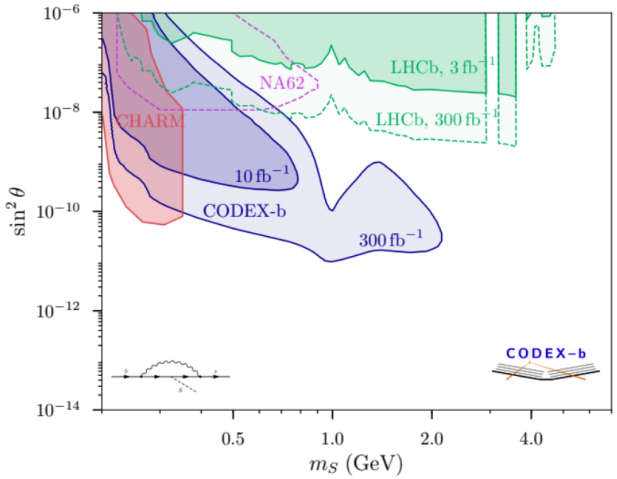
For the third model, an HNL that couples to the SM via a lepton Yukawa portal is introduced. This is mediated via the $\tilde{H}$ operator, and some models imply that there can be a mix between a sterile, non-interacting neutrino and the SM-like active neutrino. This is given by $\nu_\ell = U_{\ell j}\nu_j + U_{\ell N}N$ where N and ν_j are the mass eigenstates and $U_{\ell N}$ is the active-sterile mixing strength and an element of the extended PMNS neutrino mixing matrix. This mixing can be small leading to a long-lived HNL, N, which in the considered models can be assumed to couple predominantly to a single flavour of the active neutrinos [76] *i.e.* μ (see Figure 4.11a) or τ (see Figure 4.11b). Of the parameter space that CODEX-b can probe, a significant portion has already been covered for the μ case, but the τ case is much less explored with less excluded area. For both cases the coverage of CODEX-b is compatible with a number of other proposed and existing detectors and so would serve to reinforce their results.

The final model considered is the ALP portal where ALPs couple to the SM from operators that arise from many BSM models. Axions themselves are particles that arise when the Peccei-Quinn symmetry is broken and are often used to explain the strong CP problem [51], however there are approximate symmetries like the Peccei-Quinn one that can also be broken to produce particles like axions generically. ALPs tend to couple to quarks and/or gluons so can be produced in large quantities at the LHC via hadronisation or emission in a parton shower. This latter case applies to gluon-coupled ALPs, which are particularly notable for transverse LLP experiments. Figures 4.11c and 4.11d show the sensitivity of CODEX-b to a fermion- or gluon-coupled ALP respectively, each showing good additional sensitivity compared to the already excluded regions and a large amount of overlap with existing or proposed detectors.

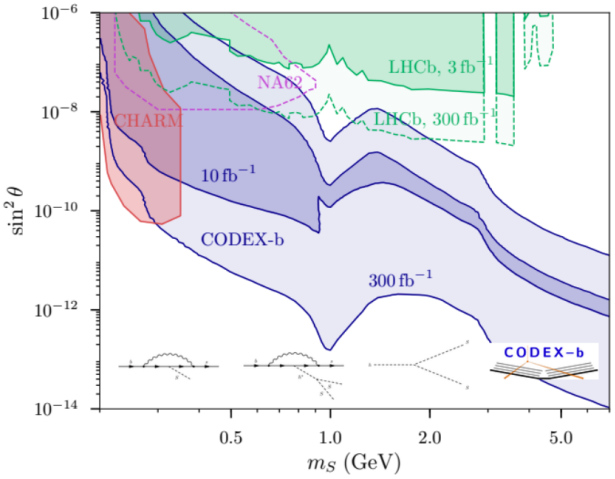
Due to the lack of magnetic field or calorimeter CODEX-b observes straight, charged tracks. As such, for each of these minimal models the experimental signature CODEX-b can observe is appearing charged tracks that can be associated with a displaced vertex that points towards IP8. The boost of these tracks will be associated with the mass of the charged particle, but without PID this will be difficult to determine. For most of the minimal models, the expected signature will be an LLP decay to charged fermions, either as a pair *e.g.* a dilepton or pair of quarks (abelian hidden sector, dark Higgs, ALPs, HNLs) or potentially as a displaced jet of fermions (dark Higgs, HNLs). The HNL model may also provide disappearing track signatures if it decays to neutrinos that CODEX-b would not be sensitive to.



(a)



(b)



(c)

Figure 4.10: The projected sensitivity of CODEX-b for a collected luminosity of 300 fb^{-1} for minimal LLP models: (a) Abelian hidden sector, with expected limits for ATLAS and CMS (b & c) Dark Higgs model with the mixed quartic with the SM Higgs boson set such that $\mathcal{B}(h \to SS)$=0 (0.01) in b (c) [76].

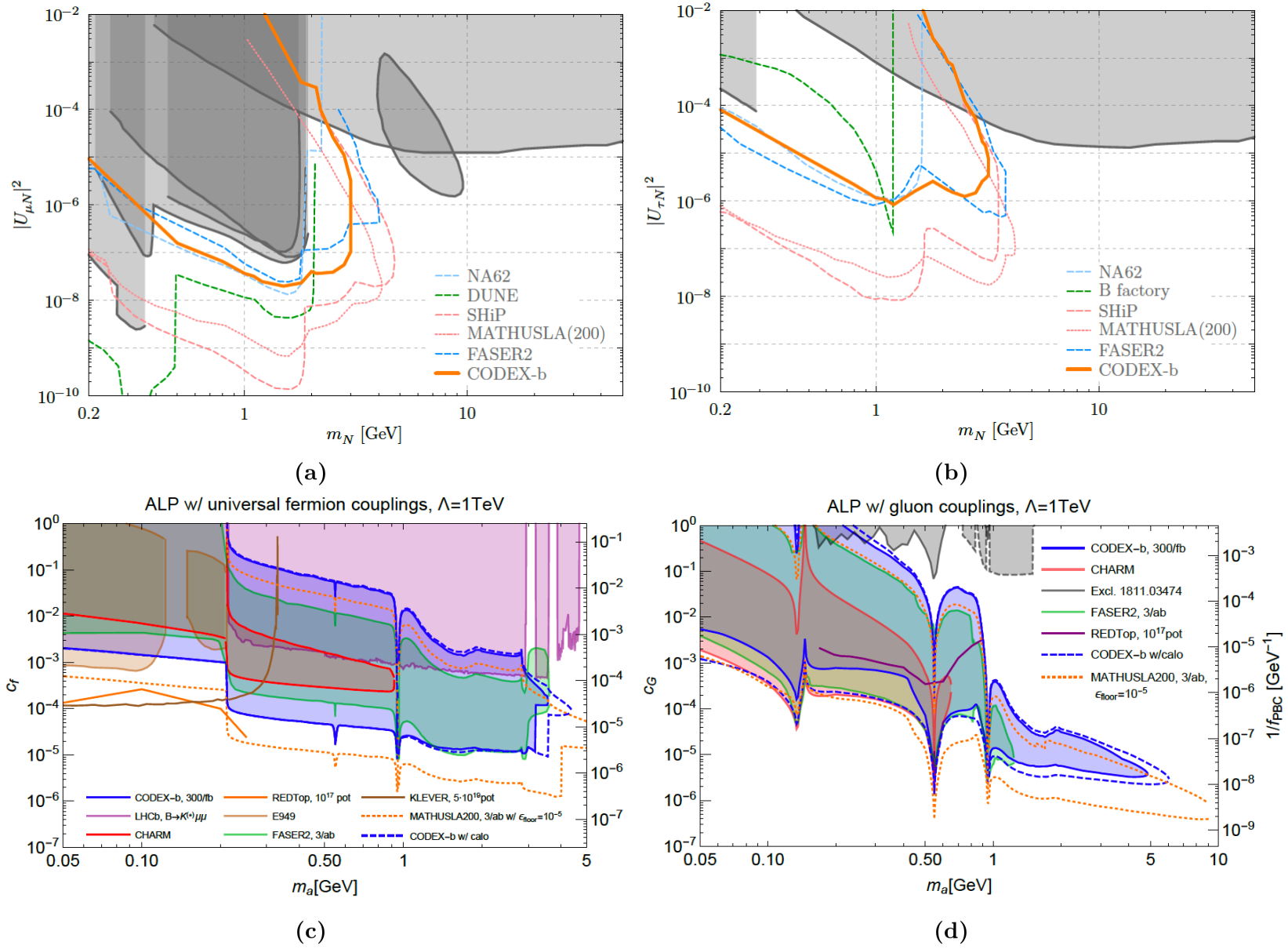


Figure 4.11: The projected sensitivity of CODEX-b for a collected luminosity of 300 fb^{-1} for minimal LLP models: (a & b) A HNL model coupled to the SM μ (a) and τ (b) flavoured neutrinos. (c & d) An ALP model for fermion-coupled (c) or gluon-coupled (d) ALPs [76].

Chapter 5

Performance Studies for the CODEX-b Demonstrator: CODEX-β

5.1 Performing Quality Assurance Tests on Front-End Electronics

In order to assemble the RPC modules that comprise CODEX-β, an individual singlet needs to be constructed first. This includes the front-end electronics that provide the ability to read out hit information for that singlet. As described in Section 4.2, each singlet has 64 and 32 readout strips in the Eta and Phi directions respectively. This requires eight Phi and four Eta boards to be mounted on one of the equivalent Phi or Eta edges, as in Figure 4.3a. Therefore, each RPC module consists of a triplet of RPCs and requires 24 Phi and 12 Eta boards for a total of 36. For the CODEX-β installation there are 14 RPC modules corresponding to a total of $36 \times 14 = 504$ total front-end readout boards (336 Phi and 168 Eta). For CODEX-b 800 modules are required assuming the same design is used[1] corresponding to 28800 readout boards (19200 Phi and 9600 Eta). The exact details of the CODEX-b installation have not been determined, unlike CODEX-β, and so the latter will be the focus of this section. The quality of each of the boards installed must be checked to ensure that the detector is working optimally and that any unexpected behaviour can either be eliminated or controlled before data taking begins. This section will detail the work performed using a bespoke set-up to test the front-end readout boards to be used for CODEX-β.

The proposed LLP experiment: AN Underground Belayed In-Shaft experiment

[1]There are 50 modules per face instrumented with triplets, each external face is a sextet giving $100 \times 6 = 600$ plus the four interior triplets $4 \times 50 = 200$.

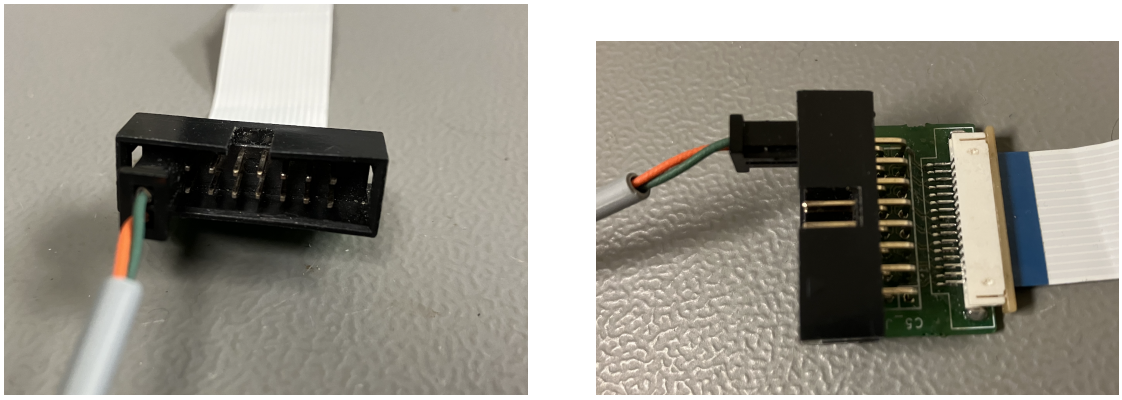
(ANUBIS) [128], is also installing a demonstrator unit but on an earlier timescale than CODEX-β. As ANUBIS will instrument their detector using the same BIS78 RPCs as CODEX-β in triplets with the same dimensions, a set of CODEX-β readout boards were made available to ANUBIS for their prototype detector, to be returned in time for installation of CODEX-β. These boards have already been tested and shown to be of good quality before mounting, and so 84 boards (56 Phi and 28 Eta) have already been examined leaving a total of 420 required to fully instrument CODEX-β. Ideally, there would be an excess of available boards beyond those required in the construction of the detector elements, to ensure that faulty boards can be easily replaced. A conservative excess of $\sim$10% ($\sim$50 boards) is assumed, the exact configuration of these excess boards (Phi or Eta) does not matter as a skilled technician can convert one configuration to another by moving the capacitor highlighted in Figure 4.4 for each of the eight channels on the board. Although this procedure does not take long per board, and assuming any failure rate is distributed uniformly between Phi and Eta boards, the excess should reflect the 2:1 ratio of Phi to Eta boards in each singlet. In actuality, for the quality assurance tests performed there were only 440 available boards to test and therefore this threshold of an excess of 10% could not be currently achieved even if none of the tested boards had shown any defects or erroneous behaviour. Additionally, 69 of those available boards had previously been tested, but the documentation for these tests could not be located and so were retested on the same set-up for consistency.

5.1.1 Equipment and Methodology

A pulse generator[2] was used to inject signals of $\mathcal{O}(5)$ns as input to a multiplexer[3] which allows the pulse to be output to channels 1–8. The outputs of the multiplexer channels were connected to a bespoke 'test-box' with each channel having a corresponding BNC port which couples to the associated channel on the front-end readout board. A ribbon cable was connected to the board's output, and in turn, connected to a pin adaptor with eight pairs of pins corresponding to each board output that can be read by a test probe, see Figure 5.1. A comparison between the injected pulse and the board's output can be made by reading the input signal and the test probe using an oscilloscope[4] with a resolution of $\geq 1\,\text{GHz}$. The full testing set-up is shown in Figure 5.2, and the settings used for each piece of equipment are given in Tables 5.1 to 5.3.

The test-box is a custom built container designed to test the BIS78 readout boards as seen in Figure 5.3. This accommodates a single board and has a set of connection pins that correspond directly to the connections on the boards shown in Figure 4.4, *i.e.* the

[2]TTI TGF4242 arbitrary function generator.
[3]Keysight 34972A LXI Data acquisition/switch unit.
[4]Lecroy WR8104 oscilloscope.



(a) **(b)**

Figure 5.1: (a) A set of sixteen output pins, where each vertical pair corresponds to one of eight output channels, and the test probe reading the output of channel one. (b) A side view of this arrangement, showing the pin adaptor to the ribbon connector.

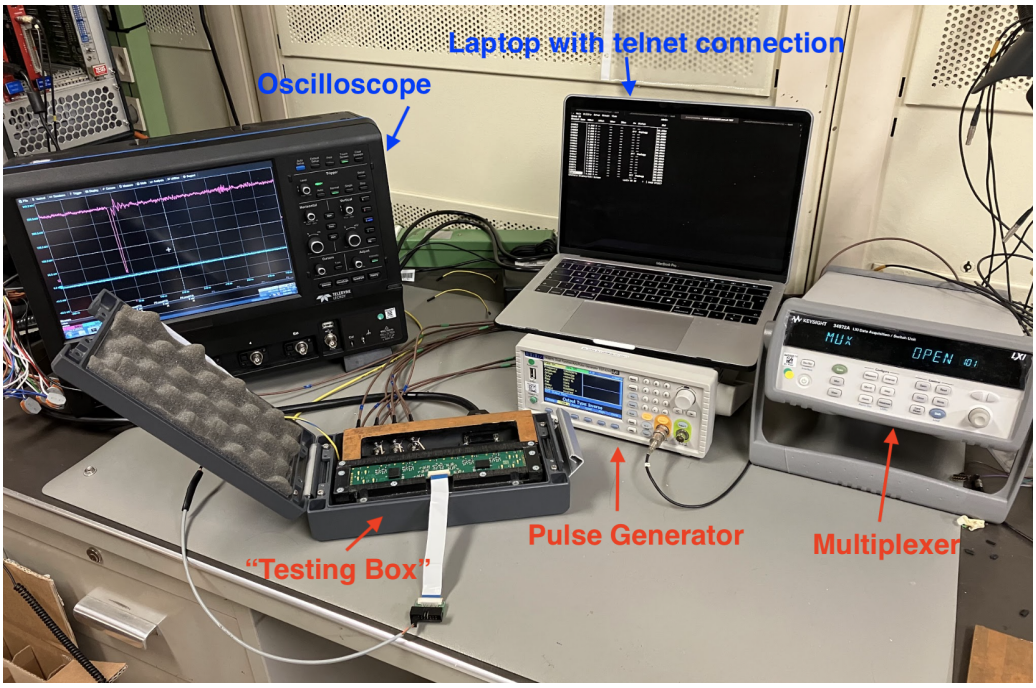


Figure 5.2: The overall BIS78 front-end electronics readout board quality assurance test set-up: A pulse generator injects a signal to a multiplexer (and channel 1 of an oscilloscope) that is connected to a bespoke test-box that houses the readout board, and its output is read in channel 2 of an oscilloscope.

Table 5.1: Settings of the pulse generator.

Setting	Value
Frequency	1 kHz
Amplitude	100 mVpp
Duty Cycle	0.001%
Rise time	2.5 ns
Fall time	2.5 ns
Delay	0 s
Shape	Pulse

Table 5.2: Settings for the voltages supplied to the BIS78 front-end electronics readout board.

Setting	Value /V
Threshold Voltage	1.60
Pull-up Voltage	0.30
Voltage Amplitude	1.30
Discriminator Voltage	2.60
LVDS Voltage	2.50

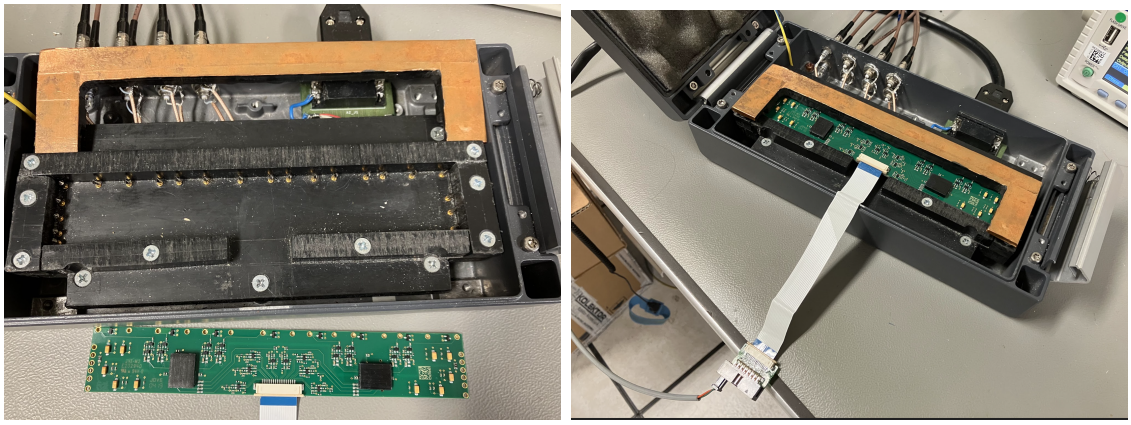
Table 5.3: The oscilloscope settings.

Setting	Value
Input Signal Channel Volts Per Division	20 mV/div
Input Signal Channel Offset	−32 mV
Output Channel Volts Per Division	50 mV/div
Output Channel Offset	−106.5 mV
Timebase	100 ns/div
Timebase Offset	−218 ns
Trigger Threshold	37.22 mV
Trigger Type	Positive, Edge

top pins are the input signal pins, and the side pins are related to the supplied voltages. These voltages are: the supply voltage; the threshold voltage of the discriminators; the pull-up voltage, which helps control the load on the board as a safety device; and the LVDS voltage. Each of these can be individually set and turned on and off via a telnet connection. The signals for the input channels are injected via eight ports at the back of the box. In the full set-up a multiplexer was used to change quickly and easily the input channel in order to refine the work-flow for a large set of boards to be tested.

The testing procedure itself is simple, robust and fast, which is useful for large-scale testing. The general procedure was the injection of a signal into one input channel, *e.g.* channel 1, and then measuring each output channel. There are four possible outcomes:

1. a strong signal peak is seen in the corresponding output channel with no signal in any other channels, which is the expected behaviour;
2. no signal is seen in the corresponding output channel, a 'dead' channel;
3. independent of whether there is signal in the corresponding output channel a signal is observed in a different unexpected channel typically adjacent to the injection channel, known as 'cross-talk'; and
4. unusual behaviour is observed in the output channels *e.g.* a spiky or truncated signal.



(a) (b)

Figure 5.3: (a) Bespoke testing box designed to provide the eight channel inputs with the top pins, and the voltages for the board on the left and right pins. A (copper-coloured) mechanical brace presses the board onto the pins to ensure a good connection when it is closed. (b) The box with the board and brace in place.

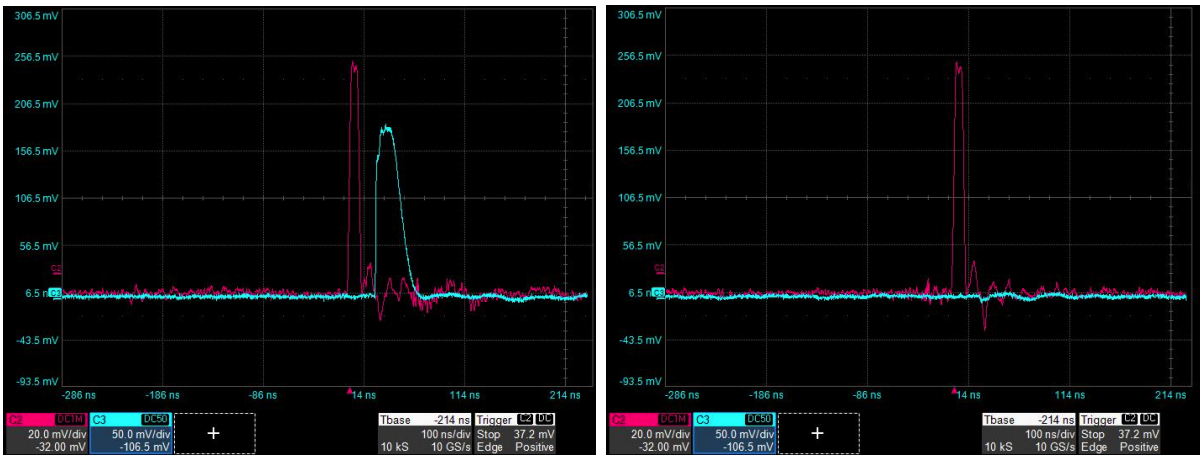
This outcome was then recorded within a database together with: an assigned testing ID consisting of an integer that was incremented after each test and a letter referring to the board's configuration *e.g.* P for Phi; a unique ID that was screen printed as a QR code on the board itself and can be read by a QR code reader; a batch ID which was either 4918 or 0419, though the two appeared to have identical circuitry; and any additional notes. Some boards had previously been tested and so these already had an assigned testing ID that was maintained during this updated quality assurance test.

5.1.2 Results

Figure 5.4a shows the injected signal peak that was applied to channel 1 and the associated channel 1 output. Overall, the characteristic shape was mostly maintained but with the output signal approximately twice the width of the input of $\lesssim 50\,\text{ns}$, with a delay between the peaks of $\sim$35 ns and $\sim$1.5 times larger amplitude. This output signal should be sufficient for CODEX-β. However, the exact nature of the output signal was not specifically studied for this quality assurance test and therefore a binary pass/fail system based on the presence of an observed signal in the expected channel (*i.e.* no cross-talk) was used. It should be noted that the Phi boards require a positive input signal and Eta boards a negative one, varied in the pulse generator settings. The board's output is positive in either case, and so the two signals for an Eta board are shown on the oscilloscope to have opposite polarity but the input and output signal shapes are as aligned as in the Phi case.

The signal of cross-talk tends to have the same characteristic as that of the actual output channel, as illustrated in Figure 5.4a, but appears in another channel *e.g.* channel 3 if the input signal is injected in channel 1.

If a 'dead channel' is encountered then no output signal is observed, which is identical to the situation where a different output channel number to the injected channel number is observed, as in Figure 5.4b. Occasionally, channels adjacent to the expected output channel show a minor peak of $\sim$10 mV, this is not particularly notable compared to the scale of some background noise but is systematically present. These small peaks are believed to be due to 'fast transients', where there is some arcing in an inductive-capacitive circuit due to air gaps that briefly interrupt the current in $\mathcal{O}(1)$ ns pulses [129]. Although this is a known effect that can be accounted for, it should not have a significant effect for CODEX-β.



(a) **(b)**

Figure 5.4: Oscilloscope readings during a quality assurance test of the readout board "2P", a Phi configuration board, with the signal injected (pink) in channel 1 of the board and the output observed (blue). (a) Observing the output of channel 1, giving a positive response. (b) Observing the output of channel 2, showing no response which is identical to the response of a 'dead channel'.

Finally, there are occasions where the output signal shows some unusual and unexpected behaviour, the origin of which is not always readily understood. Figure 5.5 shows an example of this where the peak is much sharper than that of the input peak, which implies there is perhaps a large resistance difference somewhere within this channel's electronics that affects the peak part way through the pulse or a poor connection in the test probe. However, it does appear systematically in particular channels on particular boards and not in others which implies that it is due to a property of the board and not the test set-up itself.

It should also be noted that after all the boards were tested at least once, those

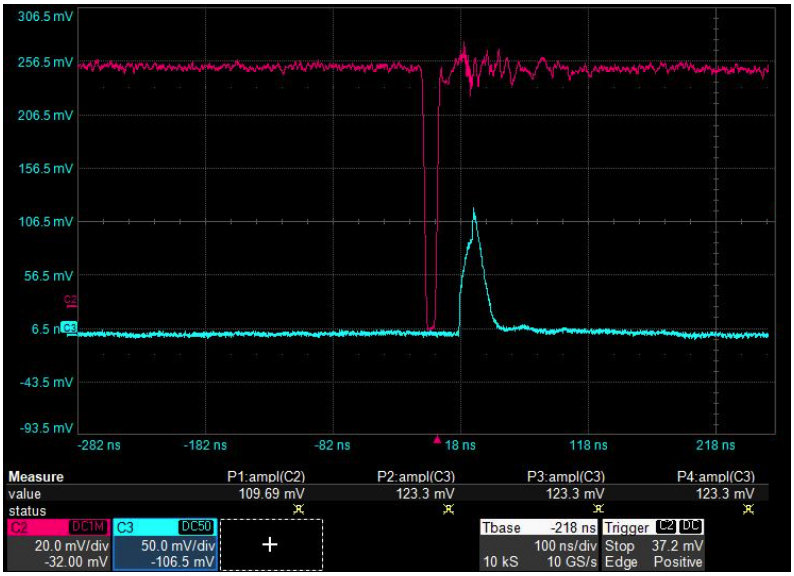


Figure 5.5: Oscilloscope reading during a quality assurance test of the readout board "164E", an Eta configuration board, with the signal injected (pink) and the output observed (blue) in channel 8. Showing an erroneous peak shape in the output signal.

that failed were subsequently retested and of these several were found to not have any observed faults. This is likely to be due to either poor contact made between the pins in the test box and the board, or the influence of loose connections in the output test probe or ribbon cable. Hence, when repeated under closer scrutiny, some of these systematic errors were removed.

Overall, there were a total of 18 boards that showed some issue (six Eta and 12 Phi), which could be separated into three categories[5]: all Channels were unresponsive (0E, 64E, 235E, 384E); one Channel was unresponsive (4P, 51P, 61P, 85P, 86P, 175P, 233E, 266P, 300P, 335P, 1000P); and one channel showed a pronounced sharp, spiky peak compared to the normal smooth peak, as in Figure 5.5 (132P, 164E, 357P). Therefore there are 143 Eta boards that are ready for installation and 279 Phi boards, which means that there is an excess of three Eta boards available and a deficit of one Phi board. It is apparent that there will be a minimal number of spare boards, but if a single Eta board is converted to a Phi board then there are enough working boards to instrument CODEX-β if none fail during installation, with two spare boards in the Eta configuration.

[5]The full testing database is available upon request.

5.2 Reconstruction Tolerances of Unknown Misalignments

As described in Chapter 4, the intention is for the CODEX-β detector to be installed as a prototype to explore potential issues that may arise during the installation and operation of the larger scale CODEX-b detector. One such issue is that of the potential backgrounds that need to be controlled or eliminated by the lead shield introduced in Section 4.4. Other fundamental aspects to be studied include the development of tracking and reconstruction algorithms to convert RPC hits to an LLP candidate event.

To ensure that tracking and reconstruction algorithms accurately depict the underlying LLP candidate event the alignment of the detector elements should be known precisely. If there were imperfections in the known alignment of an RPC, the hits in that RPC would be incorrectly reconstructed and affect the overall nature of the measured tracks. A schematic example is shown in Figure 5.6. This compares the measured hits in an aligned RPC singlet to an angular misalignment of 25°. This is a rather dramatic case but illustrates how the reconstruction of the LLP candidate vertex could be significantly affected. In contrast, if the RPC singlet were simply moved towards the LLP decay then the detected hits would be closer together and imply that the opening angle of the LLP decay is smaller, or vice versa for moving it away.

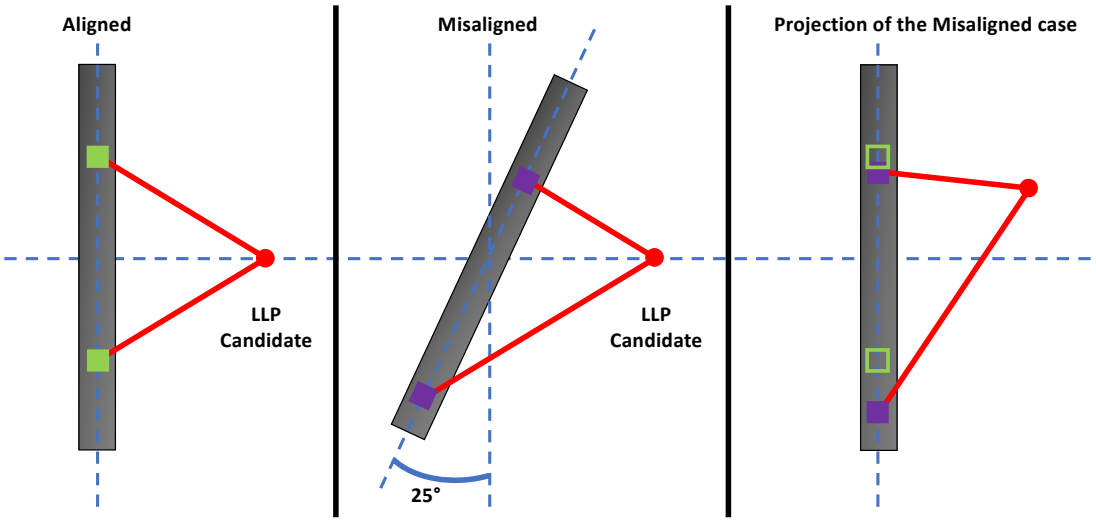


Figure 5.6: Impact of an angular misalignment of an RPC singlet on reconstruction, given a 2-body decay from an LLP candidate. (Left) The fully aligned case with the green filled squares representing the detected hits. (Middle) The misaligned case with the purple filled squares representing the hits in this case. (Right) The misaligned case projected to the expected aligned RPC position with the green outlined squared representing the hits in the aligned case.

Although it appears that certain properties such as the opening angle in the angular misalignment could be preserved in these cases, this may not be universally true as the reconstruction makes use of multiple RPC planes and so a single misaligned singlet may have a reduced impact on the overall reconstruction. Therefore, a study was performed in order to determine how large an effect would be expected for a range of different unknown misalignments. Here, 'unknown' refers to the projection of the misaligned case to the aligned case, which would occur if the reconstruction software believed the RPC singlet to be in a different place than it actually was.

In order to quantify how the overall performance of the detector would be affected by an unknown misalignment, a standard must be set for the purposes of comparison. The CODEX-β and CODEX-b detectors are not currently planned to include calorimetry or a magnetic field, and thus quantities such as the invariant mass and particle identity would be difficult to determine. The most fundamental piece of information that CODEX-β provides is tracking information in the form of hits within the RPC singlets from the passage of charged particles. Therefore, an easily accessible quantity is that of the opening angle for a decay to a pair of charged particles, which is a common LLP signature *e.g.* $A' \rightarrow e^- e^+$.

The basic procedure performed, therefore, was to simulate the CODEX-β geometry using the GEANT4 toolkit, then model the detector response from a di-electron event with a known, true opening angle, θ_t, (Section 5.2.1) and convert the produced hits to a set of tracks. From these tracks the reconstructed opening angle for the fully aligned detector, θ_a, can be determined (Section 5.2.2) through the standard equation of determining the angles between two vectors, θ_{OA},

$$\theta_{OA} = cos^{-1}\left(\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\,|\vec{b}|}\right), \tag{5.2.1}$$

where $\vec{a}$ and $\vec{b}$ are the vectors of the two reconstructed tracks in a di-electron event. The produced hits can be projected as if they were detected by a misaligned RPC singlet or set of singlets which are misaligned in a known way, and then these hits can be converted to tracks and the reconstructed opening angle for a misaligned detector, θ_{mis}, determined (Section 5.2.3) using Equation 5.2.1 in the same way as θ_a.

5.2.1 Simulated Events for a Given Detector Geometry

The full detector does not need to be simulated to study the effect of a singlet misalignment. Instead, the CODEX-β geometry can be broken down to one of its most

fundamental components: the RPC module described in Section 4.3, a stacked triplet of RPCs kept in place with the CX1 mechanical frame. However, a detailed simulation of the detector response of the BIS78 RPC was not available and so a simplifying assumption was made: the RPC singlet as seen in Figure 4.3b is replaced by a 2 cm thick silicon plane of uniform density. This thickness was selected to approximate the expected energy deposition for a charged particle passing through an RPC.

This "thick Si RPC" approximation does lead to several issues when compared to reality. Firstly, the standard RPC singlet is ∼0.7 cm thick and three are stacked on top of one another to form the triplet. If the distance between the centres of adjacent planes is t and the misalignment distance is d, the angle between the same 'pixel' on each adjacent RPC is $\theta = \tan^{-1}\frac{d}{t}$. Therefore a thicker RPC implies that the resultant angular displacement is lower and leads to an underestimation of the potential effect of misalignments. Therefore, to compensate for this underestimation the misalignment distances considered are overestimated and large compared to the mechanical assembly precision. Another consequence of the approximation is that the read out is not performed in the same way, instead of a set of orthogonal strips in two planes across a gas gap there is only a single plane that is solid silicon. This means this reproduction did not perform the most fundamental part of track reconstruction: converting the detector read out to a set of hits. Although the true coordinates of hits within the Si planes instead were used for the hit coordinates, these were digitised using 3D histograms that have bins of size $1 \times 1 \times 10\,\text{mm}^3$ in order to approximate the expected $1\,\text{mm}^2$ resolution of the RPC's strips and the RPC thickness.

The frame that the module sits within was modelled using a set of technical drawings [130]. The frame itself is a rigid aluminium structure with inner dimensions that match that of the RPC singlet so that their alignment is maintained as well as possible. To brace the outer rigid structure there is a set of six evenly spaced, hollow crossbars with four slots punched through each to allow easier cable and gas management. Unlike the module described in Section 4.3 the skins and shims are not included in this simulation as it was based on an earlier prototype design for the CX1 frame focusing on the main mechanical support, and had not yet included these details. These thin layers of aluminium could potentially introduce some material interactions but this effect is likely to be minimal. There was not a notable impact from the inclusion of the crossbars which would introduce more material than the skins themselves, and therefore the skins and shims can be safely disregarded.

The resultant simulated model of the RPC module is then saved in a GDML file to allow it to be efficiently reused in the same state during the study. As GDML is application independent this does not have to be performed within GEANT4. The parameters for the

module used are summarised in Table 5.4 and a representation of it is shown in Figure 5.7 using GEANT4's visualisation.

Table 5.4: The dimensions for different components of the CODEX-β RPC module simulated in GEANT4.

Component	**Dimensions /cm**
RPC planes	$110 \times 200 \times 2$
Aluminium Frame (inner)	$110 \times 200 \times 2.3$
Aluminium Frame (outer)	$120 \times 210 \times 2.3$
Aluminium Crossbars (inner)	$120 \times 5.4 \times 2.4$
Aluminium Crossbars (outer)	$120 \times 6 \times 3$
Crossbar Slots	15×1

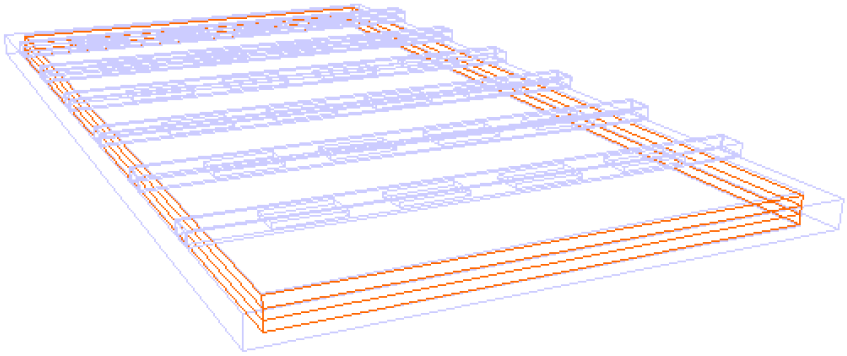


Figure 5.7: Representation of the CODEX-β RPC module created in GEANT4, with the RPC planes in red, the aluminium crossbars and frame in light purple.

As mentioned in Section 5.2, the parameter of interest examined was the opening angle of an LLP decaying to two charged tracks. The exact nature of this decay was not simulated because the goal of this study was to examine the impact of misaligned detector elements and not sensitivity to particular exotic models. Therefore, instead of generating exotic events via an external MC software that are then passed to GEANT4, the 'Particle Gun' tool in GEANT4 is used to fire a specified particle with a given momentum from a starting point. As a large proportion of detected charged particle events in CODEX-b are expected to be charged leptons, and to simplify the reconstruction (*i.e.* not a multi-body decay), a di-electron pair was selected. This was modelled using two GEANT4 particle guns that share the same starting point, with their angles set by the user to give the true opening angle, θ_t. One gun was assigned to fire electrons and the other positrons, both with identical set momenta. These particles were then propagated through the RPC module, and the hit information of the energy depositions into the Si plane was stored as

truth information if the energy deposit was larger than 10 MeV. This loose cut was used to reduce noise and represent the detector threshold of the RPC.

For this study the decay vertex of the LLP event was set below the RPC module, 10 cm from the first singlet with the momenta of each lepton in the di-electron event set to 10 GeV travelling in the positive z direction with an opening angle of $\theta_t = 0.4$ rad. This corresponds to a momentum vector of $(\pm 1.4048, \pm 1.4048, 9.80067)$ GeV where the positive (negative) case is for the electron (positron). A total of 1200 such events were generated, where the differences arise from the randomised responses of the simulated detector, providing some statistical variation that approximates the detector resolution and efficiency of detection.

5.2.2 Track Reconstruction with the 3D Hough Transform

To reconstruct potential events the CODEX-b detector needs to use the RPC hits to build sets of tracks, which can be analysed to determine the decay topology. For CODEX-b there will be no magnetic field, which simplifies the track reconstruction to straight lines. However, there are several complications to developing a track-building algorithm for CODEX-β and CODEX-b.

The main complication is the unknown decay topologies for exotic LLP events. This means there may be an arbitrary number of tracks from a single event, so the algorithm would need to handle N-tracks generically. Additionally, the hermetic nature of the detector means there is full angular coverage of decays. There would be a preferred direction for most LLPs as they would originate predominantly from IP8, but there are also potential cosmic ray events that would travel vertically down. Also, the exact rotational orientation of the tracks around the beam direction would be isotropic. Therefore the algorithm developed would need to handle an arbitrary alignment of tracks in 3D space to ensure uniformly efficient and accurate reconstruction. Once the tracks are built then analysis of an event can be performed, including the vetoing of vertical tracks that originate from cosmic rays.

The study reported here uses artificial di-lepton events, so the difficulty of track reconstruction has been simplified further as the expected tracks are defined by the user. However, a number of potential track construction algorithms were evaluated for the full CODEX-β reconstruction as well as this smaller-scale project, including a bespoke algorithm designed by a master's student [131]. The selected track reconstruction algorithm was an implementation of an iterative 3D Hough transform given in Ref. [132] that was originally developed for image processing in computer science. This takes a set of data that defines a 3D point cloud and outputs a set of reconstructed lines, which meets both

the N-track and isotropy requirements.

The Hough Transform was introduced as a method to detect lines in 2D images [133]. The general theory is to use the parameterisation of a line,

$$\rho = x\cos\theta + y\sin\theta, \tag{5.2.2}$$

where each point (x_i, y_i) corresponds to a sinusoidal curve in (ρ, θ) space. Therefore the points which all lie on the same line will have (ρ, θ) curves that cross at a particular (ρ, θ) coordinate, allowing multiple possible lines to be discerned without a biased starting point. This has been expanded to more general 2D shape detection [134], and even to 3D [132]. In the 3D case instead of the polar equation of a line, the vector form is considered,

$$\vec{r}(t) = \vec{a} + t\vec{b}, \tag{5.2.3}$$

where t is a scalar parameter, $\vec{a}$ is a point on the line and $\vec{b}$ is the direction of the line which can be parameterised in spherical polar coordinates with $\rho = 1$ as:

$$\vec{b} = \begin{pmatrix} \cos\phi\cos\theta \\ \sin\phi\cos\theta \\ \sin\theta \end{pmatrix}, \tag{5.2.4}$$

for $\theta \in [0, \frac{\pi}{2}]$ and $\phi \in [-\pi, \pi]$ to ensure $\vec{b}$ is unique. This uniqueness also requires that $b_y \geq 0$ if $b_z = 0$ and $b_x = 1$ if $b_y = b_z = 0$ in order to remove anti-parallel vector pairs. In the implementation in Ref. [132] the parameter space of $\vec{b}$ is divided into icosahedral cells, rather than a grid as in the 2D Hough Transform, to allow for efficient computation of the 3D transform. The line parameters are then determined iteratively by comparing the points in a given 3D point cloud, if these are within an icosahedral cell of the current iteration's line parameterisation then they determine the 'optimal' line by performing a least squares fit of those points. If no points are fitted then the next iteration is performed until an optimal line has been fitted, and the associated points are removed for the next cycle. This process is repeated until all points have either been associated with a line or classified as noise. The exact implementation of this algorithm is given in greater detail in Ref. [132].

For the misalignment analysis, the point cloud given to the algorithm was a digitised version of the true hit coordinates obtained from the simulation of the di-electron event's detector response. As described in Section 5.2.1 these were digitised using 3D histograms with bin sizes set to represent the expected resolution of the RPCs during data-taking. The (x, y, z) values of the bin centres of any filled bins were stored in a text file which was accessed by the iterative 3D Hough transform algorithm to obtain a set of tracks,

each of which was given a score to represent how well that track fit the point cloud. The artificial di-electron event should only have two reconstructed tracks, therefore if more than two tracks were reconstructed the two highest-scoring tracks were selected[6]. These tracks were then used to determine θ_a with Equation 5.2.1.

5.2.3 Simulating Misalignments

Two different types of misalignments are possible during the installation of the RPC singlets: translation and rotation. For the former, referred to as Translational Misalignments (TMs), the exact location of the RPC singlet compared to the expected location is shifted in (x, y, z) coordinates, while remaining in the expected orientation. For the latter, referred to as Rotational Misalignments (RMs), the location of the RPC singlet was correct, but the orientation has been rotated around a set of angles (α, β, γ), where these represent the angle of rotation around the x, y, or z axis respectively. It is also possible for a combination of these two types of misalignment to occur. These affect singlets individually, though multiple singlets within an RPC module may be misaligned.

These misalignments could be introduced during installation, where the mechanical precision may not be sufficient over the large scale of the RPC frame to ensure the singlets are installed in their documented locations. Alternatively, during the transport of the RPC frames to the detector location or during the data-taking, an incident may occur that causes the RPCs to become misaligned.

Generally, the impact and scale of these misalignments would be minimal due to the numerous checks in place, relatively high engineering precision, and robust detector design. For example, the detector design includes easily referenced survey points on the detector frame and superstructure that can be measured with reference to CERN's survey of the LHCb cavern and the installation location to provide very accurate information about the detector's location. It is also possible to measure the positions of the RPCs within the frame, in reference to these survey points, after the singlets have been installed into the frame which should minimise the possibility of unknown misalignments. Additionally, once the RPC module has been assembled a cosmic ray study can be performed which takes advantage of the vertical, straight tracks of cosmic ray muons to identify any misalignment. These studies can also identify dead channels within the RPC, time resolution and cluster size [135], and so provide a vast amount of information on each individual singlet and module. This is not able to be performed as well in-situ, unlike the survey, due to the low background environment of the nominal CODEX-b location which

[6]Although, this was only required in a minority of cases where a third, low-scoring track was reconstructed that was associated with only a few points.

leads to a low rate of cosmic rays. If any misalignment is observed during these studies then it can be quantified and corrected for within the reconstruction so that the effect is minimised.

From a mechanical perspective, significant misalignments are unlikely to arise. The frame in which the triplet is situated is designed to permit little movement of the singlets, both by tight tolerances between the edge of the RPCs and the frame and by the inclusion of cross-braces, skins and shims that act to 'clamp' the RPCs in position. Therefore, any translation shifts are likely to be due to the size of the frame itself being different than expected from the specifications. This difference would be on the order of the manufacturing precision and the resolution of the survey which is estimated to be $\leq \mathcal{O}(1)$ mm in both cases.

However, an additional complication to introducing misalignments to the simulated RPC module is the use of the 'thick Si RPC' approximation detailed in Section 5.2.1. Therefore, the introduced misalignments are large to provide conservative estimates that account for the underestimation of angular displacement from the thicker RPC singlets. For the TMs these are much larger than the expected precision limit with misalignments of 5 mm and 20 mm considered.

A RM could be introduced due to the unknown presence of material between the RPC layers that would cause a singlet to not lie flat on top of the adjacent singlet. This additional material between layers may be introduced during installation from debris possibly related to the installation itself. To approximate the size of RM that may pass unnoticed, an extreme case is considered where the angular shift is such that a singlet is diagonal along its entire short(long) side which is 1100 mm(2000 mm) long, creating a vertical spacing opposite the pivot point between that singlet and the one below it. A spacing of 1 mm is likely to be observed during the survey which would correspond to an angle of $0.05°(0.03°)$. If a rotation of a singlet in the $x-y$ plane around the middle of the plane is instead considered, the misalignment would be made obvious by the corners of the singlet no longer overlapping with those of the adjacent singlet. The distance between these corners, l, would likely be observed if it was $\sim$1 mm, which would correspond to a rotation of $0.05°$. Therefore, the RMs introduced were $1°$ and $3°$. As with the translational case these are larger than the estimated possible misalignments for the same reasons, especially as the approximations themselves are relatively conservative.

These misalignments are considered within the framework developed for this study using the same procedure that creates a 3D histogram to digitise the recorded hits from the detector simulation described in Section 5.2.2. Before passing the bin centres to the iterative 3D Hough transform algorithm, this histogram was processed to project the hits to a misaligned case via transformations. For TMs these are performed as $(x + \delta x, y +$

$\delta y, z + \delta z)$, where δx represents the TM in the x direction, with the new coordinates saved to the histogram. The RMs instead make use of a rotation matrix created using ROOT's `TRotation` class to transform the given position, where the angular rotation about each of the (x, y, z) planes is specified. The resultant (x, y, z) position is then stored in the histogram which has the same form as the original but with a set of misalignments performed. Therefore, the bin centres can be saved into a text file that the 3D Hough transform algorithm can read and in the same way as in Section 5.2.2 a set of tracks are determined. Equation 5.2.1 provides the associated opening angle reconstructed from the misaligned system, θ_{mis}.

This method of introducing the misalignments programmatically rather than generating a set of different geometries with the same considered misalignments directly introduced, has allowed greater versatility. By the nature of this study there are a large number of possibilities that can be examined. Therefore, for ease of reference, each singlet plane is identified by a number that represents the order in which they are encountered in the event where particles are travelling from below the triplet in the positive z direction; the bottom singlet is plane 0, the middle is plane 1, and the top is plane 2.

For an RPC module there are seven possible misalignments: each of the singlets is misaligned individually; each pair of singlets are misaligned; and all three singlets are misaligned. Three possible TMs $[0, 5, 20]$ mm and three RMs $[0, 1, 3]^\circ$ were considered, each of which can be performed in three orthogonal directions (either individually, in pairs, or all directions, giving another seven possibilities). This gives a total of 5103 misaligned detector geometries that would have to be created. This does not even account for if the various singlets are misaligned differently or for negative misalignments. In this programmatic way the large combinatorics are more easily handled and allow greater adaptability. Another benefit of using the same event for aligned and misaligned cases is that any observed effects in the reconstructed opening angle are entirely due to the introduced misalignment.

To reduce the number of cases considered, only the misalignment of an individual singlet or the entire triplet at once was performed. For the latter, each of the singlets was misaligned by the same amount to simulate the case where the RPC module itself was misaligned during installation. Also, as the singlets are stacked on top of each other a translation in the z direction is never possible for the middle singlet of the triplet without also causing another singlet to move. Likewise, the bottom(top) triplet cannot be moved in the positive(negative) z direction. The programmatic method used does not account for these unphysical overlapping situations, so these are removed manually.

Table 5.5 summarises the different misalignments considered and an associated 'translational/rotational ID' that is used for reference.

Table 5.5: The possible misalignments considered with an associated reference 'ID'.

Translational/ Rotational ID	**Translational misalignment** /mm	**Rotational misalignment /rad**
0	(0, 0, 0)	(0.0, 0.0, 0.0)
1	(0, 0, 5)	(0.0, 0.0, 0.0175)
2	(0, 0, 20)	(0.0, 0.0, 0.0524)
3	(0, 5, 0)	(0.0, 0.0175, 0.0)
4	(0, 5, 5)	(0.0, 0.0175, 0.0175)
5	(0, 5, 20)	(0.0, 0.0175, 0.0524)
6	(0, 20, 0)	(0.0, 0.0524, 0.0)
7	(0, 20, 5)	(0.0, 0.0524, 0.0175)
8	(0, 20, 20)	(0.0, 0.0524, 0.0524)
9	(5, 0, 0)	(0.0175, 0.0, 0.0)
10	(5, 0, 5)	(0.0175, 0.0, 0.0175)
11	(5, 0, 20)	(0.0175, 0.0, 0.0524)
12	(5, 5, 0)	(0.0175, 0.0175, 0.0)
13	(5, 5, 5)	(0.0175, 0.0175, 0.0175)
14	(5, 5, 20)	(0.0175, 0.0175, 0.0524)
15	(5, 20, 0)	(0.0175, 0.0524, 0.0)
16	(5, 20, 5)	(0.0175, 0.0524, 0.0175)
17	(5, 20, 20)	(0.0175, 0.0524, 0.0524)
18	(20, 0, 0)	(0.0524, 0.0, 0.0)
19	(20, 0, 5)	(0.0524, 0.0, 0.0175)
20	(20, 0, 20)	(0.0524, 0.0, 0.0524)
21	(20, 5, 0)	(0.0524, 0.0175, 0.0)
22	(20, 5, 5)	(0.0524, 0.0175, 0.0175)
23	(20, 5, 20)	(0.0524, 0.0175, 0.0524)
24	(20, 20, 0)	(0.0524, 0.0524, 0.0)
25	(20, 20, 5)	(0.0524, 0.0524, 0.0175)
26	(20, 20, 20)	(0.0524, 0.0524, 0.0524)

To summarise the methodology of this study discussed in Sections 5.2.1 to 5.2.3:

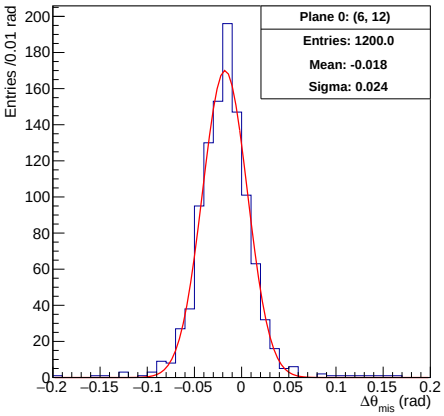
1. Model an RPC module in GEANT4 (Section 5.2.1).
2. Simulate 1200 dilepton events with a set opening angle in GEANT4 (Section 5.2.1).
3. Save the hit coordinates to a 3D histogram with bins of a size representing the position resolution (Section 5.2.2).
4. Perform a 3D Hough reconstruction on the bin centres, obtaining two reconstructed tracks (Section 5.2.2).
5. Determine the opening angle of these reconstructed tracks, θ_a (Section 5.2.2).
6. Simulate misalignments by transforming the hit coordinates via translation, rotation or a combined transformation. The transformed hits are saved in a new 3D histogram (Section 5.2.3).
7. Repeat steps 4 and 5 with this misaligned histogram, obtaining θ_{mis} (Section 5.2.3).

5.2.4 Results

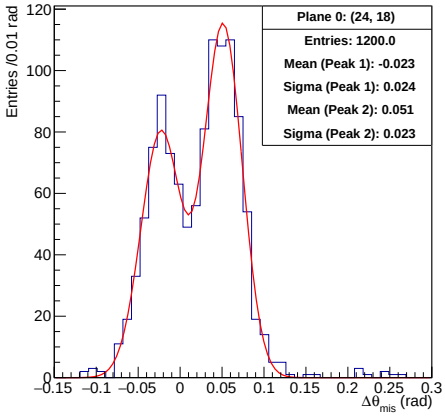
From the 1200 generated di-electron events the distributions of the opening angles in the aligned (θ_a) and misaligned (θ_{mis}) cases were determined. It can be useful to compare the differences between θ_a and θ_{mis} and the true opening angle θ_t, which in the ideal case will be distributed about zero. Any incompatibility between $\Delta\theta_a = \theta_a - \theta_t$ and $\Delta\theta_{mis} = \theta_{mis} - \theta_t$ would imply that the misalignments produce a significant effect on reconstruction. This compatibility, given as the number of standard deviations difference between them, N_σ, was taken as the primary figure of merit in this study.

The distributions of $\Delta\theta_{mis}$ over the 1200 simulated di-electron events show three main shapes: a Gaussian (Figure 5.8a), which is expected for the distribution of a random variable; a bimodal distribution (Figures 5.8b and 5.8c); and a trimodal distribution (Figures 5.8d and 5.8e). By contrast the distribution for $\Delta\theta_a$ is always a Gaussian centred on zero, implying these other distributions arise as a consequence of misalignment.

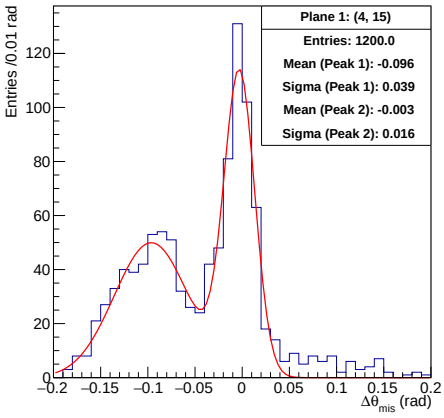
For the bimodal distribution there were two observed cases, the first is where the two peaks have approximately similar amplitudes and lie on either side of zero (Figure 5.8b), and the second is where one peak has a much larger amplitude centred around zero with a subdominant peak above or below (Figure 5.8c). For the former case, this is explained as the opening angle being reconstructed in the misaligned case either higher or lower than the expected 'true' opening angle, which over the 1200 events separates into a bimodal shape around that central point. In the latter's case, the majority of events are reconstructed around θ_t with some systematically reconstructed above or below.



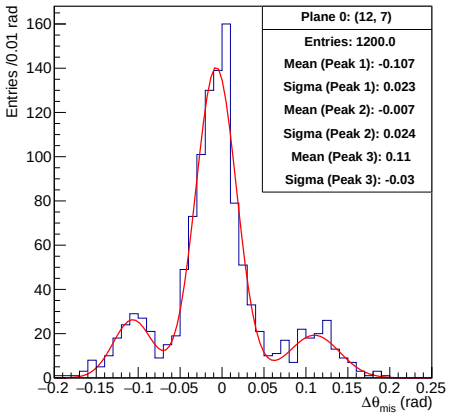
(a) (P0,6,12), Gaussian behaviour.



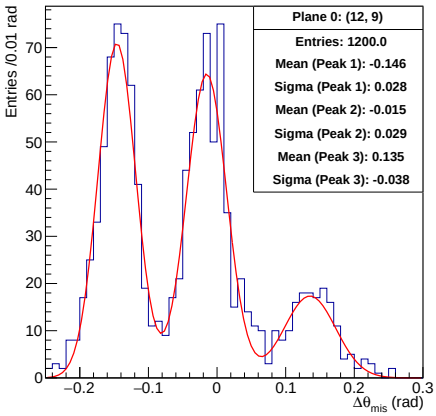
(b) (P0,24,18), Bimodal behaviour centred on zero.



(c) (P1,4,15), Bimodal behaviour with a dominant peak centred on zero.



(d) (P0,12,7), Trimodal behaviour centred on zero.



(e) (P0,12,9), Trimodal behaviour centred on zero, with multiple strong peaks.

Figure 5.8: The three main distributions observed for $\Delta\theta_{mis}$, fitted different models. The particular misalignment is given as (plane number, translation, rotation) IDs.

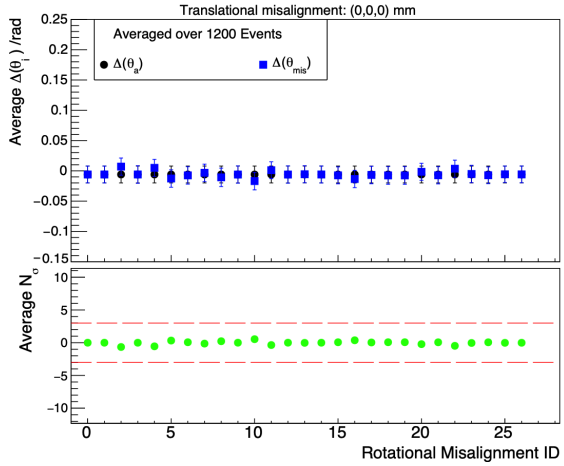
The trimodal distribution was slightly rarer than the bimodal and can likely be described as a case where the opening angle is reconstructed mostly at the true value as the central peak, but is occasionally reconstructed above or below due to the effect of misalignments. Although similar to the bimodal case, there are occasions where the central peak of the three dominates over the other two, as in Figure 5.8d, and sometimes there are multiple peaks of similar amplitudes that are not necessarily centred on zero, as in Figure 5.8e. The latter then represents a clear case where $\Delta\theta_{mis}$ would deviate from zero and therefore may be incompatible with $\Delta\theta_a$.

Out of 2916 considered distributions 72.2% (2105) are Gaussian, 18% (525) are bimodal and 9.8% (286) are trimodal. There does not appear to be a clear pattern in the RM and TM IDs where multi-modal distributions occur, which implies there is not a threshold where the non-Gaussian distributions arise for specific misalignments *e.g.* a translation in the x direction $> 5\,\text{mm}$, and is likely to be a complex effect related to the interaction between the random distribution of the misaligned hits and the 3D Hough transformation. However, a general pattern is observed when misaligning the whole RPC module, namely, the $\Delta\theta_{mis}$ distributions are entirely Gaussian. Also, planes further from the simulated decay vertex have been observed to have larger percentages of non-Gaussian distributions. Plane 0 is 85.5% Gaussian, 10.8% bimodal, and 3.7% trimodal; plane 1 is 79.3%, 12.2%, and 8.5% respectively; and plane 2 is 24%, 49%, and 27% respectively. From this a tentative conclusion is that the closer planes, 0 and 1, and the RPC module itself are likely to be closer to the aligned equivalent which is always a Gaussian but the outermost plane, plane 2, is likely to be dissimilar as non-Gaussian distributions dominate.

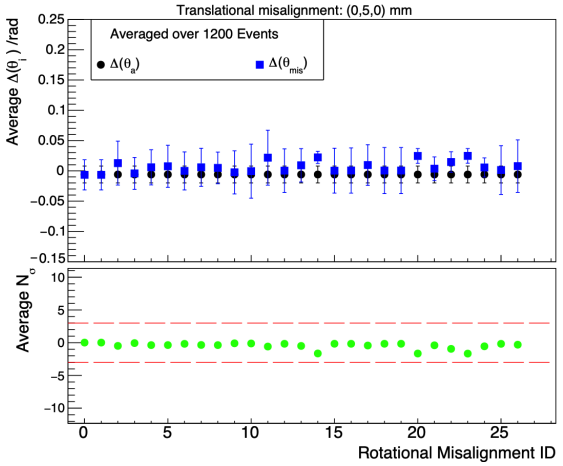
To extract the average reconstructed opening angles that could be expected for each case, these distributions are fitted: a Gaussian for the Gaussian distribution; and a sum of multiple Gaussians, each representing one of the peaks, for the bimodal and trimodal distributions. For the former, the average opening angle is taken by the mean, μ, and its error given by the peak width, σ. For the latter, the average opening angle is taken by an average of the peak positions, weighted by the approximate peak area given by the amplitude multiplied by the peak width. This is to favour the peaks of the multi-modal with more entries. The error in this case is given by the peak widths added in quadrature. From these average values of $\Delta\theta_{(a,mis)}$, the average compatibility can be found.

It should be noted that in Figures 5.9 to 5.12 each point in those plots uses the same set of 1200 events, hence their errors are highly correlated. Each point only differs in how it was misaligned before reconstruction.

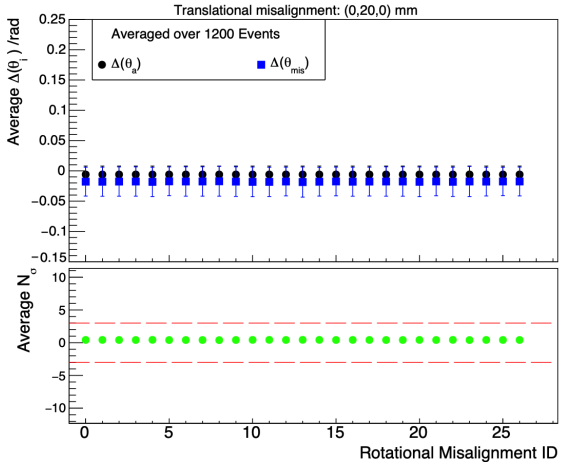
Figure 5.9 shows the average value of $\Delta\theta_{(a,mis)}$ as a function of the RM ID, above a plot of the compatibility of $\Delta\theta_a$ and $\Delta\theta_{mis}$, N_σ, which should be < 3 to be considered compatible, both for the situation where only plane 0 is misaligned. The nominal value



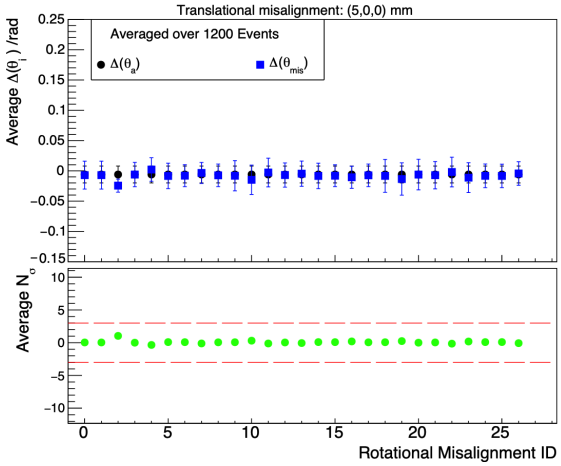
(a) (0,0,0)mm



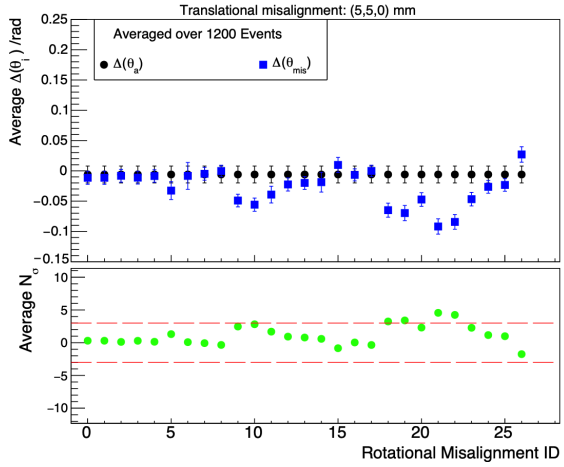
(b) (0,5,0)mm



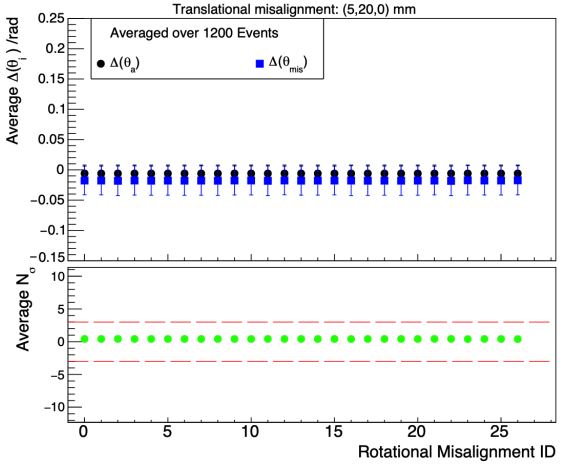
(c) (0,20,0)mm



(d) (5,0,0)mm

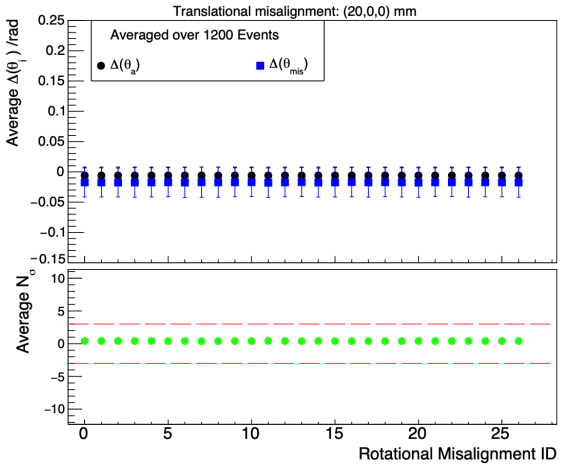


(e) (5,5,0)mm

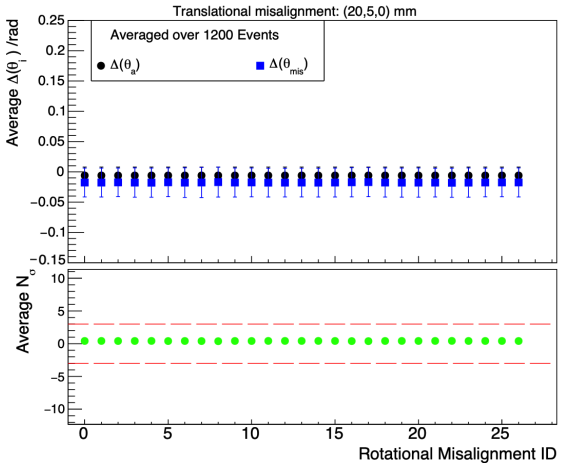


(f) (5,20,0)mm

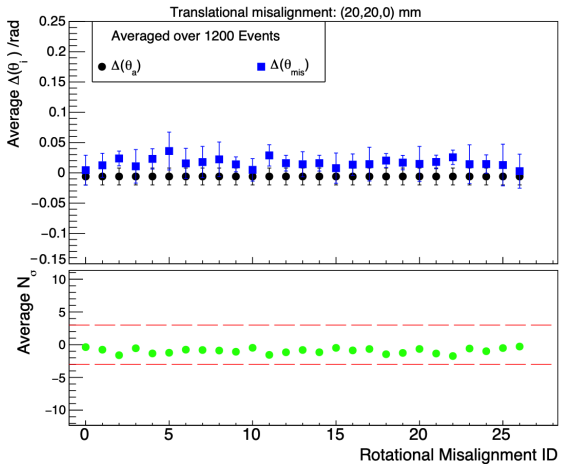
Figure 5.9



(g) (20,0,0)mm



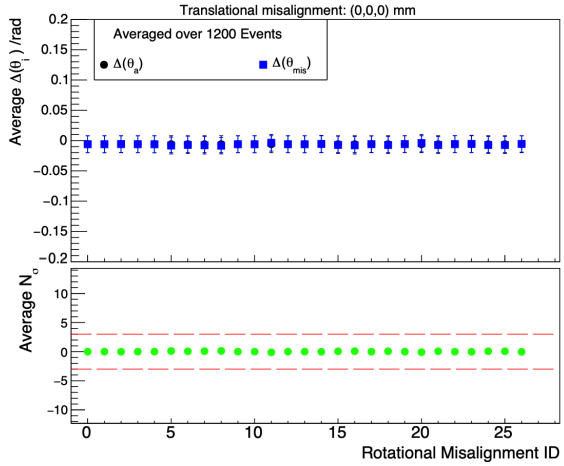
(h) (20,5,0)mm



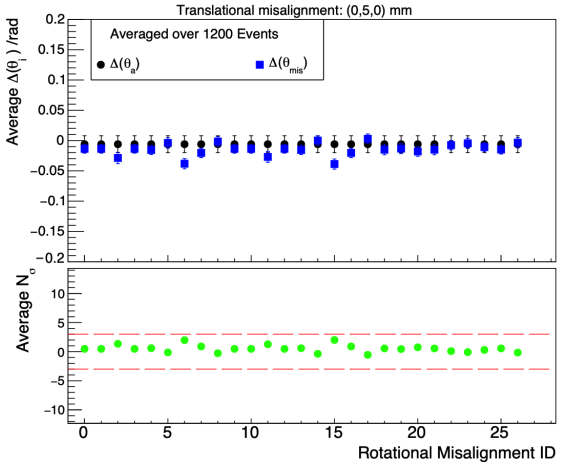
(i) (20,20,0)mm

Figure 5.9: Summary of the average $\Delta\theta_a$ and $\Delta\theta_{mis}$ when plane 0 has been misaligned by translation and/or rotation. For each translational shift (see captions), the 27 separate rotations considered are shown on the x-axis. Each point of these plots is averaged over the same 1200 simulated events, differing only in reconstruction due to the misalignment introduced.

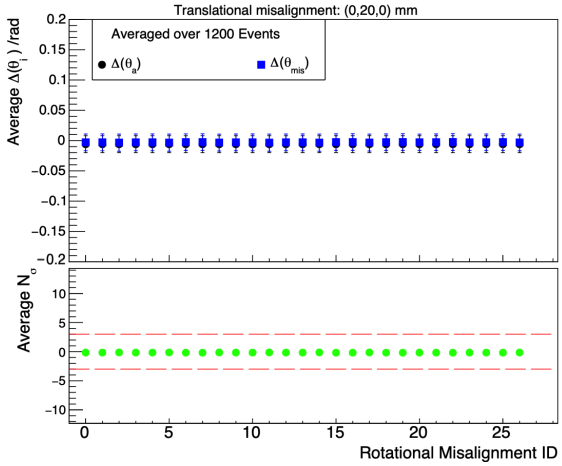
for both the aligned and misaligned cases is very close to 0, and hence to the true opening angle, with errors $< 1°$. It can also be observed that there is very little spread in values as a function of the rotational misalignment implying even the large values used in this study should not meaningfully impact the reconstruction ability of CODEX-β, this is also observed in all other plane misalignments considered in this study. Although, Figure 5.9e also shows some variation with rotational misalignment, with those with IDs > 18 showing a much greater deviation of $\sim 5°$ from the true opening angles, these are still mostly compatible with the aligned case. This is not observed for more extreme translations in either the x or y directions, so overall it can be concluded that misalignments both in translation and rotation for plane 0 should not affect CODEX-β's reconstruction.



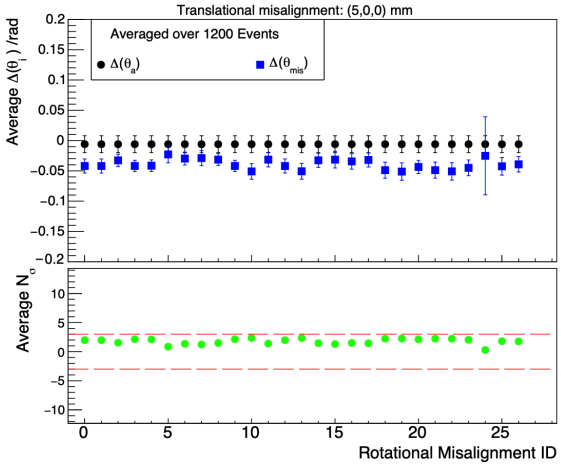
(a) (0,0,0)mm



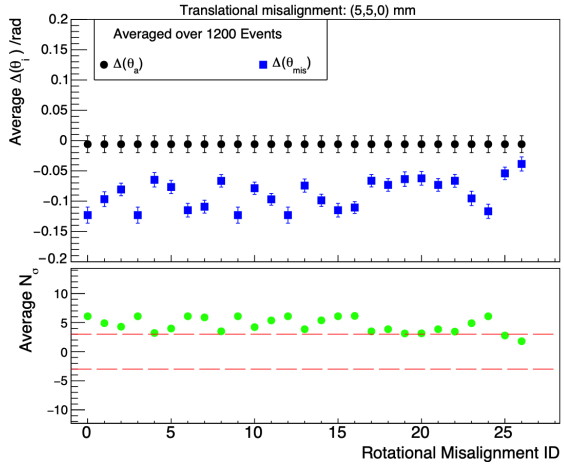
(b) (0,5,0)mm



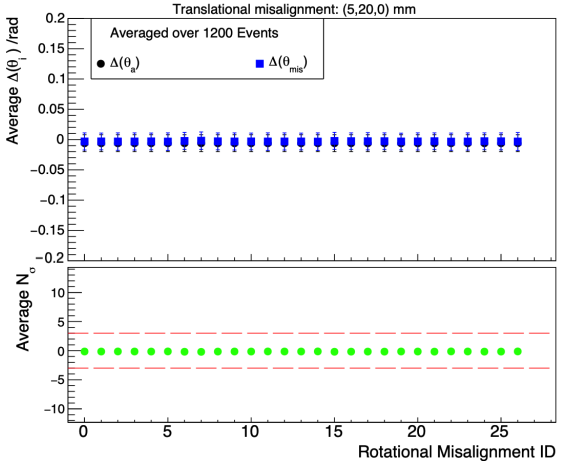
(c) (0,20,0)mm



(d) (5,0,0)mm

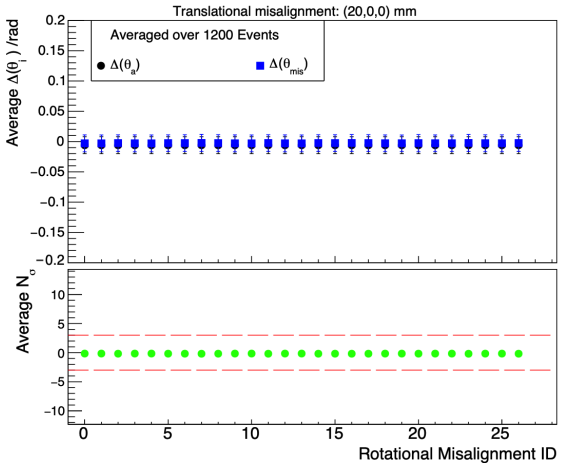


(e) (5,5,0)mm

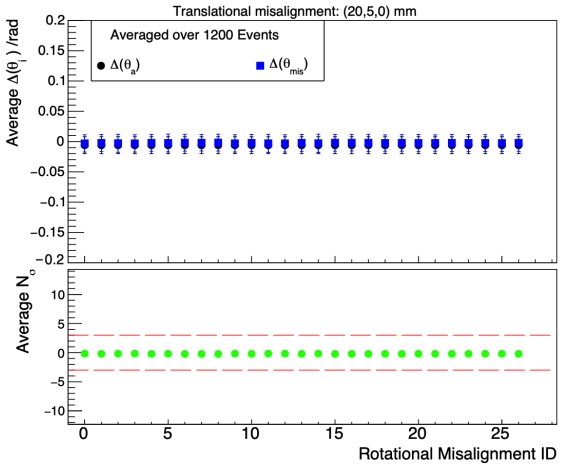


(f) (5,20,0)mm

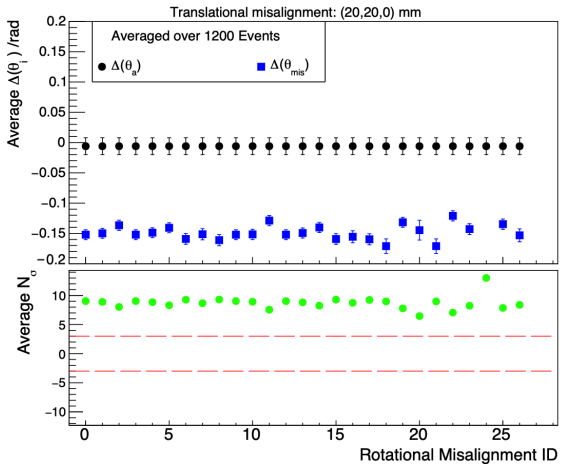
Figure 5.10



(g) (20,0,0)mm



(h) (20,5,0)mm



(i) (20,20,0)mm

Figure 5.10: Summary of the average $\Delta\theta_a$ and $\Delta\theta_{mis}$ when plane 1 has been misaligned by translation and/or rotation. For each translational shift (see captions), the 27 separate rotations considered are shown on the x-axis. Each point of these plots is averaged over the same 1200 simulated events, differing only in reconstruction due to the misalignment introduced.

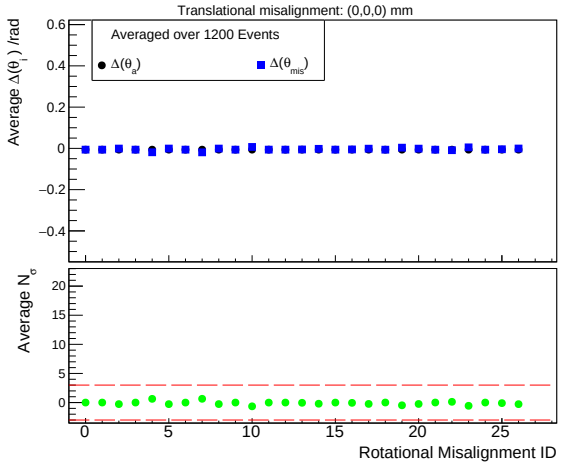
Considering now the misalignments of plane 1, given in Figure 5.10, while most of the considered cases show very good compatibility between $\Delta\theta_a$ and $\Delta\theta_{mis}$, more notable incompatibilities are seen for translations of $\geq 5\,\text{mm}$ in both the x and y directions. For the case with a TM of $(5, 5, 0)\,\text{mm}$ (Figure 5.10e), the misaligned case tends to be $\sim 4\sigma$ away from the aligned, with absolute differences from the true angle of $\sim 5°$. For a TM of $(20, 20, 0)\,\text{mm}$ (Figure 5.10i) these numbers are almost twice as large at $\sim 9\sigma$ and $\sim 8.5°$ respectively. However, if a mixture of 5 mm and 20 mm misalignments in (x, y) are made (Figures 5.10f and 5.10h) it appears that this does not show similar behaviour, implying that it is the combined effect that so strongly affects reconstruction. This could be because plane 1 is the middle plane, so a strong combined translation in multiple directions could

lead to the 3D Hough transform algorithm potentially grouping hits in the middle singlet and associating them with points in plane 0, and likewise associating some of the plane 2 hits with some in plane 0, leading to two lines with a much greater opening angle. The middle singlet does not have as strong an effect on reconstruction generally, which is likely to be due to the start and end points in the other singlets providing a good indication of the track direction and smoothing out minor misalignments.

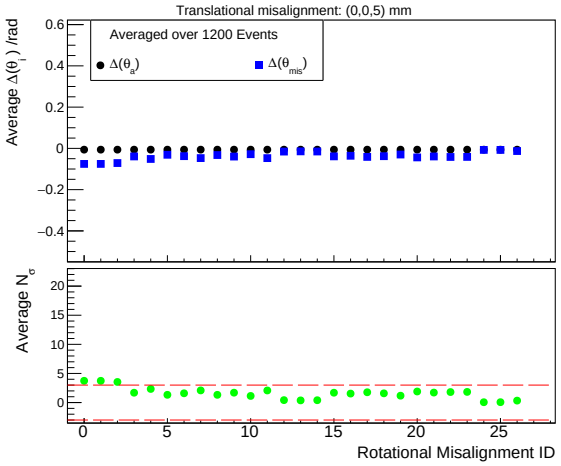
Figure 5.11 shows the effects of misaligning plane 2, which is the outermost plane relative to LHCb's interaction point. Any translational misalignment of up to 5 mm in any combination, *i.e.* TMs (0,0,0), (0,0,5), (0,5,0), (0,5,5), (5,0,0), (5,0,5), (5,5,0), (5,5,5)mm, shows good compatibility between the aligned and misaligned case, but $\Delta\theta_{mis}$ appears generally greater than previous cases $\sim$0–5°. By contrast, only one case with a TM of 20 mm shows similarly good behaviour and that is for $(20, 5, 0)$ mm, which also has two RMs that are incompatible (IDs 14 and 21). The rest of the combinations of TMs involving 20 mm are incompatible with values up to 10σ. This implies that the plane furthest from the interaction point has the strongest effect on the reconstruction for this algorithm. This could be because the two endpoints of the line will have the greatest impact on its resultant angle. The two clusters of hits in this final plane will have the greatest separation and so a smaller shift will induce a larger potential solid angle change in the point cloud arrangement for the 3D Hough transform. This is unlike the results for plane 0 where the separation between hit clusters would be smaller.

Finally, if the entire RPC module is misaligned, translational misalignments appear to have a minimal impact on the opening angle reconstruction as seen in Figure 5.12. However, there is an interesting effect that can be observed for Figures 5.12b, 5.12n and 5.12z where RMs of ID 12, 13, 14, and 24, 25, 26 are notably different to the rest and $\sim$3° away from 0. These points still show compatibility between $\Delta\theta_a$ and $\Delta\theta_{mis}$, but at a much reduced level. This implies some real effect due to the combination of a translation of 5 mm in the z direction and the same shift in x and y, coupled with an RM around both the x and y axis.

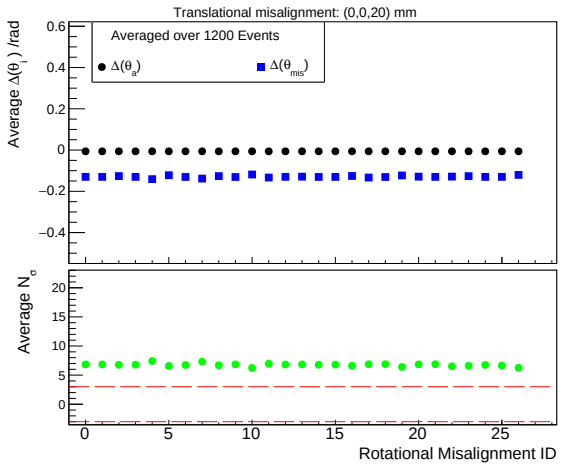
In conclusion, it appears that there is negligible dependence on rotational misalignments either of an individual RPC singlet or of the RPC module as a whole. Additionally, the two singlets closest to the interaction point mostly show good compatibility between the aligned and misaligned reconstructed opening angles for a range of translational misalignments with most being less than 3σ different. The effect of these misalignments also appears to be relatively flat across the RMs considered, implying that larger shifts could potentially be tolerated.



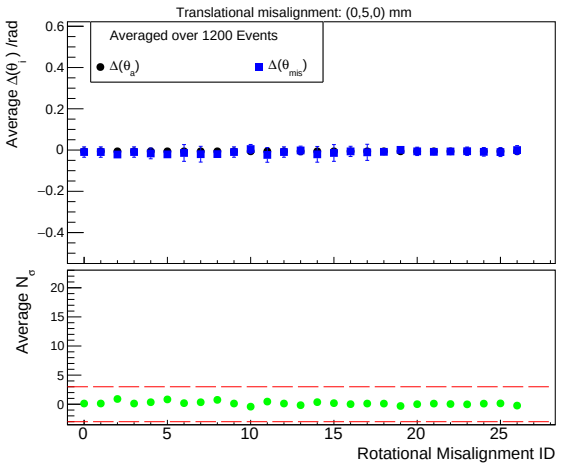
(a) (0,0,0)mm



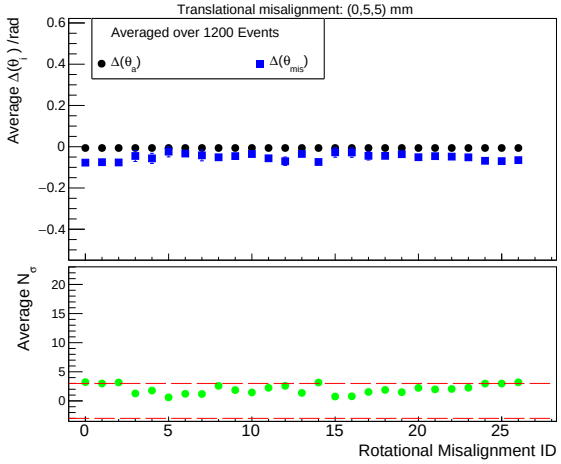
(b) (0,0,5)mm



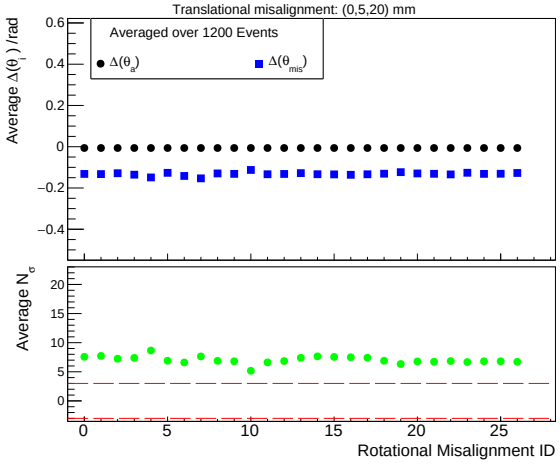
(c) (0,0,20)mm



(d) (0,5,0)mm

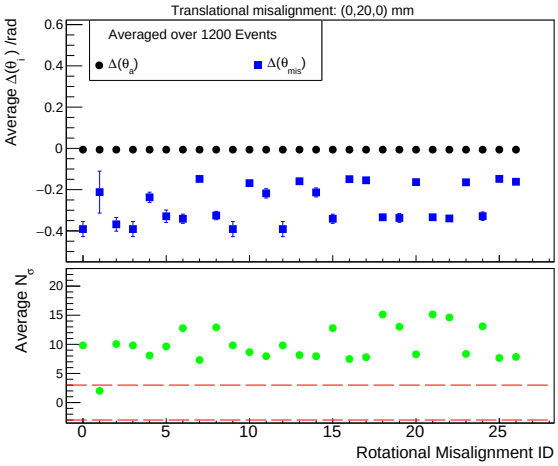


(e) (0,5,5)mm

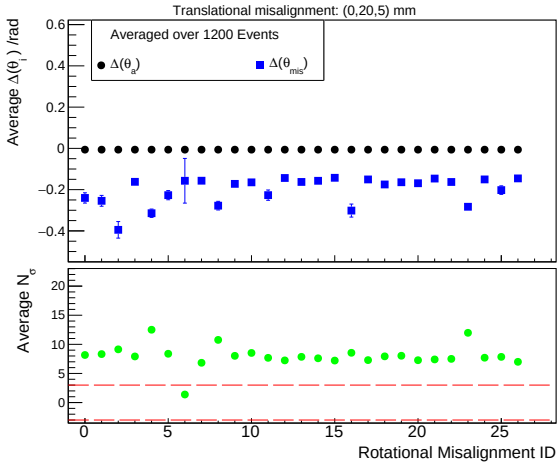


(f) (0,5,20)mm

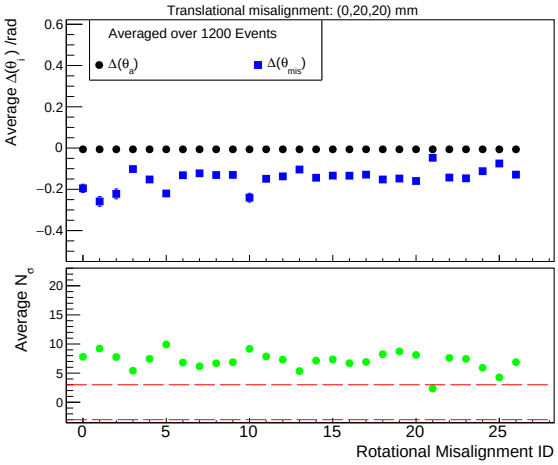
Figure 5.11



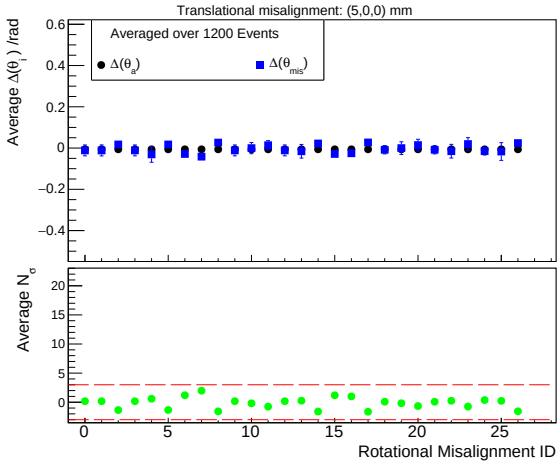
(g) (0,20,0)mm



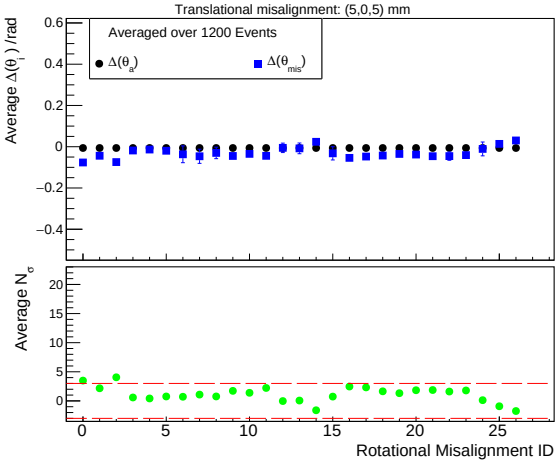
(h) (0,20,5)mm



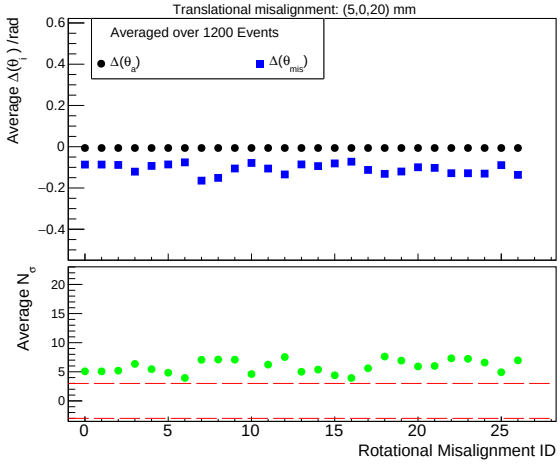
(i) (0,20,20)mm



(j) (5,0,0)mm

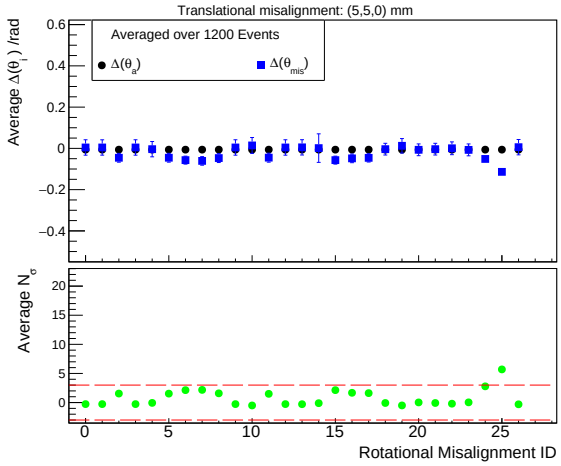


(k) (5,0,5)mm

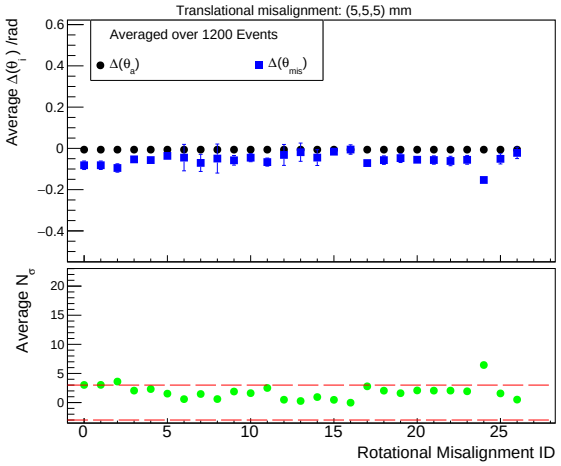


(l) (5,0,20)mm

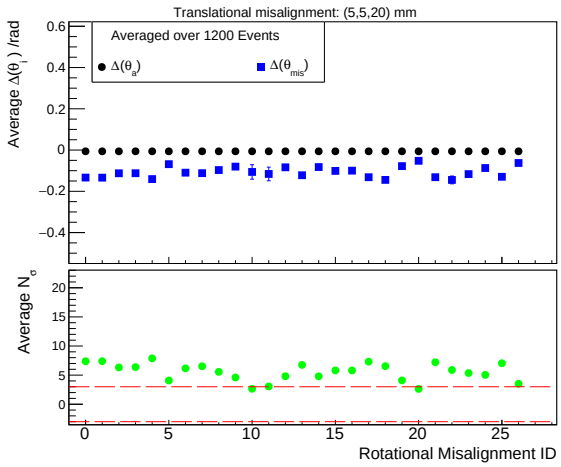
Figure 5.11



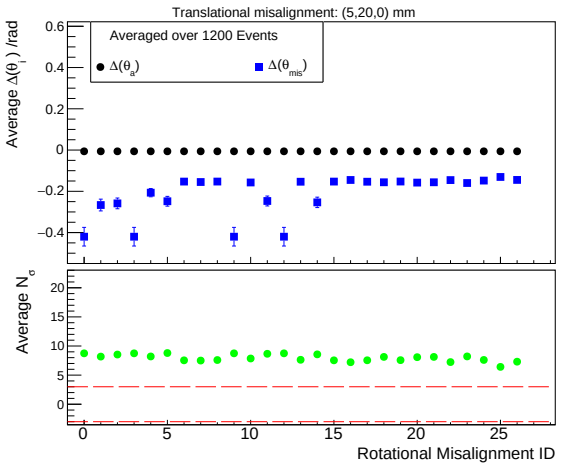
(m) (5,5,0)mm



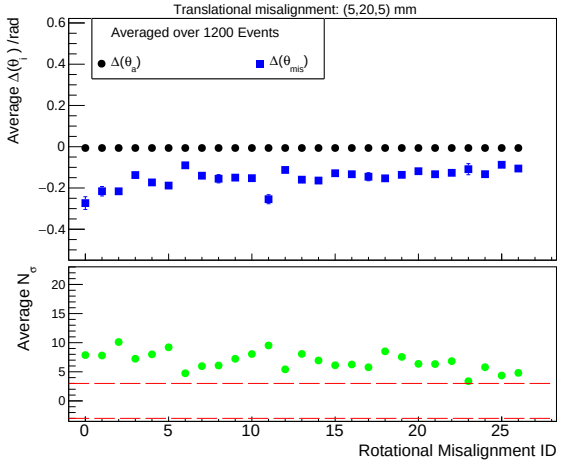
(n) (5,5,5)mm



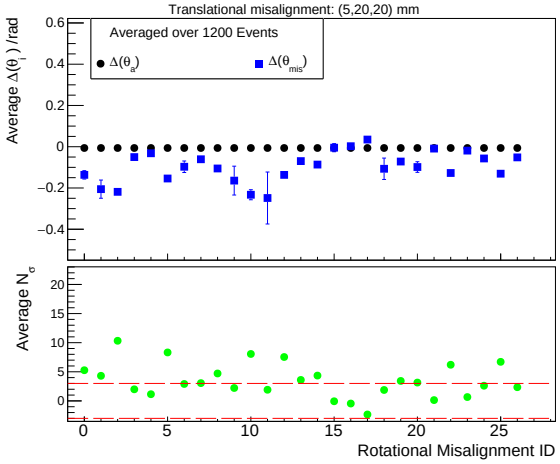
(o) (5,5,20)mm



(p) (5,20,0)mm

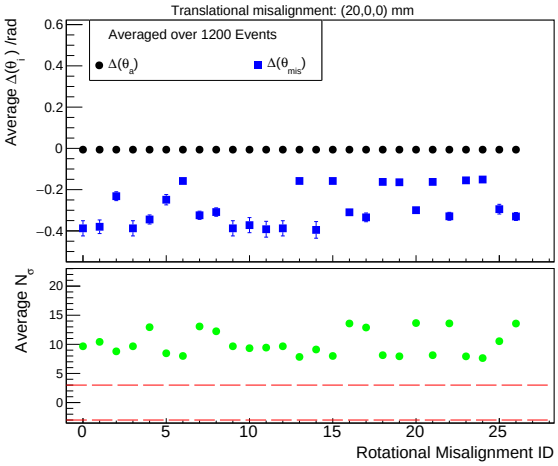


(q) (5,20,5)mm

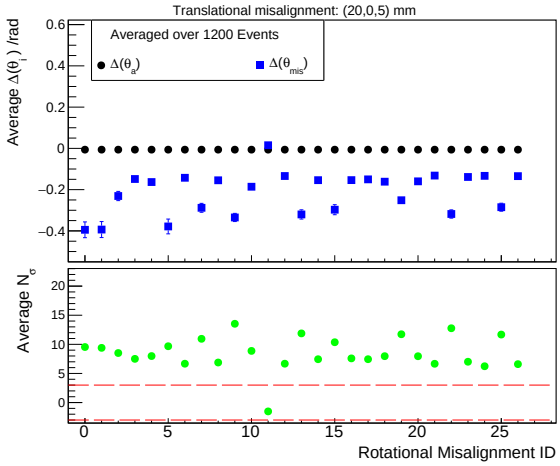


(r) (5,20,20)mm

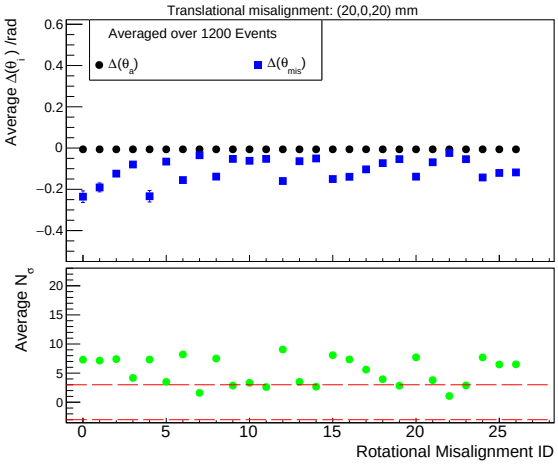
Figure 5.11



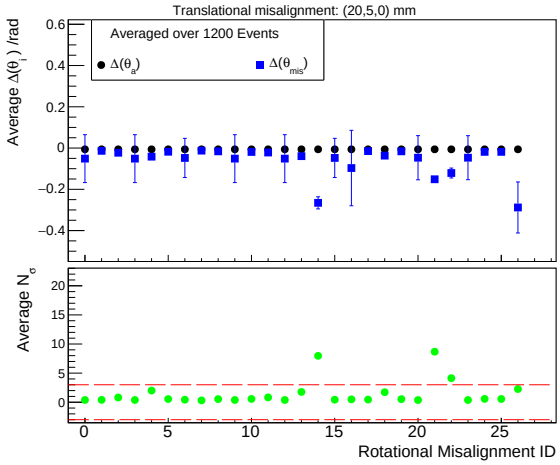
(s) (20,0,0)mm



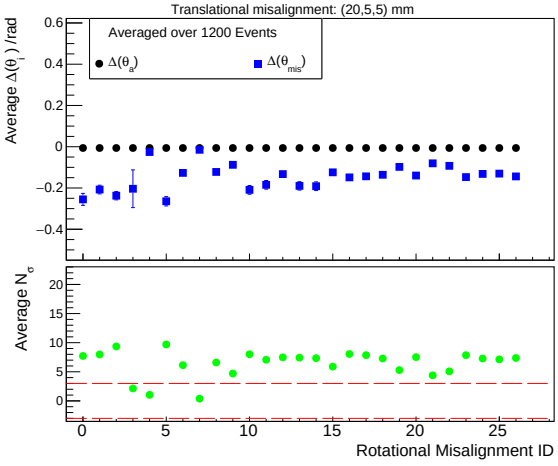
(t) (20,0,5)mm



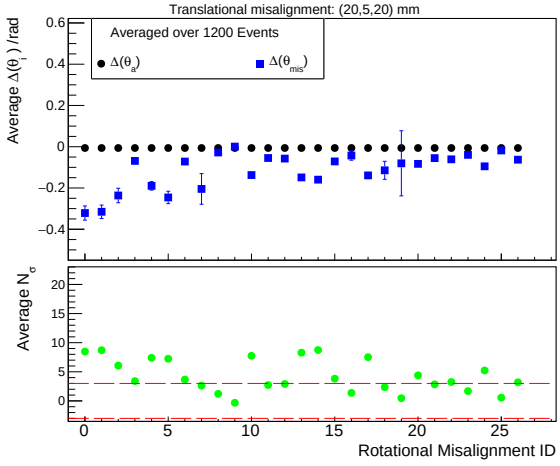
(u) (20,0,20)mm



(v) (20,5,0)mm

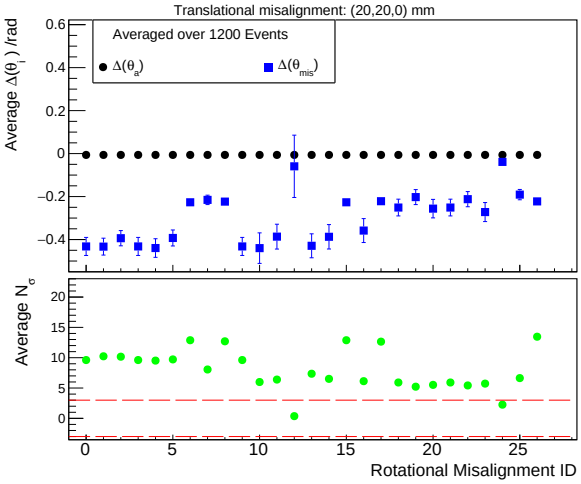


(w) (20,5,5)mm

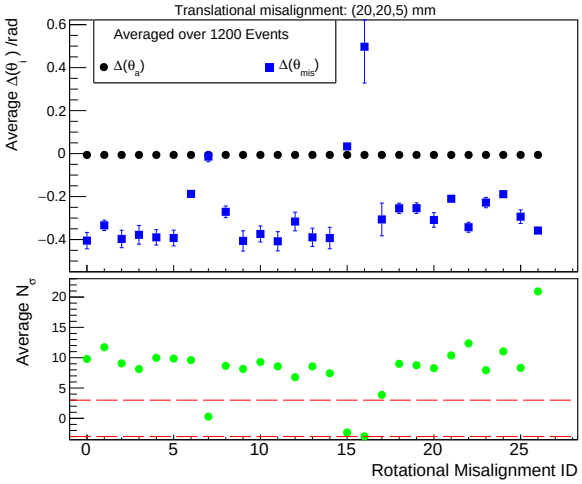


(x) (20,5,20)mm

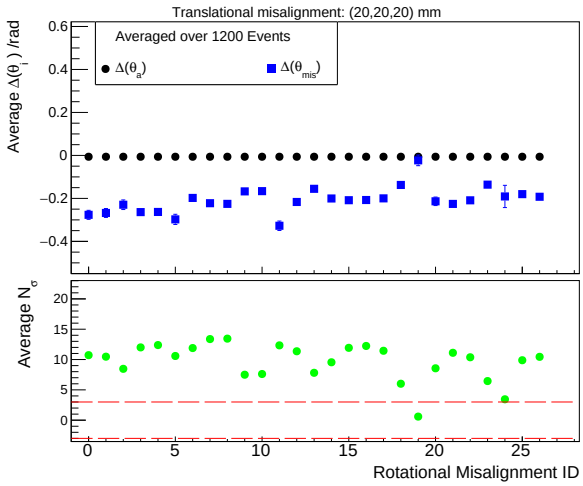
Figure 5.11



(y) (20,20,0)mm

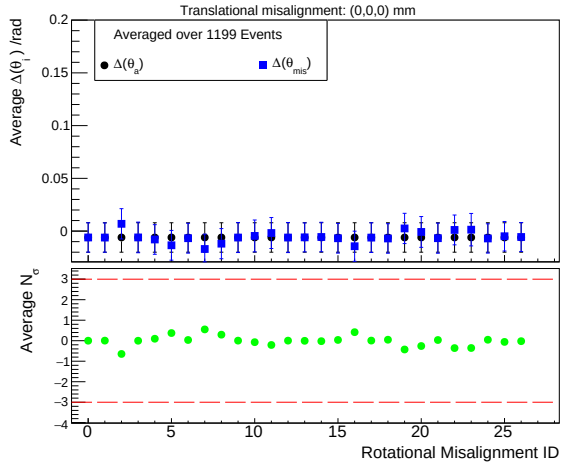


(z) (20,20,5)mm

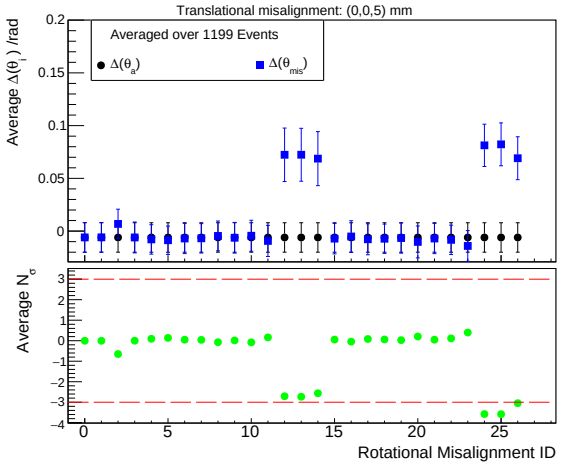


(aa) (20,20,20)mm

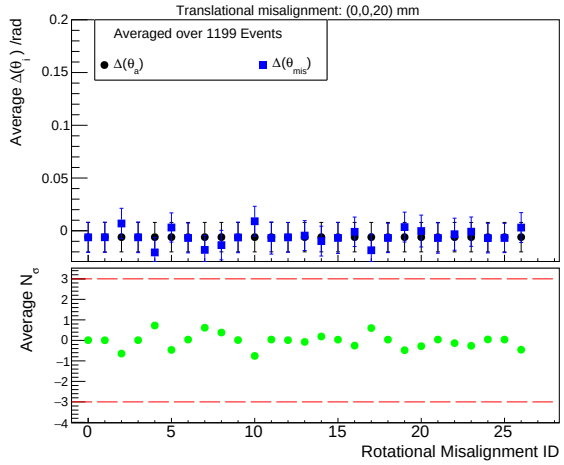
Figure 5.11: Summary of the average $\Delta\theta_a$ and $\Delta\theta_{mis}$ when plane 2 has been misaligned by translation and/or rotation. For each translational shift (see captions), the 27 separate rotations considered are shown on the x-axis. Each point of these plots is averaged over the same 1200 simulated events, differing only in reconstruction due to the misalignment introduced.



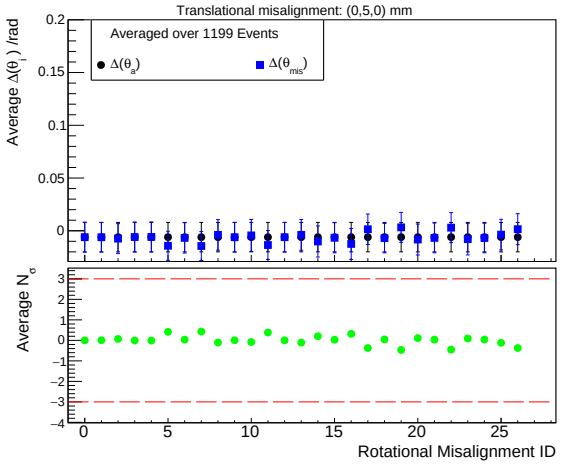
(a) (0,0,0)mm



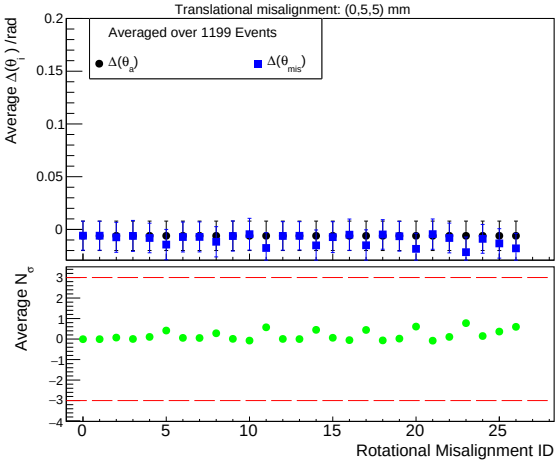
(b) (0,0,5)mm



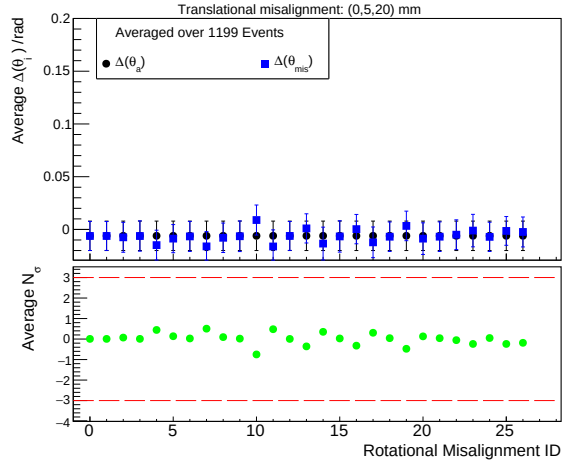
(c) (0,0,20)mm



(d) (0,5,0)mm

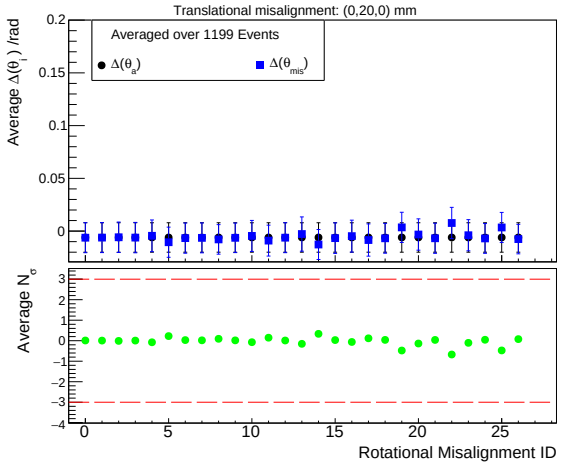


(e) (0,5,5)mm

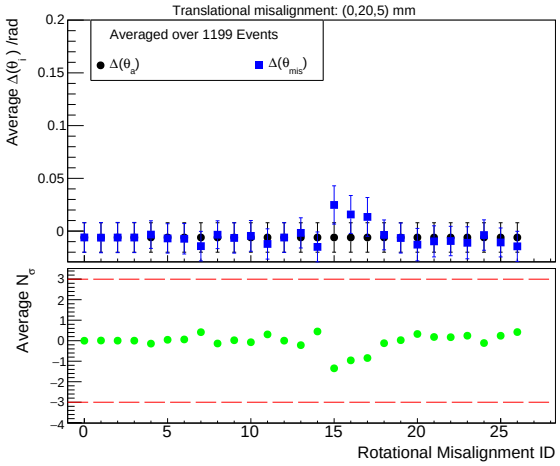


(f) (0,5,20)mm

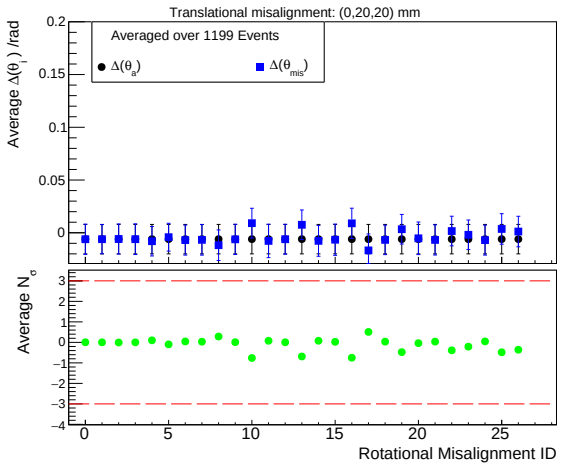
Figure 5.12



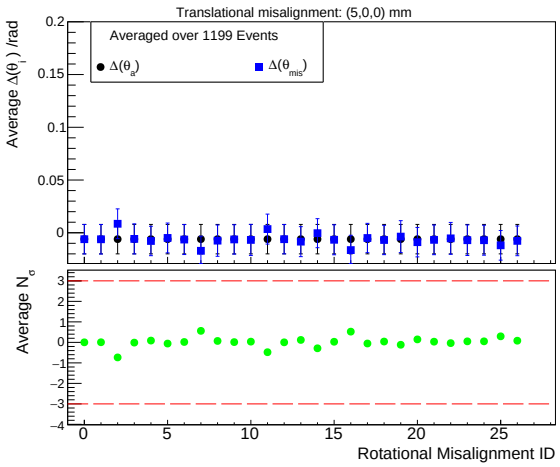
(g) (0,20,0)mm



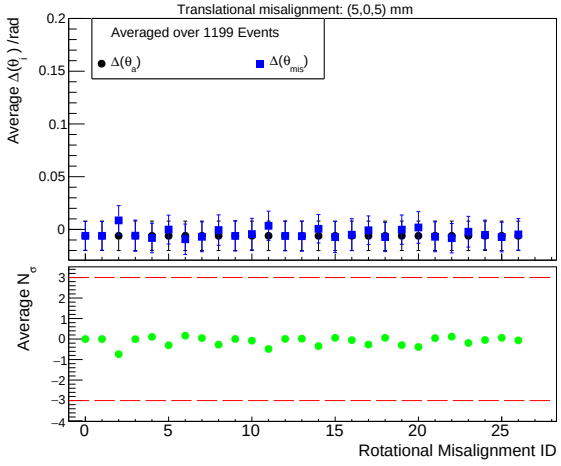
(h) (0,20,5)mm



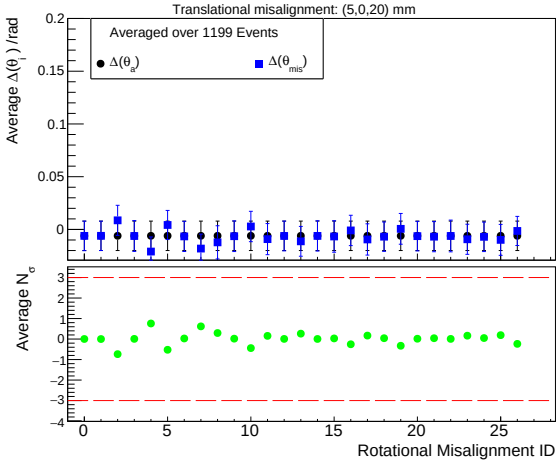
(i) (0,20,20)mm



(j) (5,0,0)mm

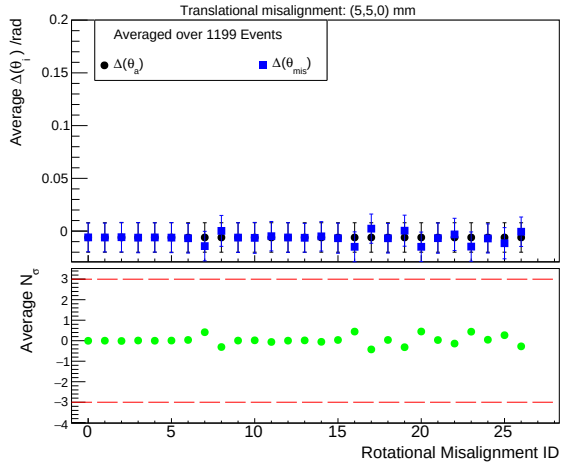


(k) (5,0,5)mm

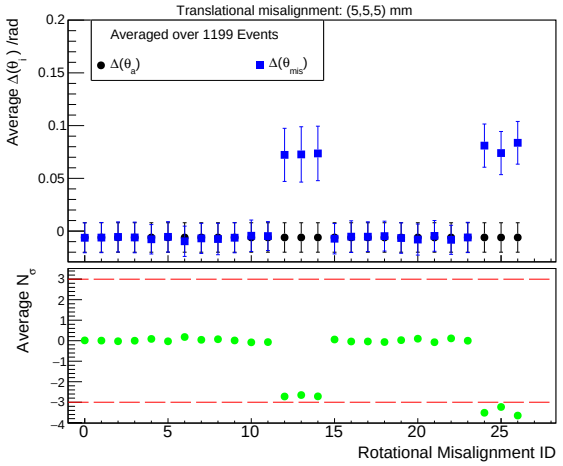


(l) (5,0,20)mm

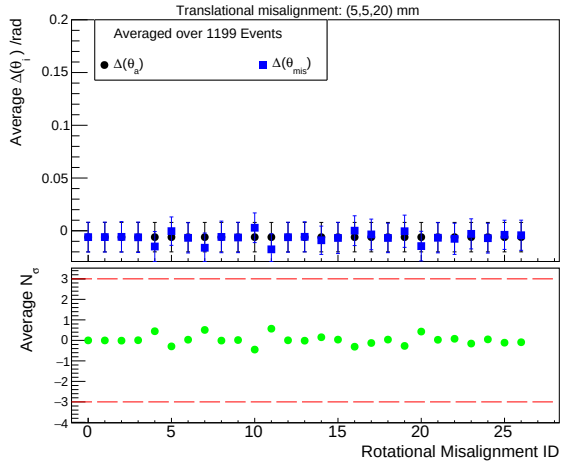
Figure 5.12



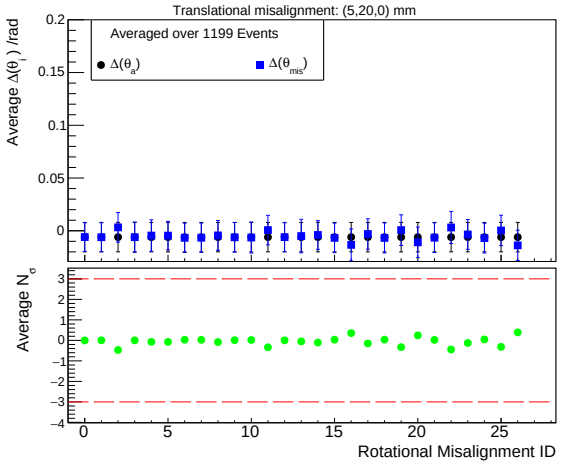
(m) (5,5,0)mm



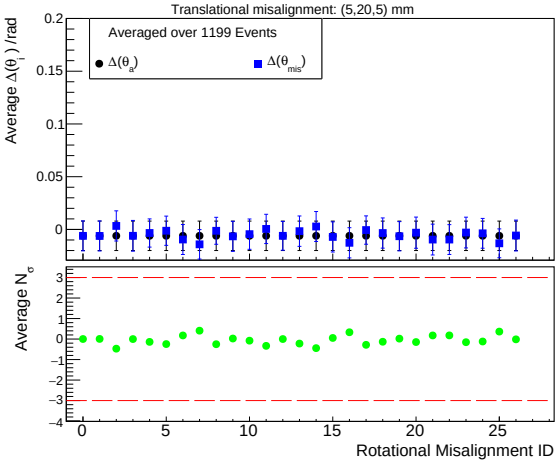
(n) (5,5,5)mm



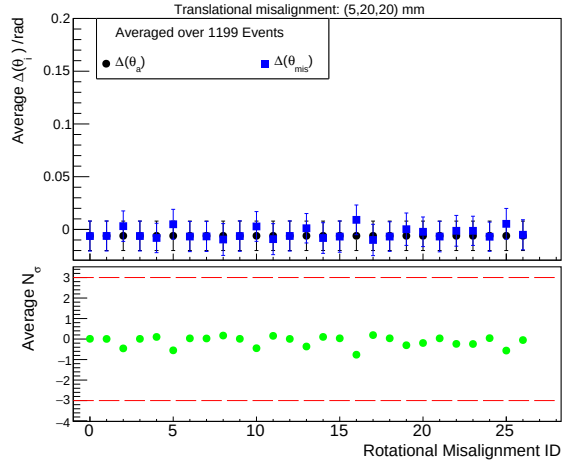
(o) (5,5,20)mm



(p) (5,20,0)mm

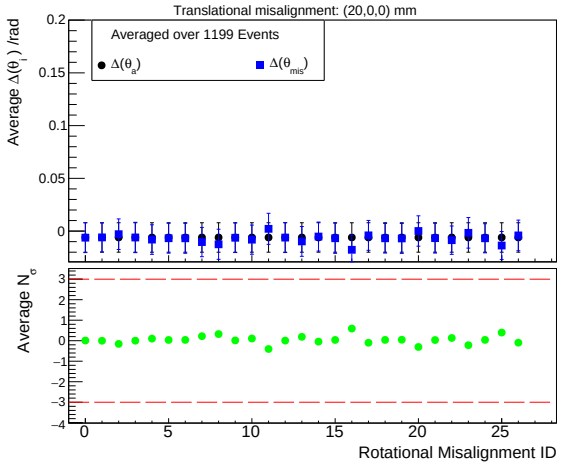


(q) (5,20,5)mm

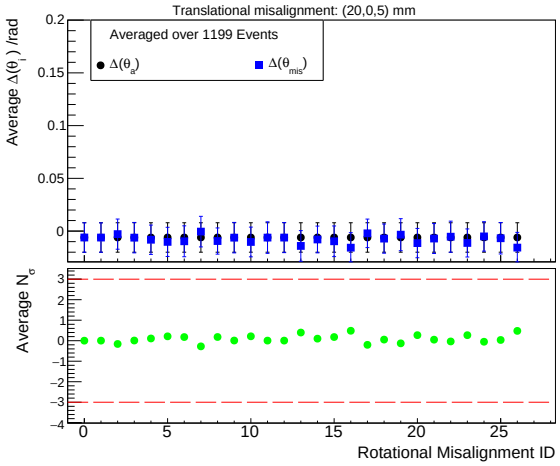


(r) (5,20,20)mm

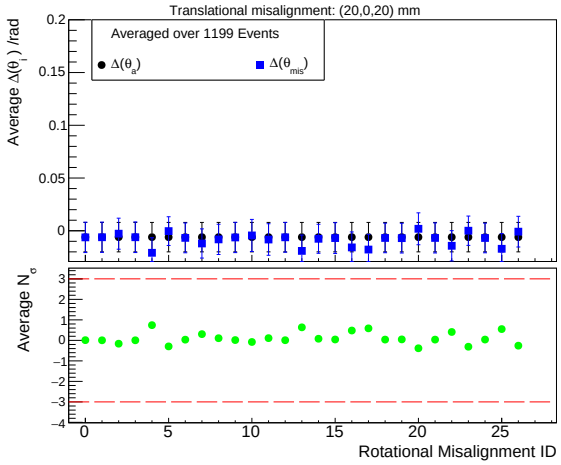
Figure 5.12



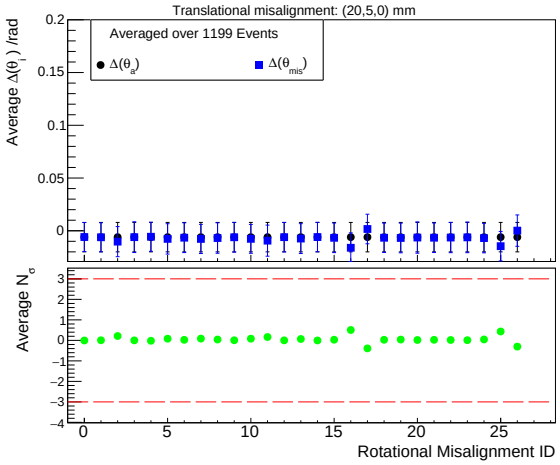
(s) (20,0,0)mm



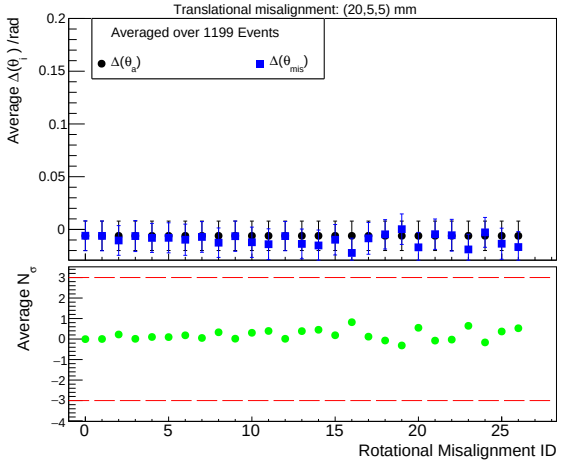
(t) (20,0,5)mm



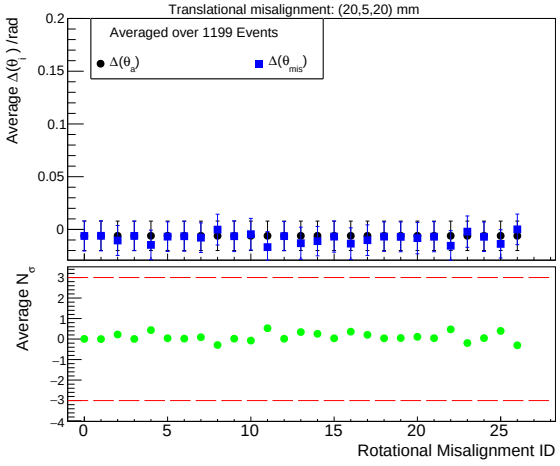
(u) (20,0,20)mm



(v) (20,5,0)mm

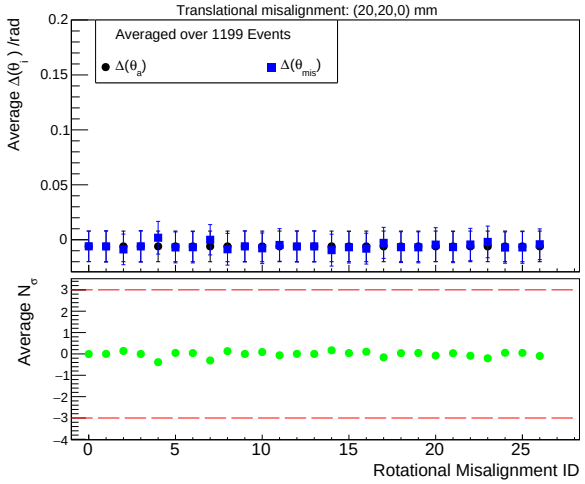


(w) (20,5,5)mm

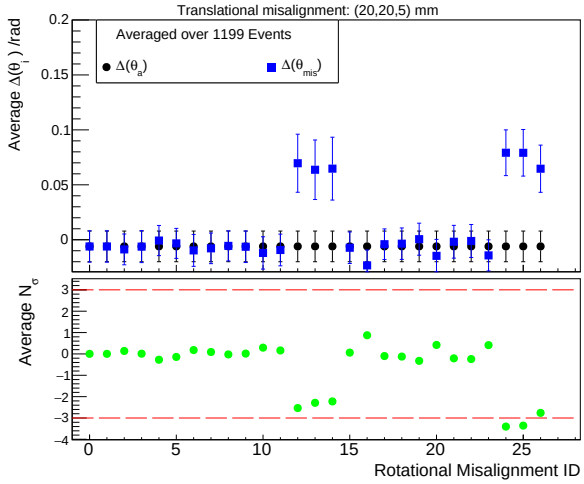


(x) (20,5,20)mm

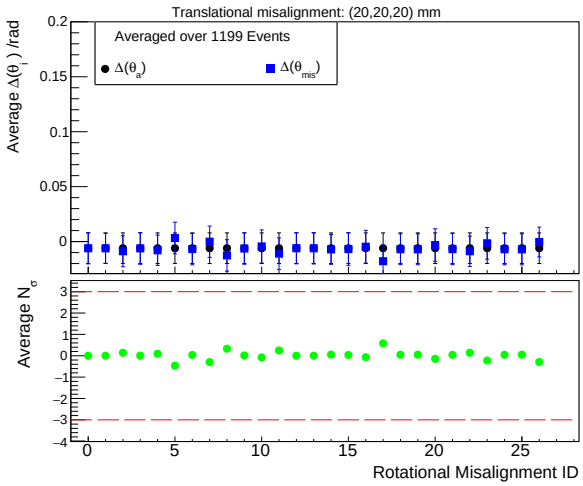
Figure 5.12



(y) (20,20,0)mm



(z) (20,20,5)mm



(aa) (20,20,20)mm

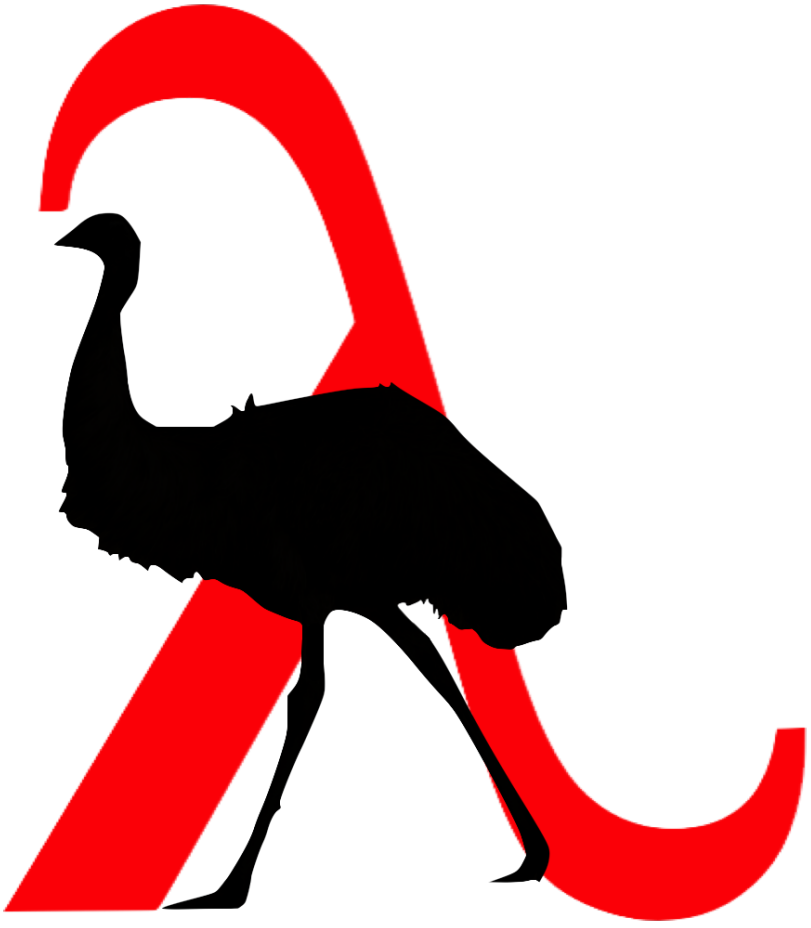
Figure 5.12: Summary of the average $\Delta\theta_a$ and $\Delta\theta_{mis}$ when all planes have been misaligned by translation and/or rotation. For each translational shift (see captions), the 27 separate rotations considered are shown on the x-axis. Each point of these plots is averaged over the same 1199 simulated events, differing only in reconstruction due to the misalignment introduced.

It should also be noted that this study represents the worst case scenario from a tracking perspective. The track reconstruction algorithm does not include any additional constraints such as requiring hits in the first and last planes, which could reduce the impact of misalignments. Also, these are for completely unknown misalignments which were not observed during the mechanical assembly of the RPC module, or the mechanical survey of the positions of the modules and planes in-situ, or a study of the tracking using cosmic ray tracks. If at either of these stages the misalignments are identified and quantified, the effect on the tracking reconstruction demonstrated by the study would be reduced or eliminated entirely.

Overall, the mechanical precision when manufacturing the frame to support the RPC triplet will satisfy the requirements of $< 5\,\text{mm}$ that this study suggests would cause no meaningful impact on tracking reconstruction for CODEX-β. Additionally, the most likely scenario for misalignment is not the misalignment of individual planes within the triplet but the misalignment of an entire triplet which likewise demonstrates no significant effect should be observed in track reconstruction in CODEX-β and hence CODEX-b as well.

Part II

Searching for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ Decay



Chapter 6

Strategy for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ Analysis

The following chapters describe a search for the cLFV decay, $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$, which was also detailed in Ref. [136]. This analysis was motivated by the fact that cLFV is explicitly forbidden in the SM. Therefore, observation of this decay would be evidence of BSM physics. The ultimate goal of this analysis was to make the world's first observation of this channel and cLFV generally, or to set the first experimental upper limit to its branching fraction to constrain BSM models.

The search was performed for the full period of Run 1 (2011 and 2012) and Run 2 (2015–2018), corresponding to integrated luminosities of $2.98\,\text{fb}^{-1}$ and $5.36\,\text{fb}^{-1}$. The per year luminosities are shown in Table 3.1. Due to the nature of the decay, no significant signal is expected from SM processes, but some BSM models predict that the signal would be within LHCb's sensitivity. To reduce the possibility and impact of bias when developing the analysis selection, the data in the signal region was made inaccessible until studies had been concluded, known as 'blinding' the dataset. The MC covers the full invariant mass range and is not subject to blinding. The signal region is defined by the Λ_b^0 baryon's invariant mass interval $[5000, 5800]\,\text{MeV}/c^2$, determined by comparison with the Λ_b^0 mass shape in the signal MC. This is a relatively large window in comparison to the Λ_b^0 mass, $5619.60 \pm 0.17\,\text{MeV}/c^2$ [22], which is due to the presence of the electron in the final state producing a long bremsstrahlung tail on the lower side of the mass peak. Before selection the 'mass sidebands' around this peak should be dominated by backgrounds. Therefore, these sidebands can be used to test the effectiveness of the selection on those potential backgrounds and a simulated dataset of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ is used to estimate the effect on the signal. The mass sidebands are defined as $m(\Lambda e^{\mp}\mu^{\pm}) < 5000$ or $m(\Lambda e^{\mp}\mu^{\pm}) > 5800\,\text{MeV}/c^2$. Additionally, the search was performed for the q^2 region $[0.1, 20]\,\text{GeV}^2/c^4$, where q^2 is the invariant mass squared of the di-lepton pair ($e\mu$ in this case). This q^2 region was selected to remove potential

pollution from candidates at the photon pole, $q^2 < 0.1\,\mathrm{GeV}^2/c^4$.

The selection was performed in two stages, a cut based initiative known as the pre-selection and a second stage that used a multivariate classifier trained on kinematic information to reduce the remaining background. The pre-selection uses the known reconstructed information about the decay particles to remove either unphysical or dominant background data, and reducing the overall impact of potential backgrounds. This was done while ensuring these cuts did not distort the underlying mass shape, or reject an unacceptably large fraction of the potential signal. The multivariate classifier uses multivariate analysis techniques, based on kinematic information, to better isolate candidates that are more signal-like than background-like. From this, an optimised cut removes a significant number of background-like candidates while preserving as high a proportion of signal-like candidates as possible.

The $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ decay topology includes a long-lived Λ baryon which allows most potential backgrounds to be suppressed despite the abundance of Λ_b^0 baryons produced at LHCb. This is because there are only two known particles with a decay length of around 1 m, the Λ and the K_S^0. Therefore, a displaced vertex at this associated lifetime provides a very clean signal for this decay. Additionally, an optimised leptonic PID cut is included in the data samples to reduce the potential misidentification of final-state leptons so that the $\Lambda_b^0 \to \Lambda \mu^- \mu^+$ decay is unlikely to act as a potential background.

The data was sub-divided into a total of eight independent categories to ensure that the analysis could be performed optimally. These categories were formed based on three binary choices of selection: track type, bremsstrahlung reconstruction and data period. For the track type case, this refers to the decay products of the Λ baryon, *i.e.* $\Lambda \to p\pi^-$, where the tracks of the p and π^- are either both 'Long' tracks or both 'Down' tracks as defined in Section 3.2. These two cases are referred to by the shorthand: Long-Long (LL) and Down-Down (DD). The overall track resolution of these two differs due to the LL tracks having the additional information from the VELO, so they should be considered independently. Additionally, some mis-modelling has been previously observed, such as in Ref. [137] where the DD efficiency for $K_S^0 \to \pi^+\pi^-$ is $\sim 20\%$ different than expected to the LL case in Run 1. Therefore, by measuring these categories separately any potential effect from this is reduced.

The bremsstrahlung reconstruction classification indicates whether or not bremsstrahlung has been reconstructed for the final state electron for the Λ_b^0 candidate considered ('Brem' and 'No Brem' respectively). This affects the mass resolution as described in Section 3.2 due to the potential missing energy when an electron undergoes bremsstrahlung before the magnet. Additionally, separating the two categories provides an additional check for exclusive backgrounds as the Brem categories would have two well-

defined final state leptons and so an inconsistency with the No Brem equivalent categories could highlight an additional background to be considered.

Finally, the categories were split by the data-taking period, Run 1 or Run 2. The LHCb detector hardware did not change between these two periods as described in Chapter 3, however the trigger strategy did (see Section 3.6). This different strategy as well as the different run conditions (see Table 3.1) mean that the two data periods should be considered separately. The resultant analysis categories will be identified throughout this part in the form: 'Bremsstrahlung reconstruction status - Track type - Run period' *e.g.* Brem-LL-Run 2.

In order to translate an observed (or excluded) number of signal candidates to a measured branching fraction, a normalisation channel decay, $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ was used. This has a similar topology to the 'rare' mode decay, $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$, but is very abundant at LHCb due to the presence of a charmonium resonance in the final state. The branching fraction of the rare mode is then given by

$$\mathcal{B}_{rare} = \mathcal{B}_{norm} \cdot \frac{\varepsilon_{norm}}{\varepsilon_{rare}} \cdot \frac{N_{rare}}{N_{norm}} = \alpha \cdot N_{rare}\,, \tag{6.0.1}$$

where $\mathcal{B}_{norm}$ is the branching fraction of the normalisation channel, N indicates the number of observed signal candidates obtained from the mass fits in data, ε indicates the total selection and reconstruction efficiency, α is the expected rare mode branching fraction for a single signal candidate for the rare mode, and the subscript indicates whether these quantities are referring to the rare or normalisation channel.

The current known value for the normalisation mode's branching fraction has a high relative uncertainty, due to $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi\,)$ having a high relative uncertainty of $\sim$16% [22],

$$\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi\,) = \frac{(5.8 \pm 0.8) \times 10^{-5}}{f_{\Lambda_b^0}} = (3.2 \pm 0.51) \times 10^{-4},$$

where $f_{\Lambda_b^0}$ is the production fraction of Λ_b^0 baryons with a Run 2 value of 0.179 ± 0.013 as determined from Refs. [138, 139]. However, an updated and more precise measurement of $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi\,)$ is being performed, see Ref. [140]. Therefore, the more precise value of

$$\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi\,) = (3.63 \pm 0.03\,(\text{stat}) \pm 0.20\,(\text{syst}) \pm 0.27(frag)) \times 10^{-4},$$

has been used. Here the first uncertainty is statistical; the second, a systematic related to the LHCb detector; and the third, a systematic based on the fraction of Λ_b^0 baryons produced compared to B-mesons, $f_{\Lambda_b^0}/(f_u + f_d)$ [138], where $f_{u/b}$ is the fraction of B^-

and $\overline{B}^0$ produced respectively. This value has a much smaller relative uncertainty of $\sim 9\%$. However, the final results can also be decoupled entirely from the $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ measurement by directly determining the upper limit from a ratio,

$$r = \frac{\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})}{\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)}. \tag{6.0.2}$$

This would also allow the value of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ to be re-evaluated for any future measurement of the normalisation mode's branching fraction. This is due to r being calculable without explicit knowledge on $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$, since Equation 6.0.1 becomes,

$$r = \mathcal{B}(J/\psi \to \mu^+\mu^-) \cdot \frac{N_{rare}}{N_{norm}} \cdot \frac{\varepsilon_{norm}}{\varepsilon_{rare}}. \tag{6.0.3}$$

The signal yield of the rare mode, N_{rare}, will ultimately be determined by a simultaneous fit to the invariant mass of the Λ_b^0 candidate over the range $[4720, 6400]\,\mathrm{MeV}/c^2$ for the eight analysis categories. This used a Double-Sided Crystal Ball (DSCB) function [141] to describe the signal mass shape, and the combination of two RooJohnson functions, which are implementations of Johnson's S_U-distribution [142] and an exponential to describe the background. The two RooJohnsons represent the contribution of exclusive backgrounds that remain after selection, in particular, $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ and $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$, and the exponential describes the background that arises from random combinations of tracks, which is known as the combinatorial background. The normalisation channel was also simultaneously fitted in similar analysis categories as $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ in order to obtain the associated yields, N_{norm}. However $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ should not be strongly affected by bremsstrahlung and so is not split by bremsstrahlung category. This fit used the same shapes for signal and combinatorial background as the fits to $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$: a DSCB and an exponential. However, the fit model also included a particle misidentification background from the decay $B^0 \to K_S^0 J/\psi\,(\to \mu^-\mu^+)$ that is described by a RooJohnson function.

The mass variables that are used throughout the analysis, *e.g.* in the fits, are produced by Decay Tree Fitter (DTF), a DaVinci reconstruction algorithm that uses additional constraints to improve the mass resolution. For the signal mode these additional constraints ensure that the event's tracks align with the primary vertex and that the reconstructed mass of the Λ baryon used to determine the Λ_b^0 mass is constrained to the known world average value. Unless specified, "constrained $m(p\pi e\mu)$" refers to this case. The normalisation mode includes these constraints alongside an additional one where the dimuon mass is constrained to the world average value for the J/ψ mass.

This analysis was still blinded at the time of writing. Therefore, an expected upper limit of the branching fraction of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ in the absence of signal in the currently

blinded region was determined using the CL_s method [143, 144] with a background-only proxy dataset and presented in Section 9.4. Additionally, in Section 9.6 this limit was evaluated for a case where a Λ_c^+ veto was included to eliminate the exclusive backgrounds in the rare mode's fit model, leaving only the combinatorial background. This provides a much cleaner dataset, but strongly reduces the sample size of the remaining data.

Chapter 7

Event Selection

7.1 Stripping Selection

The $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ analysis makes use of LHCb's stripping lines, which impose loose selection criteria to identify candidates that are consistent with the rare mode or the normalisation mode, $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$. These selections are very similar to one another and are summarised in Table 7.1.

These stripping selections make use of low-level reconstructed kinematic variables such as the p_T of the leptons, either in isolation or as a reconstructed di-lepton pair. Additionally, the selection uses a variety of χ^2 variables which measure the goodness-of-fit for certain aspects of the reconstruction. For example, χ^2_{IP} is defined as the difference in χ^2 relative to the primary pp collision vertex with and without the considered particle. The χ^2_{FD} also gives a measure of track-vertex separation through the flight distance of a particle with respect to a vertex, as does χ^2_{vtx} for the di-lepton pair, indicating the extent to which the leptons originate from the same decay vertex using their Distance Of Closest Approach (DOCA). Therefore, these χ^2 cuts in Table 7.1 ensure that the particle chain of a b-hadron to another b-hadron and two leptons has been reconstructed appropriately. Additionally, the reconstructed mass of the Λ baryon candidate is required to be close to its known value (see Table 7.1). This, in combination with a mass cut on the di-lepton pair, implies a loose requirement on the mass of the primary b-hadron of $\lesssim 6600\,\mathrm{MeV}/c^2$.

Candidate tracks should point to the Primary Vertex (PV) with which they are associated. This is ensured by the minimum requirement on its DIRection Angle (DIRA), the cosine of the angle between the momentum vector of the b-hadron candidate (the Λ_b^0) and the vector between the PV and the b-hadron's decay vertex. Finally, there are also some loose PID requirements, the presence of an electron for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ using

$DLLe > 0$ (see Section 3.5) or the presence of a muon for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ using `isMuon` and `HasMuon` (see Section 3.2.3).

Table 7.1: Requirements of the `Bu2LLK_meLine` and `Bu2LLK_mmLine` stripping lines used for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ respectively. When a cut is different, the cut for DD candidates is reported in parentheses.

Particle	**Requirement**
e, μ	$p_{\mathrm{T}} > 350\,\mathrm{MeV}/c$
	$\chi^2_{IP} > 9$
e	$\mathrm{DLL}e > 0$
μ	`HasMuon` & `isMuon`
$e\mu, \mu^+\mu^-$	$p_{\mathrm{T}} > 0\,\mathrm{MeV}/c$
	$m <5500\,\mathrm{MeV}/c^2$
	$\chi^2_{vtx} < 9$
	$\chi^2_{FD} > 16$
	$\chi^2_{\mathrm{IP}} > 0$
p, π	$p > 2\,\mathrm{GeV}/c$
	$p_{\mathrm{T}} > 0\,\mathrm{MeV}/c$
	$\chi^2_{IP} > 9(4)$
Λ	$\lvert m - m_\Lambda\rvert < 35(64)\ \mathrm{MeV}/c^2$
	$\chi^2_{vtx} < 30(25)$
	$p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$
Λ_b^0	$\chi^2_{vtx} < 9$
	$\chi^2_{IP} < 25$
	$\mathrm{DIRA} > 0.9995$
	$\chi^2_{FD} > 100$

The stripping selections provide data samples consisting of a b-hadron that decays to a Λ and a lepton pair containing an electron for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ or a muon for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$. These datasets are reconstructed fully for the particular decay specified using DaVinci, and then processed by applying the pre-selection described in Section 7.2 to obtain the first stage of the analysis tuples. These tuples were used to develop the dedicated selection for both modes, as described in the remainder of this chapter. Afterwards, the finalised data and MC tuples with the full selection already applied were created from these first stage tuples, which are used for fitting (Chapter 8) and limit determination (Chapter 9).

7.2 Pre-selection

The pre-selection uses a series of loose cuts on top of the stripping requirements to reject unphysical backgrounds arising from mis-reconstructed tracks that could mimic the signal but are not physically possible in reality. For example, if the two final state leptons have the same charge this would violate charge conservation. This Same Sign (SS) lepton data sample can only arise from a random combination of tracks, which is the combinatorial background by definition. Therefore, the SS case can be selected and used to cross-check the effect of the pre-selection on the combinatorial background compared to the Opposite Sign (OS) lepton data sample. Additionally, the pre-selection rejects correctly reconstructed decays that are sufficiently similar to the signal channel, and have a production branching fraction and selection rate such that they represent a significant source of background. For example, the normalisation mode acts as a very strong background to $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ due to the high number of $J/\psi \to \mu^+\mu^-$ decays that occur at LHCb. This can be clearly shown through the q^2 spectrum in Figure 7.1 that has strong peaks around resonances such as the J/ψ and $\psi(2S)$. For this reason in the pre-selection for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ the J/ψ q^2 region is vetoed, and for the normalisation mode the same region is instead selected. In this way, relatively pure $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$ samples are obtained. Figure 7.2 shows the J/ψ veto relative to the J/ψ mass peak in $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$ MC. The veto window fully encapsulates the J/ψ peak and so should effectively eliminate the contribution of the normalisation mode in the signal channel.

Table 7.2 shows the pre-selection cuts that are used for the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ analysis. The same selection is applied to the normalisation mode, $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$, to ensure the two decays are treated as similarly as possible so that many potential sources of systematic uncertainty cancel in the ratio of Equation 6.0.1. There are three exceptions to this, the first is the lack of inclusion of the J/ψ veto as discussed previously. Secondly, the normalisation mode contains only muons in the final state, which means that a cut developed based on the HOP mass would not remove partly reconstructed background as the α_{HOP} should always be one by definition in that case (see Section 3.2). Finally, the detector acceptance cuts, which act as a form of PID, differ between the signal and normalisation modes. The $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ decay detector acceptance cuts refer to there being an electron detected within the calorimeters (`e_InCaloAcc` and `e_HasCalo`), whereas the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$ equivalent requires a muon detected in the muon stations (μ`_InMuonAcc` and μ`_HasMuon`).

The Λ fiducial cuts are meant to ensure a well-reconstructed Λ candidate. A lower cut on the Λ_b^0 invariant mass allows the removal of combinatorial background in a region

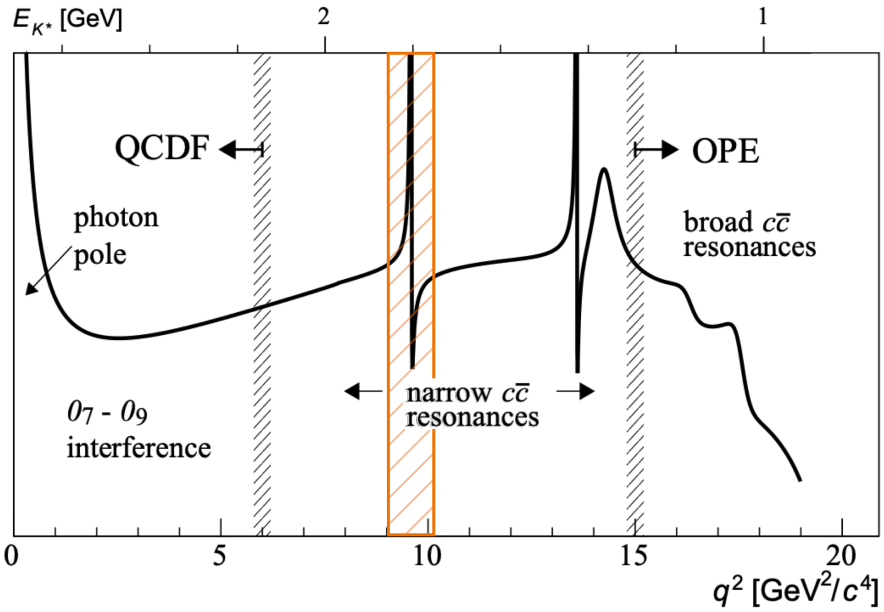


Figure 7.1: A schematic representation of the differential decay rate of a $b \to s\ell^-\ell^+$ decay, *e.g.* $B^0 \to K^*\mu^-\mu^+$, as a function of q^2. Highlighting several key features of the q^2 spectrum: The photon pole, where a di-electron is pair-produced by a photon; the narrow $c\bar{c}$ resonances, J/ψ and $\psi(2S)$ peaks; and the broader $c\bar{c}$ resonances at higher energies [145]. Overlaid in orange, the central hatching marks the window defining the J/ψ candidates to either be selected or vetoed for the rare and normalisation mode respectively.

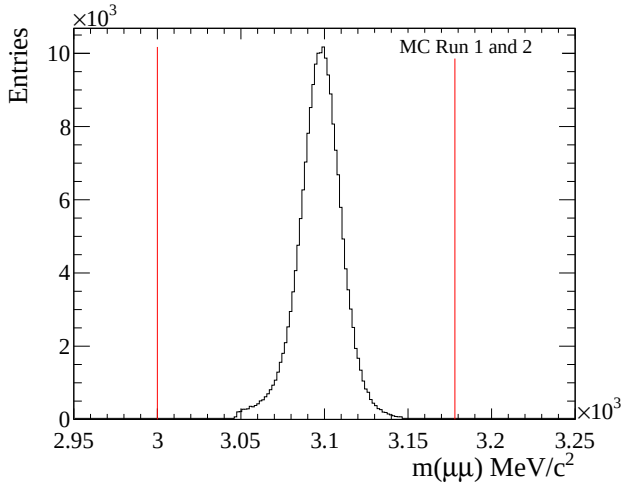


Figure 7.2: Matched MC $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ candidates for the combination of Run 1 and Run 2 samples passing the selection used in the normalisation fits. The red vertical lines indicate the J/ψ veto boundaries.

Table 7.2: Offline pre-selection cuts applied on $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$. The cuts that differ for $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ are given in the braces, or the description highlights whether that category of cuts apply solely to the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ case.

Description	**Cut**
HOP (for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$)	$M_{HOP}(\Lambda_b^0) > 2900 + 170 \cdot \log(\chi^2_{FD}(\Lambda_b^0))\,\text{MeV}/c^2$
J/ψ veto (for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$)	$m(e^+\mu^-) \notin \left[\sqrt{9.0}, \sqrt{10.1}\right]\,\text{GeV}/c^2$
$\Lambda\ell^+\ell^-$ mass	$4720\,\text{MeV}/c^2 < m(\Lambda\ell^+\ell^-) < 6400\,\text{MeV}/c^2$
Λ mass	$\lvert m(p\pi) - 1115.68\rvert < 10\,\text{MeV}/c^2$
Λ Fiducial	$\chi^2_{FD}(\Lambda) > 0$ DIRA $(\Lambda) > 0$ $0\,\text{mm} < Z_{OWN\,PV}(\Lambda) < 2250\,\text{mm}$ $0.5\,\text{ns} < t(\Lambda) < 2000\,\text{ns}$
Detector Acceptance	e_InCaloAcc{μ_InMuonAcc} e_HasCalo{μ_HasMuon}
PID fiducial cuts	$p_T(\mu, e) > 500\,\text{MeV}/c$ $p(\mu, e) > 3\,\text{GeV}/c$
Optimised PID cuts	μ_ProbNNmu > 0.2 e_ProbNNe > 0.2

which is not needed for its estimation, being too far from the rare mode's mass peak. The Λ mass cut improves isolation of decays with a resonant hadronic structure. Detector acceptance criteria require evidence of activity from particles in specific sub-detectors that are used to compute quantities such as the PID variables. Lastly, the HOP cut exploits the correlation between the flight distance of the Λ_b^0 and M_{HOP} [104, 146], as mentioned in Section 3.2, to isolate efficiently the signal from potential combinatorial or partly reconstructed backgrounds. This last cut was derived empirically and then re-examined using the procedure described in Section 7.2.1.

7.2.1 Optimisation of the HOP Cut

Figure 7.3 shows the distribution of M_{HOP} against $\log(\chi^2_{FD}(\Lambda_b^0))$ for the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ MC and the background from data sidebands, for both Run 1 and Run 2. It can be observed from these figures that the background and signal populate distinct regions in this plane, which can be exploited by the inclusion of a simple linear cut that maximises the retained signal candidates while reducing the amount of background significantly.

In order to optimise the background rejection and the signal efficiency of this linear

cut, a cut of the form $M_{HOP}(\Lambda_b^0) > \alpha + \beta \cdot log(\chi^2_{FD}(\Lambda_b^0))$ was performed. For pragmatic reasons a single cut for Run 1 and Run 2 was desired, however, to explore the impact of this choice separate optimisation for both periods was used. To do so the values of α and β were varied while monitoring the resultant background rejection, signal efficiency and signal to background ratio. The scan range of the values were $\alpha \in [50, 5000]\text{MeV}/c^2$ in $50\,\text{MeV}/c^2$ increments and $\beta \in [5, 500]\text{MeV}/c^2$ in $5\,\text{MeV}/c^2$ increments, and were performed alongside the other pre-selection cuts as summarised in Table 7.2, where a single value for α and β was selected for both Run 1 and Run 2. The results of this study are summarised in Table 7.3, demonstrating that this current cut gives a signal efficiency of 98% and background rejection of ∼25% for both Run 1 and Run 2. If Run 1 and Run 2 are considered separately a $\sim 3-4\%$ improvement in background rejection can be found for just a 0.3% decrease in signal efficiency. However, this would also introduce a difference in the pre-selection applied to both Run 1 and Run 2 which is not present for any of the other cuts in the pre-selection. The improvements to performance are relatively minor and the cut already provides a clean dataset for training a multivariate classifier without any obvious signs of partly reconstructed backgrounds. Therefore, to avoid introducing Run 1/Run 2 differences in the pre-selection, it was decided to proceed with the cut with $\alpha = 2900\,\text{MeV}/c^2$ and $\beta = 170\,\text{MeV}/c^2$.

Table 7.3: HOP cuts that provide 99.9%, 99.8%, 99.5%, 99.0% and 98.0% signal efficiencies for Run 1 and Run 2 separately alongside its background efficiency. The values for the currently selected cut in the pre-selection are highlighted in green.

	HOP-FD Cut	Signal Efficiency	Background Efficiency	Signal/Background Ratio
Run 1	$M_{HOP}(\Lambda_b) > 2900 + 170 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9831	0.7597	1.294
	$M_{HOP}(\Lambda_b) > 1050 + 270 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9990	0.9880	1.011
	$M_{HOP}(\Lambda_b) > 1200 + 275 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9980	0.9769	1.022
	$M_{HOP}(\Lambda_b) > 1550 + 275 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9951	0.9338	1.066
	$M_{HOP}(\Lambda_b) > 2200 + 230 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9900	0.8575	1.155
	$M_{HOP}(\Lambda_b) > 3100 + 155 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9800	0.7204	1.360
Run 2	$M_{HOP}(\Lambda_b) > 2900 + 170 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9834	0.7668	1.282
	$M_{HOP}(\Lambda_b) > 2900 + 5 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9990	0.9870	1.012
	$M_{HOP}(\Lambda_b) > 1100 + 300 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9980	0.9746	1.024
	$M_{HOP}(\Lambda_b) > 2350 + 170 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9950	0.9323	1.067
	$M_{HOP}(\Lambda_b) > 3700 + 10 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9901	0.8594	1.152
	$M_{HOP}(\Lambda_b) > 3300 + 125 \cdot \log(\chi^2_{FD}(\Lambda_b))$	0.9800	0.7254	1.351

An additional check was performed to ensure that this cut also does not distort the remaining background significantly, by comparing the mass shape of the Λ_b^0 in data sidebands and on same-sign data when the full selection is applied with and without the inclusion of the optimised M_{HOP} cut. Figures 7.4 and 7.5 show that this cut has minimal

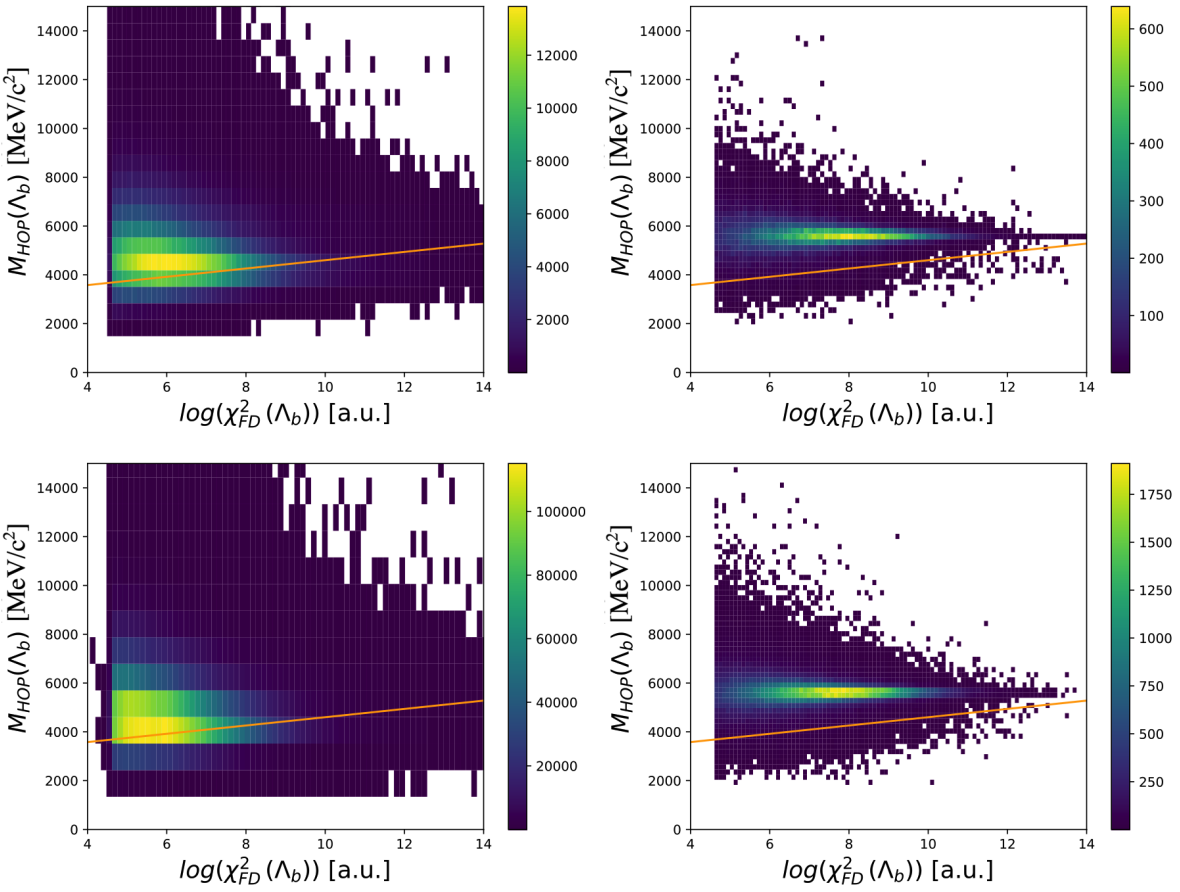


Figure 7.3: 2D histograms of the χ^2 of Λ_b^0 flight distance versus M_{HOP} for the background (Left) and $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ MC (Right) for Run 1 (Top) and Run 2 (Bottom). The HOP cut used in the pre-selection is overlaid as the solid orange line.

effect on the background shape, with the lower mass sideband only partially affected for both the SS and OS datasets.

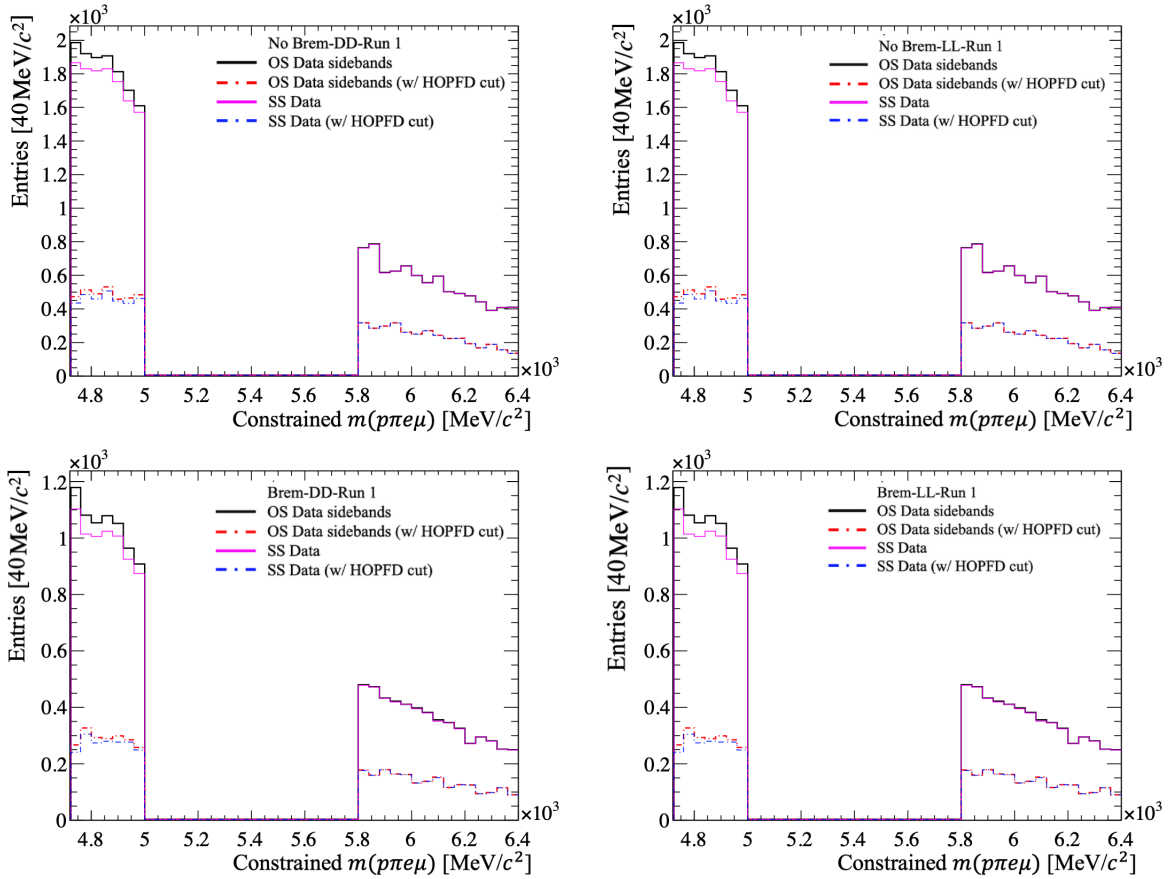


Figure 7.4: The Λ_b^0 mass shape for SS and OS and datasets with the full selection applied with and without the inclusion of the M_{HOP} cut for Run 1 DD (LL) on the left(right).

7.2.2 PID Selection

As discussed in Section 7.1 the stripping line itself includes loose PID cuts that ensure the final samples are consistent with the required decay topology, as summarised in Table 7.1. Detector acceptance criteria that require activity in specific sub-detectors contribute implicitly to PID requirements. Further PID cuts were also developed as part of the pre-selection to identify signal mode decays. For data, PID cuts are imposed directly, in contrast to simulated data where PID weights are applied based on efficiencies calculated using the PIDCalib software package. These PID weights are used for selection alongside additional weights *e.g.* trigger weights, discussed in Sections 7.2.3 and 7.3.

The PID selection was optimised by performing a 2D scan over cuts to the ProbNN variables (see Section 3.5) for both leptons to find the maximum value of $S/\sqrt{B}$. In this

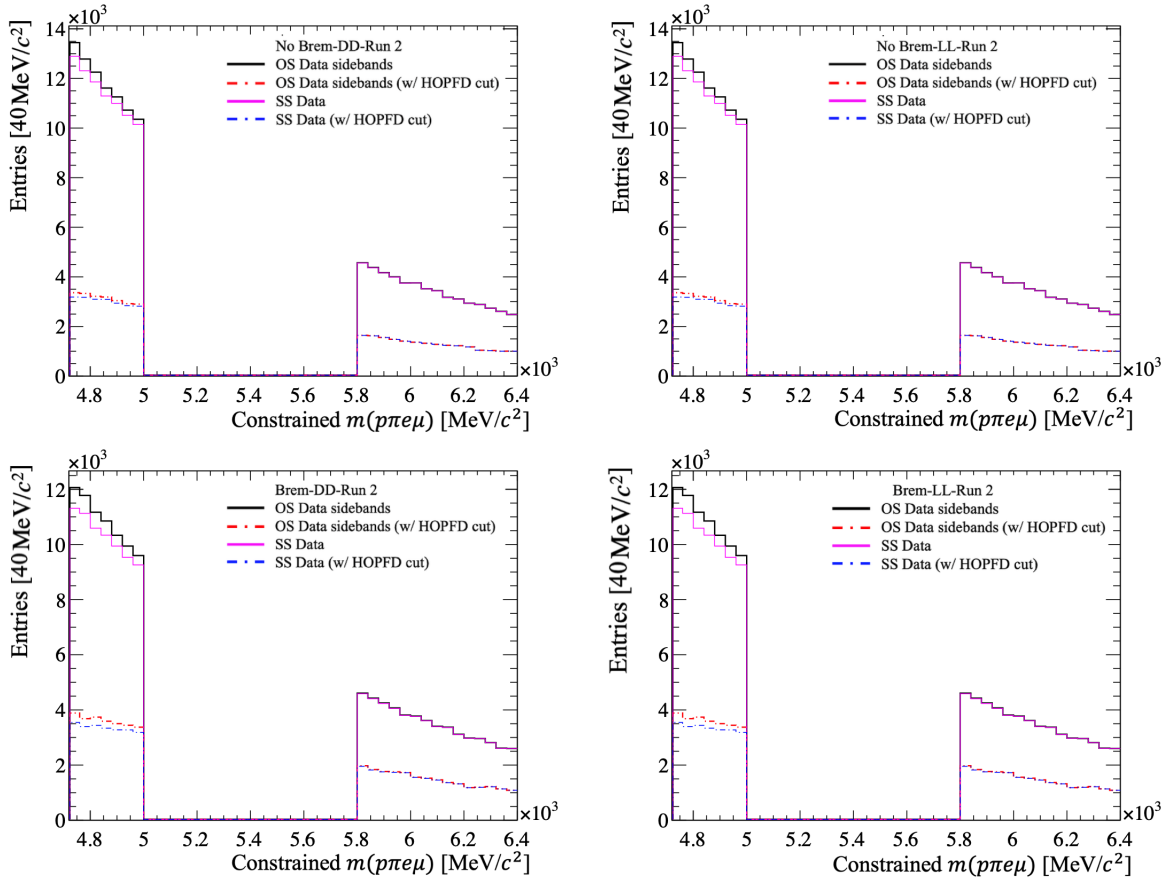


Figure 7.5: The Λ_b^0 mass shape for SS and OS datasets with the full selection applied with and without the inclusion of the M_{HOP} cut for Run 2 DD (LL) on the left(right).

figure of merit, the signal yield, S is evaluated using a MC sample and the background yield, B, is obtained from the upper mass data sideband ($m(\Lambda_b^0) > 5800\,\text{MeV}/c^2$). The ProbNN variables were used because the alternative cuts with the DLLe variable gave a lower sensitivity. Extending this treatment to a 3D scan in which proton PID was also used did not improve signal sensitivity significantly. Additionally, no sample was available for DD Λ candidates to fully optimise this and so no proton PID cut was used for the PID selection. The resulting values for the optimised PID cuts only differ from Table 7.2 for the muon PID cut, which was found to be 0.1. Subsequent to these studies, a more stringent muon PID requirement of μ_ProbNNmu > 0.2 was adopted, for uniformity with a similar rare decay measurement [40, 41]. This is the value that was used for the pre-selection cuts in Table 7.2. This tighter cut value also has the benefit of aligning with HLT2 trigger requirements, and effectively decoupling the PID and trigger from one another in data.

The normalisation channel contains two muons and therefore differs from the signal channel by requiring that both leptons satisfy the optimised PID criteria for muons. However, the normalisation channel has an observed background from a misidentified

particle, $B^0 \to K^0_S J/\psi(\to \mu^-\mu^+)$, where a pion from the $K^0_S \to \pi^+\pi^-$ decay has been mistakenly identified as a proton and therefore mimics the normalisation channel's final state. As this source of potential background is separable from genuine normalisation mode decays and can be modelled well, see Section 8.2 for details, no additional PID cuts were required for the normalisation mode.

7.2.3 Trigger Selection

For both $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ decays the trigger selection only includes muons in data and MC. This is due to the clean signal that a muon provides and has the benefit of strongly reducing the impact of di-electron backgrounds, as seen in Section 8.1. The trigger lines used for the signal and normalisation modes are shown in Tables 7.4 and 7.5 respectively. The information of the selection of trigger lines can be obtained from the analysis of the Trigger Configuration Keys used for the sample. These TCKs are unique identifiers that tag the particular settings and selections used at that particular time, which are varied occasionally to account for changes in data-taking conditions. Simulated samples are typically generated using a single TCK that is representative of the conditions in a given year. The selections for the TCKs for different years are given in Appendix A. It is essential for the evaluation of selection efficiencies that the conditions in simulation model those in recorded data. In cases where the data trigger thresholds were more restrictive than those for the MC, the simulation thresholds were increased for an equivalent fraction of the MC. Conversely, for data where the MC has a more restrictive threshold, the corresponding thresholds are imposed on data during pre-selection, leading to 1.4% of the data being rejected.

Table 7.4: Trigger lines used for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ candidates.

Trigger Level	**Trigger Lines**		
	Run 1	**2015**	**2016, 2017, 2018**
L0	Muon	Muon	Muon
HLT1	TrackAllL0 TrackMuon	TrackMVA TrackMuon	TrackMVA TrackMuon TrackMuonMVA
HLT2	Topo(E/Mu)2BodyBBDT Topo(E/Mu)3BodyBBDT	Topo2Body Topo3Body	Topo(E/Mu/MuE)2Body Topo(E/Mu/MuE)3Body

Table 7.5: Trigger lines used for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ candidates. The L0 and HLT1 lines are identical to those in Table 7.4 and so are not replicated here.

Trigger	**Trigger Lines**		
Level	**Run 1**	**2015**	**2016, 2017, 2018**
HLT2	Topo(Mu)2BodyBBDT	Topo2Body	Topo(Mu/MuMu)2Body
	Topo(Mu)3BodyBBDT	Topo3Body	Topo(Mu/MuMu)3Body

At the hardware trigger level L0Muon is required, meaning a high p_{T} muon candidate is present in the event. In particular L0Muon requires the largest p_{T} of the eight candidates[1] to exceed a p_{T} threshold, these thresholds are given in Table A.1.

For HLT1 the triggers Hlt1TrackAllL0 and Hlt1TrackMVA are required for the leptons in Run 1 and Run 2 respectively. The Hlt1TrackAllL0 line triggers when there is a displaced muon track, as indicated by their p_{T} and χ^2_{IP} values [147]. The exact values are given in Table A.2. The Hlt1TrackMVA line supersedes Hlt1TrackAllL0, by using a 2D selection in χ^2_{IP} and p_{T} (in GeV/c) [147],

$$\log(\chi^2_{IP}) > \frac{1}{(p_{\mathrm{T}}-1)^2} \cdot \frac{b}{25}(25 - p_{\mathrm{T}}) \cdot \log(7.4), \tag{7.2.1}$$

where the tracks have $1\,\mathrm{GeV}/c^2 \leq p_{\mathrm{T}} \leq 25\,\mathrm{GeV}/c^2$ and $\chi^2_{IP} > 7.4$, and b is a parameter used to define either a looser or tighter cut. The values for p_{T}, χ^2_{IP} and b are given in Table A.3. The Hlt1TrackMuonMVA line uses a similar selection to Hlt1TrackMVA but with more relaxed requirements, and was used throughout 2016 [147]. Table A.5 gives the exact values for the selection. The Hlt1TrackMuon trigger line is also used, this accepts candidates in a similar way to Hlt1TrackAllL0 with looser requirements as given in Table A.4.

For HLT2 the trigger lines involve the use of multi-body track reconstruction algorithms to select candidates with the correct topology. These 'Topo' lines are discussed in Section 3.6. These too have additional selections in the trigger lines, but instead of kinematic cuts these are PID cuts, given in Table 7.6. These PID requirements are relatively loose and align well with the optimised PID cuts described in Section 7.2.2.

Figures 7.6 and 7.7 show the distribution of the invariant mass of the reconstructed Λ_b^0 candidates after all the pre-selection cuts described in Table 7.2 have been applied. At this stage, this appears to follow an exponential shape and thus be dominated by combinatorial background.

[1] There are four L0 muon processors trying to identify two muon tracks leading to eight candidates [108].

Table 7.6: PID selection overview of the used trigger lines for both Run 1 and Run 2.

Trigger line	PID requirement	Datasets
TopoE(2/3)BodyBBDT	PIDe> −2	2011, 2012
TopoMu(2/3)BodyBBDT	`isMuon` & `HasMuon`	2011, 2012
TopoE(2/3)Body	ProbNNe> 0.2	2016–2018
TopoMu(2/3)Body	`isMuon` & `HasMuon`	2015–2018
TopoMuMu(2/3)Body	2x (ProbNNMu> 0.2 & `isMuon` & `HasMuon`)	2016–2018
TopoMuE(2/3)Body	(ProbNNMu> 0.2 & `isMuon` & `HasMuon`) & ProbNNe> 0.1	2016–2018

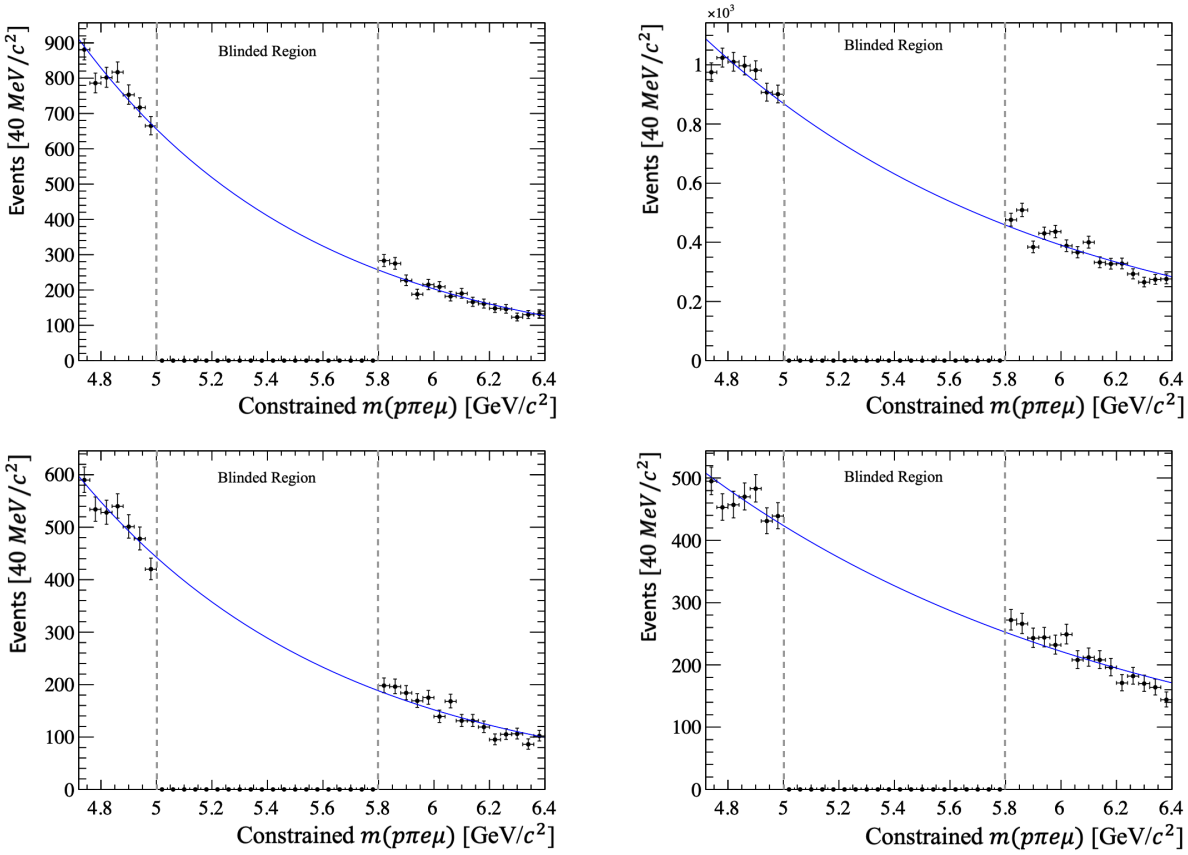


Figure 7.6: Invariant mass distribution of the reconstructed Λ_b^0 for Run 1 candidates with the pre-selection cuts applied and fitted to an exponential. The categories without (with) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column.

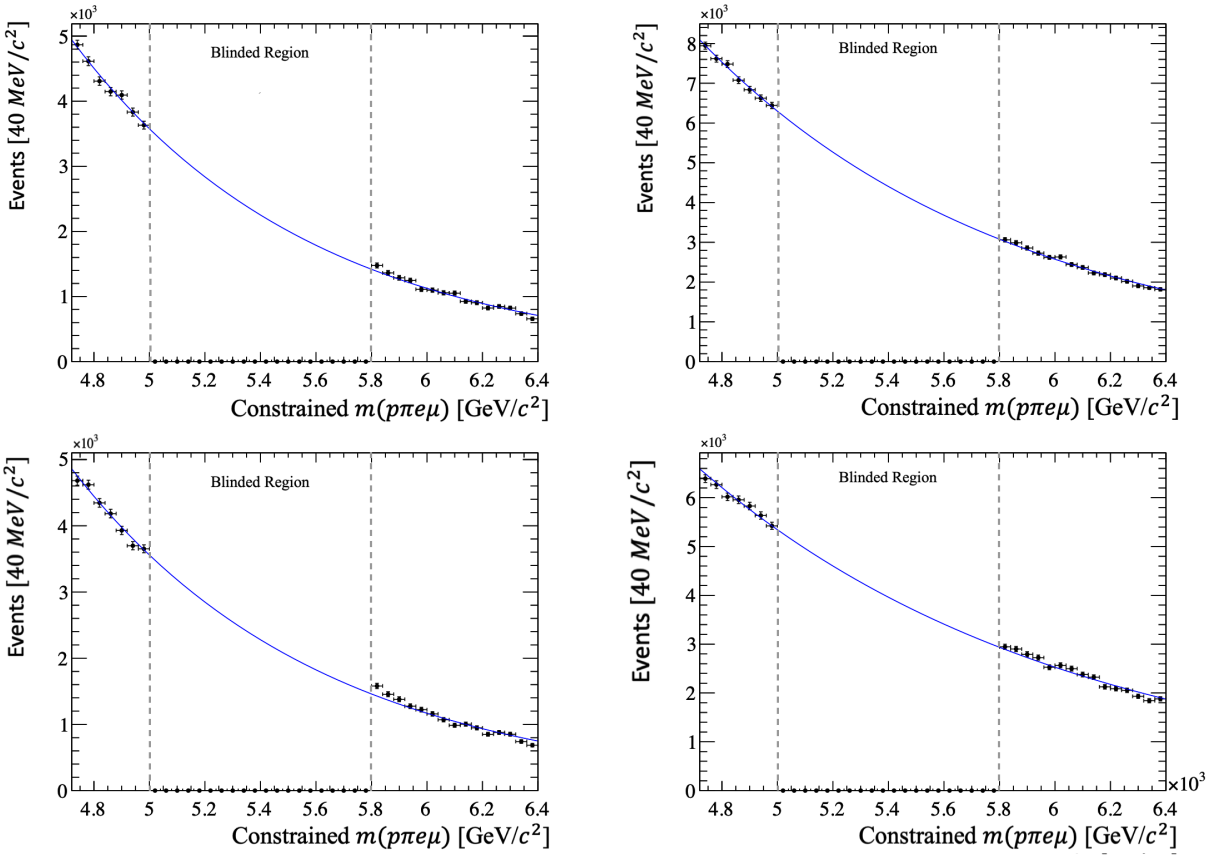


Figure 7.7: Invariant mass distribution of the reconstructed Λ_b^0 for Run 2 candidates with the pre-selection cuts applied and fitted to an exponential. The categories without (with) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column.

7.3 Re-weighting the Simulated Samples

The MC samples were used in three main ways in this analysis: to provide an example of 'true' signal mode candidates when training a multivariate classifier (see Section 7.4); to estimate the selection efficiencies of the rare and normalisation channels; and to enable characterisation of the mass distributions of signal and normalisation modes. However, the simulations do not always provide an accurate representation of data. These data-MC differences may arise from information that was not available when the simulation was performed or detector effects, such as kinematics, tracking and PID, not being accurately modelled. Therefore, to improve the agreement between data and MC a set of weights were applied to both $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ and $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ MC samples. These weights were determined in an iterative way, where the distribution of the N^{th} variable is weighted successively to the previous $N-1$ weights already applied.

Six weights were used relating to: the Λ_b^0 baryon lifetime; the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ angular distribution; the lepton PID efficiency; tracking efficiency; trigger efficiency; and Λ_b^0 production kinematics. These shall be discussed in further detail for the remainder of this section. The $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ angular weights are only included for the normalisation channel, and the tracking efficiency weights are only included for long tracks.

In order to evaluate the validity of this re-weighting procedure, the unweighted and weighted MC were compared to a signal-weighted (sWeighted) version of the data. These sWeighted datasets are produced through the sPlot method [148], which uses an extended likelihood analysis of a data sample to attempt to discriminate the signal and background components. By providing Probability Density Functions (PDFs) for each component when performing the sPlot, a set of weights can be determined for each and are applied to the original dataset. If the signal weights are used, this produces a signal-like distribution as the candidates that are more signal-like according to the sPlots are given a greater weighting. This is a useful baseline for comparison with the MC as this sWeighted data does not have any of the mis-modelling effects, to which the MC would be prone. Therefore, if the re-weighted MC and the sWeighted data are in good agreement, then so too should the re-weighted MC and the unweighted data.

7.3.1 Λ_b^0 Lifetime Weights

These weights were required as an improved world average value of the Λ_b^0 lifetime became available after the simulation samples had been generated and so these used an older value of $\tau(\Lambda_b^0)_{gen} = 1.451\,\text{ps}$ [149]. Therefore, the Λ_b^0 lifetime was then weighted to the updated world average value of $\tau(\Lambda_b^0)_{wa} = 1.470\,\text{ps}$ [150]. The associated weights are given analytically as,

$$w(t(\Lambda_b^0)) = \frac{\tau(\Lambda_b^0)_{gen}}{\tau(\Lambda_b^0)_{wa}} \cdot \exp\left[t(\Lambda_b^0) \cdot \frac{\tau(\Lambda_b^0)_{wa} - \tau(\Lambda_b^0)_{gen}}{\tau(\Lambda_b^0)_{wa} \cdot \tau(\Lambda_b^0)_{gen}} \right], \tag{7.3.1}$$

where $t(\Lambda_b^0)$ is the measured value for the Λ_b^0's decay time. Figure 7.8 shows the distribution of $t(\Lambda_b^0)$ as the weights are applied successively. This re-weighting has a relatively minor effect that improves the description of data by the simulation.

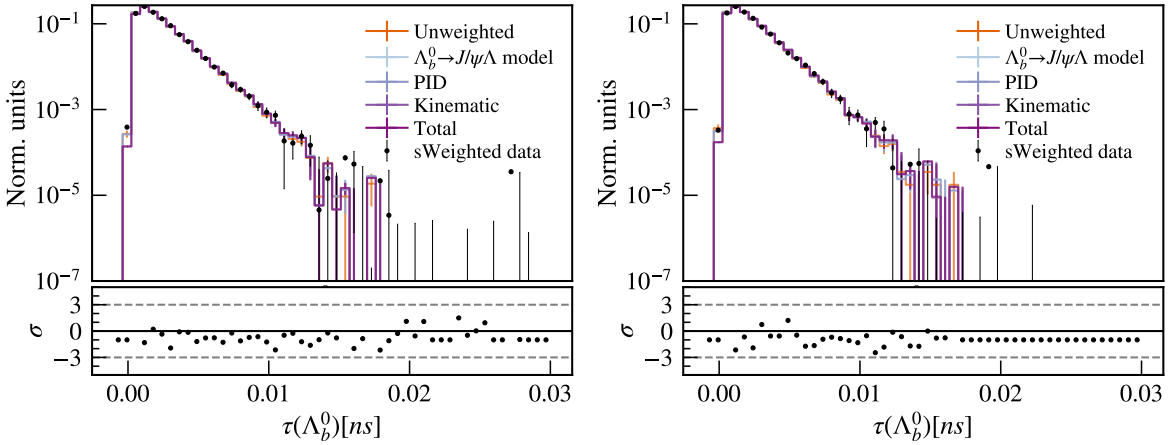


Figure 7.8: Distribution of $\tau(\Lambda_b^0)$ on sWeighted data (black points) and on simulation before (orange lines) and after (dark purple lines) the re-weighting. The left plot describes DD samples, while the right one describes the LL samples, both using Run 2 data. The pull distribution below each plot compares the sWeighted data to the fully re-weighted MC.

7.3.2 $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ Angular Distribution Weights

The rare and normalisation channels' MC were simulated by only considering the available phase space of the invariant mass of the Λ_b^0 decay. This was sufficient for the rare mode as the decay structure is unknown. BSM models could be used later to re-weight the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ MC and the impact of the model estimated by comparison with the unweighted MC sample.

However, for the normalisation channel there is a known angular distribution that affects its decay structure. This was measured in Ref. [151] by parameterising the decay with 20 angular moments. This angular distribution was then used to obtain a set of weights that remodel the $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ decay. Figure 7.9 shows the effect of the re-weighting procedure on the helicity angles[2] of $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$. This highlights the disagreement between data and MC before these weights, particularly in the angles θ_b and θ_l. The re-weighting procedure then ensures very good agreement between MC and data for these parameters.

[2]Full definitions of these Helicity angles and the 20 angular moments can be found in Ref. [151]. $\hat{n}$ is the unit vector of the vector product of the beam direction and the Λ_b^0 momentum vector; θ is the angle between $\hat{n}$ and the Λ direction in the Λ_b^0's rest frame; $\theta_{b,l}$ and $\phi_{b,l}$ are the polar and azimuthal angles for the Λ decay or the di-lepton system; and $\Delta\phi$ is the difference between ϕ_l and ϕ_b.

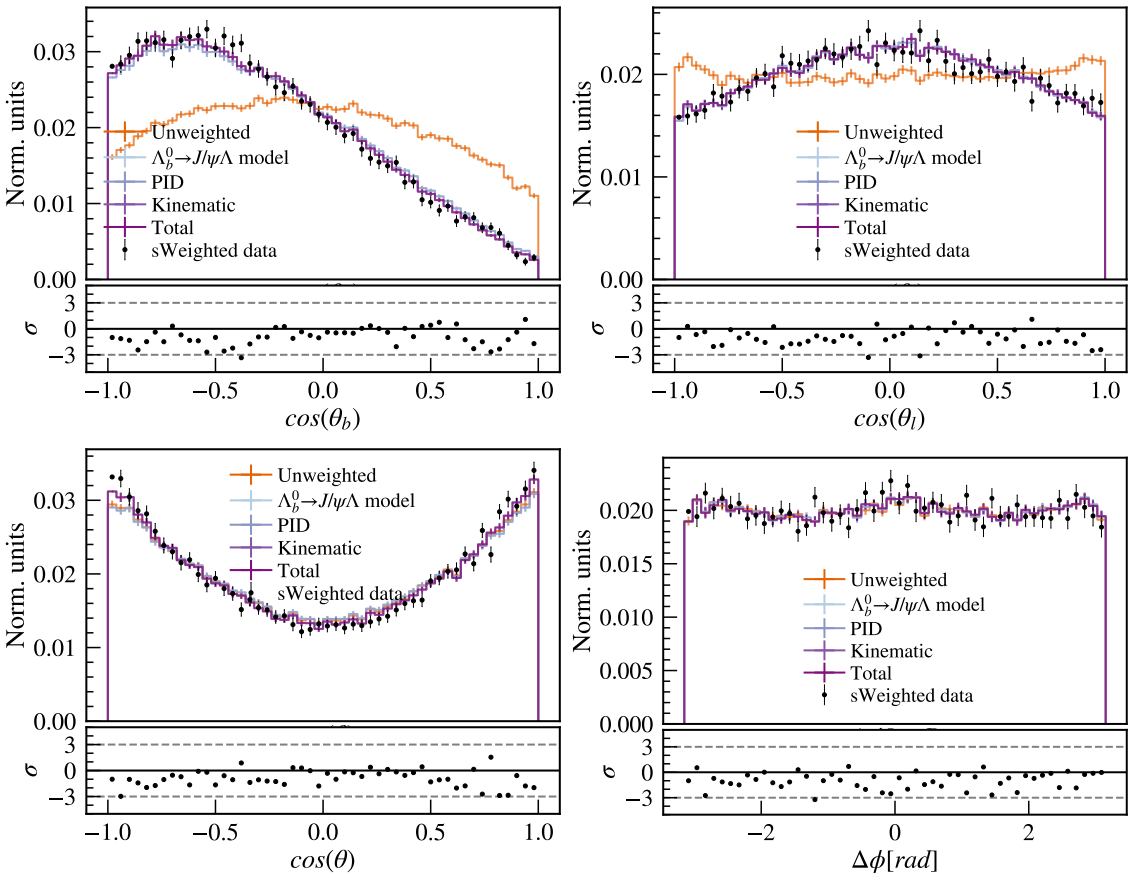


Figure 7.9: Distributions of the four helicity angles of the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$ decay, showing Run 2 data and re-weighted MC. From top-left to bottom-right: $cos(\theta_b)$, $cos(\theta_l)$, $cos(\theta)$, and $\Delta\phi$. The pull distribution below each plot compares the sWeighted data to the fully re-weighted MC.

7.3.3 Lepton PID Efficiency Weights

The software package PIDCalib [152] was used to obtain the efficiency of the PID selection outlined in Section 7.2.2. This uses large calibration samples for various particles *e.g.* electrons and muons, to determine the efficiency of the PID selection using data. These efficiencies are stored as histograms that were used to compute a per-candidate weight for each event in the MC. The resulting 'PID weights' were then included in the MC tuples with the pre-selection applied instead of specifically applying the PID cuts[3]. Figure 7.10 shows an example efficiency histogram for muon and electron PID. These used a binning scheme with isopopulated bins in p and p_{T}. Adjacent bins where the efficiency values differ by less than two standard deviations were merged[4].

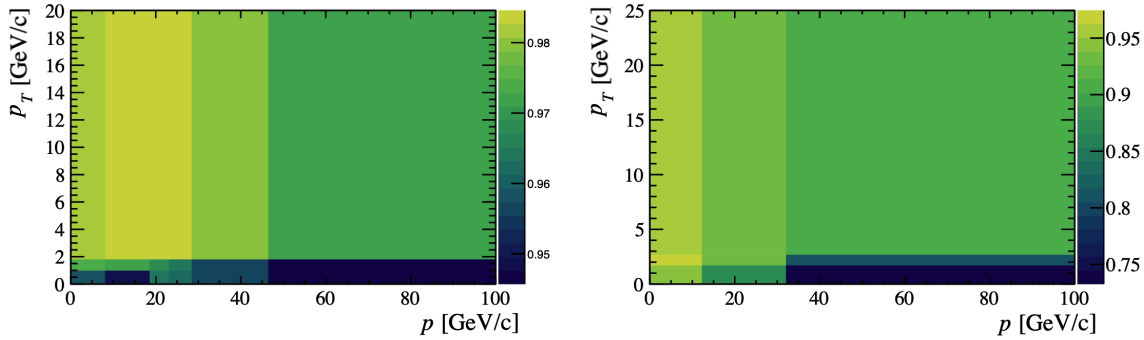


Figure 7.10: PID weight histograms for muons (Left) and electrons with bremsstrahlung recovered (Right) for 2018 MD data.

However, PIDCalib uses sWeights to produce its calibration samples, which are not optimal for electrons that have lower mass resolution due to bremsstrahlung effects. Therefore, the electron PID efficiencies were instead obtained using the 'Fit and Count' method. For this method, data samples from $B^+ \to K^+ J/\psi (\to e^+e^-)$ were used with a selection as close as possible to that given in Table 7.2 without the PID selection. These samples were split in bins of p and p_{T}, using the same binning scheme as the PIDCalib muons, with the relevant PID selection then applied. The candidates that get rejected by the selection are classified as 'fail' candidates, and the accepted ones as 'pass'. The pass and fail candidates are separately plotted and fitted, accounting for the potential background where pions are misidentified as kaons. These fits are shown in Figure 7.11,

[3]It was later found that the PID weights where bremsstrahlung was not reconstructed were incorrectly applied to both the Brem and No Brem categories. This is expected to have a small effect on the PID efficiencies that have been determined but this should not strongly alter the final results obtained.

[4]The final determined efficiencies were shown to be relatively insensitive to the choice of binning scheme.

and the corresponding yields were used to calculate the PID selection efficiency using,

$$\varepsilon^{e}_{PID} = \frac{N_{pass}}{N_{pass} + N_{fail}}. \tag{7.3.2}$$

These PID efficiencies can then be used in the same way as the efficiency histograms for the other samples to determine a set of PID weights for the electrons.

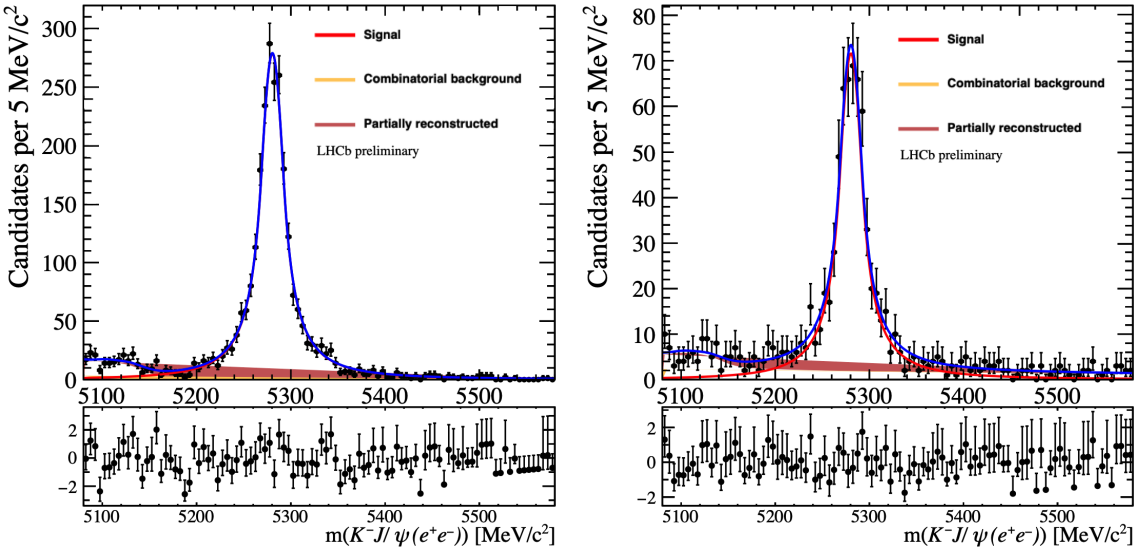


Figure 7.11: Invariant mass distribution for $B^+ \to K^+ J/\psi (\to e^+e^-)$ with fits overlaid for the pass (Left) and fail (Right) samples used to determine the electron PID efficiency, for electrons with bremsstrahlung recovered for 2018 MU.

7.3.4 Tracking Efficiency Weights

Small differences are observed in the tracking efficiencies estimated for data and simulation. To correct for these, the TrackCalib software package [103] was used to obtain weights for each Long-type track. Similarly to the PIDCalib software package, TrackCalib uses calibration samples to determine the tracking efficiency. The weights themselves are determined using 2D histograms of the track p and η, examples of these histograms are shown in Figure 7.12. In the case of electrons, the tracking efficiencies are obtained from the measured LHCb tracking efficiencies [153] using 3D histograms of the track angles ϕ and η, and the generator-level track p_{T}.

7.3.5 Trigger Efficiency Weights

Trigger efficiencies are determined using the TISTOS method on a $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ sample, as described in Section 3.6. In the associated fits the $B^+ \to K^+ J/\psi$ shape is de-

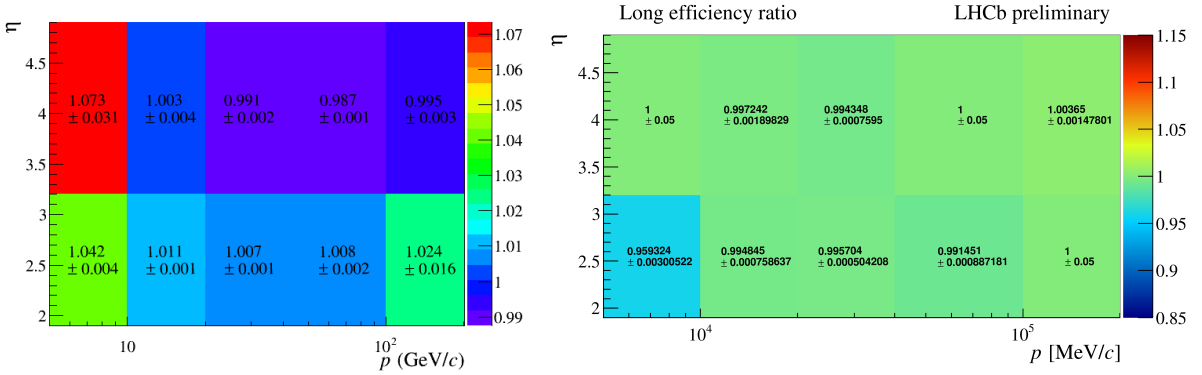


Figure 7.12: TrackCalib histograms showing tracking efficiency weights for long tracks in data for 2012 (Left) and 2016 (Right) [103].

scribed by a Gaussian function with power-law tails. The parameters of these tails are fixed to the values obtained from fits to samples of simulated data. The fits are performed on both MC and data with the data/MC ratio used to derive the trigger weights. These are stored in 2D histograms of $\max(p_T^{\ell_1}, p_T^{\ell_2})$ and $p_T^{\ell_1} \times p_T^{\ell_2}$, where ℓ_1, $\ell_2 = \mu$ for the normalisation mode and for the rare mode, $\ell_1 = e, \ell_2 = \mu$. These histograms are then used in the same way as the PID and Trigger efficiency histograms to determine the per-candidate Trigger efficiency weights to correct the MC. Figure 7.13 shows example histograms for 2015 DD data and MC.

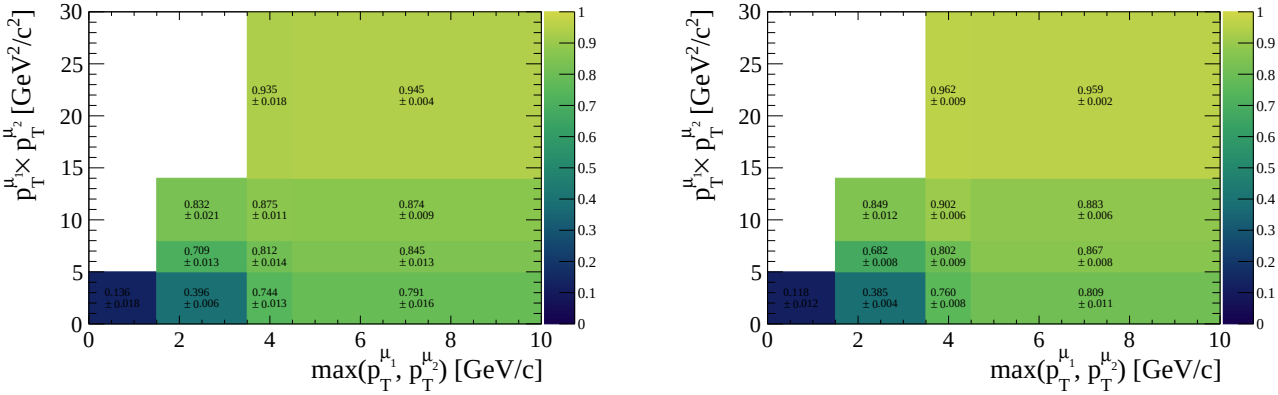


Figure 7.13: TISTOS efficiency histograms for the 2015 trigger selection for data (Left) and MC (Right).

7.3.6 Λ_b^0 Production Kinematic Weights

The production kinematics of b-hadrons are not correctly modelled in the MC samples, and so a re-weighting procedure was applied to correct this. A Gradient Boosted Reweighter (GBR) machine learning algorithm from the `hep_ml` package [154] was used to

determine the required kinematics weights. This was trained[5] based on sWeighted data and $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$ MC with the pre-selection and trigger requirements previously described applied. The variables $p_{\mathrm{T}}(\Lambda_b^0)$ and $\eta(\Lambda_b^0)$ were used as 'training' variables to attempt to discriminate between the data and MC candidates. This was done by providing known samples that were tagged as data or MC to an algorithm. Specified training variables are then used by the algorithm to determine if there was a relationship between them and allow for data and MC to be discriminated. This 'training' dataset, used to prime the algorithm, does not contain the full sample, and so, when given the full mixed dataset, there are novel candidates to be examined to separate the data and MC. The output of the GBR for data and MC can then determine a set of data-MC weights that correct for the kinematic differences of the Λ_b^0, as the GBR was explicitly trained on kinematic variables.

The GBR training was performed 100 times to account for any systematic fluctuations that may arise, showing a minimal effect ($\mathcal{O}(1)\%$). Additionally, the effect of re-weighting on other variables that were used in the analysis was examined to ensure that they are correctly aligned between data and MC. Figure 7.14 shows a set of kinematic variables with and without the re-weighting procedure performed, showing good agreement.

7.4 Selection from a Multivariate Analysis

A multivariate classifier, specifically a Gradient Boosted Decision Tree (GBDT) implemented as XGBOOST [155], was used to reduce the remaining (predominantly combinatorial) background while retaining as much signal as possible. This was performed with a selection that was optimised independently in each of the eight analysis categories.

A GBDT is a particular machine learning method that uses a 'tree' ensemble model with a set of functions that use information from the dataset to predict an output. Each of these trees represents a series of binary queries applied to an event from a dataset that assigns a score characterising the extent to which the event resembles signal or background. These binary queries that discriminate between signal and background are created using machine learning due to the highly complex nature of considering multiple variable distributions simultaneously. The samples used for training were chosen in a similar way to those described in Section 7.3, providing a training sample that tags whether an event is signal or background. The signal was represented by the weighted $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ MC, and the background from the corresponding data sidebands, with all analysis categories com-

[5]The configuration of the GBR was: 50 estimators; a learning rate of 0.1; a maximum depth of 3; 1000 minimum samples per leaf; subsample 0.4; and using k-folding with 2 folds.

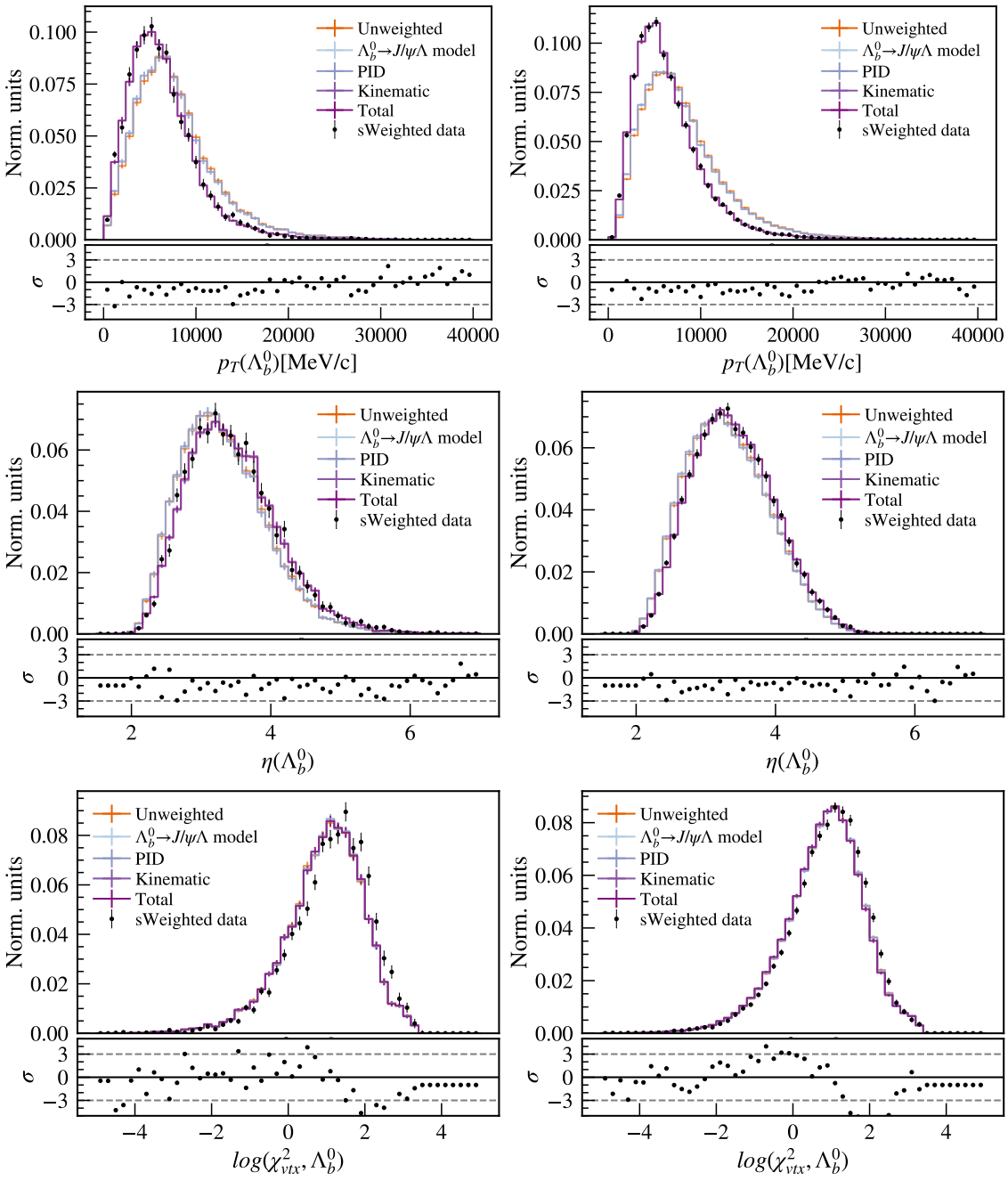


Figure 7.14: Distribution of $p_{\mathrm{T}}(\Lambda_b^0)$ (Top), $\eta(\Lambda_b^0)$ (Middle) and $\chi^2_{VTX}(\Lambda_b^0)$ (Bottom) on sWeighted data (black points) and on simulation before (orange lines) and after (dark purple line) the re-weighting. The left graphs show Run 1 DD samples, while the right shows Run 2 DD. The pull distribution below each plot compares the sWeighted data to the fully re-weighted MC.

bined in both cases. The trained XGBOOST algorithm then includes a bespoke classifier trained for this particular purpose that can score a particular unseen event on whether it is more like a $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ signal event or some form of background. A selection can then be applied to this classifier output to ensure that only candidates above a certain 'signal-like' threshold are retained and the rest rejected. In this way, the background can be significantly reduced.

The classifier used to distinguish the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ signal, which henceforth will be referred to as "The GBDT", used a total of 11 topological features as the discriminatory variables for the classifier. These variables were chosen because they are characteristic of the kinematics of the rare mode, whilst also being well modelled in the MC that is being used as the reference signal. These variables also scored highly in their individual ability to discriminate between signal and background, which is measured as a lower 'entropy'. Here entropy refers to a measure of the 'impurity' of a dataset, *i.e.* how much signal is miscategorised as background. For GBDT the scoring of the decision trees is done by minimising the entropy for each 'branch' and 'leaf' from the set of binary queries to minimise the entropy of the overall tree. Therefore, variables which themselves have a low entropy will be better able to discriminate signal and background. Additionally, the distributions of the 11 discrimination variables are themselves similar between the eight different analysis categories. This ensures that the classifier response is roughly uniform across the different categories, even if the optimised GBDT cuts are not. In XGBOOST multiple decision trees are combined into an ensemble that performs better than each individual tree, where the use of 'weakly' performant decision trees are combined sequentially to form an overall 'strongly' performant classifier. This is known as gradient boosting, as each tree in the sequence of weaker decision trees minimises the errors of the previous trees in an iterative way [155].

The 11 variables used in the classifier were[6]:

1. **Leptons vertex detachment**: The χ^2_{FD} of the final state leptons, a measure of the significance with which the production vertex of the two leptons is separated from the Λ_b^0 decay vertex.

2. $\boldsymbol{p_T(\Lambda_b^0)}$: The transverse momentum of the Λ_b^0.

3. $\boldsymbol{\chi^2_{IP}(\Lambda_b^0)}$: The χ^2_{IP} of the Λ_b^0.

4. $\boldsymbol{\alpha_{HOP}}$: The HOP scale factor[7].

[6] The definitions for χ^2_{IP}, χ^2_{FD}, and DIRA are given in Section 7.1
[7] As defined in Refs. [104, 146], see Section 3.2.

5. **Cone p_{T} asymmetry (sum)**: The sum of the transverse momentum asymmetries of the four final-state tracks, defined as $\sum_i \mathcal{A}_{p_{\mathrm{T},i}} = \sum_i \frac{p_{\mathrm{T}_i} - p_{\mathrm{T}_{cone,i}}}{p_{\mathrm{T}_i} + p_{\mathrm{T}_{cone,i}}}$, where p_{T_i} is the p_{T} of a final state particle, and $p_{\mathrm{T}_{cone,i}}$ is the sum of the p_{T} of other final state particles found within a cone of half-angle 0.25 rad around the track of particle i.

6. **DOCA(leptons)**: The distance of closest approach of the two leptons in 3D.

7. $\boldsymbol{\chi^2_{IP}}$**(leptons)**: The sum of the χ^2_{IP} of the two leptons.

8. **DIRA($\boldsymbol{\Lambda}$)**: The direction angle between the flight direction of the Λ and the direction of its momentum.

9. **Track isolation (sum)**: The sum of the track isolations of the four final-state tracks, defined as the number of other tracks found within a cone of half-angle 0.25 rad around the considered track.

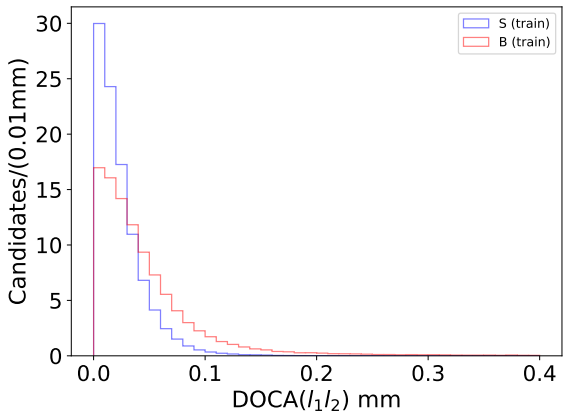
10. $\boldsymbol{\chi^2_{ORIVX}}$**($\boldsymbol{\Lambda}$)**: The difference in χ^2 relative to the reconstructed Λ origin vertex with and without the Λ.

11. $\boldsymbol{\Delta\eta}$**(hadrons)**: The difference in pseudorapidity of the two hadron tracks.

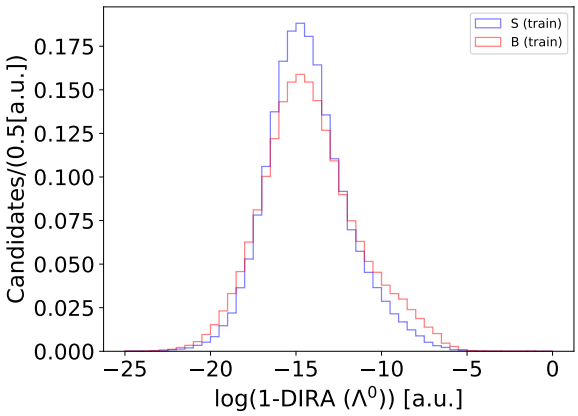
The distributions of these variables in the data sidebands (background) and the MC (signal) are shown in Figure 7.15. These distributions show that the signal and background can be distinguished visually in most cases, and therefore these were likely to be also performant from the perspective of the GBDT. Additionally, Figure 7.16 shows the linear correlations between each variable for both signal and background. There is minimal correlation between variables, except between the cone p_{T} asymmetry (sum) and track isolation (sum) variables that show a negative correlation of around -0.9. The relative importance of each of the variables to the classifier is shown in Figure 7.17.

The GBDT was trained[8] on a simulated signal sample against combinatorial background data taken from the mass sidebands around the expected signal. Both samples had the full pre-selection reported in this chapter applied to them, except the MC sample did not use the explicit PID selection and instead used PID weights. The MC was also re-weighted as described in Section 7.3, as these weights are accounted for during the training. Overall, the training sample consists of about $200,000$ candidates, half of signal and half of background. This size is dictated by the sample size of the available MC: the data candidates, being more abundant, are drawn randomly from the full samples without replacement to match the size of the MC. The ratio of the number of Run 1 to Run 2 candidates is scaled in MC to match the ratio observed in the data sample, in order to

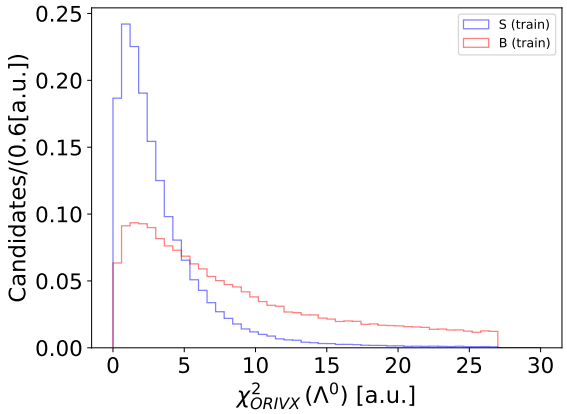
[8]The training uses a configuration with: a learning rate of 0.02; a minimum child weight of 2; a maximum depth of 3; and 511 trees in the ensemble.



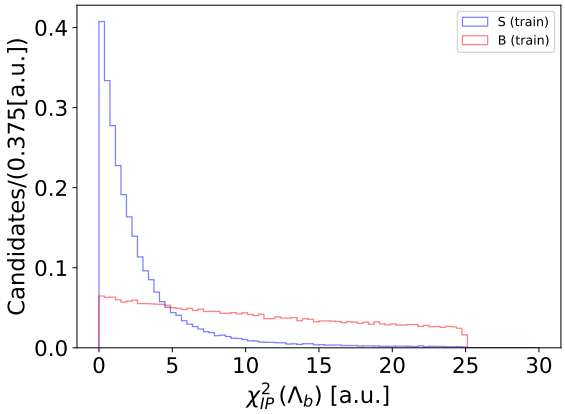
(a) DOCA(leptons)



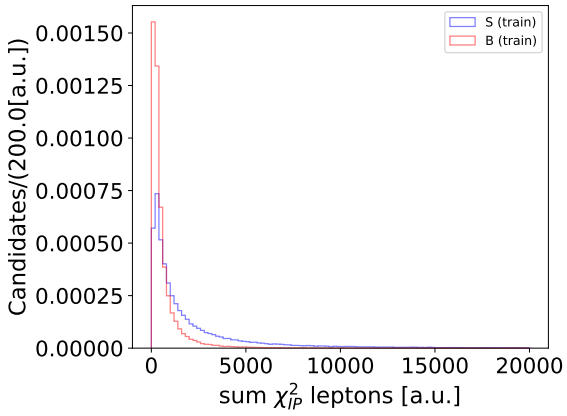
(b) DIRA(Λ)



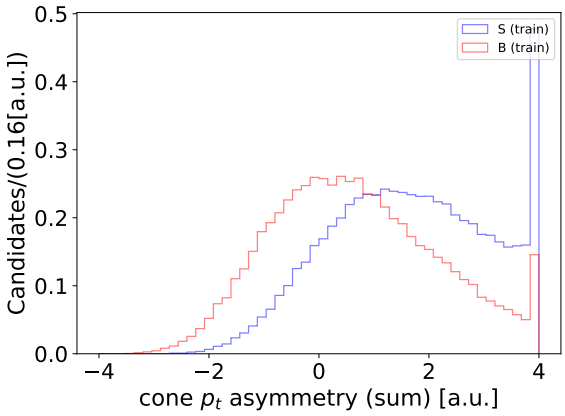
(c) $\chi^2_{ORIVX}(\Lambda)$



(d) $\chi^2_{IP}(\Lambda^0_b)$

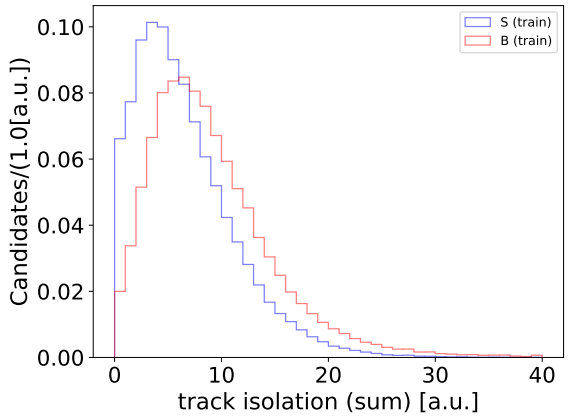


(e) χ^2_{IP}(leptons)

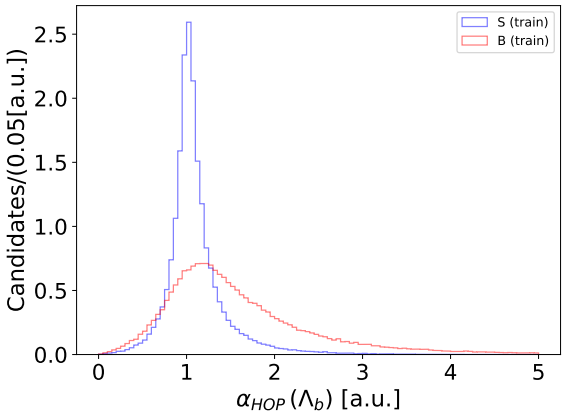


(f) cone p_{T} asymmetry (sum)

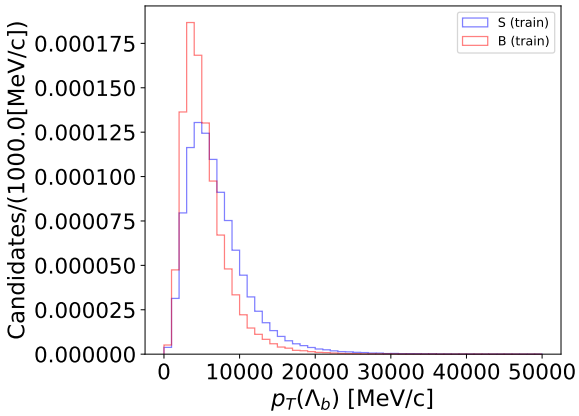
Figure 7.15



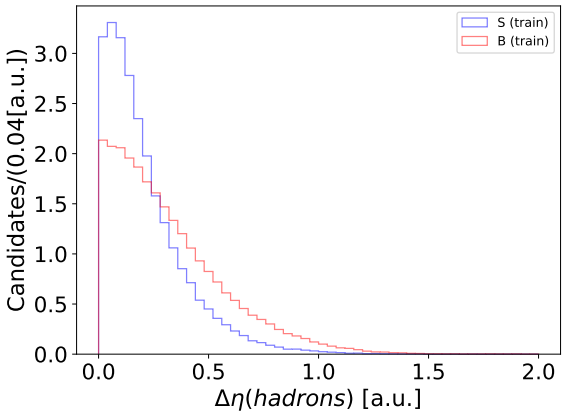
(g) track isolation (sum)



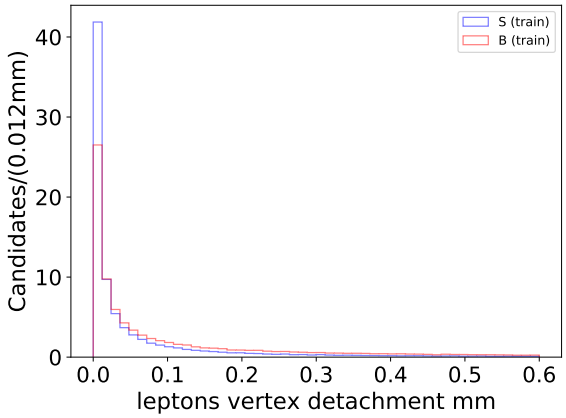
(h) α_{HOP}



(i) $p_{\mathrm{T}}(\Lambda_b^0)$



(j) $\Delta\eta$(hadrons)



(k) leptons vertex detachment

Figure 7.15: Distributions of the variables used in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ GBDT training for signal MC candidates (blue) and background candidates from data sidebands (red).

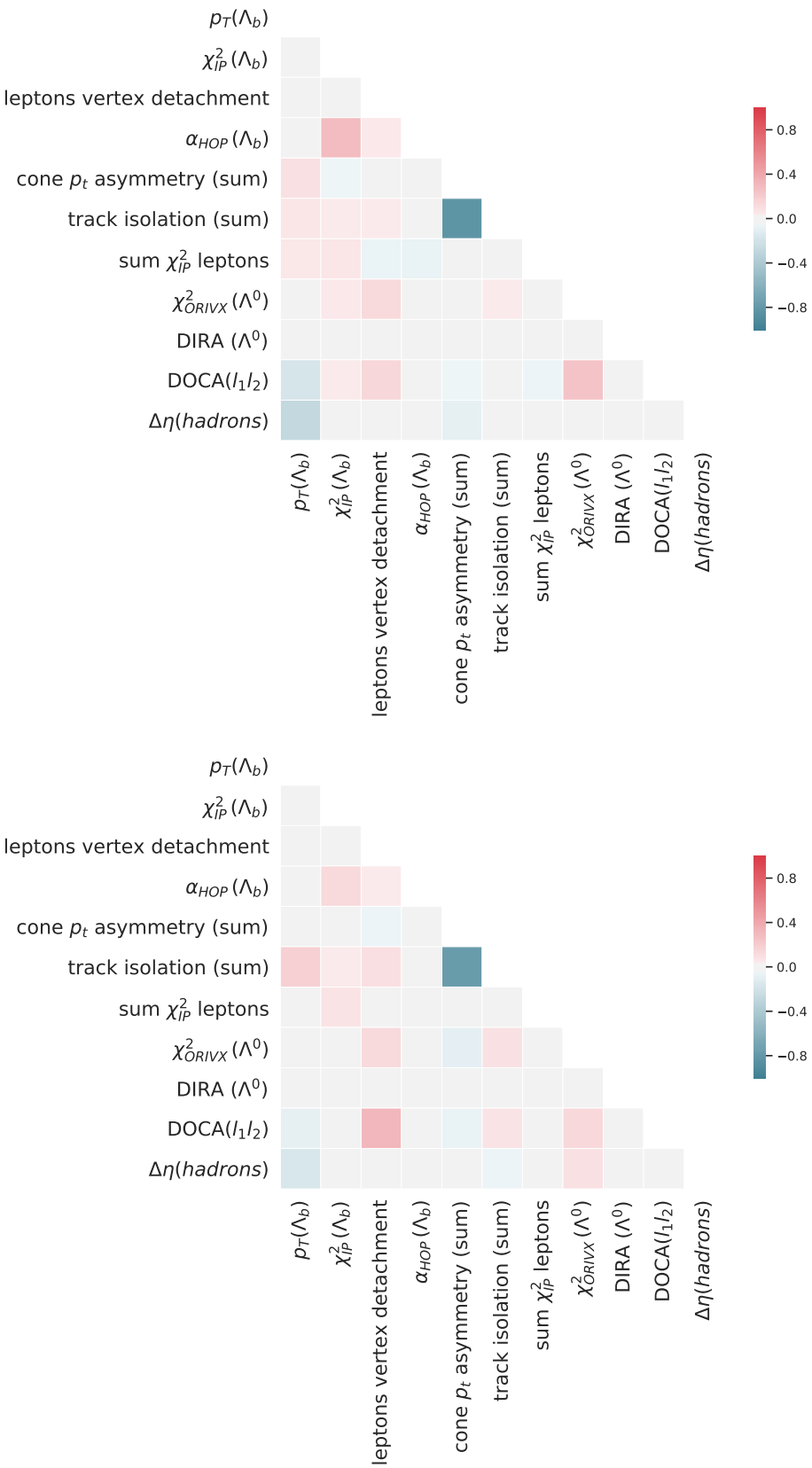


Figure 7.16: Linear correlation factors between the GBDT training variables in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis in signal (Top) and background (Bottom).

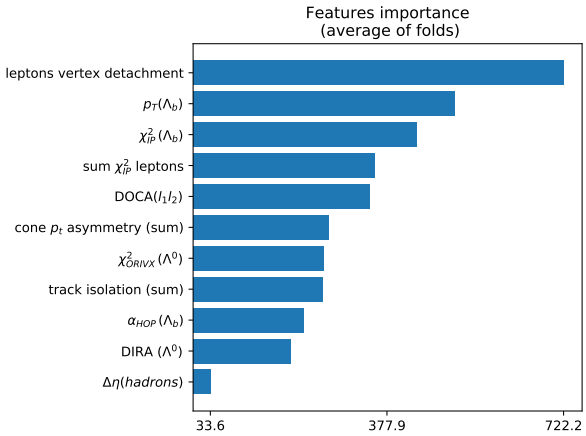


Figure 7.17: The relative importance of individual variables to the GBDT classifier, averaged over the k-folds.

avoid introducing bias into the training sample. The avoidance of bias is also why an equal amount of data and MC was used.

To exploit all of the available MC sample and avoid performance estimation bias, the k-folding cross-validation technique [156] was used with a total of five folds. Its use is particularly important to utilise the information available optimally, given the limited size of the data samples. The dataset was randomly split into k folds, five in this case, and then each fold, in turn, was used as the testing dataset to evaluate the ability of the classifier to distinguish signal and background. The remaining folds act as the training dataset. The results of the five trained classifiers are then averaged, and the k-folds form an ensemble of trees that can be evaluated with the GBDT to get a single classifier.

To examine the performance of the GBDT the figure of merit used was the integral of the Receiver Operator Characteristic (ROC) curve, known as the Area Under Curve (AUC). The ROC curve here defines the relationship between the background rejection (1− the false positive rate) and the signal efficiency (the true positive rate). For an ideal case this should be a step function with 100% background rejection with 100% signal efficiency. In reality, this will be a curve but a larger area (or AUC) corresponds to better rejection with high signal efficiency. The ideal case would then give an AUC of 1. Figure 7.18 shows the ROC curves for the five k-folds used in training, each evaluated on the respective testing fold. The five classifiers corresponding to each fold show compatible performances, with an AUC of 0.94 for all folds which shows reasonably high potential performance.

Another equally important facet of the GBDT training to consider alongside the performance is the examination of whether it has been 'overtrained'. This is where the classifier appears to be very effective in discriminating signal from background, but the

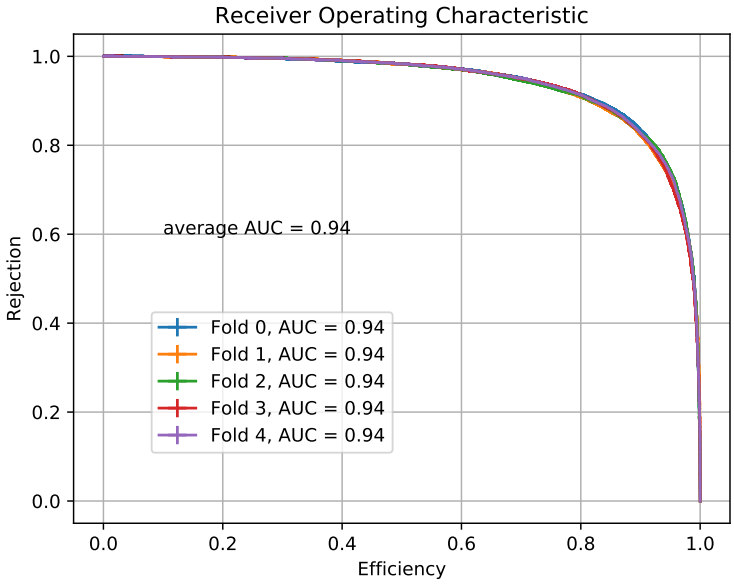


Figure 7.18: Receiver Operating Characteristics curves for the five cross-validation k-folds of the GBDT classifier. The thickness of the lines represents the measurement uncertainty. The AUC is also indicated for each fold.

classifier has instead become very effective at discrimination of a specific dataset. If the overtrained classifier were provided with new data it may perform worse than expected as it is biased to particular properties in the training dataset. This is one reason why the testing dataset that was used to evaluate the performance of the GBDT was not included in the training dataset. The presence of overtraining was assessed by comparing the GBDT response distributions on the testing and training samples, summing over the k-folds. Here the GBDT response refers to the output score of the classifier where -1 is more background-like and $+1$ is more signal-like. Figure 7.19 shows the distributions for the testing and training samples. The Kolmogorov-Smirnov test [157, 158] was used to discern whether the testing and training samples are consistent with having been sampled from the same parent distribution. The corresponding p-values were much greater than 0.05 for both signal and background. This implies that in both cases the GBDT response to the test and training samples are compatible, and therefore no overtraining was observed.

In a similar way the full testing sample can be examined by comparing the distributions of the GBDT response for each of the binary choices that define the analysis categories, *e.g.* No Brem and Brem. Figure 7.20 shows these distributions for the signal and background components and highlights that in each instance the binary choice can be separated using the trained classifier. Therefore, the combined analysis categories are likely to also have good discrimination of signal and background with this GBDT.

The ROC curves corresponding to the eight analysis categories are compared in

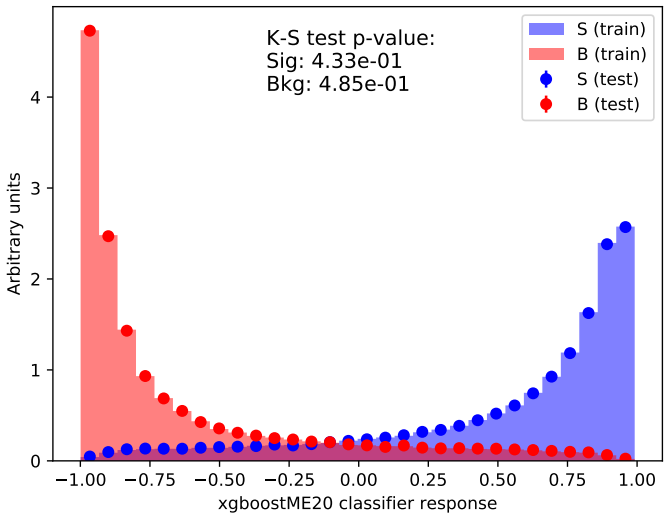


Figure 7.19: Response of the GBDT used in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis for the signal (blue) and background (red) samples used for training (points) and testing (filled histograms). The Kolmogorov-Smirnov test's p-value is overlaid.

Figure 7.21. The curves show similar but not identical performances, which is likely to be due to the small differences in the resolutions of the training variables when considered in these exclusive categories. The bremsstrahlung radiation, for example, smears the distribution of momenta, while the two Λ track types have different reconstruction resolutions because of the different quantity of information available for the hadron tracks. The small differences in performance between the Run 1 and Run 2 samples are also due to differences in the reconstruction and selection of the candidates. The eight analysis categories are therefore evaluated separately to enhance the GBDT background rejection.

7.4.1 Optimisation of the GBDT Selection

The GBDT selection was optimised by maximising the Punzi figure of merit [159], defined as

$$\mathrm{P} = \frac{\varepsilon(\mathrm{t})}{\mathrm{a}/2 + \sqrt{\mathrm{B(t)}}}, \tag{7.4.1}$$

where t is the cut value, ε is the signal efficiency, B is the expected number of background candidates and a is the number of standard deviations of significance required for the observation of the signal. Specifically, a is fixed to 3 for this case. B is obtained by fitting the data sidebands with an exponential function and extrapolating into the signal region to determine the fraction of candidates in the blinded region compared to the candidates in the sidebands. This fraction is then used to scale the total candidates remaining in

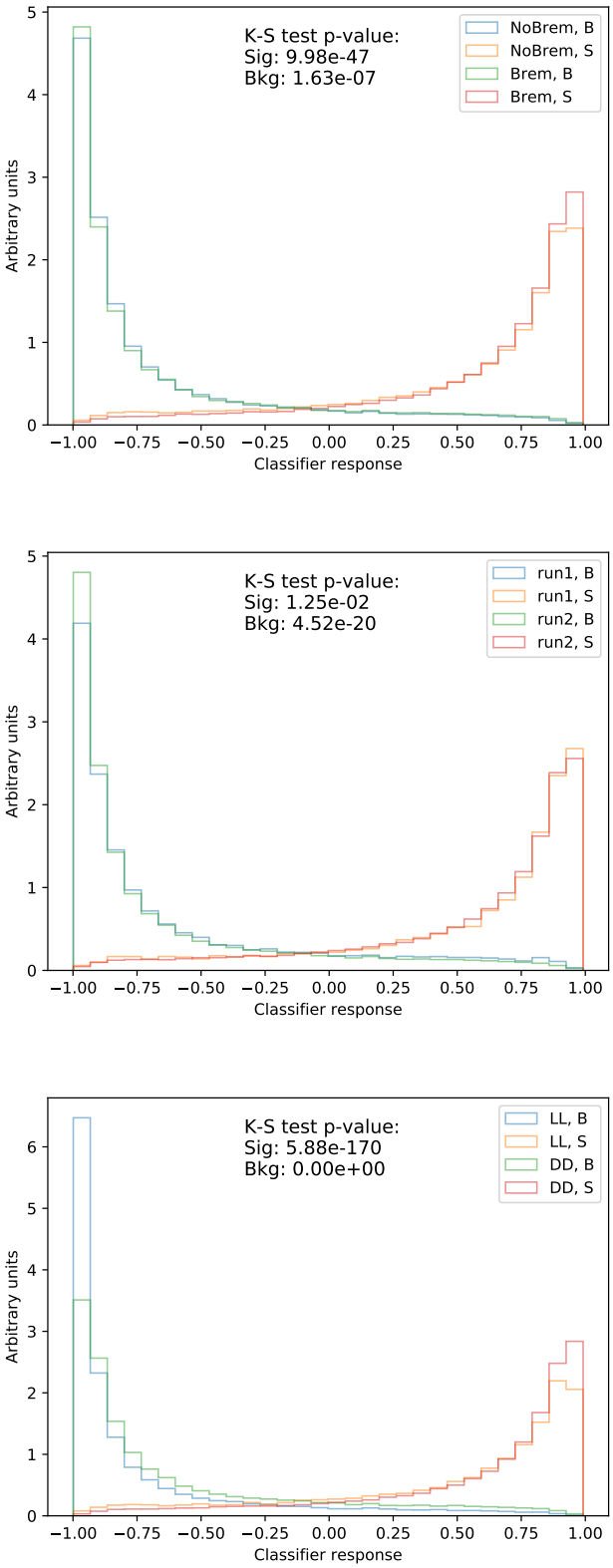


Figure 7.20: Response of the GBDT classifier used in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis, compared for different bremsstrahlung categories (Top), LHC run periods (Middle) and Λ track types (Bottom). In each plot, the sample is integrated over the other categories.

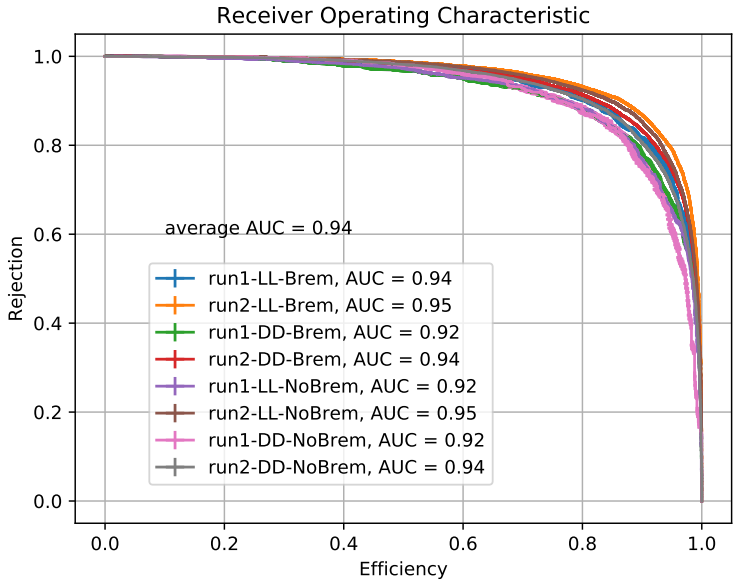


Figure 7.21: Receiver Operating Characteristic curves of the GBDT classifier used in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis, for the eight categories in which the analysis is split. The thickness of the lines represents the measurement uncertainty.

sidebands after each cut value of t. Finally, ε is computed from the sum of weighted signal simulated candidates[9] before and after the cut.

The cut, t, is varied over the GBDT response from -1 (background-like) to $+1$ (signal-like). Therefore, the general form of the optimal cut is a classifier output > 0 where a GBDT response of 0 is likely to represent an event that cannot be easily classified as either signal or background. The resulting ε and P plots are shown in Figures 7.22 and 7.23. To reduce the impact of statistical fluctuations, a moving average over five adjacent points was used for the ε and P curves. Additionally, the value of P is required not to vary by more than 90% within five adjacent points of the optimal cut.

In order to calculate the value of B used in Equation 7.4.1, the fraction of candidates in the blinded region compared to the sidebands is obtained before applying a GBDT cut. This assumes that the overall background shape will vary smoothly with the applied GBDT cut, such that the initial fit to determine B is valid for all cut values. Therefore, a check was made where the background is re-fitted with an exponential after each GBDT cut, so this fraction can be determined with the true mass shape that would be present after that cut. This should give a more reliable value of the 'observed' B, but it also increases the statistical uncertainty of the estimated background in the signal region as the GBDT cut is increased, due to the smaller sample size on which the exponential fit is

[9]The candidates are weighted as described in Section 7.3.

performed. Overall, it was found that the optimal cut using 're-fitting' is similar to that obtained without. This implies the overall combinatorial shape responds smoothly to the GBDT cuts. Thus, the optimal cut without re-fitting was taken to avoid potential issues with a reduced sample size in the optimisation.

A summary of the selected, optimal GBDT cuts from the moving average Punzi scan is given in Table 7.7. This demonstrates that the GBDT provides a background rejection of $\gtrsim 98.5\%$ with signal efficiencies of $\gtrsim 31\%$ and some as high as 45%, which gives a very clean environment for the remaining data. This can be seen in the fits performed on the signal mode in Section 8.3. The trained classifier was applied to both the rare and normalisation modes, and the same optimised cuts are applied to both to ensure as much parity between the selections for each mode as possible. The normalisation mode does not include any potential bremsstrahlung photons to be recovered. Therefore, the optimised GBDT cuts for the Brem categories were used, *e.g.* the cut for 'Brem-LL-Run 2' was used for 'LL-Run 2'. This was decided due to the fact that Brem categories in $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ data had larger sample sizes than the No Brem categories, and the corresponding cuts are tighter than the No Brem equivalents, which would give a more conservative and accurate value of total efficiency for the normalisation mode.

Table 7.7: The optimal GBDT cuts when optimised by a Punzi figure of merit with a moving average.

Category	Optimal Cut	Signal Efficiency	Rejection	Figure of Merit Value
Brem-LL-Run 1	0.84	0.3425±0.0137	0.9930±0.0011	0.0421
Brem-LL-Run 2	0.75	0.4528±0.0080	0.9886±0.0005	0.0183
Brem-DD-Run 1	0.88	0.3495±0.0102	0.9912±0.0012	0.0381
Brem-DD-Run 2	0.89	0.3101±0.0046	0.9952±0.0002	0.0148
No Brem-LL-Run 1	0.74	0.4223±0.0143	0.9847±0.0013	0.0330
No Brem-LL-Run 2	0.82	0.3408±0.0059	0.9947±0.0003	0.0196
No Brem-DD-Run 1	0.85	0.3664±0.0094	0.9884±0.0010	0.0263
No Brem-DD-Run 2	0.87	0.3163±0.0040	0.9948±0.0002	0.0137

Figures 7.24 and 7.25 show the mass distribution of the reconstructed Λ_b^0 after applying both the pre-selection and the optimal GBDT cuts. Overall, the background is strongly reduced by the GBDT but there are candidates remaining in the lower mass sideband that were found to include both combinatorial candidates and some exclusive backgrounds, $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$, see Section 8.1 for more detail.

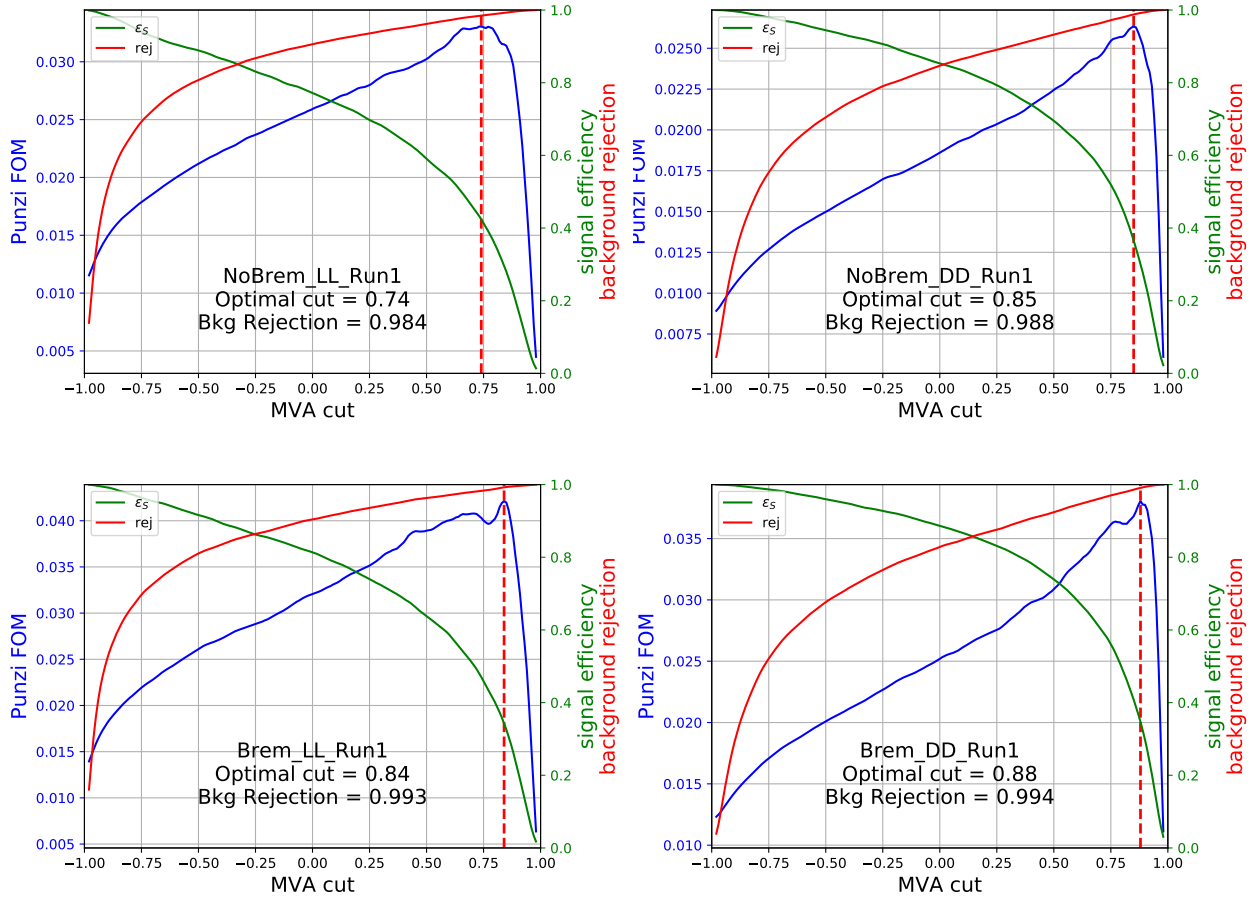


Figure 7.22: Punzi figure of merit and signal selection efficiencies as a function of the cut on GBDT for Run 1 candidates. The categories without (with) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column. The red dashed lines indicate the optimal cuts.

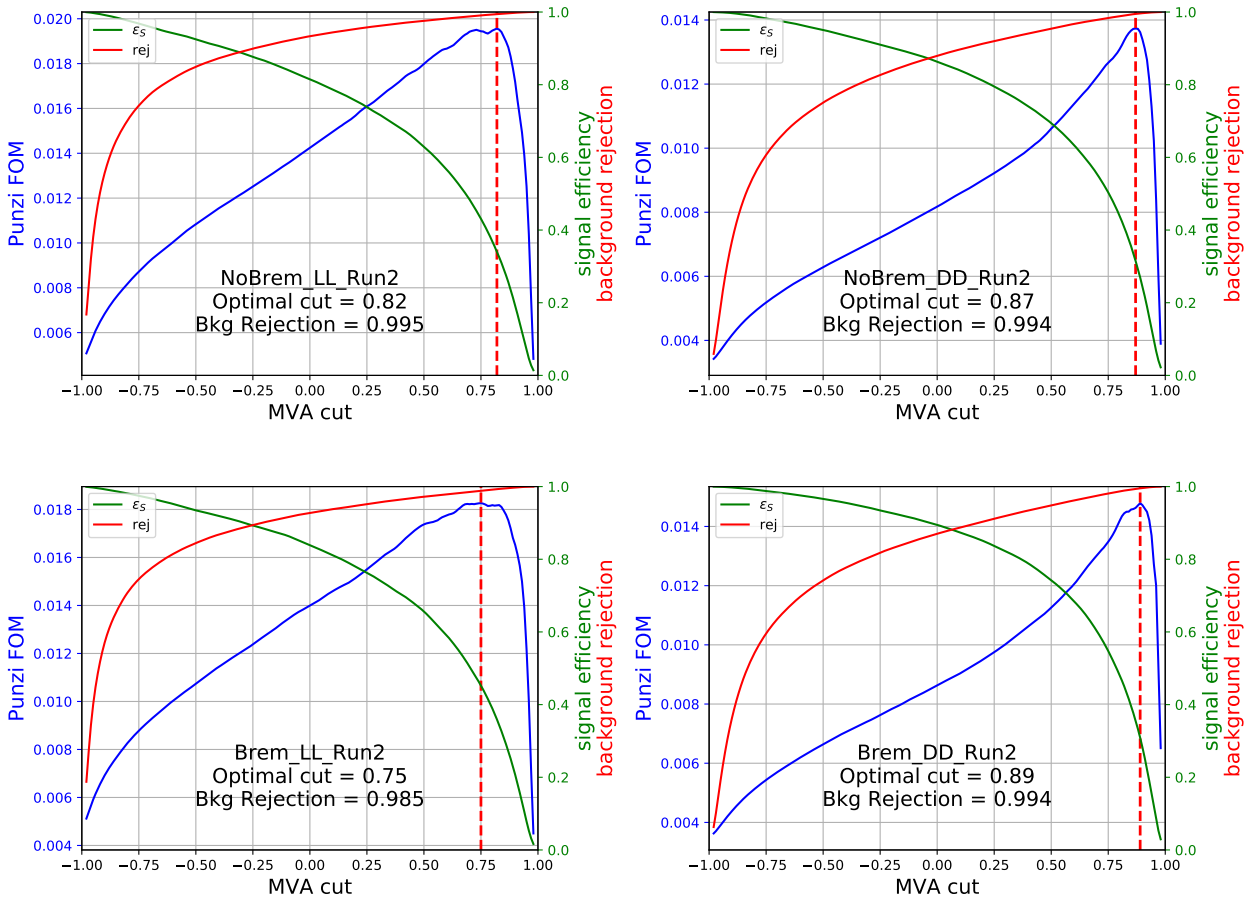


Figure 7.23: Punzi figure of merit and signal selection efficiencies as a function of the cut on GBDT for Run 2 candidates. The categories without (with) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column. The red dashed lines indicate the optimal cuts.

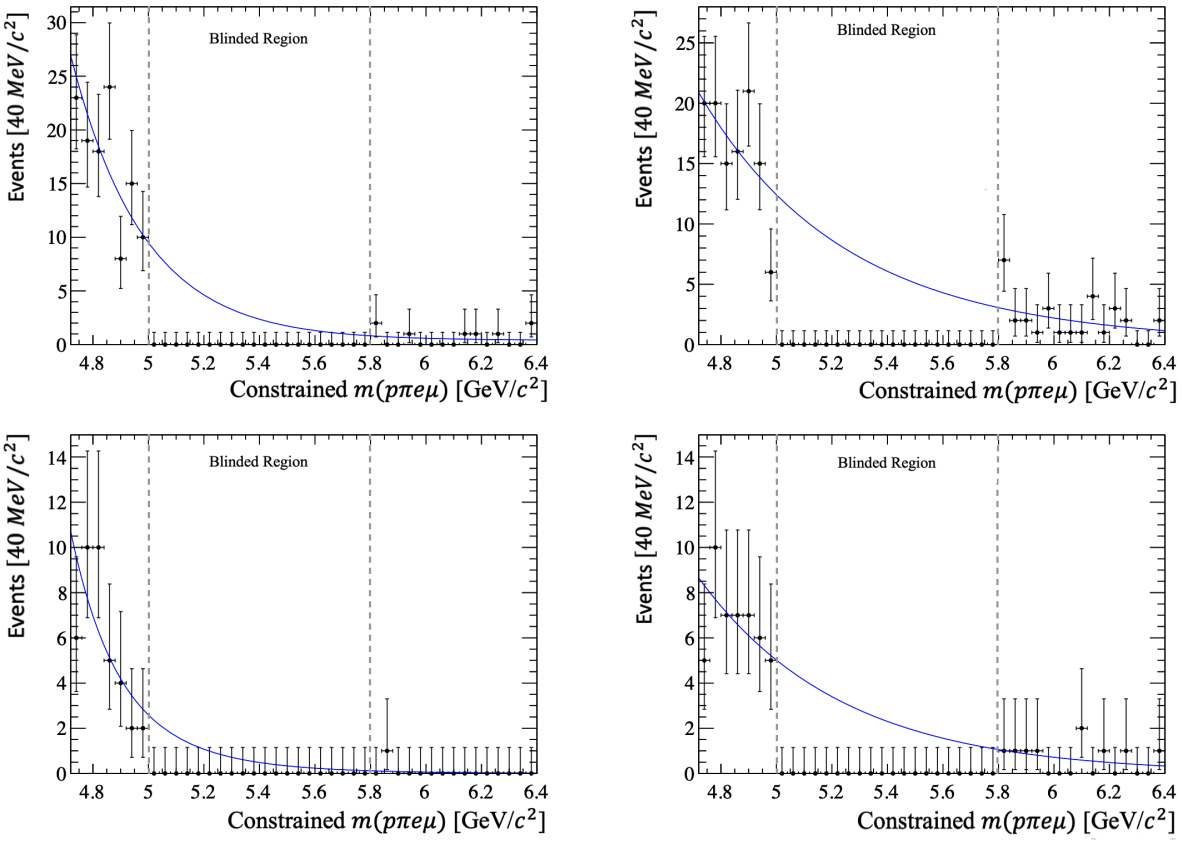


Figure 7.24: Invariant mass distribution of the reconstructed Λ_b^0 for Run 1 candidates with the pre-selection and GBDT cuts applied, that is fitted to an exponential. The categories without (with) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column.

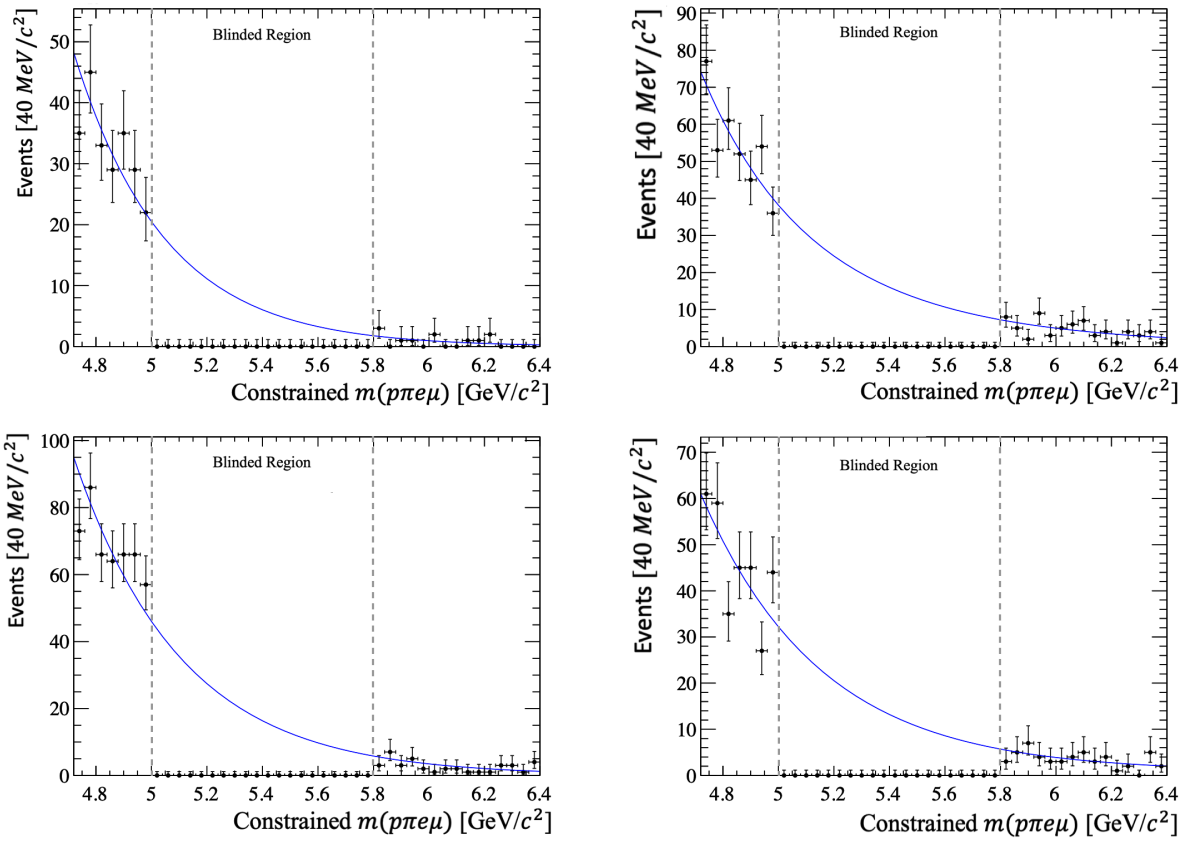


Figure 7.25: Invariant mass distribution of the reconstructed Λ_b^0 for Run 2 candidates with the pre-selection and GBDT cuts applied, that is fitted to an exponential. The categories without (with) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column.

7.4.2 Cross-checking the GBDT Response on Data

As seen in Section 7.3 there can be mis-modelling present within the simulation that is used to train the GBDT. Although this is largely controlled by the efforts of re-weighting to align with the expected response from data there could still remain some minor imperfections that bias the GBDT's performance. Therefore, the classifier's response was cross-checked on an alternative dataset. However, no SM process can be used as a direct proxy dataset, due to the unique signature of a LFV channel. As a viable alternative, the classifier response was compared between sWeighted data and simulation on the same known SM process: the $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ decay. This channel is known to have a clean signal and an abundant yield, hence its use as the normalisation channel and validating the GBDT response for both the rare and normalisation channels.

It should be noted that the α_{HOP} variable, present as one of the features on the GBDT, should have a value of 1 for the normalisation channel as it does not contain electrons. Therefore, α_{HOP} is assigned the default value of 1 for all the candidates at the moment of computing the GBDT response on $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$. The α_{HOP} distribution is then checked separately on samples of $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ sWeighted data and MC to verify the distribution is also well-modelled between sWeighted data and MC. The GBDT response with α_{HOP} fixed to 1 is referred to as GBDT$^{fix-hop}$.

Figure 7.26 shows the efficiency of a running cut on the GBDT$^{fix-hop}$ distribution of $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ in Run 1 and Run 2, for LL and DD Λ tracks. This was determined as the sum of the data/MC weights for the MC, or the sWeights for the sWeighted data after the GBDT$^{fix-hop}$ cut, compared to the sum of weights before any GBDT$^{fix-hop}$ cut was applied. The response of the GBDT appears to be well reproduced, especially in Run 2 candidates.

The relative difference between the response on data and simulation in each category is then given by

$$\varepsilon_{sys,GBDT} = \frac{1}{2}\frac{\varepsilon_{sW} - \varepsilon_{MC}}{\varepsilon_{MC}}, \tag{7.4.2}$$

where ε_{sW} and ε_{MC} are the efficiencies of the sWeighted data and the MC respectively. The value of this relative difference of efficiencies at the optimal GBDT$^{fix-hop}$ cut was assigned as a systematic uncertainty on the nominal response of the normalisation mode MC from signal GBDT, and propagated to the final result as seen in Tables 9.3 and 9.4. For the LL case this uncertainty is mostly below 1% in each bin of the GBDT$^{fix-hop}$ cut efficiency for the MC. In the DD case the uncertainty is mostly below 2% but does reach almost 3% for the last few bins of DD Run 1.

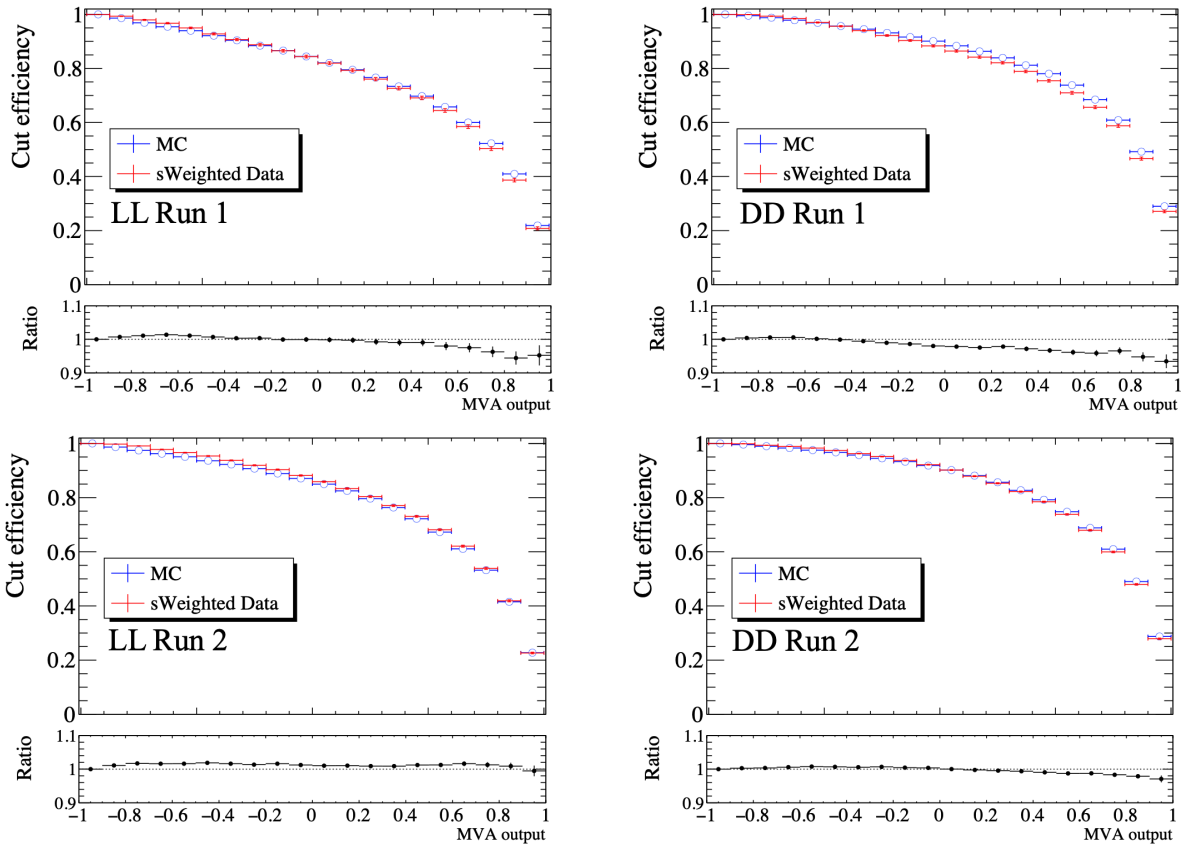


Figure 7.26: Running cut efficiency on GBDT$^{fix-hop}$ evaluated on $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ data and simulated samples. The left plots correspond to LL Λ tracks and the right plots to DD Λ tracks. The top row corresponds to Run 1 candidates and the bottom row corresponds to Run 2.

Figure 7.27 shows the comparison of the α_{HOP} variable's distribution between data and simulation, using combined Run 1 and Run 2 $\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$ sWeighted data and the corresponding simulated samples. In order to maximise the available data sample size no split was performed in the hadron track types since α_{HOP} mainly depends on the lepton tracks. The agreement was found to be satisfactory, and no additional systematic error was assigned. Thus, the response of the GBDT on data was verified and controlled.

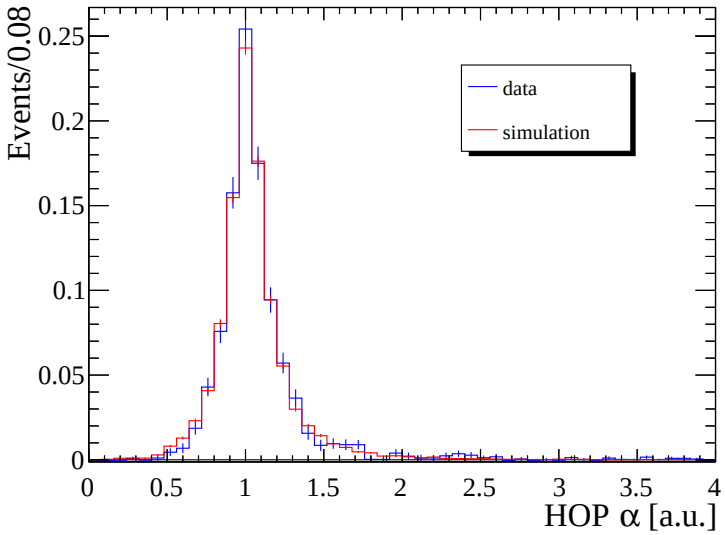


Figure 7.27: α_{HOP} distribution on $\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$ data and MC for combined Run 1 and Run 2 candidates, in both Λ track categories.

An additional check of the GBDT response was performed to investigate potential correlations with particular mass variables. This was done to ensure no peaking structure is created artificially in its distribution that could potentially mimic a signal or unexpected background. Figure 7.28 shows the two-dimensional histograms of the GBDT classifier output versus the reconstructed invariant mass of the Λ_b^0 candidates from the left and right data sideband separately, for these two sidebands 'folded' around the centre of the blinding window[10], and for the SS data. To quantify the potential correlation between the mass and GBDT response the Pearson correlation factor is determined. This measures the linear correlation between two variables where a value of ± 1 indicates a direct correlation and 0, no correlation. Figure 7.28 shows that this correlation factor is always below 1%, and thus is well under control with no unexpected behaviour. This check can also be performed by specifically looking at the invariant mass of the leptons via the q^2 value. Figure 7.29 shows the equivalent plots checking the q^2 correlation with the GBDT classifier response, these have a higher Pearson correlation factor but still are below 4%. Therefore, there is also no unexpected peaking behaviour in the reconstructed di-lepton invariant mass.

[10]Specifically this used the translated mass variable, $|m(\Lambda_b^0) - 5400\,\mathrm{MeV}/c^2|$.

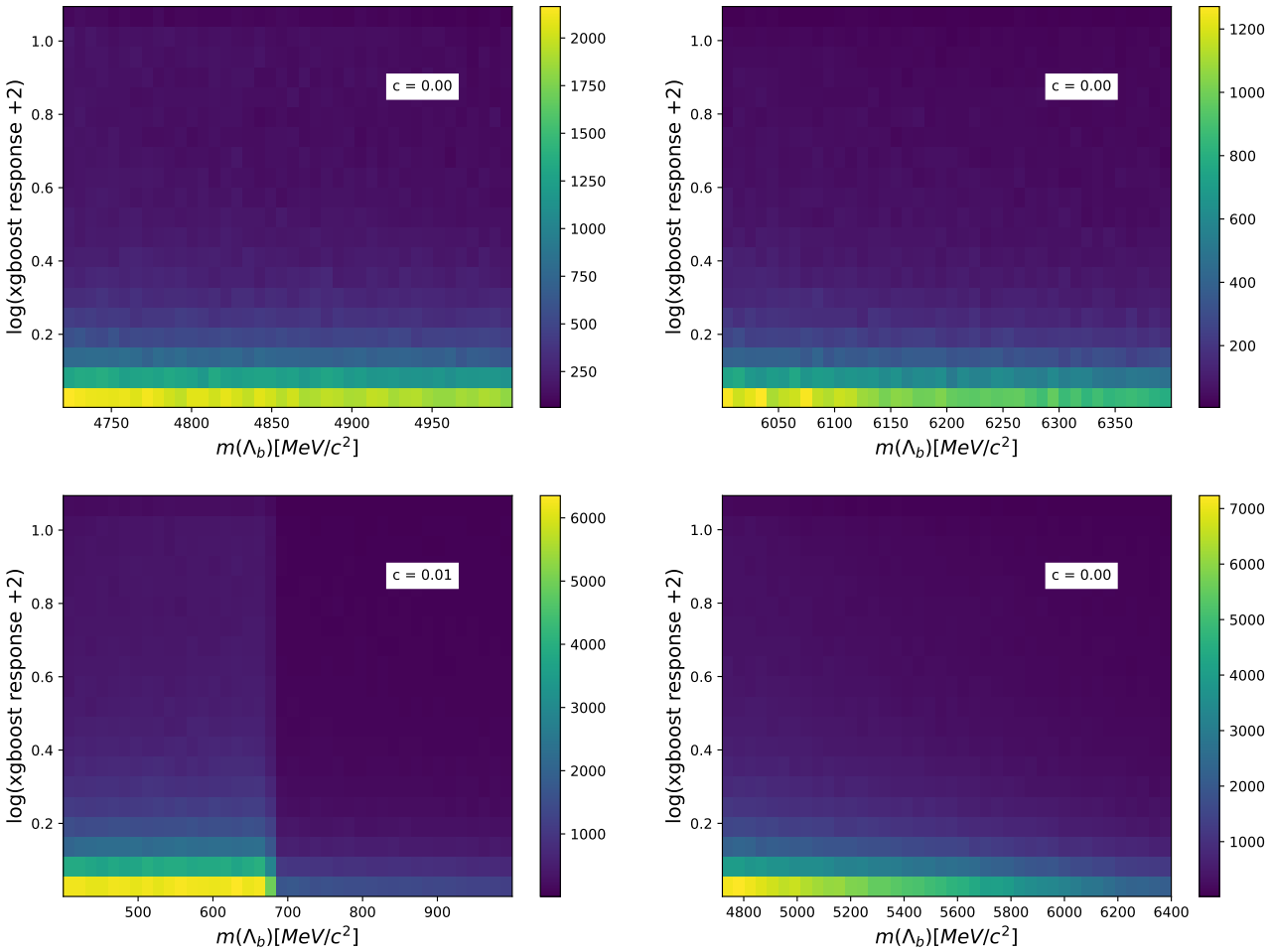


Figure 7.28: GBDT output vs mass distribution for data candidates in the lower mass sideband (Top Left), upper mass sidebands (Top Right), both mass sidebands folded together (Bottom Left), and same-sign sample (Bottom Right). The Pearson correlation coefficients are overlaid.

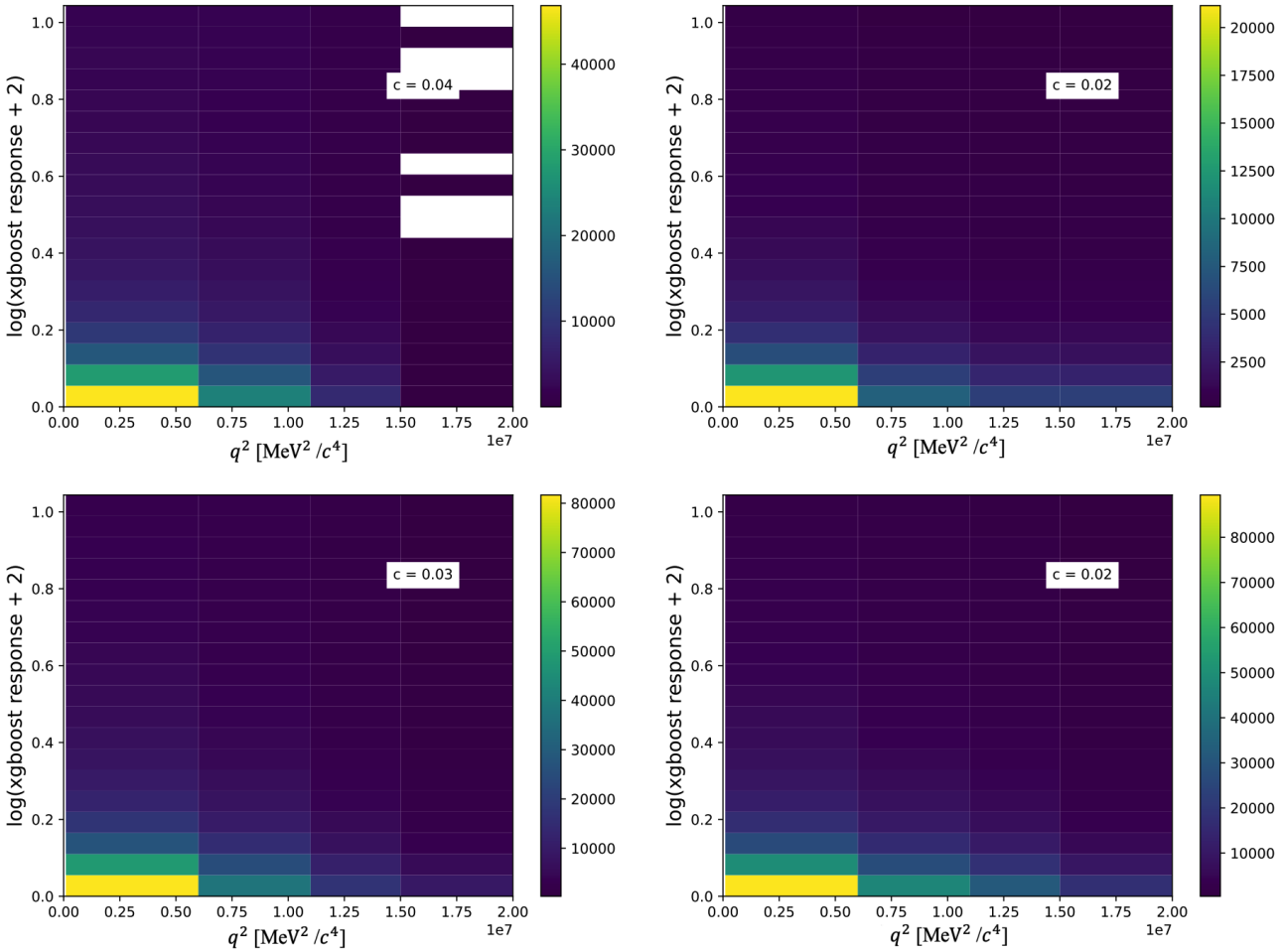


Figure 7.29: GBDT output vs q^2 distribution for data candidates in the lower mass sideband (Top Left), upper mass sidebands (Top Right), both mass sidebands folded together (Bottom Left), and same-sign sample (Bottom Right). The Pearson correlation coefficients are overlaid.

7.5 Selection to Remove Multiple Candidates

During the reconstruction of an event, multiple candidates of a signal decay can be reconstructed and selected for the same collision event. Therefore, an event can be erroneously duplicated and potentially introduce a small bias and so a selection was determined to ensure only a single candidate remained for each event. The multiple candidates were identified as candidates that shared both the same unique ID (event number $\times$ run number) and the total number of candidates in that event, given by DaVinci as 'totCandidates'. This would ensure that events with two physical decays that occurred within the same event are not mistaken for multiple reconstruction candidates and hence removed.

Tables 7.8 and 7.9 show the fraction of events having multiple candidates for the modes that contribute to the rare and normalisation channel fits after the full selection has been applied. For data these fractions are relatively low at typically $< 1\%$ for both the rare and normalisation modes. However, the signal mode data has much lower sample size and tends to be mostly zero, with a greater fraction of multiple candidates in the surviving events. For MC this tends to be similar to or greater than the fraction of multiple candidates in data, typically, at $< 1\%$ for both the rare and normalisation modes. These were considered significant enough to warrant the inclusion of a selection to remove the multiple candidates.

Therefore, within the data and MC for each of the presented samples in which an event passing the full selection contains multiple candidates, one of the candidates was randomly selected and the others were rejected. This random selection ensures a lack of selection bias for the remaining candidate [160].

Table 7.8: Percentage of events with multiple candidates, for each analysis category without bremsstrahlung reconstructed for data and MC with full selection applied for various samples.

Category		**No Brem Events with > 1 Candidates /%**			
		LL-Run 1	DD-Run 1	LL-Run 2	DD-Run 2
Data	Signal mode	0.81±0.81	0.70±0.71	0.42±0.42	0.00±0.00
MC	$\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$	0.00±0.00	0.25±0.10	0.02±0.02	0.17±0.04
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda \mu^+ \nu_\mu) e^- \overline{\nu}_e$	0.00±0.00	0.00±0.00	0.00±0.00	0.63±0.63
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda e^+ \nu_e) \mu^- \overline{\nu}_\mu$	0.00±0.00	0.00±0.00	0.00±0.00	0.45±0.45

Table 7.9: Percentage of events with multiple candidates, for each analysis category with bremsstrahlung reconstructed, for data and MC with full selection applied for various samples. The normalisation mode and other modes used for normalisation fits are presented in the bremsstrahlung categories though no bremsstrahlung selection was performed for them, but the Brem version of the GBDT cut was applied to these samples.

Category		**Brem Events with >1 Candidates /%**			
		LL-Run 1	DD-Run 1	LL-Run 2	DD-Run 2
Data	Normalisation mode	0.25±0.09	0.28±0.09	0.35±0.05	0.28±0.04
	Signal mode	0.00±0.00	0.00±0.00	0.19±0.19	0.00±0.00
MC	$\Lambda_b^0 \to \Lambda J/\psi (\to \mu^-\mu^+)$	0.11±0.05	0.31±0.06	0.27±0.03	0.24±0.03
	$B^0 \to K_S^0 J/\psi (\to \mu^-\mu^+)$	0.00±0.00	0.22±0.04	0.02±0.01	0.12±0.03
	$\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$	0.00±0.00	0.14±0.08	0.00±0.00	0.24±0.05
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda \mu^+ \nu_\mu) e^- \overline{\nu}_e$	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda e^+ \nu_e) \mu^- \overline{\nu}_\mu$	0.00±0.00	0.00±0.00	0.00±0.00	0.37±0.37

7.6 Determination of Efficiencies

To determine the expected signal yield for a given branching fraction, as used in Equation 6.0.1, the efficiency of all selection and reconstruction requirements was determined. These efficiencies were calculated by applying successively the full selection as described throughout this chapter to a weighted MC sample. The ratio of the sum of the weights of the candidates before and after the full selection was then taken as the final efficiency, given as

$$\begin{aligned}\varepsilon_{\text{tot}} =& \varepsilon_{\text{gen}} \cdot \frac{\sum\limits_{\text{gen, rec}} w_{kin} \times w_{\tau_{\Lambda_b^0}}(w_{ang,\Lambda_b^0\to\Lambda J/\psi}) \times w^{PID} \times w^{trk} \times w^{trig}}{\sum\limits_{\text{gen, all}} w_{kin} \times w_{\tau_{\Lambda_b^0}}(w_{ang,\Lambda_b^0\to\Lambda J/\psi})}, \\ =& \varepsilon_{\text{gen}} \cdot \varepsilon_{\text{sel.+rec.}}\end{aligned} \tag{7.6.1}$$

where ε_{gen} is the generator-level cut efficiency. This corresponds to the efficiency for candidates to satisfy criteria, such as loose angular cuts which ensure the produced particles are generally within LHCb's acceptance, before specifically simulating the effects of finite detector acceptance and resolution. $\varepsilon_{\text{sel.+rec.}}$ is the total efficiency to reconstruct and select candidates, determined from the simulated sample based on two sums: the sum over all generator-level candidates $\sum\limits_{\text{gen, all}}$; and the sum over reconstruction-level candidates, $\sum\limits_{\text{gen, rec}}$, including the full selection chain. The weights are given by w with a sub- or super-script that identifies the source of the weight. The $\Lambda_b^0 \to \Lambda J/\psi$ angular weight is included in

brackets as it is only included for the normalisation mode. Additionally, weights related to reconstruction-level properties (w^{PID}, w^{trk}, w^{trig}) are only applied to the numerator as they are not defined for the denominator.

The efficiencies were evaluated in the same eight analysis categories introduced previously, from which yields are extracted. Combining these with the efficiencies allows the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ branching fraction to be determined. Due to the analysis being blinded an estimated upper limit was determined using these values, see Chapter 9.

Tables 7.10 to 7.12 show the efficiencies determined for the signal mode with and without recovered bremsstrahlung photons, and for the normalisation mode. In these tables, the efficiencies are dependent on the previous row's selection, where the uncertainty is the weighted statistical uncertainty from MC. The generator efficiency refers to the first term in Equation 7.6.1. Stripping+rec refers to the total efficiency of stripping and reconstruction (see Section 7.1). Pre-selection refers to all cuts described in Table 7.2, excluding the q^2 cut (which removes candidates at $q^2 < 0.1\,\mathrm{GeV}^2/c^4$, shown separately), and the J/ψ veto (labelled as ψ resonance veto). The requirement of the DTF algorithm to converge to provide a cleaner invariant mass variable due to additional constraints is also included as 'DTF converged'. The penultimate row then is for the selection to ensure only a single candidate is present for each event (see Section 7.5). The product of all the efficiencies in these tables except for the generator efficiency defines the efficiency, $\varepsilon_{\text{sel.+rec.}}$, with the final row giving the total efficiency.

In Tables 7.10 and 7.11 there are some differences between Run 1 and Run 2 category values. These are due to a combination of differing trigger conditions; PID variable changes *i.e.* for ProbNN; and the separate optimisation of the GBDT selection which is dependent on the other changes. Therefore, some divergence between Run 1 and Run 2 categories is to be expected.

It should also be noted that in Tables 7.10 and 7.11 the Brem and No Brem efficiencies differ, whereas there should be a roughly equal amount of Brem and No Brem candidates. This is because the PID weight for the No Brem case was incorrectly assigned to the Brem categories as well. Although this affects the total efficiencies, the overall impact on the final upper limit is relatively small. Additionally, these tables show a larger stripping efficiency and lower trigger efficiency for the No Brem case compared to Brem. This is due to the PID for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ MC being introduced by weights rather than cuts, which means that every event has at least two signal candidates corresponding to the instance where the lepton IDs are assigned correctly and one where the muon is assigned to be an electron and vice versa. For candidates where at least one track cannot be matched to a true particle, known as background category 60, the mis-assigned lepton candidate is rejected either by the requirement to have bremsstrahlung reconstructed or the L0Muon

Table 7.10: Efficiencies for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ for different components of the selection for each analysis category without bremsstrahlung. Each efficiency shown is conditional based on the one directly above.

Efficiency Categories	**Weighted Efficiency for No Brem Categories (%)**			
	LL-Run 1	**DD-Run 1**	**LL-Run 2**	**DD-Run 2**
Generator	19.290 ± 0.026	19.290 ± 0.026	20.519 ± 0.027	20.519 ± 0.027
Stripping + rec.	2.045 ± 0.009	2.112 ± 0.009	1.251 ± 0.004	1.630 ± 0.004
PID	96.589 ± 0.155	96.777 ± 0.149	96.564 ± 0.103	96.734 ± 0.088
L0 trigger	19.910 ± 0.172	29.105 ± 0.186	26.985 ± 0.126	32.520 ± 0.114
HLT1 trigger	76.510 ± 0.403	80.382 ± 0.294	83.625 ± 0.201	85.848 ± 0.148
HLT2 trigger	54.722 ± 0.546	51.905 ± 0.414	77.841 ± 0.263	75.688 ± 0.209
Pre-selection	66.089 ± 0.681	74.374 ± 0.493	66.117 ± 0.313	70.815 ± 0.231
q^2 selection	99.430 ± 0.137	99.774 ± 0.060	99.435 ± 0.062	99.404 ± 0.048
HOP selection	100.000 ± 0.000	99.964 ± 0.026	99.874 ± 0.029	99.909 ± 0.017
ψ resonance Veto	92.880 ± 0.451	93.312 ± 0.323	92.890 ± 0.205	93.091 ± 0.152
GBDT selection	41.551 ± 0.884	36.150 ± 0.631	32.672 ± 0.383	30.922 ± 0.279
DTF converged	99.682 ± 0.132	99.821 ± 0.068	99.817 ± 0.047	99.732 ± 0.043
Single Candidate	100.000 ± 0.000	99.604 ± 0.139	99.691 ± 0.080	99.729 ± 0.056
Total efficiency	0.041 ± 0.001	0.060 ± 0.001	0.042 ± 0.001	0.067 ± 0.001

Table 7.11: Efficiencies for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ for different components of the selection for each analysis category with bremsstrahlung. Each efficiency shown is conditional based on the one directly above.

Efficiency Categories	**Weighted Efficiency for Brem Categories (%)**			
	LL-Run 1	**DD-Run 1**	**LL-Run 2**	**DD-Run 2**
Generator	19.290 ± 0.026	19.290 ± 0.026	20.519 ± 0.027	20.519 ± 0.027
Stripping + rec.	1.001 ± 0.006	1.387 ± 0.007	0.736 ± 0.003	1.110 ± 0.003
PID	87.442 ± 0.215	88.149 ± 0.178	74.653 ± 0.129	76.047 ± 0.103
L0 trigger	34.162 ± 0.284	40.895 ± 0.244	38.670 ± 0.179	41.281 ± 0.144
HLT1 trigger	80.712 ± 0.395	82.340 ± 0.287	84.818 ± 0.212	86.742 ± 0.154
HLT2 trigger	50.131 ± 0.561	45.985 ± 0.408	79.161 ± 0.276	75.461 ± 0.221
Pre-Selection	75.747 ± 0.658	78.787 ± 0.483	75.690 ± 0.300	77.833 ± 0.223
q^2 selection	98.716 ± 0.208	98.873 ± 0.146	98.538 ± 0.099	98.675 ± 0.070
HOP selection	99.951 ± 0.035	99.941 ± 0.027	99.952 ± 0.018	99.955 ± 0.013
ψ resonance Veto	93.253 ± 0.438	93.960 ± 0.304	93.421 ± 0.197	93.794 ± 0.144
GBDT selection	33.297 ± 0.833	34.155 ± 0.634	43.610 ± 0.405	30.161 ± 0.276
DTF converged	99.756 ± 0.144	99.894 ± 0.065	99.762 ± 0.058	99.881 ± 0.029
Single Candidate	100.000 ± 0.000	99.872 ± 0.093	99.580 ± 0.091	99.566 ± 0.084
Total efficiency	0.028 ± 0.001	0.047 ± 0.001	0.043 ± 0.001	0.049 ± 0.001

Table 7.12: Efficiencies for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ for different components of the selection for each analysis category. Each efficiency shown is conditional based on the one directly above.

Efficiency Categories	**Weighted Efficiency (%)**			
	LL-Run 1	**DD-Run 1**	**LL-Run 2**	**DD-Run 2**
Generator	18.566 ± 0.018	18.566 ± 0.018	19.815 ± 0.024	19.815 ± 0.024
Stripping + rec.	1.542 ± 0.008	2.440 ± 0.010	1.555 ± 0.005	2.550 ± 0.006
PID	93.965 ± 0.193	94.268 ± 0.150	93.653 ± 0.110	93.943 ± 0.085
L0 trigger	68.614 ± 0.256	75.248 ± 0.189	70.607 ± 0.143	73.165 ± 0.109
HLT1 trigger	86.003 ± 0.228	87.941 ± 0.155	92.362 ± 0.097	93.301 ± 0.070
HLT2 trigger	60.801 ± 0.348	57.262 ± 0.254	80.033 ± 0.163	76.104 ± 0.130
Pre-selection	74.597 ± 0.388	79.509 ± 0.265	74.619 ± 0.183	77.675 ± 0.135
q^2 selection	94.772 ± 0.229	94.720 ± 0.163	94.064 ± 0.116	94.145 ± 0.086
ψ resonance Veto	100.000 ± 0.000	100.000 ± 0.000	100.000 ± 0.000	100.000 ± 0.000
GBDT selection	33.883 ± 0.493	34.310 ± 0.350	48.545 ± 0.252	31.870 ± 0.172
DTF converged	99.439 ± 0.123	99.901 ± 0.031	99.791 ± 0.030	99.892 ± 0.016
Single Candidate	99.894 ± 0.047	99.666 ± 0.066	99.751 ± 0.033	99.786 ± 0.026
Total efficiency	0.125 ± 0.002	0.221 ± 0.003	0.260 ± 0.002	0.291 ± 0.002

trigger. The former rejection case leads to the Brem category stripping efficiencies being comparable to those expected for the No Brem stripping efficiencies, whereas, the trigger requirement appeared to cause the No Brem categories to have a lower trigger efficiency than expected. However, neither of these apparently inconsistent efficiencies impact the efficiencies of the different aspects of the selection from the HLT1 trigger row onwards and the total efficiency is likewise not affected.

Chapter 8

Fitting the Signal and Normalisation Modes

The development of a fit model in data for a particular decay requires two main components: signal and background. The former can be determined through examination and fitting to MC samples that model the signal structure and are described by a single PDF. The latter can contain multiple models that each contribute to the total background PDF model. For example, the combinatorial background, which is the random combination of tracks, is typically described by an exponential PDF as there are fewer high-energy tracks that can be combined to obtain the Λ_b^0 mass. However, there are also other backgrounds that require their own model, such as partially-reconstructed or misidentified backgrounds. The partially-reconstructed background refers to instances where the whole event is not reconstructed but provides a signal-like final state. There is inherently some missing mass that is unaccounted for which biases the overall kinematic properties of the background. Misidentified backgrounds occur when the wrong identity is assigned to a reconstructed particle. For example, the $\Lambda_b^0 \to \Lambda\mu^-\mu^+$ decay has a very similar decay topology to $\Lambda_b^0 \to \Lambda e^\mp\mu^\pm$, so were a muon to be misidentified as an electron, then it would act as a background to the rare mode. Additionally, there may be exclusive backgrounds which have the same observable final state as the signal and with a sufficiently high branching fraction that even the full selection cannot eliminate entirely. Therefore, before the development of the rare and normalisation fit models a study of the potential backgrounds was performed. For the normalisation mode this had previously been studied [151, 161], with the only identified contributing backgrounds being a misidentification background from $B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$ and the combinatorial background. The study performed for the backgrounds that could contribute to $\Lambda_b^0 \to \Lambda e^\mp\mu^\pm$ is discussed in Section 8.1. The fit models and their results are then discussed in Section 8.2 and Section 8.3 for the normalisation and rare mode respectively.

8.1 Study of Potential Backgrounds for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$

A variety of background sources were considered for the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ decay, which are summarised in Table 8.1. The majority of these were potential misidentified backgrounds from di-muon or di-electron modes that are abundant in LHCb. To estimate the contribution from individual decays, the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ selection was applied to the simulated MC for each background, which were weighted as described in Section 7.3. Additionally, a set of lepton misidentification weights are included to simulate the effect of misidentifying the final state of the MC to give an $e\mu$ final state that mimics the signal mode. The sum of weights before and after the selection was then used to determine the equivalent selection efficiency for the potential backgrounds. This allowed the normalised yield of that background expected in the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ data to be determined as

$$N_{normalised} = \varepsilon_{BG} \cdot \mathcal{B}_{BG} \cdot N_{\Lambda_b^0} = \varepsilon_{BG} \cdot \mathcal{B}_{BG} \cdot \frac{N_{norm}}{\mathcal{B}_{norm} \cdot \varepsilon_{norm}}, \tag{8.1.1}$$

where ε_{BG} is the total efficiency of the background MC, the product of the selection efficiency and the generator efficiency for that MC sample; $\mathcal{B}_{BG}$ is the branching fraction of the considered background mode; $N_{\Lambda_b^0}$ is the expected number of Λ_b^0 candidates in the dataset; N_{norm}, $\mathcal{B}_{norm}$, and ε_{norm} are the known yield, branching fraction, and efficiency of the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ mode respectively. This normalised yield of the background samples can then be used to show the expected relative contribution of the different background samples in the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ data.

A summary of the normalised expected background yields can be seen in Tables 8.2 and 8.3. The uncertainties are entirely statistical and arise from the uncertainty on weighted candidates. For cases where the remaining yield is zero, the uncertainty is propagated as if a single event remained.

The mass shapes of surviving backgrounds plotted relative to the blinded data and are shown in Figures 8.1 and 8.2. These use the background MC with the analysis selection applied, which is then scaled to the normalised yields shown in Tables 8.2 and 8.3. The total sum of the background contributions is also shown. This can then be directly compared to the data to see the relative effect of the potential remaining backgrounds on the data sidebands. This shows that any effect would be localised to the lower mass sideband, and that the only significant contributions are from the double semileptonic Λ_b^0 decays with $e\mu$ final states.

Figures 8.1 and 8.2 also show that the combined background contributions were always systematically higher than the level of the actual data even without considering the combinatorial background. This may be due to the fact that the simulated background

samples are reweighted using the rare mode weights and so the estimated level of these backgrounds may not be a reliable prediction of their contribution in data. This does not impact the signal fit model as the yields of the semileptonic modes, the only significant remaining background, are allowed to vary within loose constraints based on the absolute predicted yields determined here. Additionally, in each bin the data and predicted background contributions are consistent at a $< 3\sigma$ level, both when the background yields are

Table 8.1: The background samples considered with their branching fractions. Errors are combined unless specified: (r_f) is the systematic on the $f_{\Lambda_b^0}/f_d$ ratio; $(\mathcal{B})$ is the systematic based on the $\Lambda_b^0 \to \Lambda J/\psi$ branching fraction. Decay modes in **bold** are based on a preliminary value of $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ from Ref. [140].

Background Sample	Branching Fraction
$\Lambda_b^0 \to \Lambda e^- e^+$	$(1.08 \pm 0.28) \times 10^{-6}$ [22].
$\boldsymbol{\Lambda_b^0 \to \Lambda J/\psi (\to e^- e^+)}$	$(216.7 \pm 2.1(\text{stat}) \pm 16.1(r_f) \pm 11.9(\mathcal{B})) \times 10^{-7}$ [22, 140].
$\boldsymbol{\Lambda_b^0 \to \Lambda \psi(2S)(\to e^- e^+)}$	$(146.2 \pm 7.4(\text{stat}) \pm 10.9(r_f) \pm 8.1(\mathcal{B})) \times 10^{-8}$ [22, 140].
$\Lambda_b^0 \to \Lambda \mu^- \mu^+$	$(1.08 \pm 0.28) \times 10^{-6}$ [22].
$\boldsymbol{\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)}$	$(216.4 \pm 2.2(\text{stat}) \pm 16.1(r_f) \pm 11.9(\mathcal{B})) \times 10^{-7}$ [22, 140].
$\boldsymbol{\Lambda_b^0 \to \Lambda \psi(2S)(\to \mu^- \mu^+)}$	$(147.5 \pm 13.0(\text{stat}) \pm 11.0(r_f) \pm 8.1(\mathcal{B})) \times 10^{-8}$ [22, 140].
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+ \nu_e)\pi^-$	$(8.1^{+1.9}_{-1.7}) \times 10^{-4}$ [22].
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+ \nu_\mu)\pi^-$	$(8.1^{+1.9}_{-1.7}) \times 10^{-4}$ [22].
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+ \nu_e) e^- \overline{\nu}_e$	$(2.21^{+0.50}_{-0.47}(\text{stat}) \pm 0.04(\text{sys})) \times 10^{-3}$ [22, 162].
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+ \nu_\mu) \mu^- \overline{\nu}_\mu$	$(2.16^{+0.59}_{-0.54}(\text{stat}) \pm 0.17(\text{sys})) \times 10^{-3}$ [22].
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+ \nu_e) \mu^- \overline{\nu}_\mu$	$(2.21^{+0.50}_{-0.47}(\text{stat}) \pm 0.04(\text{sys})) \times 10^{-3}$ [22, 162].
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+ \nu_\mu) e^- \overline{\nu}_e$	$(2.16^{+0.59}_{-0.54}(\text{stat}) \pm 0.17(\text{sys})) \times 10^{-3}$ [22].
$B^0 \to K^0_S e^- e^+$	$(2.5^{+1.1}_{-0.9}) \times 10^{-7}$ [22].
$B^0 \to K^0_S J/\psi (\to e^- e^+)$	$(53.2 \pm 1.3) \times 10^{-6}$ [22].
$B^0 \to K^0_S \psi(2S) (\to e^- e^+)$	$(4.60 \pm 0.41) \times 10^{-6}$ [22].
$B^0 \to K^0_S \mu^- \mu^+$	$(3.4 \pm 0.4) \times 10^{-7}$ [22].
$B^0 \to K^0_S J/\psi (\to \mu^- \mu^+)$	$(53.1 \pm 1.3) \times 10^{-6}$ [22].
$B^0 \to K^0_S \psi(2S) (\to \mu^- \mu^+)$	$(4.64 \pm 0.53) \times 10^{-6}$ [22].
$B^+ \to K^+ J/\psi (\to e^- e^+)$	$(6.09 \pm 0.12) \times 10^{-5}$ [22].
$B^+ \to K^+ J/\psi (\to \mu^- \mu^+)$	$(6.08 \pm 0.12) \times 10^{-5}$ [22].
$\Lambda_b^0 \to \Lambda \pi^- \pi^+$	$(4.7 \pm 2.0(\text{stat}) \pm 0.2(\text{sys})) \times 10^{-6}$ [22].
$\Lambda_b^0 \to \Lambda K^- \pi^+$	$(5.8 \pm 1.3(\text{stat}) \pm 0.3(\text{sys})) \times 10^{-6}$ [22].
$\Lambda_b^0 \to \Lambda K^- K^+$	$(1.64 \pm 0.24(\text{stat}) \pm 0.08(\text{sys})) \times 10^{-5}$ [22].
$\Lambda_b^0 \to \Lambda(1405) J/\psi (\to \mu^- \mu^+)$	$(2.8 \pm 0.1(\text{stat}) \pm 1.0(\text{sys})) \times 10^{-6}$ [22].
$\Lambda_b^0 \to \Lambda(1520) J/\psi (\to \mu^- \mu^+)$	$(3.6 \pm 0.1(\text{stat}) \pm 0.9(\text{sys})) \times 10^{-6}$ [22].
$\Lambda_b^0 \to \Lambda(1600) J/\psi (\to \mu^- \mu^+)$	$(4.4 \pm 0.3(\text{stat}) \pm 0.6(\text{sys})) \times 10^{-6}$ [22].

estimated as previously described and when determined through an alternative method, as shown in Table B.1 in Appendix B. Therefore, these expected yields were considered to be accurate and could be used to determine the backgrounds that required controlling.

Table 8.2: Yields for background samples in Run 1 with misidentification and PID weights. Events with no MC events remaining are shown in the lower portion of the table below the double ruled lines, and the uncertainty on those yields is estimated based on propagating as if there was a single remaining MC event.

Background Sample	Weighted Normalised Yields			
	No Brem LL Run 1	**No Brem DD Run 1**	**Brem LL Run 1**	**Brem DD Run 1**
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$	$(1.42 \pm 0.46) \times 10^2$	$(1.91 \pm 0.48) \times 10^2$	$(6.28 \pm 2.98) \times 10^1$	$(1.25 \pm 0.39) \times 10^2$
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+\nu_\mu)e^-\overline{\nu}_e$	$(7.62 \pm 3.16) \times 10^1$	$(1.16 \pm 0.33) \times 10^2$	$(5.42 \pm 2.54) \times 10^1$	$(1.07 \pm 0.31) \times 10^2$
$\Lambda_b^0 \to \Lambda\mu^-\mu^+$	$(8.71 \pm 1.57) \times 10^{-3}$	$(2.41 \pm 0.51) \times 10^{-3}$	$(2.77 \pm 2.86) \times 10^{-5}$	$(1.23 \pm 1.57) \times 10^{-7}$
$\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$	$(8.53 \pm 6.85) \times 10^{-3}$	$(4.72 \pm 7.01) \times 10^{-4}$	$(4.24 \pm 4.09) \times 10^{-3}$	$(4.34 \pm 6.73) \times 10^{-5}$
$\Lambda_b^0 \to \Lambda\psi(2S)(\to \mu^-\mu^+)$	$(1.09 \pm 0.13) \times 10^{-2}$	$(5.11 \pm 0.84) \times 10^{-3}$	$(1.22 \pm 1.92) \times 10^{-4}$	$(8.85 \pm 9.57) \times 10^{-6}$
$\Lambda_b^0 \to \Lambda(1405)J/\psi(\to \mu^-\mu^+)$	$(0.58 \pm 1.10) \times 10^{-2}$	$(0.75 \pm 1.06) \times 10^{-2}$	$(3.27 \pm 9.23) \times 10^{-3}$	$(3.76 \pm 8.51) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda(1520)J/\psi(\to \mu^-\mu^+)$	$(0.49 \pm 1.17) \times 10^{-2}$	$(0.57 \pm 1.04) \times 10^{-2}$	$(0.33 \pm 1.07) \times 10^{-2}$	$(0.46 \pm 1.07) \times 10^{-2}$
$\Lambda_b^0 \to \Lambda(1600)J/\psi(\to \mu^-\mu^+)$	$(2.28 \pm 8.86) \times 10^{-3}$	$(2.49 \pm 7.79) \times 10^{-3}$	$(1.19 \pm 7.28) \times 10^{-3}$	$(1.56 \pm 6.95) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda K^-\pi^+$	$(0.22 \pm 2.42) \times 10^{-2}$	$(0.20 \pm 2.01) \times 10^{-2}$	(0 ± 1.37)	$(0.06 \pm 9.61) \times 10^{-4}$
$\Lambda_b^0 \to \Lambda\pi^-\pi^+$	$(0.32 \pm 3.34) \times 10^{-2}$	$(0.46 \pm 3.39) \times 10^{-2}$	$(0.05 \pm 4.43) \times 10^{-3}$	$(0.02 \pm 2.37) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda K^-K^+$	$(0.15 \pm 6.09) \times 10^{-3}$	$(0.35 \pm 7.79) \times 10^{-3}$	(0 ± 0.56)	(0 ± 0.38)
$B^0 \to K_S^0\mu^-\mu^+$	$(3.62 \pm 3.15) \times 10^{-3}$	$(4.84 \pm 3.05) \times 10^{-3}$	$(0.56 \pm 3.64) \times 10^{-4}$	$(0.34 \pm 2.72) \times 10^{-4}$
$B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$	$(1.13 \pm 0.14) \times 10^{-2}$	$(1.61 \pm 0.15) \times 10^{-2}$	$(4.57 \pm 0.95) \times 10^{-3}$	$(5.91 \pm 0.78) \times 10^{-3}$
$B^0 \to K_S^0\psi(2S)(\to \mu^-\mu^+)$	$(5.94 \pm 6.91) \times 10^{-2}$	$(8.01 \pm 6.32) \times 10^{-2}$	$(0.93 \pm 7.28) \times 10^{-3}$	$(0.68 \pm 5.83) \times 10^{-3}$
$B^0 \to K_S^0\psi(2S)(\to e^-e^+)$	$(0.30 \pm 2.59) \times 10^{-3}$	$(0.39 \pm 2.21) \times 10^{-3}$	(0 ± 0.08)	(0 ± 0.06)
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	$(0.43 \pm 4.08) \times 10^{-2}$	(0 ± 0.07)	(0 ± 0.12)	(0 ± 0.07)
$\Lambda_b^0 \to \Lambda e^-e^+$	$(0 \pm 0.81) \times 10^{-2}$	$(0 \pm 0.58) \times 10^{-2}$	$(0 \pm 0.81) \times 10^{-2}$	$(0 \pm 0.58) \times 10^{-2}$
$\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$	(0 ± 0.04)	$(0 \pm 0.24) \times 10^{-1}$	(0 ± 0.04)	$(0 \pm 0.24) \times 10^{-1}$
$\Lambda_b^0 \to \Lambda\psi(2S)(\to e^-e^+)$	$(0 \pm 0.34) \times 10^{-2}$	$(0 \pm 0.20) \times 10^{-2}$	$(0 \pm 0.34) \times 10^{-2}$	$(0 \pm 0.20) \times 10^{-2}$
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)e^-\overline{\nu}_e$	(0 ± 0)	(0 ± 0)	(0 ± 0)	(0 ± 0)
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+\nu_\mu)\mu^-\overline{\nu}_\mu$	(0 ± 0)	(0 ± 0)	(0 ± 0)	(0 ± 0)
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\pi^-$	(0 ± 0)	(0 ± 0)	(0 ± 0)	(0 ± 0)
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+\nu_\mu)\pi^-$	(0 ± 0)	(0 ± 0)	(0 ± 0)	(0 ± 0)
$B^0 \to K_S^0 e^-e^+$	$(0 \pm 2.03) \times 10^{-3}$	$(0 \pm 1.47) \times 10^{-3}$	$(0 \pm 2.03) \times 10^{-3}$	$(0 \pm 1.47) \times 10^{-3}$
$B^0 \to K_S^0 J/\psi(\to e^-e^+)$	(0 ± 0.17)	(0 ± 0.10)	(0 ± 0.17)	(0 ± 0.10)
$B^+ \to K^+ J/\psi(\to e^+e^-)$	$(0 \pm 0.18) \times 10^{-1}$	$(0 \pm 0.10) \times 10^{-1}$	$(0 \pm 0.18) \times 10^{-1}$	$(0 \pm 0.10) \times 10^{-1}$

Table 8.3: Yields for background samples in Run 2 with misidentification and PID weights. Events with no MC events remaining are shown in the lower portion of the table below the double ruled lines, and the uncertainty on those yields is estimated based on propagating as if there was a single remaining MC event.

Background Sample	**Weighted Normalised Yield**			
	No Brem LL Run 2	**No Brem DD Run 2**	**Brem LL Run 2**	**Brem DD Run 2**
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$	$(4.68 \pm 0.90) \times 10^2$	$(6.78 \pm 1.14) \times 10^2$	$(6.96 \pm 1.23) \times 10^2$	$(7.24 \pm 1.22) \times 10^2$
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \mu^+\nu_\mu)e^-\overline{\nu}_e$	$(2.49 \pm 0.62) \times 10^2$	$(4.15 \pm 0.82) \times 10^2$	$(4.41 \pm 0.93) \times 10^2$	$(5.62 \pm 1.12) \times 10^2$
$\Lambda_b^0 \to \Lambda\mu^-\mu^+$	$(9.33 \pm 1.17) \times 10^{-3}$	$(5.54 \pm 0.76) \times 10^{-3}$	$(8.97 \pm 4.86) \times 10^{-5}$	$(3.44 \pm 3.22) \times 10^{-5}$
$\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$	$(2.80 \pm 1.41) \times 10^{-3}$	$(1.68 \pm 0.92) \times 10^{-3}$	$(3.13 \pm 2.02) \times 10^{-3}$	$(1.09 \pm 2.78) \times 10^{-5}$
$\Lambda_b^0 \to \Lambda\psi(2S)(\to \mu^-\mu^+)$	$(1.48 \pm 0.11) \times 10^{-2}$	$(7.86 \pm 0.75) \times 10^{-3}$	$(2.35 \pm 1.08) \times 10^{-4}$	$(8.01 \pm 7.22) \times 10^{-5}$
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)\mu^-\overline{\nu}_\mu$	$(0.41 \pm 2.72) \times 10^1$	$(1.04 \pm 4.04) \times 10^1$	(0.09 ± 4.41)	(0.31 ± 8.19)
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)\pi^-$	$(0.55 \pm 1.00) \times 10^2$	$(1.60 \pm 1.72) \times 10^2$	$(0.67 \pm 3.50) \times 10^1$	$(0.15 \pm 1.60) \times 10^1$
$\Lambda_b^0 \to \Lambda(1405)J/\psi(\to \mu^-\mu^+)$	$(1.74 \pm 6.50) \times 10^{-3}$	$(2.82 \pm 7.90) \times 10^{-3}$	$(1.02 \pm 5.36) \times 10^{-3}$	$(1.25 \pm 5.70) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda(1520)J/\psi(\to \mu^-\mu^+)$	$(0.85 \pm 5.24) \times 10^{-3}$	$(1.84 \pm 7.15) \times 10^{-3}$	$(0.81 \pm 5.30) \times 10^{-3}$	$(0.30 \pm 3.11) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda(1600)J/\psi(\to \mu^-\mu^+)$	$(0.30 \pm 3.48) \times 10^{-3}$	$(0.38 \pm 3.52) \times 10^{-3}$	$(0.48 \pm 4.56) \times 10^{-3}$	$(0.49 \pm 4.66) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda K^-\pi^+$	$(0.07 \pm 1.49) \times 10^{-2}$	$(0.14 \pm 2.09) \times 10^{-2}$	$(0.06 \pm 8.78) \times 10^{-4}$	$(0.02 \pm 1.39) \times 10^{-3}$
$\Lambda_b^0 \to \Lambda\pi^-\pi^+$	$(0.32 \pm 5.17) \times 10^{-3}$	$(0.56 \pm 6.92) \times 10^{-3}$	(0 ± 0.83)	$(0.01 \pm 1.45) \times 10^{-4}$
$\Lambda_b^0 \to \Lambda K^-K^+$	$(0.36 \pm 7.69) \times 10^{-3}$	$(0.06 \pm 1.00) \times 10^{-2}$	$(0.02 \pm 3.84) \times 10^{-4}$	$(0.01 \pm 1.11) \times 10^{-3}$
$B^0 \to K_S^0\mu^-\mu^+$	$(2.32 \pm 2.59) \times 10^{-3}$	$(3.42 \pm 2.82) \times 10^{-3}$	$(0.62 \pm 5.62) \times 10^{-4}$	$(0.39 \pm 3.27) \times 10^{-4}$
$B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$	$(5.60 \pm 0.88) \times 10^{-3}$	$(8.52 \pm 1.15) \times 10^{-3}$	$(3.14 \pm 0.70) \times 10^{-3}$	$(3.49 \pm 0.87) \times 10^{-3}$
$B^0 \to K_S^0\psi(2S)(\to \mu^-\mu^+)$	$(4.11 \pm 2.83) \times 10^{-2}$	$(6.60 \pm 3.16) \times 10^{-2}$	$(0.52 \pm 2.77) \times 10^{-3}$	$(0.52 \pm 2.21) \times 10^{-3}$
$B^0 \to K_S^0 e^-e^+$	(0 ± 0.85)	(0 ± 0.81)	$(0.80 \pm 9.61) \times 10^{-5}$	(0 ± 0.81)
$B^0 \to K_S^0\psi(2S)(\to e^-e^+)$	$(0.00 \pm 4.42) \times 10^{-5}$	(0 ± 1.78)	(0 ± 2.10)	(0 ± 1.78)
$B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$	$(0.41 \pm 5.15) \times 10^{-3}$	(0 ± 0.07)	(0 ± 0.07)	(0 ± 0.07)
$\Lambda_b^0 \to \Lambda e^-e^+$	$(0 \pm 0.34) \times 10^{-2}$	$(0 \pm 0.31) \times 10^{-2}$	$(0 \pm 0.34) \times 10^{-2}$	$(0 \pm 0.31) \times 10^{-2}$
$\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$	(0 ± 0.05)	(0 ± 0.04)	(0 ± 0.05)	(0 ± 0.04)
$\Lambda_b^0 \to \Lambda\psi(2S)(\to e^-e^+)$	$(0 \pm 0.04) \times 10^{-1}$	$(0 \pm 0.04) \times 10^{-1}$	$(0 \pm 0.04) \times 10^{-1}$	$(0 \pm 0.04) \times 10^{-1}$
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\pi^-$	$(0 \pm 0.87) \times 10^2$	$(0 \pm 0.81) \times 10^2$	$(0 \pm 0.87) \times 10^2$	$(0 \pm 0.81) \times 10^2$
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)e^-\overline{\nu}_e$	$(0 \pm 0.49) \times 10^2$	$(0 \pm 0.46) \times 10^2$	$(0 \pm 0.49) \times 10^2$	$(0 \pm 0.46) \times 10^2$
$B^0 \to K_S^0 J/\psi(\to e^-e^+)$	(0 ± 0.16)	(0 ± 0.14)	(0 ± 0.16)	(0 ± 0.14)
$B^+ \to K^+ J/\psi(\to e^+e^-)$	(0 ± 0.21)	(0 ± 0.18)	(0 ± 0.21)	(0 ± 0.18)

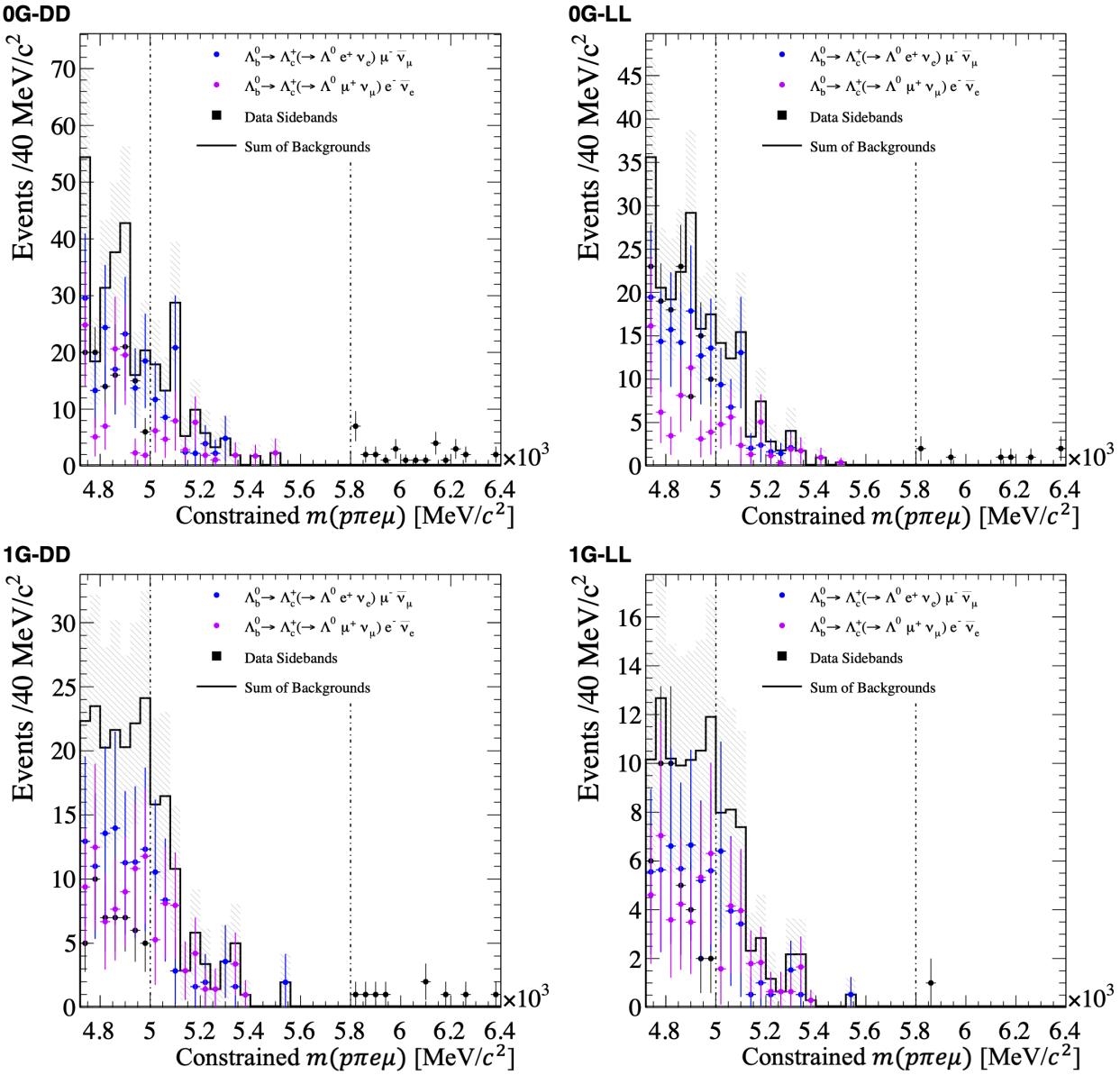


Figure 8.1: Mass distributions of remaining normalised background modes after selection is applied and using misidentification and PID weights, the combined background contribution and blinded data for Run 1 with DD (LL) categories on the left (right) and categories without (with) bremsstrahlung on the top (bottom).

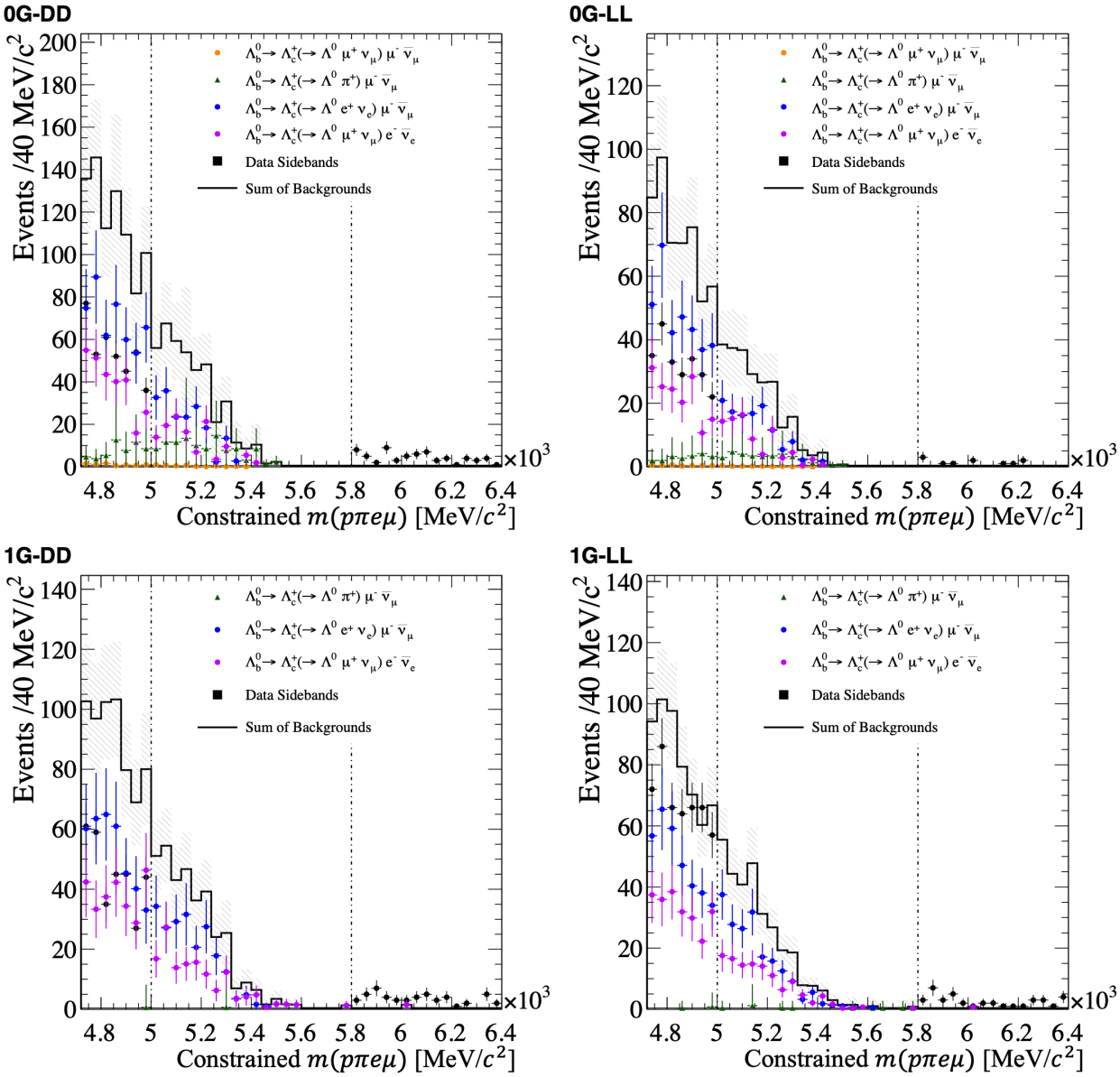


Figure 8.2: Mass distributions of remaining normalised background modes after selection is applied and using misidentification and PID weights, the combined background contribution and blinded data for Run 2 with DD (LL) categories on the left (right) and categories without (with) bremsstrahlung on the top (bottom).

8.1.1 Controlling the Potential Backgrounds

The resonant mode, $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$, used to normalise the number of rare mode candidates, can contribute directly below the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ signal peak where one of the two muons is misidentified as an electron. Even though the misidentification probability is at the level of 0.07%, the branching fraction is large enough ($\mathcal{O}(10^{-4})$) that it may contribute significantly to the background. For this reason, the region of q^2 in the proximity of the J/ψ resonance was vetoed, as anticipated in Section 7.2, to completely remove this component.

In contrast, $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ decays have a broader invariant mass distribution due to the poorer electron momentum resolution. The probability for an electron to pass the `isMuon` criterion is at the level of 1%, as evaluated from PID calibration samples. Aside from small phase-space differences, the branching fraction is the same as the di-muon mode. Therefore, this component has the potential to survive the full selection. However, in practice the requirement to have a muon in the trigger lines effectively removes not only this mode but effectively all di-electron backgrounds, as reported in Tables 8.2 and 8.3. In addition, for this background it is likely that the J/ψ veto, coupled with the 1% misidentification probability, would have lead to a negligible remaining background even without the effect of the trigger.

The $B^0 \to K_S^0 \ell^- \ell^+$ decay could be observed in the situation where one pion from the K_S^0 decay is misidentified as a proton. In the event that a muon is misidentified as an electron (or vice versa), a shoulder component might arise in the vicinity of the signal region for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$. However, the very low branching fraction ($\mathcal{B}(B^0 \to K_S^0 \mu^-\mu^+) = (3.39 \pm 0.35) \times 10^{-7}$ [22]), the requirement of one misidentified particle and stringent selection allow this source of background to be neglected.

The $B^0 \to K_S^0 J/\psi$ and $B^+ \to K^+ J/\psi$ modes have branching fractions of $(8.91 \pm 0.21) \times 10^{-4}$ and $(10.20 \pm 0.19) \times 10^{-4}$ [22] respectively, which is at a similar magnitude to $\Lambda_b^0 \to \Lambda J/\psi$. However, the tight invariant mass cut on the $p\pi$ pair included in the pre-selection, $m_{p\pi} \in [1105.68, 1125.68]\,\mathrm{MeV}/c^2$, forces the hadrons to originate from a Λ. This cut has a very high efficiency on the rare mode signal and reduces the B^0 modes by a factor of three in the DD categories. For this reason, this component is reduced to a fraction of the resonant Λ_b^0 modes, and can therefore be incorporated within them, even if the enhanced production of B^0 and B^+ mesons over Λ_b^0 baryons at LHCb is taken into account. Therefore, these B meson resonant backgrounds are also under control and accounted for in the normalisation fit by using the mass shape determined from MC, as described in Section 8.2.

Double semileptonic Λ_b^0 decays, $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \ell_i^+ \nu_{\ell i})\ell_j^- \overline{\nu}_{\ell j}$, with one electron and

one muon have the same observable final state as the rare mode. Since the muon and the electron have slightly different kinematic regimes depending on whether they are the decay products of the Λ_c^+ or Λ_b^0 baryon, with the latter having a wider phase-space, these two final states are treated separately. The branching fraction of this mode can be derived for both lepton combinations presented in Table 8.1, and are $\mathcal{O}(2\times10^{-3})$. This is a significant branching fraction and as the observed final state particles are identical to the rare mode, this background contributed significantly. The neutrinos from these decays would not be detected in LHCb so the Λ_b^0 mass peak is shifted to a lower mass, as seen in Figures 8.1 and 8.2. These figures and Tables 8.2 and 8.3 show that these backgrounds are likely to account for the majority of the background in the lower mass sideband after the full selection has been applied. Therefore, these decays were controlled by modelling them as an exclusive background component in the signal model fit. Additionally, as these are the dominant backgrounds for the rare mode a Λ_c^+ veto was considered to remove these backgrounds entirely, this is discussed in Section 9.6 where the upper limit for the rare mode is also evaluated.

Double semileptonic modes with pions, such as $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell^+\nu_\ell)\pi^-$ and $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\pi^+)\pi^-$, have lower branching fractions and require at least one misidentified particle in order to act as a potential background to $\Lambda_b^0 \to \Lambda e^\mp\mu^\pm$. Their yield, therefore, is subdominant with respect to the main $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ decays, so there was no need to include them as individual components. Although Tables 8.2 and 8.3 can show relatively large nominal yields for these decays, especially in Run 2, their uncertainty is often larger than the nominal value and hence they are highly compatible with zero.

Double semileptonic modes with ee or $\mu\mu$ final states have the same branching fractions as the $e\mu$ final state cases. However, the misidentification of the lepton to match the signal state appears to reduce significantly the expected yield for these cases, with Tables 8.2 and 8.3 showing < 1 candidates remaining for these samples in each analysis category.

The hadronic modes $\Lambda_b^0 \to \Lambda\pi^-\pi^+$, $\Lambda_b^0 \to \Lambda K^-\pi^+$, and $\Lambda_b^0 \to \Lambda K^-K^+$ may have branching fractions at a similar level to the normalisation mode, however they require two misidentified particles in order to be treated as signal. The misidentification of a pion or kaon to a muon is $\lesssim 2\%$ [106], which would be required to satisfy trigger conditions. Therefore these decays were not likely to contribute significantly to the data in the signal region.

The decays to excited Λ baryon modes: $\Lambda_b^0 \to \Lambda(1405)J/\psi$, $\Lambda_b^0 \to \Lambda(1520)J/\psi$, and $\Lambda_b^0 \to \Lambda(1600)J/\psi$ (all with $J/\psi \to \mu^+\mu^-$), have branching fractions at a similar level to the normalisation mode. However, the constraint for the $p\pi$ pair to come from a Λ combined with the $e\mu$ misidentification means that these modes are not likely to contribute

significantly in the signal region. The di-electron equivalents of these modes were not explicitly examined, but they would also have to pass the muon trigger and so could be even more strongly rejected as a potential background.

8.2 Fitting the Normalisation Mode

The $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ mode used the same selection as in the rare mode (see Chapter 7), except for the absence of the J/ψ veto and HOP cut. This includes the same GBDT cut, because the multivariate classifier applied to the normalisation mode was the same one that was trained on the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ decay. However, the α_{HOP} value was fixed to one and hence this variable would not contribute for the normalisation mode. Additionally, the GBDT was optimised in both Brem and No Brem categories but the normalisation mode does not include any potential bremsstrahlung photons to be recovered. Therefore, the normalisation mode is only considered in four fit categories split by the track type and data-taking period. It was decided to use the equivalent Brem category's optimised GBDT cut for the normalisation mode's selection to GBDT$^{fix-hop}$, *i.e.* the value for 'Brem-LL-Run 2' is used for 'LL-Run 2'. The reasoning for this is given in Section 7.4.1. The alignment of the selection between the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ modes also ensured that the majority of potential systematic errors could be reduced.

In each of the four normalisation analysis categories, an independent maximum-likelihood fit was performed to the DTF invariant mass of the Λ_b^0. The additional constraints present with the DTF mass, as discussed in Chapter 6, ensured a clean signal peak with good resolution. The yields of the normalisation mode were extracted from the fits, which, when combined with the rare mode's yields and the efficiencies (see Section 7.6), allow the rare mode's branching fraction to be determined with Equation 6.0.1.

The selected mass window around the signal peak was found to be very clean, with only two background components used to create the background PDF. The total fit model for the normalisation mode was:

- **Signal**: This used a DSCB shape [141], which uses the standard Crystal Ball (CB) PDF [163, 164] with exponential tails on either side of the peak. The parameters of the model are fixed from fitting to a simulated $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ sample, except for the mean and peak width parameters that were allowed to vary freely.
- **Combinatorial background**: This used an exponential function to model its contribution, with separate parameters used for each category.
- **Misidentified background**: This is from the $B^0 \to K_S^0 J/\psi (\to \mu^- \mu^+)$ mode, and

is present below the mass peak. A RooJohnson [142] function was used to model this background. This function has four parameters that were fixed from fitting a simulated $B^0 \to K^0_S J/\psi(\to \mu^-\mu^+)$ sample, except for the mean value that was allowed to vary to account for data/MC differences.

The fits used to fix the misidentified background are given in Figure 8.3, which demonstrate that this component was well modelled in the MC. Also, it shows the fit shape developed is appropriate for use in the fit to data for the normalisation mode. This is also true for the fits used to determine the fit shape for the $\Lambda^0_b \to \Lambda J/\psi(\to \mu^-\mu^+)$ signal, which are given in Figure 8.4.

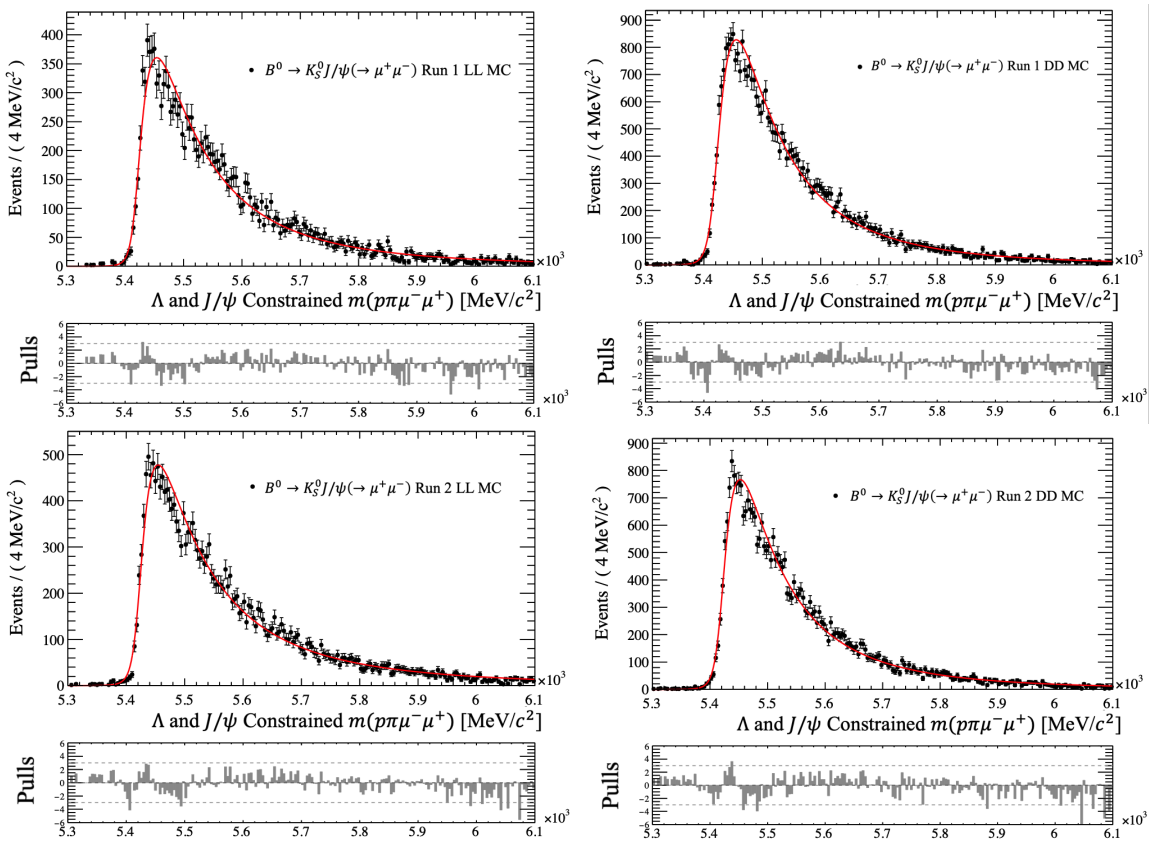


Figure 8.3: Fits to the invariant mass of $B^0 \to K^0_S J/\psi(\to \mu^-\mu^+)$ simulated candidates built from LL (Left) and DD (Right) tracks, for Run 1 (Top) and Run 2 (Bottom).

Finally, the fit performed on data over the full mass window is shown in Figure 8.5. The obtained yields, used to calculate the normalisation factors, are listed in Table 8.4. This includes the raw yields from the fits as well as luminosity and efficiency weighted yields, and a ratio of Run 1 to Run 2 yields given as

$$R = \frac{N_1}{N_2} \cdot \frac{\varepsilon_2 \cdot \mathcal{L}_2 \cdot \sigma_2 \cdot f_{\Lambda^0_b,2}}{\varepsilon_1 \cdot \mathcal{L}_1 \cdot \sigma_1 \cdot f_{\Lambda^0_b,1}}, \tag{8.2.1}$$

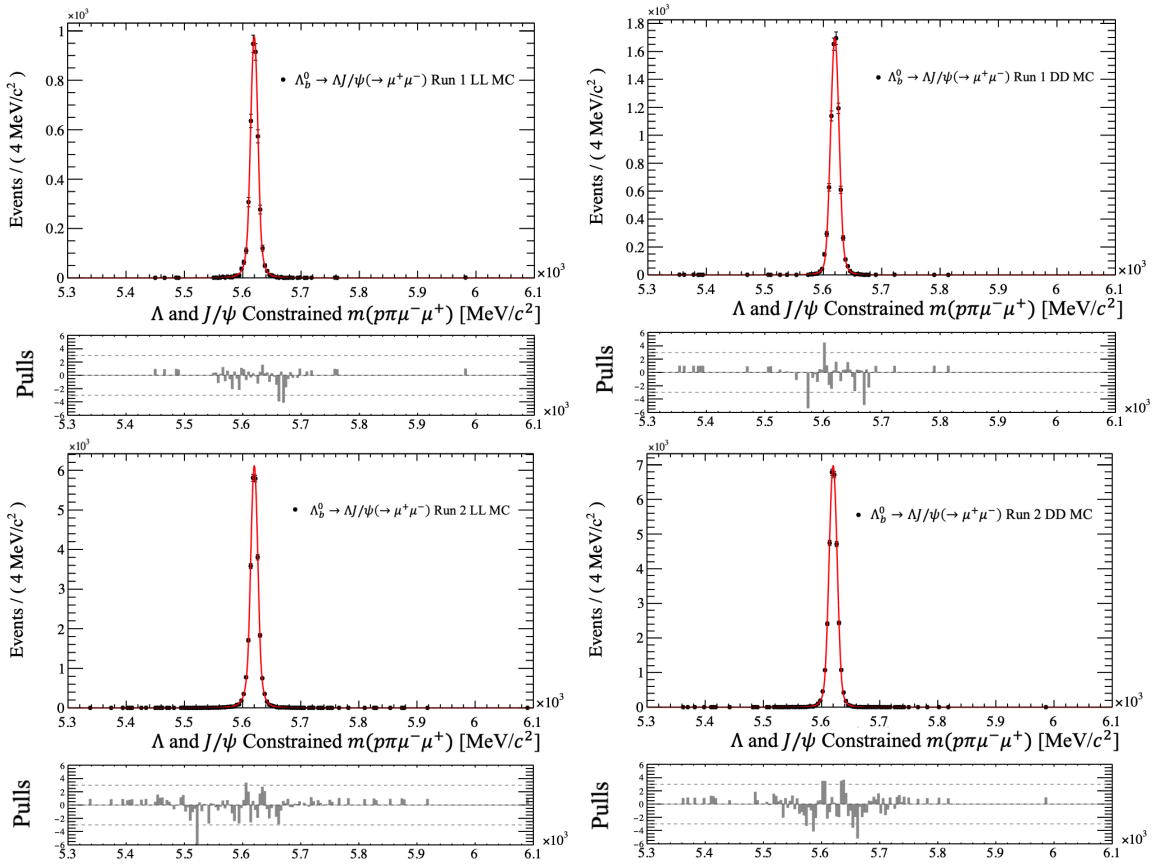


Figure 8.4: Fits to the invariant mass of $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ simulated candidates built from LL (Left) and DD (Right) tracks, for Run 1 (Top) and Run 2 (Bottom).

where the ratio, R, should be $\sim$1; the subscript 1 or 2 refers to the run period; N is the raw yield, ε is the total efficiency; $\mathcal{L}$ is the integrated luminosity; σ is the cross-section and $f_{\Lambda_b^0}$ is the Λ_b^0 fragmentation fraction which is included due to the use of the inclusive production cross-section, $\sigma(pp \to b\bar{b}X)$. This was measured to be $(298 \pm 0.2(\text{stat}) \pm 36(\text{syst}))$ µb at 8 TeV and $(495 \pm 0.2(\text{stat}) \pm 52(\text{syst}))$ µb at 13 TeV [165, 166]. The latter is used directly as the Run 2 value, while the 8 TeV number is scaled by a fraction 23/24 to obtain the full Run 1 value. This factor accounts for the expected linear behaviour of the cross-section with energy and that approximately two-thirds of the integrated luminosity of Run 1 was at 8 TeV. The value for $f_{\Lambda_b^0}$ is 0.185 ± 0.014 and 0.179 ± 0.013 for Run 1 and Run 2 respectively, as derived using the Run 2 value for $\frac{f_{\Lambda_b^0}}{f_u+f_d} = 0.259 \pm 0.018$ [138] for both runs and $f_u + f_d = 0.713 \pm 0.026$ or 0.692 ± 0.015 [139] for Run 1 and Run 2 respectively. The uncertainties within the table for the yields and efficiencies are purely statistical, those on the cross-section include systematic uncertainties that are independent between the two measurements.

The difference between the observed ratio in DD and LL is due to the over-estimation of the DD efficiencies in Run 1 at the level of 20% for K_S^0 and Λ production [137]; when

this effect is corrected for, the DD and LL ratios were consistent at the level of 6% with a 7% uncertainty.

Table 8.4: Normalisation mode yields for different applied corrections. The raw yield is the direct result from the fit, efficiency corrected values are the yield divided by the total efficiency, luminosity corrected values are the yields for $1\,\text{fb}^{-1}$, and the 'all corrections' row shows the full corrections including the cross-section *i.e.* $\frac{N}{\varepsilon\mathcal{L}\sigma}$. The ratio of Run 1 to Run 2 yields is also shown. The uncertainties are purely statistical.

Yield Type	**Track Type**	**Run 1**	**Run 2**
Raw Yield	LL	1762 ± 44	8719 ± 98
	DD	2353 ± 50	9203 ± 99
Efficiency Corrected	LL	4670240 ± 178671	12993187 ± 210631
	DD	3468157 ± 102760	12328418 ± 185352
Luminosity Corrected ($/\text{fb}^{-1}$)	LL	590 ± 15	1627 ± 18
	DD	789 ± 17	1717 ± 19
All Corrections ($/10^{-5}\ \text{fb}^{-1}$)	LL	2.97 ± 0.26	2.73 ± 0.35
	DD	2.20 ± 0.19	2.59 ± 0.33
Ratio $\left(\frac{\text{Run 1}}{\text{Run 2}}\right)$	LL	1.09 ± 0.13	
	DD	0.85 ± 0.10	
	DD (corrected)	1.02 ± 0.12	
	Total	0.93 ± 0.11	
	Total (corrected)	1.05 ± 0.12	

8.3 Fitting the Signal Mode

The $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ mode uses the selection specifed in Chapter 7. A set of maximum likelihood fits were performed simultaneously in the eight analysis categories outlined in Chapter 6. Initially, these fits were performed on blinded datasets to ensure that the background model could adequately describe the mass shape without biasing the signal yield determination. Section 9.1 describes the procedure to validate the analysis strategy before and after unblinding, including some preliminary estimates.

The fits were performed on the DTF invariant mass of the Λ_b^0. The fit model used was:

- **Signal**: This was modelled by a DSCB shape [141], with parameters fixed from fits to simulated $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ samples, including the parameter related to the mass peak's resolution.

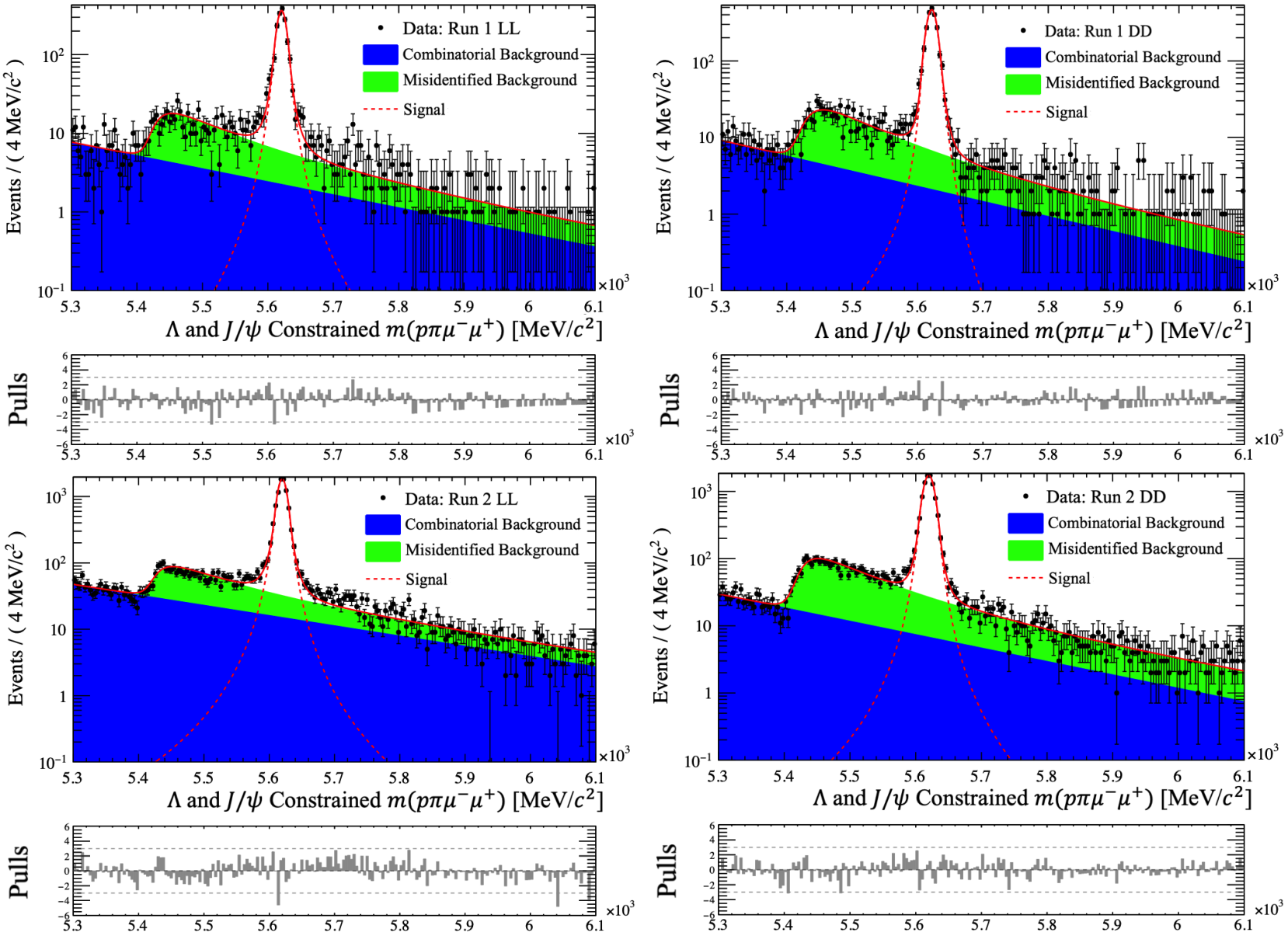


Figure 8.5: Fits to the invariant mass of $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ data candidates built from LL (Left) and DD (Right) tracks; Run 1 (Top) and Run 2 (Bottom). The misidentified background refers to the $B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$ mode.

- **Combinatorial Background**: This is handled identically to the normalisation mode, modelled as an exponential with each independent category having different parameters. Its shape was constrained based on the slope parameters obtained from fitting an exponential to SS $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ data.

- **Exclusive Backgrounds**: This accounts for the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ and $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ backgrounds discussed in Section 8.1.1. These are considered separately and included individually. For both, a RooJohnson [142] shape is used to describe them, the parameters of which are fixed by fits to the MC samples.

As previously discussed in Section 7.3, MC simulation can incorrectly model certain aspects of the sample. This could include the resolution for the invariant mass such that $\sigma_{\text{MC}} \neq \sigma_{\text{data}}$. Therefore as the signal PDF parameters are fixed from fitting to this MC sample the signal shape for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ could be incorrectly modelled. To account for this, the differences in the resolution between data and simulation can be estimated using a data/MC factor, $\mathcal{C}$. However, no SM cLFV decays have been observed, so there are no calibration channels with an electron and a muon in the final state that can be used to determine this $\mathcal{C}$ factor directly. The information on the resolution must instead be extracted from the combination of a muonic and an electronic decay channel. Appendix C shows how the $\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$ and $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ modes can be used to estimate this correction factor for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$. This study ultimately produced values of $\mathcal{C}$ between 0.8 and 1.4 depending on the analysis category (See Table C.1), but with very high uncertainties such that they were all essentially compatible with a value of one. Additionally, the estimated upper limit was evaluated when varying this $\mathcal{C}$ factor by 40% in each category with no obvious difference compared to the nominal $\mathcal{C}$ factor value. This implies that the upper limit itself is not sensitive to the corrected signal peak. Thus the standard MC value is used and fixed in the signal fit, to avoid introducing additional large uncertainties unnecessarily[1].

The signal yield of the rare mode is determined via a signal branching fraction variable defined in the fit procedure that aligns with the expression in Equation 6.0.1 and multiplies the signal PDF. The corresponding values of efficiencies, the normalisation mode's branching fraction and yields are included and define a set of Gaussian constraints specified by their total uncertainties. The rare mode's branching fraction is allowed to vary freely, such that the yield of the constrained fit simultaneously provides a measurement of the rare mode's branching fraction with all the inputs and their uncertainties accounted for. In this way, the systematic uncertainties are mostly incorporated via those

[1]This lack of sensitivity in the case of the estimated upper limit may be due to the fact that, by construction, this limit is performed on a background-only dataset (see Chapter 9). Once the dataset is unblinded this may require re-evaluation.

of the efficiency ratio, with the systematic uncertainties of the branching fractions of the normalisation mode and the exclusive backgrounds being included separately.

The background PDF consists of three components that were constrained to better reflect the expected contribution of each in data. For the combinatorial background, this was done by constraining the shape of the exponential to that obtained from fits to the SS $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ data, as this should be purely combinatorial and have a similar sample size to the equivalent combinatorial in OS data. For the exclusive backgrounds, their yields were constrained based on the expected yields determined previously in Tables 8.2 and 8.3. However, the specific formulation of these yields in the fit model was made relative to α in Equation 6.0.1 to account for the correlations that arise from many shared variables. Additionally, the relative fraction of $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ to $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ candidates are also constrained. The background-only, signal-only and combined signal and background PDFs are then given by Equations 8.3.1 to 8.3.3.

$$\begin{aligned}\text{PDF}_{BG} &= N_{comb}\cdot\text{PDF}_{exp} + \frac{(\varepsilon_{\Lambda e}\mathcal{B}_{\Lambda e} + \varepsilon_{\Lambda\mu}\mathcal{B}_{\Lambda\mu})}{\varepsilon_{rare}\cdot\alpha}\cdot\left(\frac{N_{\Lambda e}\cdot\text{PDF}_{\Lambda e} + N_{\Lambda\mu}\cdot\text{PDF}_{\Lambda\mu}}{N_{\Lambda e}+N_{\Lambda\mu}}\right), \\ &= N_{comb}\cdot\text{PDF}_{exp} + N_{excl}\cdot\text{PDF}_{excl}, \end{aligned} \tag{8.3.1}$$

$$\text{PDF}_{rare} = \frac{\varepsilon_{rare}}{\varepsilon_{norm}}\cdot\frac{N_{norm}}{\mathcal{B}_{norm}}\cdot\mathcal{B}_{rare}\cdot\text{PDF}_{DSCB} = \frac{\mathcal{B}_{rare}}{\alpha}\cdot\text{PDF}_{DSCB}, \tag{8.3.2}$$

$$\text{PDF}_{sig+BG} = \text{PDF}_{rare} + \text{PDF}_{BG}, \tag{8.3.3}$$

where Λe here refers to the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ sample and $\Lambda\mu$ to the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ sample; $N_{\Lambda e,\Lambda\mu}$ are the expected background yields as determined in Section 8.1; PDF_{exp} is an exponential function; $\text{PDF}_{\Lambda e,\Lambda\mu}$ are RooJohnson functions; PDF_{excl} is the sum of the exclusive backgrounds; and PDF_{DSCB} is a DSCB function. The nature of these PDFs and their parameters are discussed at the start of this section.

Figures 8.6 and 8.7 show the fits performed on the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ simulated candidates, divided into the eight analysis categories, which were used to fix the mass shape of the signal model. Figure 8.8 shows the fits performed to the blinded data of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$. The good description of the data suggests that all significant sources of background have been taken into account. Therefore, an estimated upper limit was determined for the signal branching fraction.

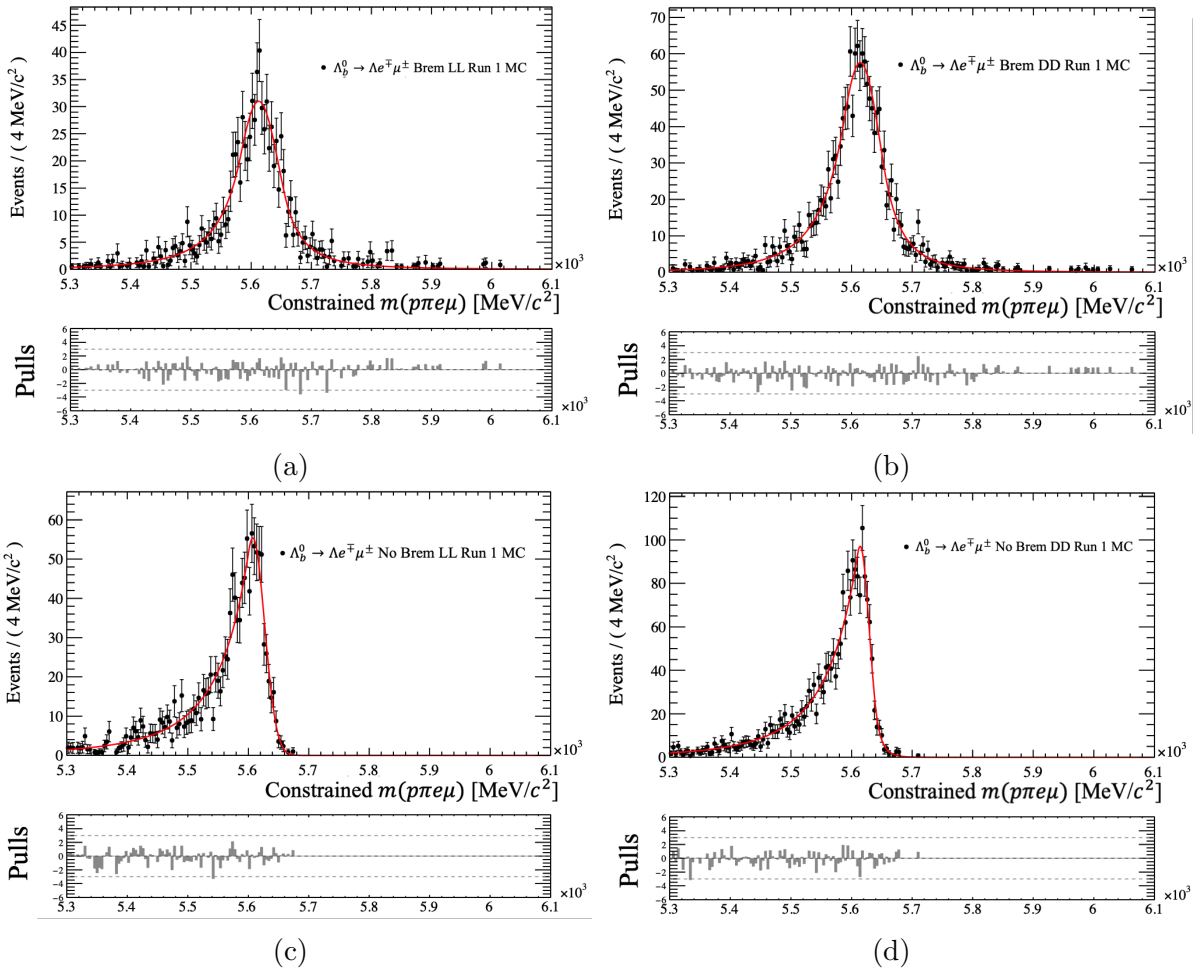


Figure 8.6: Fits to the invariant mass of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ simulated candidates to determine the parameters for the signal model in Run 1. The categories with (without) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column.

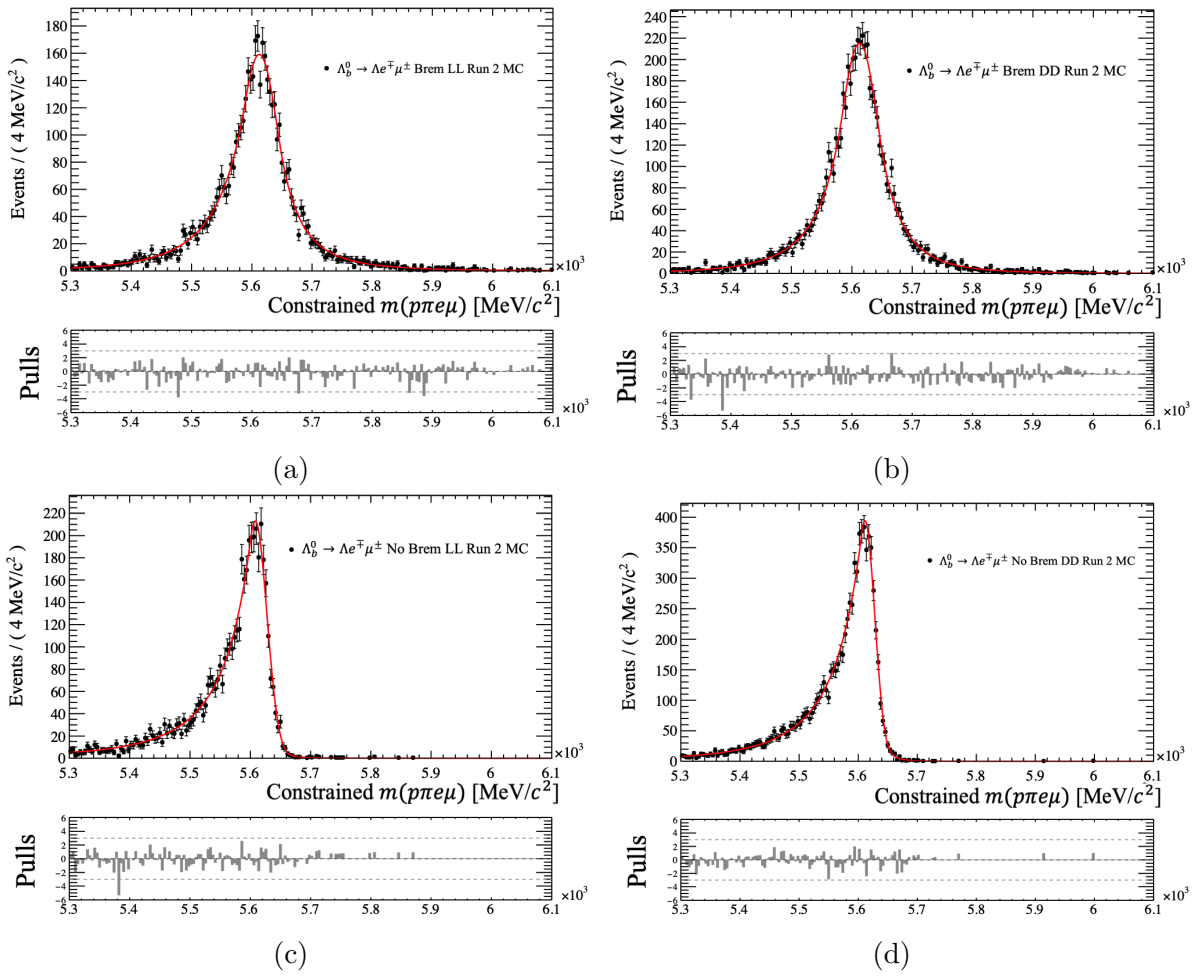


Figure 8.7: Fits to the invariant mass of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ simulated candidates to determine the parameters for the signal model in Run 2. The categories with (without) bremsstrahlung photons reconstructed are shown in the top (bottom) row, and those with LL (DD) tracks are shown in the left (right) column.

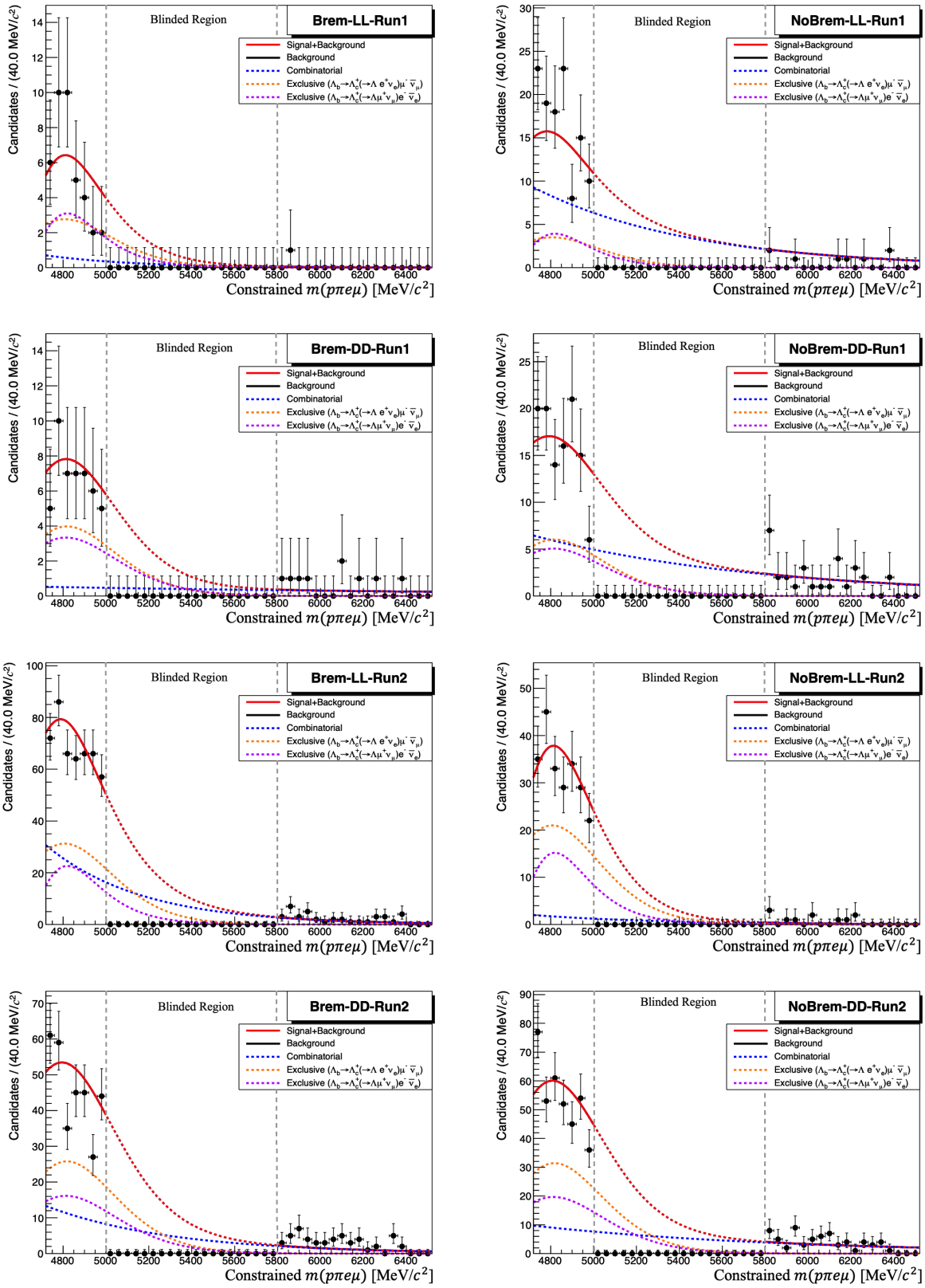


Figure 8.8: Fits to the invariant mass for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ data in each analysis category as labelled on each plot. Here the combined signal and background, and the background-only PDFs are identical due to the signal region being blinded.

Chapter 9

Determination of an Estimated Upper Limit to $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$

The analysis was blinded at the time of writing, and therefore an absolute determination of the branching fraction of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ or an upper limit to it cannot be performed as the signal yield is unknown. Instead, an estimated upper limit of the branching fraction of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ in the absence of signal in the currently blinded region is determined. This used a background-only proxy dataset, henceforth referred to as the 'BG proxy dataset'. This BG proxy dataset used a fit to the data sidebands with no signal component, where the fit was extrapolated into the blinded region, to generate a new dataset which matches the shape of the data sidebands over the entire data range (See Figure 9.1). The expected number of candidates in the data sidebands aligns with a background-only unblinded sample. This BG proxy dataset can then be evaluated with the 'CL_s' method (see Section 9.3) to obtain an upper limit to the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ branching fraction. This also provides an approximate signal yield of the rare mode that could be observed in the fit even without a signal present and hence a statistical benchmark that the unblinded result would have to exceed to be considered a significant observation.

Within this chapter, the strategy for unblinding the analysis is described in Sections 9.1 and 9.2 together with the cross-checks to reduce bias when doing so. The methodology behind the estimated upper limit determination is discussed in Section 9.3, with the results of this shown in Section 9.4. A discussion of the potential systematics, how they are controlled or determined and their relative influence, is given in Section 9.5. Finally, in Section 9.6 the estimated upper limit for the rare mode is evaluated with the inclusion of a Λ_c^+ veto to remove the dominant remaining background.

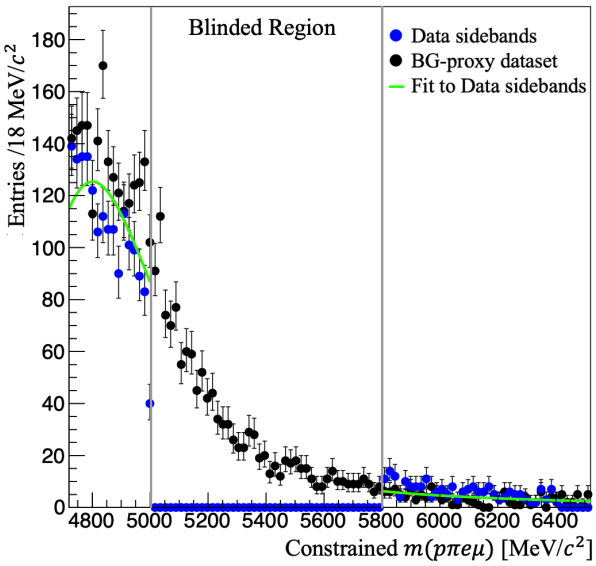


Figure 9.1: The simulated background-proxy dataset (black) compared to the blinded data sidebands (blue) and the data sideband fit used to generate it (green line).

9.1 Unblinding Cross-checks

Before the dataset is unblinded, the background model used within the fits should be validated. This ensures that the background is under control, while not biasing the particular signal model as the effect of adjustments to that model is not known when the signal region is blinded. Figure 8.8 shows the fits to the blinded invariant mass distribution. Pull distributions of these plots were observed and showed that in each of the eight analysis categories both the upper and lower mass sidebands are very consistent with the background model. Additionally, the combined signal and background model does not deviate from the background-only case in the data sidebands and so there is unlikely to be any signal that leaks outside the blinded region.

Overall, this already provided a reasonable validation for the background model but additional checks for potential missing backgrounds or unexpected behaviour were performed. These included the evaluation of different aspects of the selection to ensure that they do not create artificial peaking structures whilst removing backgrounds, such as described in Sections 7.2.1 and 7.4. The former relates to the HOP cut, and the latter to the correlations between the invariant mass or q^2 variable and the GBDT response. Another cross-check performed was to examine the different combinations of the four final-state particles that together are reconstructed to give the invariant mass of the Λ_b^0 used as the fit variable after the full selection was applied. In particular, the $m(p\pi e)$, and $m(p\pi\mu)$ mass distributions, which were evaluated while still blinding the signal region in $m(p\pi e\mu)$. Figure 9.2 shows these distributions, and there are no unexpected mass struc-

tures or resonances within these distributions which could represent some unaccounted background. The differences between the OS and SS distributions can be fully explained by the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \ell_i^+ \nu_{\ell i}) \ell_j^- \overline{\nu}_{\ell j}$ backgrounds. This is shown by the overlaid mass distributions from the MC samples for the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda \ell_i^+ \nu_{\ell i}) \ell_j^- \overline{\nu}_{\ell j}$ backgrounds, which peak in the same places that the OS data sidebands do, without the same peak present in the SS data sidebands.

Additional distributions where the mass of final state particles was exchanged for potential misidentified particles, *e.g.* $e \to \pi$, were also performed. The presence of a mass peak would indicate the existence of a misidentification background which would not have been accounted for. However, no such features were found so it was therefore concluded that the backgrounds were fully described by the developed background model described in Section 8.3.

9.2 Unblinding Strategy

Although the SM predictions suggest no observable signal for the rare mode within the current data sample, this was not assumed. After performing the pre-unblinding cross-checks described in the previous section and ensuring that the data fits converge with an accurate error matrix, the data can be unblinded. The presence of a signal and its significance would then be determined using the $p-$value of the $\mathrm{CL_s}$ method, see Section 9.3. This compares the fraction of pseudo-experiments that perform worse, *i.e.* larger fit χ^2 value, than the blinded data sample for both the scenario where there are and where there are not any signal candidates. This $p-$value will first be determined without including the blinded regions of the data distributions themselves. If the observed result differs from the expected background-only scenario by a value exceeding 2.5σ then additional checks will be performed to verify the background model. For example, the distributions for the invariant mass combinations will be studied as they were before unblinding, however without the Λ_b^0 mass being blinded. This would highlight any new backgrounds without providing direct information on the Λ_b^0 mass distribution that is being fitted, and thus maintain minimal biasing if fit models need to be corrected. If the signal remains significant after these checks have been performed and the data is fully unblinded, then the associated $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ yields and branching fraction will be published. If, instead, the signal is not significant an upper limit to the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ branching fraction will be determined. This would use the same method as the estimated upper limit determination reported in this thesis, but with the actual unblinded dataset.

Prior to unblinding, as part of the cross-checks the estimated upper limit to

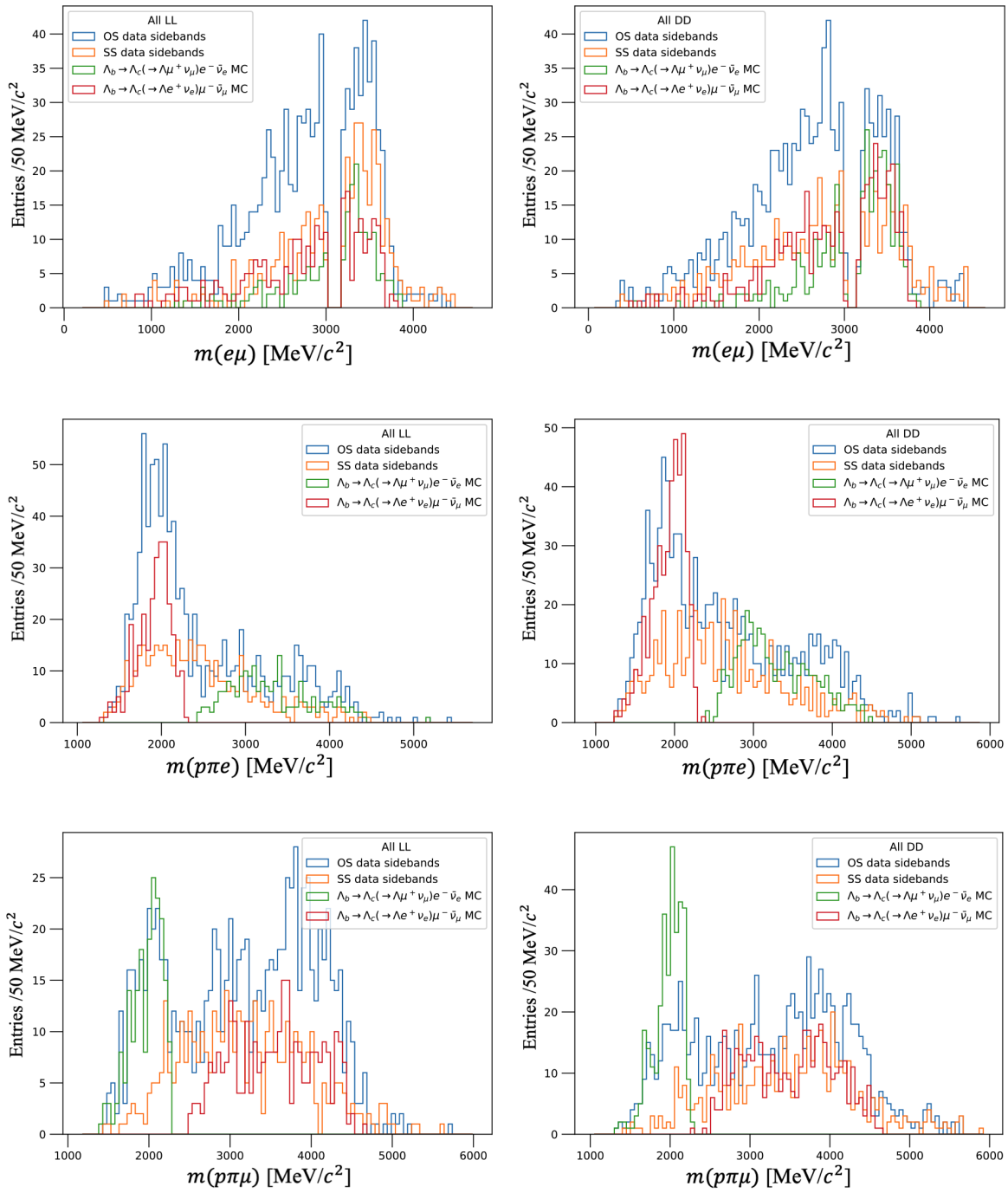


Figure 9.2: Reconstructed invariant mass combinations of (Top) $e\mu$, (Middle) $p\pi e$, and (Bottom) $p\pi\mu$ in the OS and SS data sidebands for the combined (Left) LL and (Right) DD categories. The simulated samples for the exclusive backgrounds are overlaid to show their contributions to the OS data.

$\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ in the absence of any signal can be determined using the CL_s method with the BG proxy dataset. This dataset by construction has no signal present and covers the full mass range, representing a scenario where no observable signal is present after unblinding. Therefore, this provides the value for the null hypothesis to compare to the true situation post-unblinding. As an intermediate stage just before unblinding, a fit to the full unblinded mass range including signal and background could be performed while blinding the signal parameters and the resulting plots, but not the dataset itself. From the background model of this fit, another BG proxy dataset could be generated in the same way as the one based on the blinded dataset, and then used to determine an estimated branching fraction upper limit as well. This ensures that the signal and background model has sufficient fit stability by insisting that the fits converge with an accurate error matrix. If the fits are not stable, then adjustments can be made to the model whilst it is still blinded, which would not strongly bias the final determined yield in the process. Additionally, the estimated upper limits in both cases should be compatible if there are no additional backgrounds or structures within the blinded region as both represent the background-only scenario. If the fits cannot converge or the upper limits are significantly different, then this could imply the existence of some additional background present within the blinded region. In this situation, a partial unblinding could be performed, and if a secondary peak or extended tail is observed, the background model will be reconsidered. Then the checks described both in this section and the preceding one would be repeated. If these checks cannot control the fit stability then a full unblinding may need to be performed to get a complete picture of the mass distribution in the blinded region, modelling and understanding any additional backgrounds.

Consistency checks can also be performed by comparing the measurement of the single-candidate sensitivity of the rare mode branching fraction, given as α in Equation 6.0.1, for the different analysis categories.

9.3 The CL_s Method

The CL_s method [143, 144] was originally developed to determine the bounds on the Higgs boson mass at the Large Electron-Positron collider (LEP). It was designed to allow for frequentist-like analysis of searches that have low signal yields and background present, to allow for accurate and rigorous confidence levels to be placed on limits. The 'confidence levels' that are calculated with this method cannot be interpreted in the same way as typical frequentist confidence levels or the bayesian credible intervals, but shall be referred to as confidence levels for ease of reference.

When attempting to distinguish between two possible non-composite hypotheses, the Neyman-Pearson Lemma [167] shows the most powerful test statistic to discern the null hypothesis from the alternative one is a ratio of likelihood functions. These likelihood functions represent the probability density of observing the dataset given a theoretical model. The null hypothesis is defined in this case as the background-only scenario, and the alternative hypothesis as the presence of signal as well as background. From Wilks' theorem [168], a distribution of the form $-2\ln\mathcal{L}$ is distributed as a χ^2 function. Therefore, when performing maximum likelihood fits to determine the signal yield, and thus the rare mode's branching fraction, this will minimise the resultant χ^2 value. This leads to a variable that can be used as a test statistic,

$$Q_{(s+b,\,b)}(\vec{\alpha}) = \chi^2_{(s+b,\,b)}(\vec{\alpha}) = -2\ln\mathcal{L}_{(s+b,\,b)}(\vec{\alpha}), \tag{9.3.1}$$

where $\vec{\alpha}$ is a vector of parameters that describe the likelihood model and the subscript refers to whether the fit model is the combined signal and background $(s+b)$ or background-only (b) scenario. This can be converted into a p-value corresponding to the probability that the value, Q_0, for a test statistic such as χ^2_{s+b}, is exceeded in a χ^2 distribution with a single degree of freedom. This corresponds to

$$p(Q < Q_0) = 1 - \mathrm{CL} = \frac{1}{\sqrt{(2)}\Gamma\left(\frac{1}{2}\right)}\int_{Q_0}^{\infty} e^{\frac{t}{2}}t^{-\frac{1}{2}}\mathrm{d}t, \tag{9.3.2}$$

where CL is the confidence level of this scenario; and Γ is the gamma function. Equation 9.3.2 can be evaluated for both the minimised $\chi^2_{min,s+b}$ and $\chi^2_{min,b}$ values to determine a p_{s+b} and p_b respectively. From this the $\mathrm{CL_s}$ p-value is defined,

$$p_{\mathrm{CL_s}} = 1 - \mathrm{CL_s} = \frac{p_{s+b}}{1-p_b}. \tag{9.3.3}$$

If a considered region has a CLs value less than a specified type-II error threshold, *i.e.* $\mathrm{CL_s} < 1 - \mathrm{CL}$, the alternative, combined signal and background hypothesis will be excluded at a confidence of CL (See Figure 9.3), allowing for an upper limit to be determined [143, 144]. A confidence level of 95% is reported in this thesis.

9.3.1 Implementation of the $\mathrm{CL_s}$ Method in GammaCombo

The $\mathrm{CL_s}$ method itself can be implemented in code in different ways, through explicit construction using statistics modules such as `RooStats` or as a feature from more dedicated software packages. The GammaCombo framework [169, 170] is a statistical framework developed by members of the LHCb collaboration that was originally designed to combine

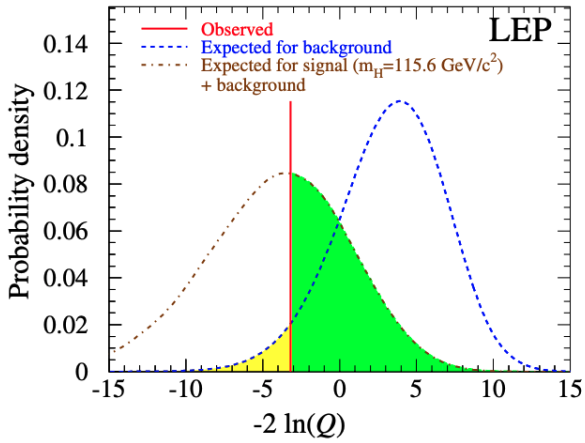


Figure 9.3: The PDFs of the combined Higgs search at LEP for the different hypotheses [143] as an example of the CL_s method where the yellow area on the left represents $1 - CL_b$ and the green area on the right is CL_{s+b}

measurements of the γ angle of the standard unitarity triangle. It has expanded to a generic framework that combines measurements as well as fitting datasets with specified likelihood inputs and dependencies on the parameter of interest. Additionally, it provides several methods that allow for 1D or 2D confidence intervals to be constructed and limits obtained, including the CL_s method. It takes as input a collection of: the input dataset; the fit models for the null and alternative hypotheses; the associated variables with an initial fit to the dataset in both fit models; the variables within the dataset that are used as observables; any constraints applied to the variables; and a defined parameter of interest. These are all stored within a `RooWorkspace` object from the `RooFit` library.

To obtain confidence levels within GammaCombo two main methods are utilised. The first is the profile likelihood method, which is referred to as the 'Prob' method. The second is the pseudo-experiment based frequentist method, which is referred to as the 'Plugin' method. The former is less computationally intensive but in many models the Prob method has poorer frequentist coverage [169]. The Plugin method requires the profile likelihood examined from the Prob method as a starting point, so both are used and can be compared. In both methods the parameter of interest examined is the rare mode branching fraction, given by Equation 6.0.1, or r, given by Equation 6.0.3, where the value of this parameter is fixed to a series of values within a given range while fitting the given dataset.

The Profile Likelihood (Prob) Method: An initial, global fit is performed with the combined signal and background fit model where the value of the rare mode branching fraction, $\mathcal{B}_{rare}$, is not limited to any particular value. These fits are maximum likelihood fits, where the 'negative log-likelihood', $-\ln \mathcal{L}(\vec{\alpha})$, is minimised. This can be converted

into the test statistic in Equation 9.3.1 by multiplying the resultant negative log-likelihood by two. This initial fit serves to determine a global minimum without fixing the value of $\mathcal{B}_{rare}$, which has an associated test statistic of $\chi^2_{min}(\vec{\alpha}_{global})$. The user defines a range of values over which to examine $\mathcal{B}_{rare}$ and a number of scan points. At each of these points, i, the value of $\mathcal{B}_{rare}$ is fixed to $\mathcal{B}_{rare,i}$. The fits are then repeated under this additional constraint and for each scan point the value of the test statistic, $\chi^2_{min}(\vec{\alpha}_i)$, is obtained. These scan point test statistics then follow the relationship,

$$\Delta\chi^2 = \chi^2_{min}(\vec{\alpha}_i) - \chi^2_{min}(\vec{\alpha}_{global}) \geq 0, \tag{9.3.4}$$

which in itself can serve as a test statistic distributed as a χ^2 function to evaluate an upper limit by using $\Delta\chi^2$ in place of Q_0 in Equation 9.3.2. The upper limit can be determined by the value of $\mathcal{B}_{rare}$ where this p-value corresponds to a 3σ confidence level. This is the standard result obtained with the Prob method, which just uses the value from the combined signal and background model. An equivalent background-only relation is given by

$$\Delta\chi^2_{bkg} = \chi^2_{min}(\vec{\alpha}_{bkg}) - \chi^2_{min}(\vec{\alpha}_{global}) \geq 0, \tag{9.3.5}$$

where $\chi^2_{min}(\vec{\alpha}_{bkg})$ is obtained by performing a fit to the dataset with the value of $\mathcal{B}_{rare}$ set to zero. This is then used to obtain p_{bkg} through Equation 9.3.2 in the same way as the previous case. From which the $\mathrm{CL_s}$ p-value can be calculated from Equation 9.3.3, for each of the scan points. In the same way as the standard Prob method, this can then be used to determine an upper limit of $\mathcal{B}_{rare}$ by requiring the p-value to exceed a threshold confidence level, *i.e.* 3σ.

The Pseudo-experiment Based Frequentist (Plugin) Method: The previous method includes one major assumption, that the $\Delta\chi^2$ test statistic that is constructed is also distributed as a χ^2 function. This method, known as the Plugin method, builds upon a technique given in Ref. [171], which GammaCombo expands to include so-called 'nuisance' parameters, to account for the unknown distribution of $\Delta\chi^2$. These are parameters that are not the parameter of interest but are required in order to perform a hypothesis test. The Plugin method uses a collection of pseudo-experiments, which are random simulated events based on the original dataset and parameters, with variation arising from the uncertainties of these parameters. In the same way as the Prob method, the value of $\mathcal{B}_{rare}$ is varied within a particular set range and the value of $\Delta\chi^2$ is determined using the global fit value from the Prob method's scan. At each scan point, a number of pseudo-experiments can be generated from the fit model parameters at that point, including the particular fixed value for $\mathcal{B}_{rare}$. For each pseudo-experiment, a global

fit is performed where $\mathcal{B}_{rare}$ is allowed to vary to obtain $\chi^2_{min}(\vec{\alpha'}_{global})$, where $\vec{\alpha'}$ refers to the vector of parameters for the pseudo-experiment fit. Then a background-only fit where $\mathcal{B}_{rare}$ is zero, and a fit with the full combined signal and background model with $\mathcal{B}_{rare}$ set to the scan point value are both performed. These are then used to determine $\chi^{2\prime}_{min}(\vec{\alpha'}_{bkg})$ and $\chi^{2\prime}_{min}(\vec{\alpha'}_{i})$ respectively, from which the $\Delta\chi^{2\prime}$ values can be determined from Equations 9.3.5 and 9.3.4. During this procedure, the nuisance parameters are set to the best fit values for each scan point independently. Finally, the confidence levels themselves are determined by calculating the fraction of the generated pseudo-experiments that performed worse than the measured data, $\Delta\chi^2_{(s+b,\,b)} < \Delta\chi^{2\prime}_{(s+b,\,b)}$. This can be understood as the scan point where the pseudo-experiments perform most similarly to the data, as the parameters used to generate that pseudo-experiment are also likely to accurately represent the data. Therefore, above a given statistical threshold of confidence, the value of the rare mode branching fraction should align with the value of that particular scan point, or in the absence of a signal this would set the upper limit. The associated p-value in this case is given by,

$$p_{plugin,(s+b,\,b)} = 1 - \mathrm{CL} = \frac{N(\Delta\chi^2_{(s+b,\,b)} < \Delta\chi^{2\prime}_{(s+b,\,b)})}{N_{pseudo}}, \tag{9.3.6}$$

where $N(\Delta\chi^2_{(s+b,\,b)} < \Delta\chi^{2\prime}_{(s+b,\,b)})$ is the number of pseudo-experiments that perform worse than the data in the fits, and N_{pseudo} is the total number of pseudo-experiments generated. In this way, this method does not require specific knowledge or assumptions about the particular distribution of $\Delta\chi^2$ as these pseudo-experiments should, by construction, follow this distribution.

This defines the standard Plugin method and, as with the Prob method, it can be expanded and used with the $\mathrm{CL_s}$ method. In this case, the p-values in Equation 9.3.3 are determined from Equation 9.3.6.

9.4 Results

The single-candidate sensitivity, α, corresponds to the value of the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ branching fraction for a single observed signal candidate. This, therefore, represents an order-of-magnitude estimate of the upper limit in the absence of a signal. It shows the value of the combination of the constrained parameters in the data fits that multiplies the signal yield, namely ε_{rare}, ε_{norm}, $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+))$ and N_{norm}, for the different analysis categories. As such it can show if there are any unexpected relationships that might bias

the fit. The combined single-candidate sensitivity is determined using,

$$\alpha_{tot} = \left(\sum_{cat} \frac{1}{\alpha_{cat}} \right)^{-1}, \tag{9.4.1}$$

where α_{cat} is the single-candidate sensitivity for each of the analysis categories. Table 9.1 lists these values. These show that Run 2 appears to be generally more sensitive than Run 1 and that within both run periods the sensitivities are mostly compatible at the $< 2.9\sigma$ level for the two track type and bremsstrahlung categories. This does not imply that there is a problem with the input parameters, and instead is likely to arise from the different resolutions in the analysis categories. This serves to highlight the reasoning to perform the search in these independent analysis categories. The total combined single-candidate sensitivity is better than the model-independent upper limit given in Ref. [56] and discussed in Section 2.2.2, $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}) < 1.1 \times 10^{-8}$. Therefore, the final result after unblinding could allow for more stringent constraints to be applied to BSM physics models, in the absence of any observed signal.

Table 9.1: Single-candidate Sensitivities determined for each analysis category independently, and the combined case.

Category	**Single-candidate Sensitivity**
No Brem-DD-Run 1	$(5.34 \pm 0.56) \times 10^{-8}$
No Brem-LL-Run 1	$(5.93 \pm 0.66) \times 10^{-8}$
Brem-DD-Run 1	$(6.94 \pm 0.74) \times 10^{-8}$
Brem-LL-Run 1	$(8.70 \pm 1.00) \times 10^{-8}$
No Brem-DD-Run 2	$(1.29 \pm 0.12) \times 10^{-8}$
No Brem-LL-Run 2	$(1.94 \pm 0.19) \times 10^{-8}$
Brem-DD-Run 2	$(1.75 \pm 0.17) \times 10^{-8}$
Brem-LL-Run 2	$(1.89 \pm 0.18) \times 10^{-8}$
Total	$(3.32 \pm 0.14) \times 10^{-9}$

Currently, the data are blinded and so an estimated upper limit in the absence of a signal was determined. To do so the CL_s method outlined previously is performed with the BG proxy dataset as its input. This methodology allows for two upper limits to be determined, one based on the Prob method and the other based on the Plugin method. However, there are some known limitations for the Prob method results when evaluating the BG proxy dataset. This is discussed in further detail in Appendix D alongside the results for the Prob method. Therefore, only the results for the Plugin method are quoted here as these are unaffected by these specific limitations, as well as

requiring fewer assumptions to be made about the underlying distributions and giving appropriate statistical coverage by construction.

For the Plugin method a total of 2000 pseudo-experiments were generated and used to determine the upper limit. Figure 9.4 shows the distribution of the CL_s value for the Plugin scan, with the result from the proxy dataset, given as 'Observed' in the plot. This mostly lies within 1σ of the median CL_s result of the 2000 pseudo-experiments, which is labelled as 'Expected'. Table 9.2 shows the values obtained for the estimated upper limits in each case.

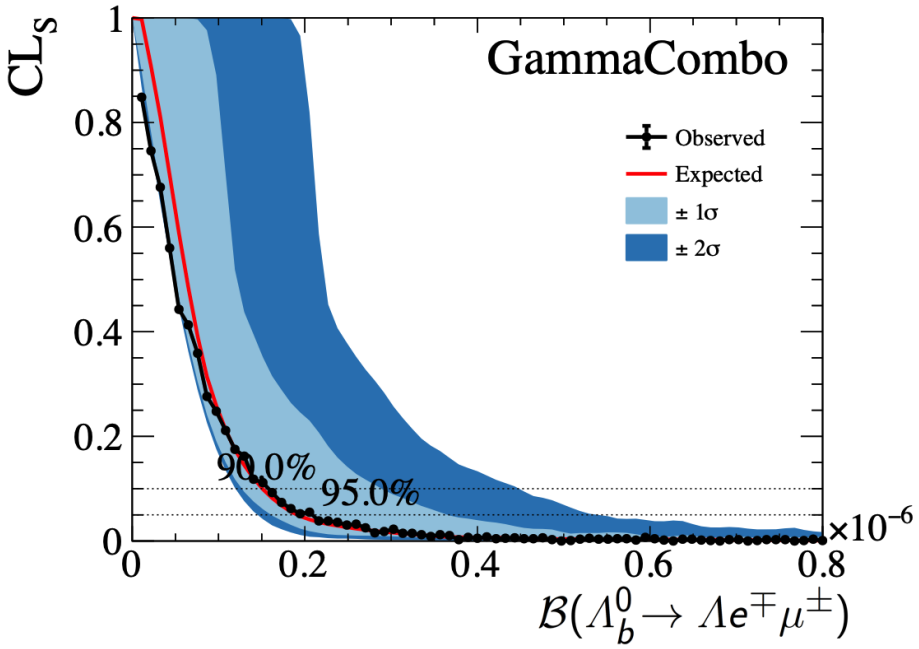


Figure 9.4: The CL_s values determined from 2000 pseudo-experiment runs for the BG proxy dataset for the parameter of interest, $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$. The 90% and 95% confidence level thresholds are overlaid.

As outlined in Chapter 6, one of the largest potential sources of uncertainty to the rare mode branching fraction value originally was the branching fraction of the normalisation mode. The current known value has a relative uncertainty of ∼16%, and was found to contribute a relative uncertainty of ∼54% to the rare mode branching fraction upper limit in the fit. However, an updated (preliminary) measurement reduces the relative uncertainty to ∼9% [140] for both the normalisation mode and its contribution to the rare mode branching fraction upper limit. This ensured the normalisation branching fraction was no longer the leading uncertainty, as seen in Section 9.5. The final results can also be quoted independently of the measured branching fraction of $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+))$ by expressing them in terms of the ratio, r, given in Equation 6.0.2. The CL_s procedure can then be repeated but with this r ratio treated as the parameter of interest instead of

$\mathcal{B}_{rare}$, calculated similarly through Equation 6.0.3.

Figure 9.5 shows the CL_s value of the Plugin scan, which is also performed using 2000 pseudo-experiments where r is the parameter of interest. Table 9.2 also shows the estimated upper limits for this method. The results can then be compared to the upper limit obtained using the full signal branching fraction expression by multiplying by the value of $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$. This is given in Table 9.2 in the row labelled $r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$. This serves as a cross-check of the upper limit derivation as this product should be compatible with the value $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$, as only a scale factor has been removed. This is observed for both the observed and expected results, where the compatibility is 1.06σ and 2.05σ respectively. The CL_s plots themselves do differ slightly in shape, where the scan over r has a less smooth error band at higher values but the size of the error band relative to the expected values are similar.

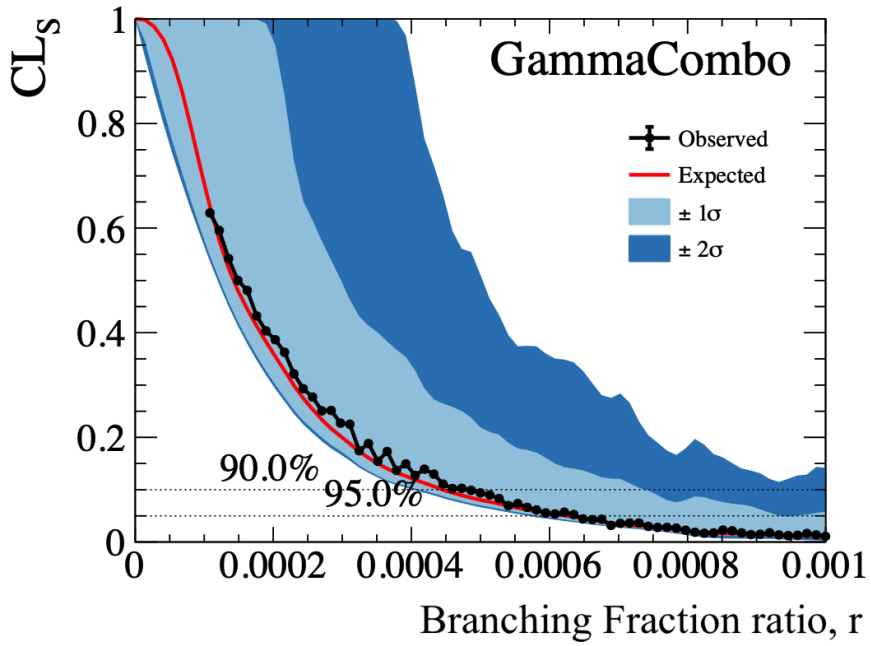


Figure 9.5: The CL_s values determined from 2000 pseudo-experiment runs for the BG proxy dataset for the parameter of interest, r. The 90% and 95% confidence level thresholds are overlaid.

Overall, by comparing the estimated limits on $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ to the single-candidate sensitivity the corresponding signal yields are determined as 63 ± 3 and 55 ± 2 for the observed and expected Plugin results respectively. This would correspond to 14–21 signal candidates in the No Brem-DD-Run 2 category, with $\mathcal{O}(3)$ candidates in Run 1 categories and $\mathcal{O}(10)$ candidates in Run 2 categories. As these numbers are derived from the BG case, any potential observation of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ would require these signal yields to be exceeded significantly. Assuming a branching fraction for the rare mode at the

Table 9.2: The estimated upper limits at a 95% confidence level obtained for different parameters of interest using GammaCombo.

Parameter of Interest	**Pseudo-experiment CL_s Method**	
(@ 95% CL)	Observed	Expected
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$	$\lesssim 2.09 \times 10^{-7}$	$\lesssim 1.83 \times 10^{-7}$
r	$\lesssim 6.39 \times 10^{-4}$	$\lesssim 6.23 \times 10^{-4}$
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	$\lesssim (2.31 \pm 0.22) \times 10^{-7}$	$\lesssim (2.26 \pm 0.21) \times 10^{-7}$

same value as that predicted for the model-independent case of Ref. [56], around three to four candidates would be expected. Therefore, this estimated limit is directly examining whether their assumptions are valid experimentally. However, these limits would not be able to provide stronger constraints on the existing scalar leptoquark models outlined in Section 2.2.2. This new measurement would still represent both LHCb's first and the world's most precise experimental upper limit placed on the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ branching fraction. This will generally aid theorists in their model development by providing additional constraints in the Λ_b^0 baryon sector. It should also be noted that this is an estimated limit, based on the scenario that no signal is observed, and that future developments during the unblinding process may lead to a different final observed upper limit or even a potential first observation of a cLFV decay.

9.5 Determination of Systematic Uncertainties

There are uncertainties associated with the procedure used to carry out an experimental analysis. The systematic uncertainties on the obtained estimated upper limits in Table 9.2 arise from a variety of sources contributing in varying degrees to the total. This section discusses these sources of systematic uncertainties, their size and the attempts to control their effect where possible. At the time of writing, final quantitative estimates for some of these sources are still in preparation.

Generally, the effect of the total systematic uncertainty is propagated to the final result as the Gaussian-constrained variables in the fits to the dataset to determine the upper limit, which include this systematic uncertainty when defining the width of that Gaussian distribution. Figures 8.8, 9.4 and 9.5 are derived with its inclusion. Three main types of systematics are evaluated, depending on the parameter they contribute to: the systematic on the efficiency ratio, $\varepsilon_{norm}/\varepsilon_{rare}$; systematics related to the shape of the fit model considered; and systematics related to the branching fractions of the normalisation

mode, and the Λ_c^+ semileptonic modes used as exclusive backgrounds. Overall, the total systematic uncertainty is found to be: 3–7% depending on the analysis category for the efficiency ratio[1]; negligible for the signal shape; and up to 25% for the branching fractions. The uncertainty on the efficiency ratio is dominated by the finite size of the MC samples used for its determination, as well as the data/MC difference for the GBDT selection efficiency. Tables 9.3 and 9.4 summarise the individual contributions to the total systematic uncertainties for each of the analysis categories.

Table 9.3: The systematic and total uncertainties that contribute to the determination of the upper limit of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ for the No Brem categories. For the efficiency ratio, $\frac{\varepsilon_{sig}}{\varepsilon_{BG}}$, the fit shapes, or branching fractions included in the fit model.

Systematic Source		**Category (No Brem)**			
		LL-Run 1	LL-Run 2	DD-Run 1	DD-Run 2
efficiency ratio, $\frac{\varepsilon_{sig}}{\varepsilon_{BG}}$	GBDT	2.25%	0.35%	3.01%	1.44%
	MC statistics	5.66%	2.63%	4.30%	2.18%
	Kinematic corrections	0.02%	0.66%	0.06%	0.02%
	Angular weight	1.34%	0.80%	1.34%	0.80%
	PID weight	0.32%	0.22%	0.31%	0.21%
	PID binning		$\mathcal{O}(1)$% estimated		
	PID data/MC		$\mathcal{O}(1)$% estimated		
	Tracking weight	0.21%	0.25%	0.10%	0.13%
	Trigger weight		negligible		
	Trigger binning		$\mathcal{O}(1)$% estimated		
	Trigger data/MC	0.45%	0.23%	0.50%	0.08%
	Total	6.50%	3.36%	5.72%	3.25%
Signal shapes			negligible		
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e)$ (total)			21.3%		
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu)$ (total)			25.9%		
$\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+))$ (total)			9.31%		

[1]This includes several $\mathcal{O}(1)$% estimations as summarised in Tables 9.3 and 9.4.

Table 9.4: The systematic and total uncertainties that contribute to the determination of the upper limit of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ for the Brem Categories. For the efficiency ratio, $\frac{\varepsilon_{sig}}{\varepsilon_{BG}}$, the fit shapes, or other branching fractions included in the fit model.

Uncertainty Source		**Category (Brem)**			
		LL-Run 1	LL-Run 2	DD-Run 1	DD-Run 2
efficiency ratio, $\frac{\varepsilon_{sig}}{\varepsilon_{BG}}$	GBDT	1.44%	0.53%	3.22%	1.64%
	MC statistics	6.42%	2.58%	4.54%	2.27%
	Kinematic corrections	0.00%	0.87%	0.91%	0.15%
	Angular weight	1.35%	0.80%	1.34%	0.80%
	PID weight	0.26%	0.15%	0.22%	0.14%
	PID binning		$\mathcal{O}(1)$% estimated		
	PID data/MC		$\mathcal{O}(1)$% expected		
	Tracking weight	0.20%	0.22%	0.15%	0.07%
	Trigger weight		negligible		
	Trigger binning		$\mathcal{O}(1)$% estimated		
	Trigger data/MC	0.94%	0.06%	1.19%	0.15%
	Total	7.01%	3.38%	6.17%	3.40%
Signal shapes			negligible		
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e)$ (total)			21.3%		
$\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu)$ (total)			25.9%		
$\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+))$ (total)			9.31%		

9.5.1 Potential Sources of Systematic Uncertainty

GBDT: As discussed in Section 7.4.2, there is a systematic uncertainty related to the potential mis-modelling of the GBDT. To evaluate the potential impact this would have the classifier response was compared between sWeighted data and simulation of the normalisation mode because of the lack of a direct proxy dataset for the rare mode due to its LFV nature. The efficiency from the GBDT cut was considered on both the re-weighted $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ MC and associated sWeighted data. The relative difference between them (Equation 7.4.2) was then used to quantify the systematic uncertainty of the potential mis-modelling of the GBDT. This is found to be between 0.4% and 3.2% depending on the analysis category.

Additionally, the mass shape in data has been shown not to be strongly affected by the GBDT selection, due to the lack of correlation between the mass variable and the GBDT output, and the misidentified background from $B^0 \to K_S^0 J/\psi$ can still be clearly seen and modelled within the normalisation mode fits.

MC sample size: As many aspects of the analysis rely on MC samples, *e.g.* signal fit shapes and efficiencies, the statistical size of the MC itself can impact the analysis in a variety of ways. The statistical uncertainty on the MC after the full selection for both $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ is then included in the systematic uncertainty for the associated efficiency. This ranges from 2.2% to 6.4% based on the sample and the analysis category.

MC Re-weighting: The statistical uncertainty for each of the weights discussed in Section 7.3 was propagated to the efficiency ratio through the use of a pseudo-experiment method. An ensemble of 100 sets of weights were generated, where variations arise due to the uncertainties on the individual weights, which was then used to evaluate the efficiency ratio for each pseudo-experiment. The overall systematic uncertainty relating to the MC re-weighting was then estimated to be ∼2%. This is dominated by the $\Lambda_b^0 \to \Lambda J/\psi$ angular weights which contribute approximately ∼0.8% to the MC re-weighting systematic uncertainty, as well as the tracking from the leptons (contributing ∼0.1% for DD and ∼0.3% for LL) and the lepton PID weights (contributing ∼0.3%). It was found that the trigger weights only contribute an additional ∼0.02% to the total weight systematic uncertainty which arises due to the sample size of the MC. Therefore, the trigger weight systematic was considered to be negligible. The exact contributions for the MC re-weighting in different categories are summarised in Tables 9.3 and 9.4 as the rows labelled as 'weight'.

Additionally, a systematic uncertainty based on the improvement that the weights

would provide relative to an unweighted sample is evaluated directly by examining the effect this has on the efficiency ratio. These are given for the PID and trigger and labelled as 'data/MC' systematic uncertainties and provided some temporary conservative estimates for the potential effect of binning before exact numbers were determined.

PID: The PID dependence on event multiplicity was examined by evaluating the electron PID efficiency maps in three equally-populated bins of the `nTracks` variable, which gives the total number of tracks associated with a particle and hence the track multiplicity. These new PID efficiency maps were applied to the $\Lambda_b^0 \to \Lambda e^- e^+$ MC sample, where the presence of two electrons would give an enhanced effect from the different maps. The PID efficiencies from the low and high multiplicity bins were compatible to within 2σ, with similar results found if an alternative multiplicity variable (`nSPDHits`) was used instead. Therefore, no systematic uncertainty was assigned based on PID event multiplicity.

Another aspect of the PID method to be investigated is the choice of binning for the efficiency maps that have been used to develop the PID weights and the optimised PID cut. For the electrons, the efficiency was computed in the case where a much finer binning scheme was employed but no systematic effect was observed. However, for the muons the systematic effect has currently not been fully evaluated, but an estimated effect of at most 1% is expected.

Trigger: Similarly to the PID, the trigger efficiencies were obtained with maps that have a specified binning scheme. Therefore, a systematic uncertainty can be assigned based on the difference in the efficiency as determined from the nominal binning scheme compared to an alternative one. This has not yet been fully evaluated, but an estimated effect of at most 1% is also expected.

The trigger selection was determined after the application of the PID selection, but for MC samples that was done via weights. Therefore, there is the potential for the PID selection to affect the HLT2 trigger despite nominally being decoupled when directly applying the optimised PID cut. This is because the PID weights are used instead and as such the underlying dependence could remain, but is not expected to give a significant effect ($\lesssim 1\%$). The trigger 'data/MC' systematic uncertainty value should also provide coverage of this effect until a detailed cross-check and determination of this value is performed in the future.

Tracking: The majority of the potential systematic uncertainty contributions for the tracking efficiencies were handled through the tracking weights as discussed when consid-

ering the MC re-weighting systematic uncertainties. However, DD tracks currently have no determined tracking efficiency maps so these could not be evaluated in the same way. One potential method to investigate potential systematic differences associated with the DD tracks, is to compare the normalisation yields in DD and LL categories, given in Table 8.4. There is a known difference within Run 1 that was corrected for [137] and, after including this correction, all categories give similar results. Therefore, no systematic uncertainty was assigned related to the Λ tracks.

Exclusive Background: From the exclusive background PDF given in Equation 8.3.1, it can be shown that the branching fractions for the individual exclusive background modes are included, which themselves have uncertainties of ∼20–26%. This uncertainty is included to determine and constrain the yields of the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ backgrounds in the fit model. The systematic uncertainties related to the determination of the total efficiency ratio of the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ mode and the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ backgrounds are assumed to be negligible as these are derived in almost identical manners, differing only in the MC samples used. The statistical uncertainties have been included for each individual efficiency to account for the different sample sizes. Additionally, cross-checks of the data sidebands have not found any additional backgrounds, so no systematic uncertainties related to the exclusive background fit models were assigned.

$\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ signal shape: Although the DSCB function was selected as the preferred PDF to describe the signal peak of the $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ mode, there are alternative fit models that could have been used. These different fit models may result in different yields for the normalisation mode which could ultimately affect the determination of the upper limit to $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$. Therefore, a two-sided Hypatia [172] function was used as the signal PDF, with the resulting yields for each fit category shown in Table 9.5 alongside the nominal yields. The difference between the nominal yields and those from the Hypatia fit differ by at most 0.2% and the yields in all categories are compatible with both fit models. Therefore, no systematic uncertainty was assigned for the $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ signal shape

Table 9.5: Normalisation mode yields obtained from fits with a double-sided crystal ball, as in the nominal case, and with a Hypatia function [172].

		Run 1		**Run 2**	
		DSCB	Hypatia	DSCB	Hypatia
Raw Yield	LL	1762 ± 44	1760 ± 44	8719 ± 98	8696 ± 98
	DD	2353 ± 50	2353 ± 50	9203 ± 99	9193 ± 99
	Total	4116 ± 67	4113 ± 67	17922 ± 139	17889 ± 139

$\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ signal shape: Due to the fact that no decays with cLFV have been found, there is no preferred signal shape to use and so the actual model for the signal is not varied, in contrast with the normalisation mode. Instead, the sensitivity to variations in the resolution of that signal model was evaluated using the $\mathcal{C}$ factor discussed in Chapter 8.3 and Appendix C. The $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ signal resolution was varied by up to 40% before performing the CL_s profile likelihood scan, and no significant deviation from the nominal value for the upper limit was observed. Therefore, no systematic uncertainty was assigned for the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ signal shape.

9.5.2 Effect of the Total Systematic Uncertainty on the Results

Systematic uncertainties on parameters such as selection efficiencies, signal and background shapes and fractions of each category are propagated to the CL_s limit by allowing each parameter to fluctuate in the pseudo-experiments and in the fit PDFs. This is done by multiplying the likelihood function with a Gaussian distribution centred in the nominal value for each of the constrained parameters and having the quadratic sum of the parameter's systematic and statistical uncertainties as the standard deviation. If the CL_s procedure is repeated with each of these constrained values fixed to their central values then the effect of these constraints and hence the uncertainties can be approximately evaluated. This is not entirely representative of the reality of the fit model application as it allows no freedom in the central values of the efficiencies, the branching fractions for the different components, or the yields of the exclusive background components. This can provide an indication of how the uncertainties can generally affect the fit result.

Figures 9.6 and 9.7 show the resultant CL_s curves, and Table 9.6 gives the estimated upper limits from these plots. Overall these plots are similar to Figures 9.4 and 9.5, *i.e.* the plots with the uncertainties included, but with larger error bands. This should not strongly affect the CL_s values, which are comparable. However, there is a significant impact on the upper limits compared to those in Table 9.2, with Table 9.7 highlighting the total relative uncertainty associated with each upper limit by comparing the two results. Therefore, the upper limit itself could have a relatively large total uncertainty, $\mathcal{O}(20)\%$. This is potentially due to the more inflexible nature of the fit when the constrained values are instead fixed to their nominal results. This may also imply that the initial central values for these constrained parameters do not provide a reliable description of the data. In particular, the exclusive background PDFs have large uncertainties associated with them, such that their yields can vary dramatically and the nominal values from Tables 8.2 and 8.3 are known to be overestimated. In the case where the yields are constrained and not fixed, they are able to describe the data well.

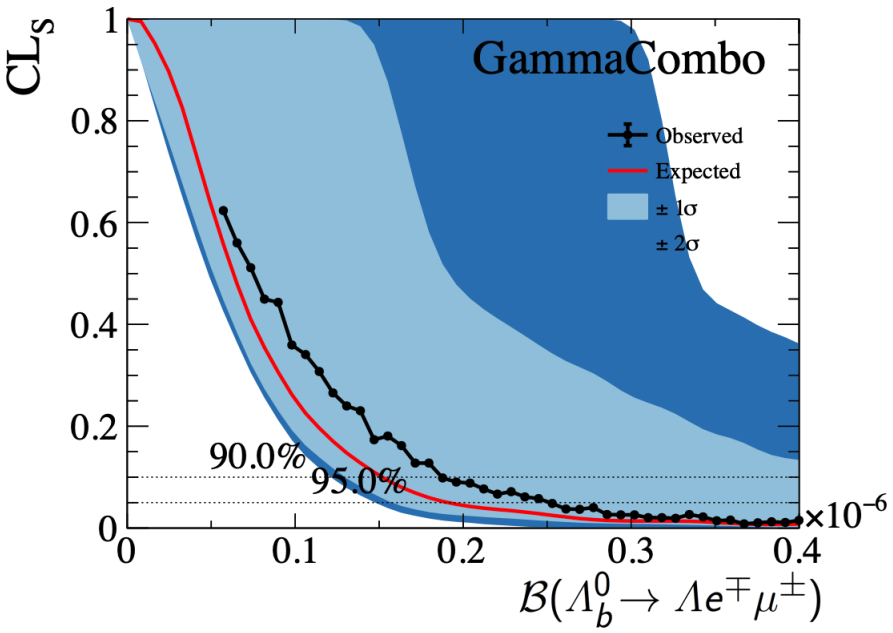


Figure 9.6: The CL_s values determined from 2000 pseudo-experiment runs for the background-only proxy dataset for the parameter of interest, $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ with constrained variables fixed to their nominal value. The 90% and 95% confidence level thresholds are overlaid.

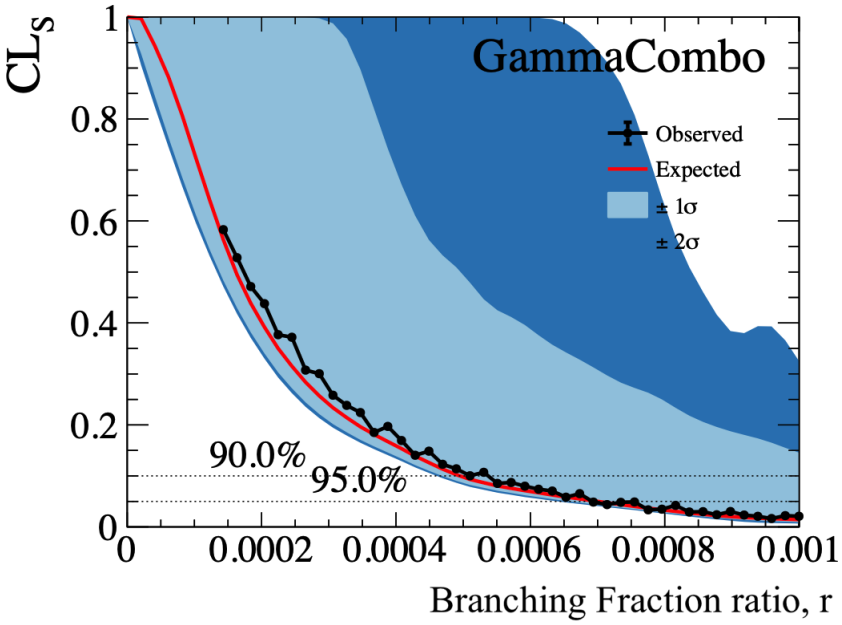


Figure 9.7: The CL_s values determined from 2000 pseudo-experiment runs for the background-only proxy dataset for the parameter of interest, r with constrained variables fixed to their nominal value. The 90% and 95% confidence level thresholds are overlaid.

Table 9.6: The estimated upper limits at a 95% confidence level obtained for different parameters of interest using GammaCombo with constrained variables in the fit model fixed to their nominal values.

Parameter of Interest	**Pseudo-experiment CL_s Method**	
(@ 95% CL)	Observed	Expected
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$	$\lesssim 2.52 \times 10^{-7}$	$\lesssim 1.84 \times 10^{-7}$
r	$\lesssim 6.9 \times 10^{-4}$	$\lesssim 7.0 \times 10^{-4}$
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	$(2.50 \pm 0.23) \times 10^{-7}$	$(2.54 \pm 0.24) \times 10^{-7}$

Table 9.7: The uncertainties derived for the estimated upper limits at a 95% confidence level obtained for different parameters of interest using GammaCombo.

Parameter of Interest	**Pseudo-experiment CL_s Method**	
(@ 95% CL)	Observed	Expected
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$	20.6%	0.5%
r	20.2%	18.1%
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	60.4%	65.9%

9.6 Evaluating the Limit with an Additional Λ_c^+ Veto

Having determined an estimated upper limit for the branching fraction of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ through the method detailed up to this point, an additional selection was considered to evaluate its impact on the sensitivity of the search. This was the inclusion of a Λ_c^+ veto, because the background model for the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ fit includes exclusive backgrounds from two specific double semileptonic backgrounds involving a Λ_c^+ baryon. Although Tables 8.2 and 8.3 suggest that these two are the only significant backgrounds, the other Λ_c^+ backgrounds may be more significant than expected as the sample sizes for the MC candidates in these channels are much smaller than those for the exclusive backgrounds considered. Additionally, the developed fit model may not be able to fully control the exclusive backgrounds once unblinding has been performed despite fitting well the blinded sample where the majority of the Λ_c^+ backgrounds are expected to contribute. A simple solution to eliminate any possible effects from these Λ_c^+ backgrounds, either those already accounted for or any additional contributions, is the inclusion of a Λ_c^+ veto, given as,

$$m(\Lambda\ell) > 2300\,\text{MeV}/c^2, \tag{9.6.1}$$

where $\Lambda\ell$ is the combination of the Λ candidate and an electron or muon in the final state.

This section outlines an initial study of the effect of an additional Λ_c^+ veto that was applied after the full selection detailed in Chapter 7. The $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ blinded dataset was significantly reduced in size by this veto, which introduces much greater uncertainty when performing the fits. The idealised implementation of this Λ_c^+ veto would be as part of the pre-selection, before training the GBDT. Then the optimisation of the GBDT selection would account for the reduction of Λ_c^+ backgrounds. Thus, the blinded dataset may overall have a looser selection than applying the Λ_c^+ veto after training, with the larger sample size reducing the statistical uncertainty in the fit. An initial study is reported here, but the full treatment of this analysis with and without the inclusion of the Λ_c^+ veto from the start will be performed in the future.

The veto could be applied to both the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ modes to reduce the systematic uncertainty relating to this selection. This preliminary investigation highlights that this is possible because this veto does not appear to distort the underlying mass distributions for either case. However, in the initial investigation, the Λ_c^+ veto was only applied to the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ data and used the same $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ yields from Table 8.4 as the normalisation data has no evidence of Λ_c^+ backgrounds present. As such the inclusion of the Λ_c^+ veto for $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ is not necessary other than to harmonise the selections of both the rare and normalisation modes as closely as possible to reduce the effect of systematic uncertainties.

The known value of the Λ_c^+ baryon mass is $2286.46 \pm 0.14\,\text{MeV}/c^2$ [22]. This informed the selection of $2300\,\text{MeV}/c^2$ for the veto's lower limit as this would contain almost all potential Λ_c^+ candidates. This was also verified by examining the veto relative to the mass peak in $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ MC samples, shown for the Brem-DD-Run 2 category in Figure 9.8, see Appendix E for the other categories. These figures show that these Λ_c^+ backgrounds are almost entirely eliminated.

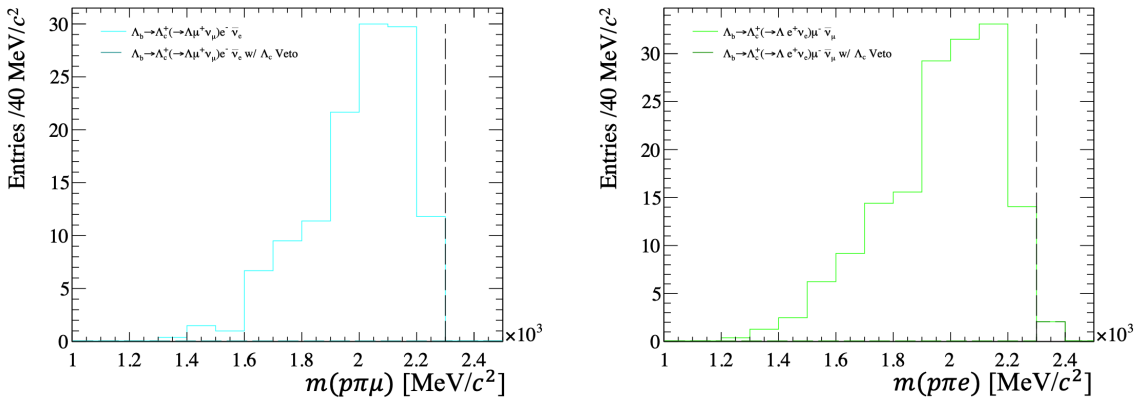


Figure 9.8: Distribution of $m(p\pi\mu)$ (Left) and $m(p\pi e)$ (Right) for the category Brem-DD-Run 2 with and without the Λ_c^+ veto applied for the different MC samples : (Left) $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ (Right) $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ relative to the Λ_c^+ veto.

The effect of the Λ_c^+ veto on the Λ_b^0 mass distribution for the OS data sidebands and SS data for the 'Brem-DD-Run 2' category is given in Figure 9.9, see Appendix E for the other categories. The veto does not introduce significant bias, appearing similar for both the opposite-sign and same-sign datasets as well as the MC samples. The lower data sideband was reduced as expected due to the removal of the Λ_c^+ backgrounds. The effect on the selected signal MC samples for the $\Lambda_b^0 \to \Lambda e^\mp\mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ fits, as well as the misidentification sample MC for the normalisation mode, $B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$, with and without the Λ_c^+ veto included was also examined, as were additional mass parameters *e.g.* $m(\Lambda\ell)$. These similarly do not show any significant evidence for shaping the mass variable. Therefore, no additional adjustment of the normalisation fit mode with the inclusion of the Λ_c^+ veto is required.

The efficiency of the Λ_c^+ veto can then be determined by comparing the number of weighted candidates in MC samples before and after the veto has been applied, following the same methodology outlined in Section 7.6. This was done for $\Lambda_b^0 \to \Lambda e^\mp\mu^\pm$, $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$, $B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$ and $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ samples to show the effect of the veto on all the signal and background decays that contribute to both the rare and normalisation mode fits. This is shown in Tables 9.8 and 9.9. Overall, for $\Lambda_b^0 \to \Lambda e^\mp\mu^\pm$ and $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ the remaining candidates are reduced by $\sim$30%, $B^0 \to K_S^0 J/\psi(\to \mu^-\mu^+)$ by $\sim$20%, and the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ modes are

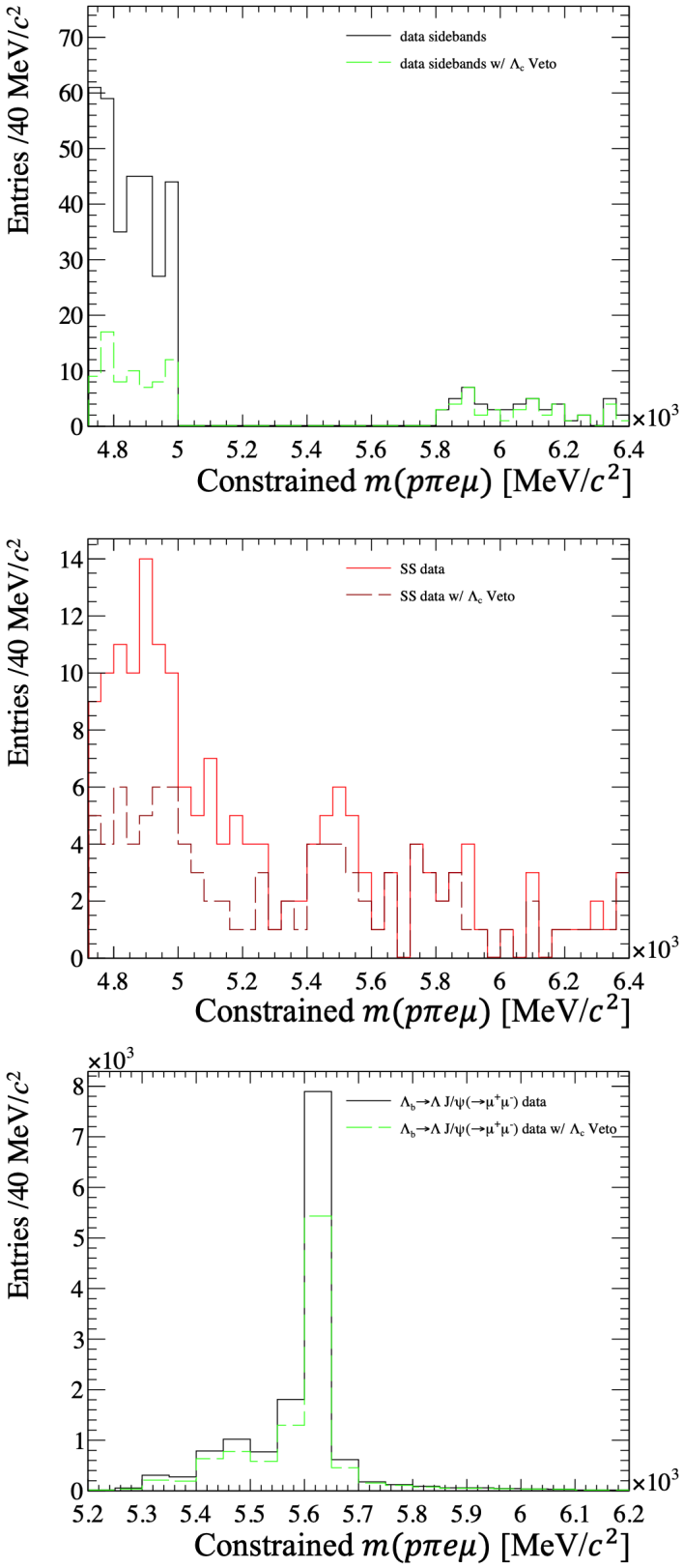


Figure 9.9: Distribution of the constrained Λ_b^0 mass for the category Brem-DD-Run 2 with and without the Λ_c^+ veto applied for the data sidebands (Top-Left), the same-sign dataset (Top-Right), and the dataset used for the normalisation mode fits (Bottom).

almost entirely eliminated. The total efficiencies for the rare and normalisation modes in this case are then taken as the product of the total efficiencies from Tables 7.10 to 7.12 and the associated Λ_c^+ veto efficiencies from Tables 9.8 and 9.9.

Table 9.8: Efficiencies for the Λ_c^+ veto on different MC samples relevant for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ without bremsstrahlung.

Efficiency Categories		**No Brem Weighted Efficiencies (%)**			
		LL-Run 1	DD-Run 1	LL-Run 2	DD-Run 2
Rare	$\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$	67 ± 3	67 ± 3	67 ± 1	67 ± 2
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda \mu^+ \nu_\mu) e^- \overline{\nu}_e$	0	0	0	0
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda e^+ \nu_e) \mu^- \overline{\nu}_\mu$	0	0	0	0

Table 9.9: Efficiencies for the Λ_c^+ veto on different MC samples relevant for the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ with bremsstrahlung and the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ modes (that use the BDT selection with bremsstrahlung).

Efficiency Categories		**Brem Weighted Efficiencies (%)**			
		LL-Run 1	DD-Run 1	LL-Run 2	DD-Run 2
Rare	$\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$	68 ± 3	71 ± 4	69 ± 1	71 ± 1
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda \mu^+ \nu_\mu) e^- \overline{\nu}_e$	1 ± 1	0	1 ± 1	6 ± 2
	$\Lambda_b^0 \to \Lambda_c^+ (\to \Lambda e^+ \nu_e) \mu^- \overline{\nu}_\mu$	0	0	0	0
Normalisation	$\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$	69 ± 2	71 ± 2	70 ± 1	72 ± 1
	$B^0 \to K_S^0 J/\psi (\to \mu^- \mu^+)$	81 ± 1	84 ± 2	82 ± 1	83 ± 1

The inclusion of the Λ_c^+ veto then modifies the PDF related to the background fit model for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ data, such that Equation 8.3.1 is simplified to

$$\text{PDF}_{BG} = N_{comb} \cdot \text{PDF}_{exp}, \tag{9.6.2}$$

i.e. only including the exponential PDF from the combinatorial contribution, which is constrained from fitting the SS dataset. The signal PDF remains unchanged from Equation 8.3.2, the signal parameters themselves are also fixed to the same values as those determined from the fits to $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ MC without the Λ_c^+ veto applied. This is because the mass shape of the $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ MC did not appear to be strongly shaped by the Λ_c^+ veto, so it was sufficient for this initial study to reuse the values from the previous case. In future, when the analysis is re-evaluated from the start to be performed with and without the Λ_c^+ veto, dedicated fits for each particular case should be used to define the signal parameters.

The blinded fits to the data sidebands for the combined signal and background, and the background-only PDFs are given in Figure 9.10. These fits illustrate the small size of

the remaining dataset, particularly in Run 1, which, overall, could bias the resultant fit shape. However, these blinded fits appear to be consistent with the data for most categories. Therefore, the estimated CL_s upper limit to the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ can be determined using the same methodology described in Section 9.4 but with this modified fit model.

The CL_s value of the Plugin scan plots for the fit model including the Λ_c^+ veto and scanning over $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ are given in Figure 9.11, with the equivalent r plots shown in Figure 9.12. These show a similar shape to the equivalent case including the exclusive backgrounds, although the CL_s scan has a much larger error band. Thus, although the nominal upper limit appears to be a factor $\sim$3 improvement over the exclusive background result, the associated uncertainty would be larger. This uncertainty is suggested to be due to the low sample size of the underlying datasets that can allow for a variation when generating the pseudo-experiments as the statistical uncertainties and fit parameter uncertainties are large. As such, this highlights both the potential advantages and disadvantages that come with the use of the Λ_c^+ veto. The fit is performed in a far cleaner environment with lower background contributions, which leads to a more stringent upper limit. Although these lead to a much lower sample size for the dataset, and hence greater statistical uncertainty in individual fits, this can be compensated for by utilising more pseudo-experiments.

The estimated upper limits for $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$, r and $r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ are given in Table 9.10. The $r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ values are compatible with the upper limit of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ at a level of 0.64σ, and 0.40σ for the observed and expected Plugin results respectively. The corresponding numbers of signal candidates can be determined using the single-candidate sensitivities in Table 9.1, for the observed and expected Plugin results, these were 20 ± 1 and 23 ± 1 respectively. This implies that a smaller number of signal candidates would have to be exceeded in order to claim observation of this cLFV mode, compared to the previous result considered with the exclusive backgrounds, showing the improved sensitivity from the inclusion of the Λ_c^+ veto.

Additionally, an approximate total relative uncertainty can be derived for the upper limits with the Λ_c^+ veto, similarly to Section 9.5.2. Figures 9.13 and 9.14 show the CL_s plots where the constrained variables are fixed to their central values. These are very similar to Figures 9.11 and 9.12. The associated upper limits are given in Table 9.11, and the approximate relative uncertainties between the constrained and fixed upper limits with the Λ_c^+ veto applied are given in Table 9.12. This implies that the total relative uncertainty on the upper limit is $\mathcal{O}(1-4)\%$, which is a significant improvement compared to the result without the Λ_c^+ veto included.

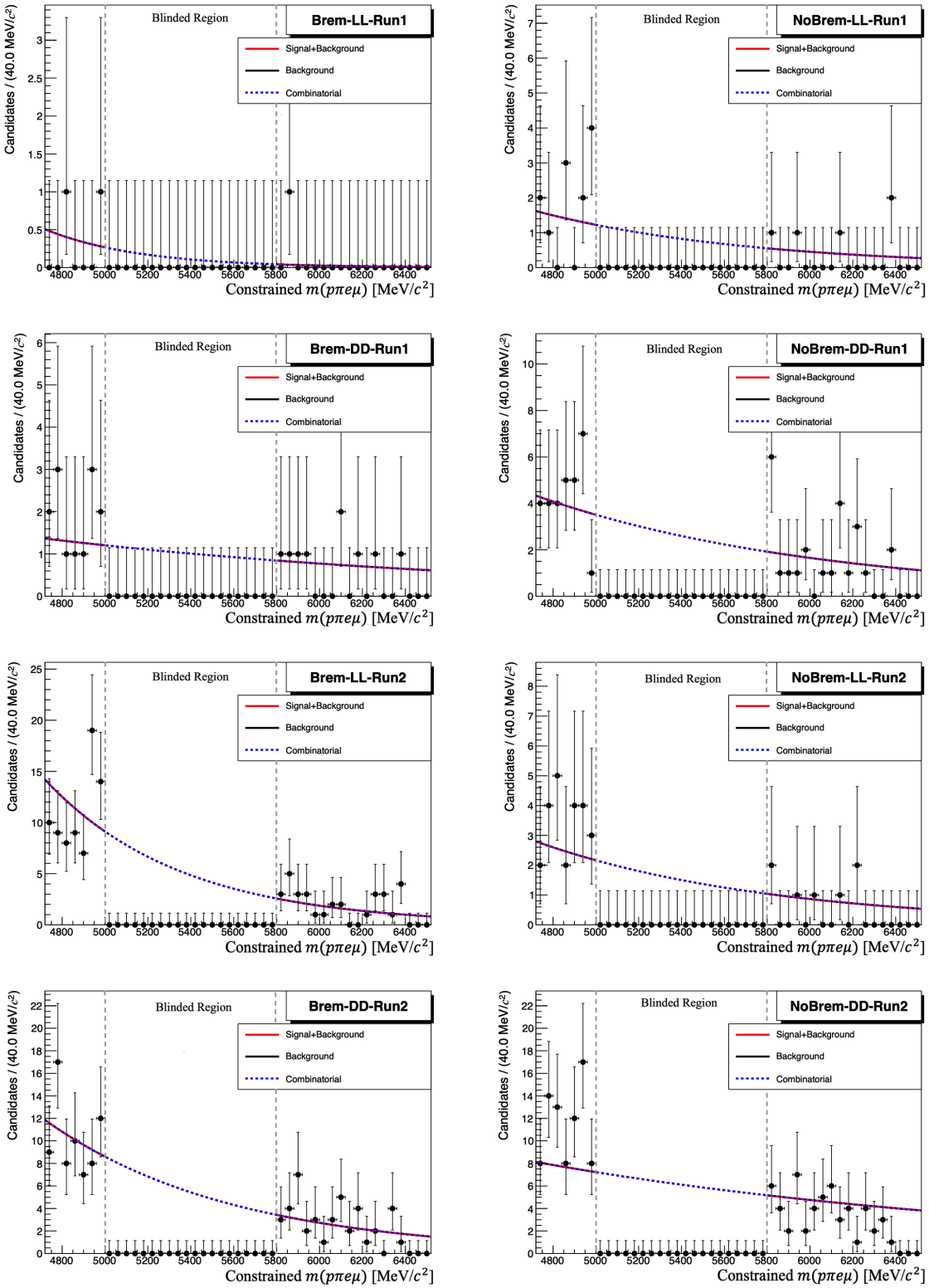


Figure 9.10: Fits of the invariant mass of the Λ_b^0 for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ data in each analysis category, as labelled on each plot, with the Λ_c^+ veto applied. Here the combined signal and background, and the background-only PDFs are identical due to the signal region being blinded.

Therefore, the conclusion of this initial study is that the Λ_c^+ veto should be considered more thoroughly from the start of the analysis and upper limits derived with and without its inclusion. It is also expected that the inclusion of the Λ_c^+ veto should serve as the default result due to the benefits it provides to both the upper limits' value and approximate uncertainty.

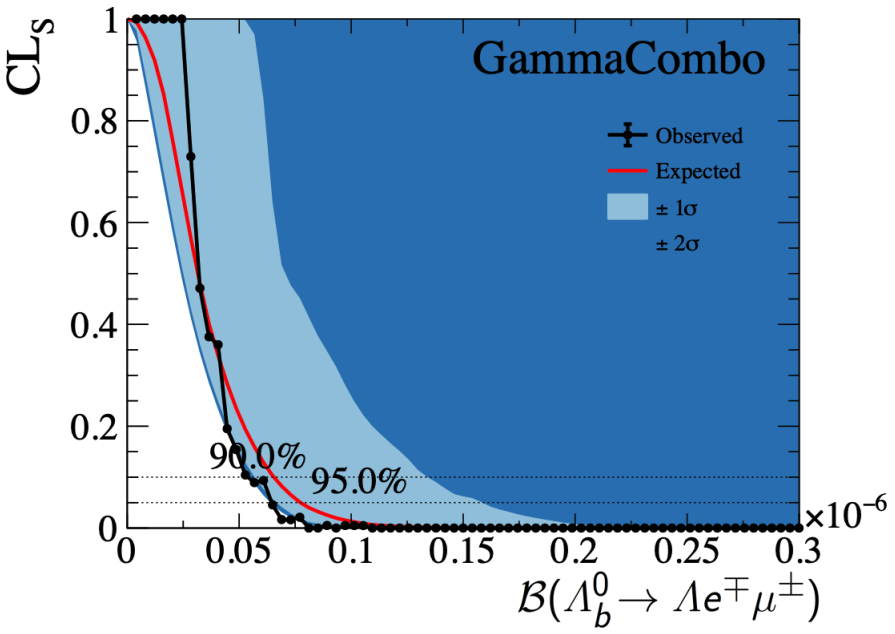


Figure 9.11: The CL_s values determined from 2000 pseudo-experiment runs for the background-only proxy dataset for the parameter of interest, $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ with the LcVeto included. The 90% and 95% confidence level thresholds are overlaid.

Table 9.10: The estimated upper limits at a 95% confidence level obtained for different parameters of interest using GammaCombo, when the Λ_c^+ veto is applied.

Parameter of Interest (@ 95% CL)	**Pseudo-experiment CL_s Method**	
	Observed	Expected
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$	$\lesssim 6.45 \times 10^{-8}$	$\lesssim 7.48 \times 10^{-8}$
r	$\lesssim 1.68 \times 10^{-4}$	$\lesssim 2.14 \times 10^{-4}$
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	$\lesssim (6.10 \pm 0.57) \times 10^{-8}$	$\lesssim (7.77 \pm 0.72) \times 10^{-8}$

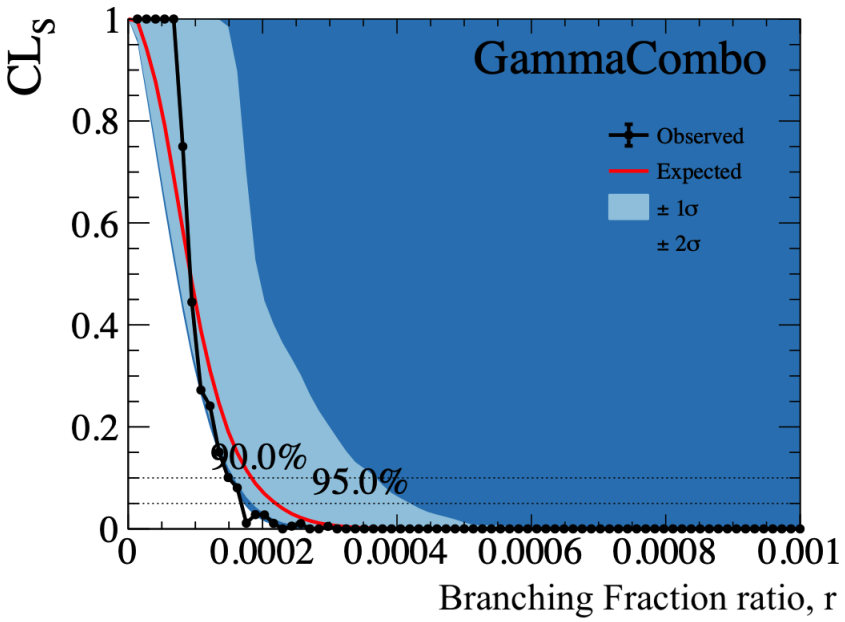


Figure 9.12: The CL_s values determined from 2000 pseudo-experiment runs for the background-only proxy dataset for the parameter of interest, r with the LcVeto included. The 90% and 95% confidence level thresholds are overlaid.

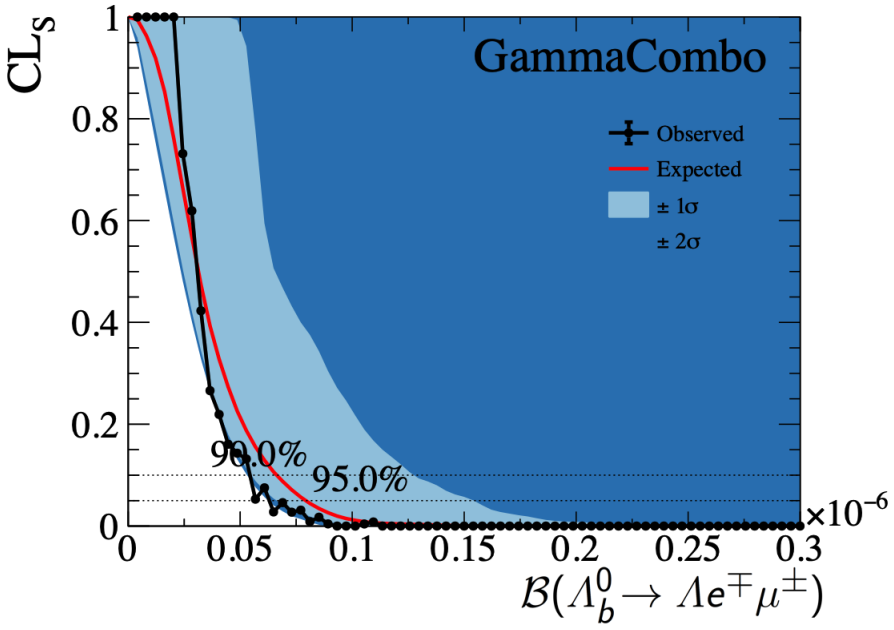


Figure 9.13: The CL_s values determined from 2000 pseudo-experiment runs for the background-only proxy dataset for the parameter of interest, $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ with constrained variables fixed to their nominal value and including the Λ_c^+ veto. The 90% and 95% confidence level thresholds are overlaid.

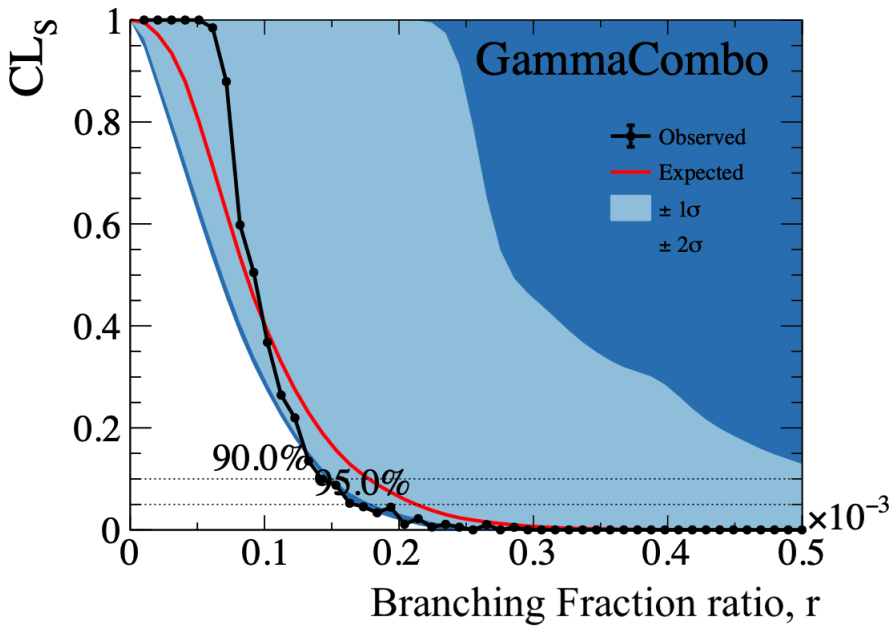


Figure 9.14: The CL_s values determined from 2000 pseudo-experiment runs for the background-only proxy dataset for the parameter of interest, r with constrained variables fixed to their nominal value and including the Λ_c^+ veto. The 90% and 95% confidence level thresholds are overlaid.

Table 9.11: The estimated upper limits at a 95% confidence level obtained for different parameters of interest using GammaCombo with constrained variables in the fit model fixed to their nominal values and the Λ_c^+ veto is applied.

Parameter of Interest	**Pseudo-experiment CL_s Method**	
(@ 95% CL)	Observed	Expected
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$	$\lesssim 6.30 \times 10^{-8}$	$\lesssim 7.60 \times 10^{-8}$
r	$\lesssim 1.67 \times 10^{-4}$	$\lesssim 2.06 \times 10^{-4}$
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	$\lesssim (6.06 \pm 0.56) \times 10^{-8}$	$\lesssim (7.48 \pm 0.69) \times 10^{-8}$

Table 9.12: The uncertainties derived for the estimated upper limits at a 95% confidence level obtained for different parameters of interest using GammaCombo when the Λ_c^+ veto is applied.

Parameter of Interest	**Pseudo-experiment CL_s Method**	
(@ 95% CL)	Observed	Expected
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$	2.2%	1.6%
r	0.6%	3.7%
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	0.7%	3.7%

Chapter 10

Conclusions

Throughout the decades, the SM has been explored in many different ways, and this work presents two such approaches. The first explores the potential of a direct search for BSM physics from physical signatures that are not possible in the SM using the proposed CODEX-b detector. The second examines a search for the forbidden cLFV process $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ using the LHCb detector.

The CODEX-b detector represents a valuable addition to the range of LLP experiments currently planned or running, due to its inexpensive nature, simple design, and the world-leading sensitivity to transverse LLPs it could provide. In order to proceed to a full installation, preparations are underway for the installation of a smaller demonstrator unit, CODEX-β, to verify the design choices made before scaling up to CODEX-b. This will also provide information about the installation environment, such as the level of SM background, which may inform future design iterations. The RPC module that defines the base-unit of the CODEX-β and CODEX-b detectors, as it is tessellated to form their surfaces, is a key aspect of the detector design which requires verification. In particular, whether misalignment of the internal triplet of the module relative to the mechanical framework would negatively affect the track reconstruction of LLP candidate events in CODEX-β and CODEX-b. Section 5.2 shows the methodology of a study of these misalignments and demonstrates that most translations of up to 20 mm and/or any rotation about any axis of up to 3° does not have a significant effect on the track reconstruction for a dilepton event. Overall, the mechanical precision when manufacturing the frame to support the RPC triplet is anticipated to be < 5 mm, and hence no meaningful impact on the tracking reconstruction for CODEX-β or CODEX-b is expected.

Another, more fundamental, aspect of CODEX-b that required examination was the hardware of the RPCs that comprise the RPC module. Specifically, the ability of the hit information of each individual RPC to be read out and stored, so that LLP candidate

tracks can be reconstructed. Those tracks could then be evaluated to determine the characteristics of the parent particle that produced it, providing constraints to many BSM models. Section 5.1 shows the procedure to test a set of RPC front-end readout boards to be installed on RPCs for the CODEX-β demonstrator unit. Overall, a total of 440 boards were evaluated and the results documented. Of these only 18 were found to be unusable in their current form, but there were enough working boards to instrument CODEX-β if none fail during installation, with two spare boards in the Eta configuration.

The LHCb detector has a varied physics programme in comparison to the more directed focus of CODEX-b, with many analyses being performed to search for BSM physics, or more precisely measure known SM processes. Part II highlighted one particular search for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$, a process that requires cLFV and is forbidden in the SM. As it is forbidden, any observation of this specific decay would be direct evidence of BSM physics, such as the existence of additional tree-level diagrams with new particles *e.g.* the leptoquark, that enhance its branching fraction to the level of LHCb's sensitivity. Although the signal region for $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ is currently blinded, estimated upper limits on its branching fraction in the absence of signal can still be determined through examination with the $\mathrm{CL_s}$ method on a constructed dataset that contains no signal. The limit from this background-only proxy dataset would represent the result in the absence of signal, and the equivalent number of signal events that would be expected to be exceeded, in order to claim observation of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$. The observed upper limit from a set of pseudo-experiments has been evaluated both in the nominal case proposed in Part II, and the case where an additional Λ_c^+ veto has been applied for a cleaner background environment, as discussed in Section 9.6. These are given as Equations 10.0.1 and 10.0.2 respectively.

$$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)_{\mathrm{std}} \lesssim 2.09 \times 10^{-7} \ @\ 95\%\ \mathrm{CL}, \tag{10.0.1}$$

$$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)_{\mathrm{veto}} \lesssim 6.45 \times 10^{-8} \ @\ 95\%\ \mathrm{CL}. \tag{10.0.2}$$

The result that naively includes a Λ_c^+ veto gives a value $\sim$3 times smaller than the nominal case and thus represents a stronger limit, but also has larger error bands on the $\mathrm{CL_s}$ value, due to greater statistical uncertainties compared to the nominal case. However, the overall relative uncertainty is approximately $\mathcal{O}(1\text{–}4)\%$, compared to $\mathcal{O}(20)\%$ without the veto. This implies that the Λ_c^+ veto should be considered more fully throughout the analysis to improve the limit and reduce the effect of statistical uncertainties on it. From these estimated limits, the model-independent theoretical prediction from Ref. [56] for $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm) < 1.1 \times 10^{-8}$ cannot yet be further constrained. However, future results with the full unblinded dataset may provide a stronger limit that could inform BSM models. Moreover, this estimated upper limit represents both LHCb's first and the world's most stringent experimental upper limit placed on $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$.

References

[1] Muon g-2, B. Abi *et al.*, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, Phys. Rev. Lett. **126**, 141801 (2021), arXiv:2104.03281.

[2] Muon g-2, D. P. Aguillard *et al.*, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm*, (2023), arXiv:2308.06230.

[3] T. Aoyama *et al.*, *The anomalous magnetic moment of the muon in the Standard Model*, Phys. Rept. **887**, 1–166 (2020), arXiv:2006.04822.

[4] G. Colangelo *et al.*, *Prospects for precise predictions of a_μ in the Standard Model*, (2022), arXiv:2203.15810.

[5] CMD-3, F. V. Ignatov *et al.*, *Measurement of the $e^+e^- \to \pi^+\pi^-$ cross section from threshold to* 1.2 GeV *with the CMD-3 detector*, (2023), arXiv:2302.08834.

[6] S. Borsanyi *et al.*, *Leading hadronic contribution to the muon magnetic moment from lattice QCD*, Nature **593**, 51–55 (2021), arXiv:2002.12347.

[7] H. Andernach and F. Zwicky, *English and Spanish Translation of Zwicky's (1933) The Redshift of Extragalactic Nebulae*, 2017, arXiv:1711.01693.

[8] E. Corbelli and P. Salucci, *The Extended Rotation Curve and the Dark Matter Halo of M33*, Mon. Not. Roy. Astron. Soc. **311**, 441–447 (2000), arXiv:astro-ph/9909252.

[9] S. M. Carroll, W. H. Press, and E. L. Turner, *The cosmological constant.* Annual Rev. Astron. Astrophys. **30**, 499–542 (1992).

[10] Supernova Search Team, A. G. Riess *et al.*, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, Astron. J. **116**, 1009–1038 (1998), arXiv:astro-ph/9805201.

[11] R. P. Feynman, *Mathematical formulation of the quantum theory of electromagnetic interaction*, Phys. Rev. **80**, 440–457 (1950).

[12] R. P. Feynman and M. Gell-Mann, *Theory of the Fermi Interaction*, Physical Review **109**, 193–198 (1958).

[13] E. C. G. Sudarshan and R. E. Marshak, *Chirality invariance and the universal fermi interaction*, Phys. Rev. **109**, 1860–1862 (1958).

[14] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2**, 1285–1292 (1970).

[15] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519**, 367–377 (1968).

[16] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22**, 579–588 (1961).

[17] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19**, 1264–1266 (1967).

[18] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13**, edited by J. C. Taylor, 321–323 (1964).

[19] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12**, 132–133 (1964).

[20] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13**, edited by J. C. Taylor, 508–509 (1964).

[21] M. Thomson, *Modern particle physics* (Cambridge University Press, New York, 2013), doi: 10.1017/CBO9781139525367.

[22] P. D. Group, R. L. Workman, *et al.*, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2022**, 10.1093/ptep/ptac097 (2022).

[23] J. C. Swallow, *Searches for rare and forbidden kaon decays at the NA62 experiment at CERN* (Birmingham U., Mar. 2021), CDS: 2766585.

[24] BESIII Collaboration, M. Ablikim *et al.*, *Observation of a charged charmoniumlike structure in* $e^+e^- \to \pi^+\pi^- J/\psi$ *at* $\sqrt{s}$=4.26 GeV, Phys. Rev. Lett. **110**, 252001 (2013).

[25] Belle Collaboration, Z. Q. Liu *et al.*, *Study of* $e^+e^- \to \pi^+\pi^- J/\psi$ *and observation of a charged charmoniumlike state at belle*, Phys. Rev. Lett. **110**, 252002 (2013).

[26] LHCb Collaboration, R. Aaij *et al.*, *Observation of the resonant character of the* $Z(4430)^-$ *state*, Phys. Rev. Lett. **112**, 222002 (2014), arXiv:1404.1903.

[27] LHCb Collaboration, R. Aaij *et al.*, *Observation of* $J/\psi p$ *resonances consistent with pentaquark states in* $\Lambda_b^0 \to J/\psi p K^-$ *decays*, Phys. Rev. Lett. **115**, 072001 (2015), arXiv:1507.03414.

[28] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97**, 31–145 (1983).

[29] T. Sjostrand, *Jet Fragmentation of Nearby Partons*, Nucl. Phys. B **248**, 469–502 (1984).

[30] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. B **238**, 492–528 (1984).

[31] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. **10**, 531–533 (1963).

[32] R. P. Feynman, *Space–time approach to quantum electrodynamics*, Phys. Rev. **76**, edited by L. M. Brown, 769–789 (1949).

[33] F. Halzen and A. D. Martin, *Quarks and leptons: an introductory course in modern particle physics* (Jan. 1984).

[34] C. P. Burgess and G. D. Moore, *The standard model: A primer* (Cambridge University Press, Dec. 2006), ISBN: 978-0-511-25485-7, 978-1-107-40426-7, 978-0-521-86036-9.

[35] B. Pontecorvo, *Inverse beta processes and nonconservation of lepton charge*, Zh. Eksp. Teor. Fiz. **34**, 247 (1957).

[36] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Progress of Theoretical Physics **28**, 870–880 (1962), doi: 10.1143/PTP.28.870.

[37] Super-Kamiokande Collaboration, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81**, 1562–1567 (1998), arXiv:hep-ex/9807003.

[38] L. S. Littenberg and R. Shrock, *Implications of improved upper bounds on $|\Delta L| = 2$ processes*, Phys. Lett. B **491**, 285–290 (2000), arXiv:hep-ph/0005285.

[39] A. Atre, T. Han, S. Pascoli, and B. Zhang, *The Search for Heavy Majorana Neutrinos*, JHEP **05**, 030 (2009), arXiv:0901.3589.

[40] LHCb Collaboration, R. Aaij *et al.*, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, (2022), arXiv:2212.09152, to appear in Phys. Rev. Lett.

[41] LHCb Collaboration, R. Aaij *et al.*, *Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays*, (2022), arXiv:2212.09153, to appear in Phys. Rev. D.

[42] Belle Collaboration, S. Choudhury *et al.*, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, JHEP **03**, 105 (2021), arXiv:1908.01848.

[43] Belle Collaboration, A. Abdesselam *et al.*, *Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle*, Phys. Rev. Lett. **126**, 161801 (2021), arXiv:1904.02440.

[44] BaBar Collaboration, J. P. Lees *et al.*, *Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)}l^+l^-$*, Phys. Rev. D **86**, 032012 (2012), arXiv:1204.3933.

[45] LHCb Collaboration, 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), arXiv:2105.14007.

[46] LHCb Collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, JHEP **06**, 115 (2015), arXiv:1503.07138.

[47] LHCb Collaboration, 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), arXiv:2003.04831.

[48] LHCb Collaboration, R. Aaij *et al.*, *Angular analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ decay*, Phys. Rev. Lett. **126**, 161802 (2021), arXiv:2012.13241.

[49] E. Gildener, *Gauge-symmetry hierarchies*, Phys. Rev. D **14**, 1667–1672 (1976).

[50] A. A. Natale and R. C. Shellard, *The gauge hierarchy problem*, Journal of Physics G: Nuclear Physics **8**, 635 (1982).

[51] R. D. Peccei and H. R. Quinn, CP *Conservation in the presence of pseudoparticles*, Phys. Rev. Lett. **38**, 1440–1443 (1977).

[52] P. Langacker, *The physics of heavy Z' gauge bosons*, Rev. Mod. Phys. **81**, 1199–1228 (2009).

[53] J. C. Pati and A. Salam, *Lepton number as the fourth "color"*, Phys. Rev. D **10**, 275–289 (1974).

[54] I. de Medeiros Varzielas and J. Talbert, *Simplified models of flavourful leptoquarks*, The European Physical Journal C **79**, 10.1140/epjc/s10052-019-7047-2 (2019).

[55] M. Neubert, *Effective field theory and heavy quark physics*, in Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in D $\geqq$ 4 (Dec. 2005), pp. 149–194, arXiv:hep-ph/0512222.

[56] M. Bordone, M. Rahimi, and K. K. Vos, *Lepton flavour violation in rare Λ_b decays*, Eur. Phys. J. C **81**, 756 (2021), arXiv:2106.05192.

[57] M. Bordone, *Rare decays: a theoretical perspective*, Dec. 2019, Private Communication, indico event: 858082.

[58] C. Greub and P. Liniger, *Calculation of next-to-leading QCD corrections to $b \to sg$*, Phys. Rev. D **63**, 054025 (2001), arXiv:hep-ph/0009144.

[59] C. Froggatt and H. Nielsen, *Hierarchy of quark masses, cabibbo angles and CP violation*, Nuclear Physics B **147**, 277–298 (1979).

[60] M. Bordone, O. Catà, T. Feldmann, and R. Mandal, *Constraining flavour patterns of scalar leptoquarks in the effective field theory*, JHEP **03**, 122 (2021), arXiv:2010.03297.

[61] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton-flavour-violating decays $B_s^0 \to e^\pm\mu^\mp$ and $B^0 \to e^\pm\mu^\mp$*, Phys. Rev. Lett. **111**, 141801 (2013), arXiv:1307.4889.

[62] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decay $D^0 \to e^\pm\mu^\mp$*, Phys. Lett. **B754**, 167 (2016), arXiv:1512.00322.

[63] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decays $B_{(s)}^0 \to e^\pm\mu^\mp$*, JHEP **03**, 078 (2018), arXiv:1710.04111.

[64] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton-flavour-violating decays $B_s^0 \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$*, Phys. Rev. Lett. **123**, 211801 (2019), arXiv:1905.06614.

[65] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decays $B^+ \to K^+\mu^\pm e^\mp$*, Phys. Rev. Lett. **123**, 231802 (2019), arXiv:1909.01010.

[66] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton-flavour violating decays $B^0 \to K^{*0}\mu^\pm e^\mp$ and $B_s^0 \to \phi\mu^\pm e^\mp$*, JHEP **06**, 073 (2023), arXiv:2207.04005.

[67] LHCb Collaboration, R. Aaij *et al.*, *Search for the lepton flavour violating decay $B^0 \to K^{*0}\tau^\pm\mu^\mp$*, (2022), arXiv:2209.09846, to appear in JHEP.

[68] Planck Collaboration, N. Aghanim *et al.*, *Planck 2018 results. VI. Cosmological parameters*, (2018), arXiv:1807.06209.

[69] G. Steigman, B. Dasgupta, and J. F. Beacom, *Precise Relic WIMP Abundance and its Impact on Searches for Dark Matter Annihilation*, Phys. Rev. D **86**, 023506 (2012), arXiv:1204.3622.

[70] M. J. Strassler and K. M. Zurek, *Echoes of a hidden valley at hadron colliders*, Phys. Lett. B **651**, 374–379 (2007), arXiv:hep-ph/0604261.

[71] Z. Liu and B. Tweedie, *The Fate of Long-Lived Superparticles with Hadronic Decays after LHC Run 1*, JHEP **06**, 042 (2015), arXiv:1503.05923.

[72] L. Lee *et al.*, *Collider Searches for Long-Lived Particles Beyond the Standard Model*, Prog. Part. Nucl. Phys. **106**, 210–255 (2019), arXiv:1810.12602.

[73] D. E. Kaplan, M. A. Luty, and K. M. Zurek, *Asymmetric Dark Matter*, Phys. Rev. D **79**, 115016 (2009), arXiv:0901.4117.

[74] R. T. Co, F. D'Eramo, L. J. Hall, and D. Pappadopulo, *Freeze-In Dark Matter with Displaced Signatures at Colliders*, JCAP **12**, 024 (2015), arXiv:1506.07532.

[75] M. Baumgart, C. Cheung, J. T. Ruderman, L.-T. Wang, and I. Yavin, *Non-Abelian Dark Sectors and Their Collider Signatures*, JHEP **04**, 014 (2009), arXiv:0901.0283.

[76] CODEX-b Collaboration, G. Aielli *et al.*, *Expression of interest for the CODEX-b detector*, Eur. Phys. J. C **80**, 1177 (2020), arXiv:1911.00481.

[77] J. Alimena *et al.*, *Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider*, J. Phys. G **47**, 090501 (2020), arXiv:1903.04497.

[78] FASER Collaboration, A. Ariga *et al.*, *Letter of Intent for FASER: ForwArd Search ExpeRiment at the LHC*, (2018), arXiv:1811.10243.

[79] C. Alpigiani *et al.*, *A Letter of Intent for MATHUSLA: a dedicated displaced vertex detector above ATLAS or CMS*. Tech. rep. (CERN, Geneva, July 2018), arXiv:1811.00927.

[80] W. Bonivento *et al.*, *Proposal to Search for Heavy Neutral Leptons at the SPS*, (2013), arXiv:1310.1762.

[81] A. Ball *et al.*, *A Letter of Intent to Install a milli-charged Particle Detector at LHC P5*, (2016), arXiv:1607.04669.

[82] CAST Collaboration, D. Anastassopoulos *et al.*, *CAST Physics Proposal to SPSC*, tech. rep. (CERN, Geneva, 2011), CDS: 1340994.

[83] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, (2022), General Photo.

[84] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl, *LHC Design Report*, CERN Yellow Reports: Monographs (CERN, Geneva, 2004), doi: 10.5170/CERN-2004-003-V-3.

[85] R. Alemany-Fernandez *et al.*, *Operation and Configuration of the LHC in Run 1*, (2013), CDS: 1631030.

[86] J. Wenninger, *Operation and Configuration of the LHC in Run 2*, (2019), CDS: 2668326.

[87] ATLAS Collaboration, G. 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), arXiv:1207.7214.

[88] CMS Collaboration, S. 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), arXiv:1207.7235.

[89] LHCb Collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3**, S08005 (2008).

[90] LHCb Collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30**, 1530022 (2015), arXiv:1412.6352.

[91] P. F. Harrison, *The physics of B factories*, 10.5170/CERN-2000-007.245 (2000).

[92] M. Pepe Altarelli and F. Teubert, *B Physics at LHCb*, Int. J. Mod. Phys. A **23**, edited by G. Kane and A. Pierce, 5117–5136 (2008), arXiv:0802.1901.

[93] LHCb collaboration, C. Elsasser, *$\bar{b}b$ Production angle plots*, %7B%5Chref%7Bhttps://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb%5C_ProductionAngles.html%7D%7Bhttps://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb%5C_ProductionAngles.html%7D%7D, Accessed: 2023-11-06.

[94] LHCb Collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030 (CERN, Geneva, 2003).

[95] W. Herr and B. Muratori, *Concept of luminosity*, (2006), doi: 10.5170/CERN-2006-002.361.

[96] A. W. Chao, K. H. Mess, M. Tigner, and F. Zimmermann, *Handbook of accelerator physics and engineering*, 2nd (WORLD SCIENTIFIC, 2013), eprint: https://www.worldscientific.com/doi/pdf/10.1142/8543.

[97] *Lhcb trigger report*, https://lbtriggerreport.cern.ch/reports/, Accessed: 2023-01-18.

[98] LHCb Collaboration, R. Aaij *et al.*, *Precision measurement of the Λ_c^+, Ξ_c^+, and Ξ_c^0 baryon lifetimes*, Phys. Rev. D **100**, 032001 (2019).

[99] HFLAV, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. C **81**, 226 (2021), arXiv:1909.12524.

[100] LHCb Collaboration, R. Aaij *et al.*, *The LHCb Upgrade I*, (2023), arXiv:2305.10515, to appear in JINST.

[101] LHCb Collaboration, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9**, P09007 (2014), arXiv:1405.7808.

[102] L. B. A. Hommels, *The LHCb Outer Tracker Detector Design and Production*, tech. rep. (CERN, Geneva, 2005), CDS: 828990.

[103] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10**, P02007 (2015), arXiv:1408.1251.

[104] M.-H. Schune, F. Polci, and M. Borsato, *HOP an additional tool for decays involving electrons*, tech. rep. LHCb-INT-2015-037. CERN-LHCb-INT-2015-037 (CERN, Geneva, Nov. 2015), CDS: 2102345.

[105] LHCb Collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010 (CERN, Geneva, 2001).

[106] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8**, P10020 (2013), arXiv:1306.0249.

[107] C. Lippmann, *Particle identification*, Nucl. Instrum. Meth. A **666**, 148–172 (2012), arXiv:1101.3276.

[108] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8**, P04022 (2013), arXiv:1211.3055.

[109] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14**, P04013 (2019), arXiv:1812.10790.

[110] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, tech. rep. (CERN, Geneva, 2014), CDS: 1701134.

[111] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178**, 852–867 (2008), arXiv:0710.3820.

[112] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05**, 026 (2006), arXiv:hep-ph/0603175.

[113] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331**, 032047 (2011).

[114] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462**, 152–155 (2001).

[115] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45**, 97–107 (2006), arXiv:hep-ph/0506026.

[116] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199**, 86 (2016), arXiv:1011.0937.

[117] Geant4 collaboration, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Dubois, *et al.*, *Geant4 developments and applications*, IEEE Trans.Nucl.Sci. **53**, 270 (2006).

[118] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506**, 250 (2003).

[119] ATLAS Collaboration, M. Aaboud *et al.*, *Search for heavy charged long-lived particles in the atlas detector in* 36.1 fb^{-1} *of proton-proton collision data at* $\sqrt{s} =$ 13 TeV, Phys. Rev. D **99**, 092007 (2019).

[120] CMS Collaboration, A. Sirunyan *et al.*, *Search for disappearing tracks in proton-proton collisions at* $\sqrt{s} = 13$ TeV, Physics Letters B **806**, 135502 (2020).

[121] CMS Collaboration, A. M. Sirunyan *et al.*, *Search for long-lived particles using displaced jets in proton-proton collisions at* $\sqrt{s} = 13$ TeV, Phys. Rev. D **104**, 012015 (2021).

[122] CODEX-b Collaboration, G. Aielli *et al.*, *The Road Ahead for CODEX-b*, (2022), arXiv:2203.07316.

[123] CODEX-b Collaboration, V. V. Gligorov, S. Knapen, M. Papucci, and D. J. Robinson, *Searching for Long-lived Particles: A Compact Detector for Exotics at LHCb*, Phys. Rev. **D97**, 015023 (2018), arXiv:1708.09395.

[124] ATLAS Collaboration, G. Aad *et al.*, *Performance of the ATLAS RPC detector and Level-1 muon barrel trigger at* $\sqrt{s} = 13\,\text{TeV}$, JINST **16**, P07029 (2021), arXiv:2103.01029.

[125] CODEX-b Collaboration, G. Aielli *et al.*, *CODEX-β Installation plan*, (2022), [Addendum: Comments from FIPPAN (Aug 2023)] (Internal submission to LHCb Technical Board).

[126] T. Gorordo, S. Knapen, B. Nachman, D. J. Robinson, and A. Suresh, *Geometry Optimization for Long-lived Particle Detectors*, (2022), arXiv:2211.08450.

[127] B. Dey, J. Lee, V. Coco, and C.-S. Moon, *Background studies for the CODEX-b experiment: measurements and simulation*, (2019), arXiv:1912.03846.

[128] M. Bauer, O. Brandt, L. Lee, and C. Ohm, *ANUBIS: Proposal to search for long-lived neutral particles in CERN service shafts*, (2019), arXiv:1909.13022.

[129] F. Martzloff and T. Leedy, *Electrical fast-transient tests : applications and limitations*, IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS. **26**, 151–159 (1990).

[130] D. Northacker, *Design drawings for RPC module support structure*, Private Communication, 2021, Email.

[131] G. Gibbons, *Preliminary framework for CODEX-b simulations*, 2019, Private Communication - this software was developed as part of G.Gibbons' master's thesis.

[132] C. Dalitz, T. Schramke, and M. Jeltsch, *Iterative Hough Transform for Line Detection in 3D Point Clouds*, Image Processing On Line **7**, 184–196 (2017), doi: 10.5201/ipol.2017.208.

[133] P. V. C. Hough, *Method and Means for Recognizing Complex Patterns,* U.S. Patent 3,069,654, Ser. No. 17,7156 Claims, 1962.

[134] A. Hassanein *et al.*, *A survey on hough transform, theory, techniques and applications*, CoRR (2015), arXiv:1502.02160.

[135] ATLAS Collaboration, L. Pizzimento, *Performance of the BIS78 RPC detectors: a new concept of electronics and detector integration for high-rate and fast timing large size RPCs*, tech. rep. (CERN, Geneva, June 2020), ATL-MUON-PROC-2020-012.

[136] G. Andreassi, F. Archilli, S. Bifani, V. Dedu, J. De Boer, S. De Keijzer, S. Ferreres-Solé, L. Greeven, M. Mulder, L. Pescatore, M. Santimaria, P. Swallow, N. Tuning, N. Sahoo, N. Watson, and C. Wang, *Search for the Lepton Flavour Violating decay* $\Lambda_b^0 \to \Lambda^0 e^{\pm}\mu^{\mp}$, (2018), LHCb-ANA-2018-013 ; CERN-LHCb-ANA-2018-013 (Unpublished), revised in 2023.

[137] E. Smith, P. Owen, and M. Patel, *Data driven reconstruction efficiency of* K_s^0 *mesons as a function of decay position*, tech. rep. (CERN, Geneva, 2014), CDS: 1752213.

[138] LHCb Collaboration, R. Aaij *et al.*, *Measurement of b-hadron fractions in* 13 TeV *pp collisions*, Phys. Rev. **D100**, 031102(R) (2019), arXiv:1902.06794.

[139] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the* B_c^- *production fraction and asymmetry in* 7 *and* 13 TeV *pp collisions*, Phys. Rev. **D100**, 112006 (2019), arXiv:1910.13404.

[140] L. M. Garcia Martin, F. Archilli, J. J. Back, S. Bifani, T. Blake, I. Carli, V.-G. Dedu, S. Ferreres, L. Greeven, S. De Keijzer, M. Kreps, M. Mulder, N. Sahoo, M. Santimaria, P. N. Swallow, N. Tuning, M. Van Veghel, and N. Watson, *Measurement of branching fraction of the* $\Lambda_b^0 \to J/\psi\Lambda$ *decay*, (2023), LHCb-ANA-2023-027; CERN-LHCb-ANA-2023-027 (Unpublished).

[141] S. Das, *A simple alternative to the Crystal Ball function*, (2016), arXiv:1603.08591.

[142] N. L. Johnson, *Systems of Frequency Curves Generated by Methods of Translation*, Biometrika **36**, 149–176 (1949).

[143] A. L. Read, *Presentation of search results: The CLs technique*, Journal of Physics G: Nuclear and Particle Physics **28**, 2693–2704 (2002), doi: 10.1088/0954-3899/28/10/313.

[144] A. L. Read, *Modified frequentist analysis of search results (the* CL_s *method)*, in Workshop on Confidence Limits: CERN, Geneva, Switzerland 17 - 18 Jan 2000. 1st Workshop on Confidence Limits (CERN, 2000), pp. 81–101.

[145] T. Blake, T. Gershon, and G. Hiller, *Rare b hadron decays at the LHC*, Ann. Rev. Nucl. Part. Sci. **65**, 113–143 (2015), arXiv:1501.03309.

[146] C. Marin Benito, G. Andreassi, V. Franco Lima, P. Stefko, A. Puig Navarro, and V. Bellee, *Introducing the HOP mass in the LHCb software*, tech. rep. LHCb-INT-2017-012. CERN-LHCb-INT-2017-012 (CERN, Geneva, May 2017), CDS: 2265969.

[147] M. Mulder, *The essence of rare beauty: Studying* $B^0_{(s)} \to \mu^+\mu^-$ *decays with the LHCb experiment*, PhD thesis (Groningen U., 2021).

[148] M. Pivk and F. R. Le Diberder, *SPlot: A Statistical tool to unfold data distributions*, Nucl. Instrum. Meth. A **555**, 356–369 (2005), arXiv:physics/0402083.

[149] Particle Data Group, K. A. Olive *et al.*, *Review of particle physics*, Chin. Phys. **C38**, 090001 (2014), and 2015 update.

[150] Particle Data Group, M. Tanabashi *et al.*, *Review of Particle Physics*, Phys. Rev. D **98**, 030001 (2018).

[151] LHCb Collaboration, R. Aaij *et al.*, *Angular moments of the decay* $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ *at low hadronic recoil*, JHEP **09**, 146 (2018), arXiv:1808.00264.

[152] L. Anderlini *et al.*, *The PIDCalib package*, tech. rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021 (CERN, Geneva, July 2016), CDS: 2202412.

[153] R. Aaij, C. Abellán Beteta, T. Ackernley, B. Adeva, M. Adinolfi, H. Afsharnia, C. Aidala, S. Aiola, Z. Ajaltouni, S. Akar, and et al., *Measurement of the electron reconstruction efficiency at lhcb*, Journal of Instrumentation **14**, P11023–P11023 (2019).

[154] A. Rogozhnikov, *Reweighting algorithms*, (2020) https://arogozhnikov.github.io/hep%5C_ml/reweight.html (visited on 09/30/2010), Accessed: 24-06-2020.

[155] T. Chen and C. Guestrin, *Xgboost: a scalable tree boosting system*, in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, KDD '16 (2016), pp. 785–794.

[156] M. W. Browne, *Cross-Validation Method*, Journal of Mathematical Psychology **44**, 108–132 (2000).

[157] A. N. Kolmogorov, *Sulla determinazione empirica di una legge didistribuzione*, Giorn Dell'inst Ital Degli Att **4**, 89–91 (1933).

[158] N. V. Smirnov, *On the estimation of the discrepancy between empirical curves of distribution for two independent samples*, Bull. Math. Univ. Moscou **2**, 3–14 (1939).

[159] G. Punzi, *Sensitivity of searches for new signals and its optimization*, in Statistical problems in particle physics, astrophysics, and cosmology, edited by L. Lyons, R. Mount, and R. Reitmeyer (2003), p. 79, arXiv:physics/0308063.

[160] P. Koppenburg, *Statistical biases in measurements with multiple candidates*, (2017), arXiv:1703.01128.

[161] LHCb Collaboration, R. Aaij *et al.*, *Isospin amplitudes in* $\Lambda_b^0 \to J/\psi\,\Lambda(\Sigma^0)$ *and* $\Xi_b^0 \to J/\psi\,\Xi^0(\Lambda)$ *decays*, Phys. Rev. Lett. **124**, 111802 (2020), arXiv:1912.02110.

[162] BESIII Collaboration, M. Ablikim *et al.*, *Study of the Semileptonic Decay* $\Lambda_c^+ \to \Lambda e^+\nu_e$, 10.1103/PhysRevLett.129.231803 (2022), arXiv:2207.14149.

[163] M. Oreglia, *A Study of the Reactions* $\psi' \to \gamma\gamma\psi$, PhD thesis (Stanford U., Dec. 1980).

[164] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis (Cracow, INP, 1986).

[165] LHCb Collaboration, R. Aaij *et al.*, *Production of* J/ψ *and* Υ *mesons in pp collisions at* $\sqrt{s}$ =8 TeV, JHEP **06**, 064 (2013), arXiv:1304.6977.

[166] LHCb Collaboration, R. Aaij *et al.*, *Measurement of forward* J/ψ *production cross-sections in pp collisions at* $\sqrt{s}$ =13 TeV, JHEP **10**, 172 (2015), arXiv:1509.00771.

[167] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231**, 289–337 (1933).

[168] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math.Stat. **9**, 60–62 (1938).

[169] GammaCombo Collaboration, M. Kenzie *et al.*, *Gammacombo user manual*, https://gammacombo.github.io/manual.pdf.

[170] LHCb Collaboration, R. Aaij *et al.*, *Measurement of the CKM angle* γ *from a combination of LHCb results*, JHEP **12**, 087 (2016), arXiv:1611.03076.

[171] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Physical Review D **57**, 3873–3889 (1998).

[172] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. A **764**, 150–155 (2014), arXiv:1312.5000.

[173] G. Andreassi, F. Archilli, A. Bay, R. Kappert, C. Onderwater, A. Pellegrino, L. Pescatore, M. Van Veghel, and G. Wiersema, *Search for the lepton flavour violating decays* $B_s^0 \to e^\pm\mu^\mp$ *and* $B^0 \to e^\pm\mu^\mp$ *with Run I data*, (2016), CDS: 2154256.

[174] J. Albrecht, J. Bhom, M. Chrzaszcz, G. Meier, T. Mombächer, M. Pikies, F. Polci, S. Reichert, and N. Serra, *Search for the lepton flavour violating decays* $B^+ \to K^+\mu^\pm e^\mp$, (2018), Internal note: LHCb-ANA-2016-066 (Unpublished).

Appendix A

Thresholds and Selection for Trigger Configuration Keys Used in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ Analysis

This section summarises the trigger thresholds and selections for L0 and HLT1 as used in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ analysis. Table A.1 contains the thresholds for L0. Table A.2 is for the HLT1 trigger line HLT1TrackAllL0. Table A.3 is for HLT1TrackMVA with the b column matching the parameter in Equation 7.2.1. Table A.4 is for HLT1TrackMuon. Finally, Table A.5 is for HLT1TrackMuonMVA. It should be noted that in these tables the luminosity percentages per year and magnet polarity do not necessarily add up to 100% as may be expected, this is because the luminosity of erroneous data not considered by the trigger is included within the total luminosity but is not shown in the tables.

Table A.1: Selections for the trigger line, L0Muon, for the 2011–2018 TCKs.

Year	TCK hex values	$E_T^{L0}(e)$ (GeV)	$p_T^{L0}(\mu)$ (GeV/c)	$p_T^{L0}(\mu_1) \cdot p_T^{L0}(\mu_2)$ (GeV$^2/c^2$)	% Luminosity MU	% Luminosity MD
2011	0x40760037 (MC)	2.50	1.48	1.68	-	-
	0x360032, 0x5A0032, 0x6D0032, 0x730035, 0x760037, 0x790037 0x790038	2.50	1.48	1.68	98.01	98.97
	0x5D0033	2.50	0.80	1.60	0.47	0.00
2012	0x409F0045 (MC)	2.96	1.76	2.56	-	-
	0x7F0040, 0x860040, 0x8C0040	2.50	1.48	1.68	0.00	6.30
	0x94003D, 0x97003D, 0x990042, 0x9A0042	2.72	1.76	2.56	65.40	47.41
	0x990044, 0x9F0045, 0xA10044, 0xA10045, 0xA20044, 0xA30044	2.96	1.76	2.56	31.57	23.91
	0xA30046, 0xA90046, 0xAB0046, 0xAC0046	2.86	1.76	2.56	2.74	21.52
2015	0x411400A2 (MC)	2.69	2.80	1.69	-	-
	0x10600A3	2.28	2.40	1.69	0.00	31.68
	0x10600A6	1.68	1.80	1.20	0.00	0.03
	0x10600A7	1.39	1.50	1.15	0.00	0.53
	0x10700A1	1.80	1.90	1.43	0.00	3.33
	0x10600A2, 0x10800A2, 0x11400A8	2.69	2.80	1.69	90.19	56.12
2016	0x5138160F (MC)	2.40	1.80	2.25	-	-
	0x11291603	2.11	1.10	1.00	0.00	3.97
	0x11291604	2.26	1.30	1.44	0.00	2.89
	0x11291605, 0x1137160E, 0x1138160E	2.59	1.50	1.69	0.00	15.48
	0x11371609, 0x11381609, 0x11321609, 0x11341609, 0x11351609, 0x11361609	2.35	1.30	1.69	83.14	9.18
	0x1138160F	2.40	1.80	2.25	0.00	67.01
	0x11381611	2.62	1.50	1.96	5.55	0.00
	0x11381612	2.62	1.60	2.25	11.29	0.00
2017	0x51611709 (MC)	2.11	1.40	1.69	-	-
	0x11541707, 0x115417A7, 0x11561707, 0x11611707, 0x11601707	2.30	1.70	3.24	40.98	55.74
	0x11611708, 0x11601708, 0x11501703, 0x114E1703	2.11	1.10	1.00	23.83	16.23
	0x11611709	2.11	1.40	1.69	0.00	27.91
	0x11501705	2.59	1.50	1.69	23.60	0.00
	0x11501704	2.26	1.30	1.44	4.80	0.00
	0x11501706	2.69	1.90	3.24	5.27	0.00
	0x114E1702	1.87	0.70	0.81	1.39	0.00
2018	0x517A18A4 (MC)	2.38	1.75	3.24	-	-
	0x11771801, 0x11751801, 0x117A18A4, 0x11741801, 0x117A18A2, 0x11711801, 0x117718A1, 0x11671801, 0x11731801	2.38	1.75	3.24	100	99.98

Table A.2: Selections for the trigger line, HLT1TrackAllL0, for the Run 1 TCKs.

Year	TCK hex values	p_T^μ (GeV/c)	p^μ (GeV/c)	χ^2_{IP}	IP (mm)	χ^2/ndf	% Luminosity MU	% Luminosity MD
2011	0x40760037 (MC)	1.7	10.0	16.0	0.1	2	-	-
	0x360032, 0x5a0032, 0x5d0033, 0x6d0032, 0x730035, 0x760037, 0x790037, 0x790038	1.7	10.0	16.0	0.1	2	98.5	99.0
2012	0x409f0045 (MC)	1.6	3.0	16.0	0.1	2.0	-	-
	0x7f0040, 0x860040	1.7	1.0	16.0	0.1	2.25	0.0	0.5
	0x8c0040	1.7	10.0	16.0	0.1	2.0	0.00	5.76
	0x94003d, 0x97003d	1.7	10.0	16.0	0.1	1.5	28.1	24.7
	0x990042, 0x990044, 0x9a0042, 0x9f0045, 0xa10044, 0xa10045, 0xa20044, 0xa30044, 0xa30046, 0xa90046, 0xab0046, 0xac0046	1.6	3.0	16.0	0.1	2.0	71.6	68.2

Table A.3: Selections for the trigger line, HLT1TrackMVA, for the Run 2 TCKs.

Year	TCK hex	p_T^μ (GeV/c)	χ^2_{IP}	b	% Luminosity MU	% Luminosity MD
2015	0x411400a2 (MC)	1.0	7.4	1.1	-	-
	0x10600a2, 0x10600a3, 0x10600a6, 0x10600a7, 0x10700a1, 0x10800a2, 0x11400a8	1.0	7.4	1.1	90.2	91.7
2016	0x5138160f (MC)	1.0	7.4	1.1	-	-
	0x11291603, 0x11291604, 0x11291605, 0x11381609, 0x1138160e, 0x1138160f, 0x11321609, 0x11381611, 0x11381612	1.0	7.4	1.1	30.3	87.5
	0x11371609, 0x1137160e, 0x11351609, 0x11361609	1.0	7.4	2.3	54.8	11.0
	0x11341609	1.0	7.4	1.6	14.9	0.00
2017	0x51611709 (MC)	1.0	7.4	1.1	-	-
	0x11541707, 0x115417a7, 0x11561707, 0x11611707, 0x11611708, 0x11611709, 0x11601708, 0x11601707, 0x11501705, 0x11501704, 0x11501706, 0x11501703, 0x114e1702, 0x114e1703	1.0	7.4	1.1	99.9	99.9
2018	0x517a18a4 (MC)	1.0	7.4	1.1	-	-
	0x11771801, 0x11751801, 0x117a18a4, 0x11741801, 0x117a18a2, 0x11711801, 0x117718a1, 0x11671801, 0x11731801	1.0	7.4	1.1	100.0	100.0

Table A.4: Selections for the trigger line, HLT1TrackMuon, for 2011–2018 TCKs.

Year	TCK hex	p_T^μ (GeV/c)	p^μ (GeV/c)	χ^2_{IP}	IP (mm)	χ^2/ndf	% Luminosity MU	% Luminosity MD
2011	0x40760037 (MC)	1.0	8.0	16	0.1	2	-	-
	0x360032, 0x5a0032, 0x5d0033, 0x6d0032, 0x730035, 0x760037, 0x790037, 0x790038	1.0	8.0	16	0.1	2	98.5	99.0
	0x409f0045 (MC)	1.0	3.0	16.0	0.1	2.5	-	-
2012	0x7f0040, 0x860040	1.0	1.0	16.0	0.1	5.0	0.0	0.5
	0x8c0040	1.0	8.0	16.0	0.1	3.0	0.00	5.8
	0x94003d, 0x97003d	1.0	8.0	16.0	0.1	2.5	28.1	24.7
	0x990042, 0x990044, 0x9a0042, 0x9f0045, 0xa10044, 0xa10045, 0xa20044, 0xa30044, 0xa30046, 0xa90046, 0xab0046, 0xac0046	1.0	3.0	16.0	0.1	2.5	71.6	68.2
2015	0x411400a2 (MC)	0.91	6.0	10.0	3.0	-	-	
	0x10600a2, 0x10600a3, 0x10600a6, 0x10600a7, 0x10700a1, 0x10800a2, 0x11400a8	0.91	6.0	10.0	3.0	90.2	91.7	
2016	0x5138160f (MC)	1.10	6.0	35.0	3.0	-	-	
	0x11291603, 0x11291604, 0x11291605, 0x11321609, 0x11341609	0.91	6.0	10.0	3.0	28.3	16.0	
	0x11371609, 0x1137160e, 0x11381609, 0x1138160e, 0x1138160f, 0x11351609, 0x11361609, 0x11381611, 0x11381612	1.10	6.0	35.0	3.0	71.7	82.5	
2017	0x51611709 (MC)	1.1	6.0	35.0	3.0	-	-	
	0x11541707, 0x115417a7, 0x11561707, 0x11611707, 0x11611708, 0x11611709, 0x11601708, 0x11601707, 0x11501705, 0x11501704, 0x11501706, 0x11501703, 0x114e1702, 0x114e1703	1.1	6.0	35.0	3.0	99.9	99.9	
2018	0x517a18a4 (MC)	1.1	6.0	35.0	3.0	-	-	
	0x11771801, 0x11751801, 0x117a18a4, 0x11741801, 0x117a18a2, 0x11711801, 0x117718a1, 0x11671801, 0x11731801	1.1	6.0	35.0	3.0	100.0	100.0	

Table A.5: Selections for the trigger line, HLT1TrackMuonMVA, for the Run 2 TCKs.

Year	**TCK hex**	p_T^μ (GeV/c)	χ^2_{IP}	b	**% Luminosity**	
					MU	**MD**
2015	0x411400a2 (MC)	1.0	7.4	1.1	-	-
	0x10600a2, 0x10600a3, 0x10600a6, 0x10600a7, 0x10700a1, 0x10800a2, 0x11400a8	1.0	7.4	1.1	90.2	91.7
2016	0x5138160f (MC)	1.0	7.4	1.1	-	-
	0x11291603, 0x11291604, 0x11291605, 0x11371609, 0x1137160e, 0x11381609, 0x1138160e, 0x1138160f, 0x11321609, 0x11351609, 0x11361609, 0x11381611, 0x11381612	1.0	7.4	1.1	85.1	98.5
	0x11341609	1.0	7.4	1.6	14.9	0.00
2017	0x51611709 (MC)	1.0	7.4	1.1	-	-
	0x11541707, 0x115417a7, 0x11561707, 0x11611707, 0x11611708, 0x11611709, 0x11601708, 0x11601707, 0x11501705, 0x11501704, 0x11501706, 0x11501703, 0x114e1702, 0x114e1703	1.0	7.4	1.1	99.9	99.9
2018	0x517a18a4 (MC)	1.0	7.4	1.1	-	-
	0x11771801, 0x11751801, 0x117a18a4, 0x11741801, 0x117a18a2, 0x11711801, 0x117718a1, 0x11671801, 0x11731801	1.0	7.4	1.1	100.0	100.0

Appendix B

An Alternative Method of Determining the Expected Background Yields for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$

In order to verify the results of Tables 8.2 and 8.3, another method of predicting the expected yields can be utilised. This uses

$$N^{\sigma}_{normalised} = 2 \cdot \sigma_{pp\to b\bar{b}X} \cdot f_{\Lambda_b^0} \cdot \mathcal{L} \cdot \mathcal{B}_{BG} \cdot \varepsilon_{BG}, \tag{B.0.1}$$

where $\sigma_{pp\to b\bar{b}X}$ is the inclusive $b\bar{b}$ production cross-section, $f_{\Lambda_b^0}$ is the production fraction of Λ_b^0 baryons from a b quark, $\mathcal{L}$ is the integrated luminosity, and the factor two arises from the fact that there are two b quarks produced that can hadronise to form a Λ_b^0. The values for $\sigma_{pp\to b\bar{b}X}$ are $298 \pm 2 \pm 36\,\mu b$ for 8 TeV [165], and $495 \pm 2 \pm 52\,\mu b$ for 13 TeV [166], where the first uncertainty is statistical and the second systematic. The latter is used for Run 2, but for the full Run 1 category the linearity of the cross-section with respect to $\sqrt{s}$ is used to scale the 8 TeV number to the full Run 1 value[1] using a factor of $\frac{23}{24}$. The value for $f_{\Lambda_b^0}$ is 0.185 ± 0.014 and 0.179 ± 0.013 for Run 1 and Run 2 respectively, as derived using the Run 2 value for $\frac{f_{\Lambda_b^0}}{f_u+f_d} = 0.259 \pm 0.018$ [138] for both runs and $f_u + f_d = 0.713 \pm 0.026$ or 0.692 ± 0.015 [139] for Run 1 and Run 2 respectively.

Equation B.0.1 can be evaluated for any of the background modes described in Tables 8.2 and 8.3, but several key modes were important to verify. One of the most important was the $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ mode that is used as the scaling factor in Equation 8.1.1, therefore any issue observed would have affected the expected yields in all other modes. Additionally, the $e\mu$ final state $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\ell_i^+\nu_{\ell i})\ell_j^-\overline{\nu}_{\ell j}$ backgrounds were checked as these were expected to be the predominant backgrounds to be controlled. Table B.1 shows the yields obtained for these modes in each analysis category using both Equations B.0.1 and 8.1.1, their compatibility and the ratio between the two. This shows that for each case the derived expected yields were compatible with one another and with a ratio of 1, except for DD Run 1 which deviates by 20%. However, this is accounted

[1] *i.e.* $\sigma_{R1} = \sigma(8\,\text{TeV}) \cdot \frac{1}{8}(7 \cdot \frac{1}{3} + 8 \cdot \frac{2}{3}) = \frac{23}{24}$

for by a known issue with the uncertainties for K^0_S and Λ baryons, given in Ref. [137]. This was performed for all the backgrounds evaluated previously, and though not shown here, they all were also highly compatible for these two methods. Therefore, the expected yields in Tables 8.2 and 8.3 have been verified, and align well with this cross-check.

Overall, the *absolute* efficiencies for the $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ mode per category are under control at a level of a few percent for the LL samples and at a level of ten percent for DD samples. The final result depends on *relative* efficiencies instead, so no significant effect to that result is expected.

Table B.1: Comparison of expected yields in data for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ in the J/ψ q^2 bin, $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ and $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ as derived from Equations (B.0.1) and (8.1.1).

Background Sample	**Category**	**Expected Yields**		**Compatibility** ($\#\sigma$)	**Ratio** $\left(\frac{N^\sigma_{normalised}}{N_{normalised}}\right)$
		$N^\sigma_{normalised}$	$N_{normalised}$		
$\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ (J/ψ q^2)	DD-Run 1	1778.4 ± 249.7	1391.7 ± 51.0	1.52	1.278 ± 0.185
	LL-Run 1	1004.9 ± 141.9	1042.0 ± 50.3	0.25	0.964 ± 0.144
	DD-Run 2	7550.2 ± 793.6	7106.5 ± 136.1	0.55	1.062 ± 0.114
	LL-Run 2	6737.2 ± 703.8	6746.1 ± 139.7	0.01	0.999 ± 0.106
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$	No Brem-DD-Run 1	145.0 ± 46.3	107.0 ± 31.2	0.68	1.355 ± 0.586
	No Brem-LL-Run 1	69.0 ± 30.2	76.2 ± 31.6	0.16	0.905 ± 0.547
	No Brem-DD-Run 2	405.9 ± 83.1	382.4 ± 70.0	0.22	1.062 ± 0.291
	No Brem-LL-Run 2	232.7 ± 59.7	229.2 ± 54.6	0.04	1.015 ± 0.356
	Brem-DD-Run 1	135.8 ± 43.3	107.0 ± 31.2	0.54	1.269 ± 0.548
	Brem-LL-Run 1	50.4 ± 24.2	54.2 ± 25.4	0.11	0.930 ± 0.624
	Brem-DD-Run 2	511.2 ± 104.2	484.1 ± 87.6	0.20	1.056 ± 0.288
	Brem-LL-Run 2	423.0 ± 95.4	416.5 ± 84.5	0.05	1.016 ± 0.308
$\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$	No Brem-DD-Run 1	238.2 ± 68.0	191.3 ± 48.3	0.56	1.245 ± 0.475
	No Brem-LL-Run 1	133.2 ± 45.3	142.5 ± 45.6	0.14	0.935 ± 0.437
	No Brem-DD-Run 2	685.3 ± 126.3	642.6 ± 102.4	0.26	1.066 ± 0.260
	No Brem-LL-Run 2	441.0 ± 91.0	433.0 ± 80.6	0.07	1.018 ± 0.283
	Brem-DD-Run 1	154.2 ± 52.5	124.9 ± 38.6	0.45	1.234 ± 0.567
	Brem-LL-Run 1	58.7 ± 28.5	62.8 ± 29.8	0.10	0.935 ± 0.635
	Brem-DD-Run 2	755.7 ± 138.3	704.7 ± 113.2	0.29	1.072 ± 0.261
	Brem-LL-Run 2	672.9 ± 129.1	665.9 ± 113.3	0.04	1.010 ± 0.259

Appendix C

Examination of a Potential Signal Mass Resolution Correction for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$

In a pure leptonic decay, such as $B^0_{(s)} \to \mu^{\pm}e^{\mp}$, it can be shown analytically that each particle resolution contributes to the total resolution quadratically [173], namely:

$$\left(\frac{\sigma_M}{M}\right) \simeq \sqrt{\left(\frac{\sigma_1}{M_1}\right)^2 + \left(\frac{\sigma_2}{M_2}\right)^2},$$

with index $i = 1, 2$ indicating the two daughter leptons. From this, the relative ratio of the resolution of data and MC can be determined as,

$$\mathcal{C}_2 = \frac{\sqrt{\left(\frac{\sigma_{ee}}{M_{ee}}\right)^2_{\text{data}} + \left(\frac{\sigma_{\mu\mu}}{M_{\mu\mu}}\right)^2_{\text{data}}}}{\sqrt{\left(\frac{\sigma_{ee}}{M_{ee}}\right)^2_{\text{MC}} + \left(\frac{\sigma_{\mu\mu}}{M_{\mu\mu}}\right)^2_{\text{MC}}}}. \tag{C.0.1}$$

This two-body correction factor, $\mathcal{C}_2$, is not valid for 4-body decays such as $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$. Therefore, the value for $\sigma^{\text{data}}_{e\mu}$ is determined by performing an expansion around $\sigma^{\text{data}}_{\mu\mu}$ (or equivalently around $\sigma^{\text{data}}_{ee}$, which gives the same result), a method outlined in Ref. [174]. This gives

$$\sigma^{\text{data}}_{e\mu} = \sigma^{\text{data}}_{\mu\mu} + (\sigma^{\text{MC}}_{e\mu} - \sigma^{\text{MC}}_{\mu\mu}) \cdot \frac{\sigma^{\text{data}}_{ee} - \sigma^{\text{data}}_{\mu\mu}}{\sigma^{\text{MC}}_{ee} - \sigma^{\text{MC}}_{\mu\mu}}, \tag{C.0.2}$$

and then the 4-body $\mathcal{C}$ factor can be calculated as

$$\mathcal{C}_4 = \frac{\sigma^{\text{data}}_{e\mu}}{\sigma^{\text{MC}}_{e\mu}}. \tag{C.0.3}$$

Chapter 7.3 shows how MC simulation can incorrectly model certain aspects of the sample. This could include the resolution for the invariant mass such that $\sigma_{\text{MC}} \neq \sigma_{\text{data}}$. Therefore, as the signal PDF parameters are fixed from fitting this MC sample, the signal shape for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ could be incorrectly modelled. To account for this, the differences in the resolution between data and simulation can be estimated using a data/MC factor, $\mathcal{C}$. However, no SM cLFV decays have been observed, so there are no calibration channels with an electron and a muon in the final state that can be used to determine this $\mathcal{C}$ factor directly. The information on the resolution must instead be extracted from the combination of a muonic and an electronic decay channel to calculate the $\mathcal{C}$ factor from Equation C.0.3.

To study the potential data/MC difference and obtain the value of the $\mathcal{C}$ factor for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$, the $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ and $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ decays can be used. This is because of their similarity with the rare mode. Fits to these modes can be performed on MC and data to obtain their resolution parameters before applying Equation C.0.2 to determine the value of the required correction factors. In both cases the J/ψ mass constraint was not used as this is not present for the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ fits and could influence the signal resolutions. This led to some additional influence of the backgrounds, which meant the signal for these two calibration modes was harder to separate. The selection described in Chapter 7 was applied to each calibration mode, except for the J/ψ veto. This too was to give parity with the mass shape of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ that is being corrected. Both $\Lambda_b^0 \to \Lambda J/\psi\,(\to \ell^-\ell^+)$ fits use a DSCB fit shape to describe them.

For $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ the main background to be controlled was the misidentification background of $B^0 \to K_{\text{S}}^0 J/\psi\,(\to e^-e^+)$, where a pion is misidentified as a proton. Unlike the $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ equivalent, the poor resolution of the electron means that this background cannot be easily modelled and controlled via a separate background PDF. Therefore, a PID cut was developed to remove this component from data. This was optimised to reduce the background and did not consider the efficiency as only the signal peak resolution was required. The cut,

$$\text{ProbNNp} \cdot (1 - \text{ProbNNpi}) > 0.1,$$

was found to reduce the misidentified background to a negligible level such that the $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ fits were not strongly influenced by it.

Another factor that needed to be considered for the $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ fits was the different possible bremsstrahlung categories. Unlike in the $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ case where only a single electron is present to undergo bremsstrahlung there are three possibilities: no bremsstrahlung photons are recovered (0γ); one photon is recovered (1γ); or more than one photon is recovered (2γ). The $\mathcal{C}$ factor for the different bremsstrahlung categories for $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ are then determined slightly differently. The 0γ $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ resolutions are used with the $\Lambda_b^0 \to \Lambda J/\psi\,(\to \mu^-\mu^+)$ resolutions where the No Brem GBDT selection was applied, to represent the No Brem categories. Equivalently, 2γ $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ resolutions are used with the $\mu\mu$ mode resolutions with the Brem GBDT selection applied, to represent the Brem categories. The small sample size of $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ in data then means that a simultaneous fit is performed to the different bremsstrahlung categories rather than being performed independently. Figures C.1

and C.2 show the fits to the $\Lambda_b^0 \to \Lambda J/\psi (\to e^- e^+)$ simulated candidates in Run 1 and Run 2 to determine the MC resolutions and fix the fit shape except for the signal peak's mean and width. Figures C.3 and C.4 show the same fits but to the data to get the data resolution values.

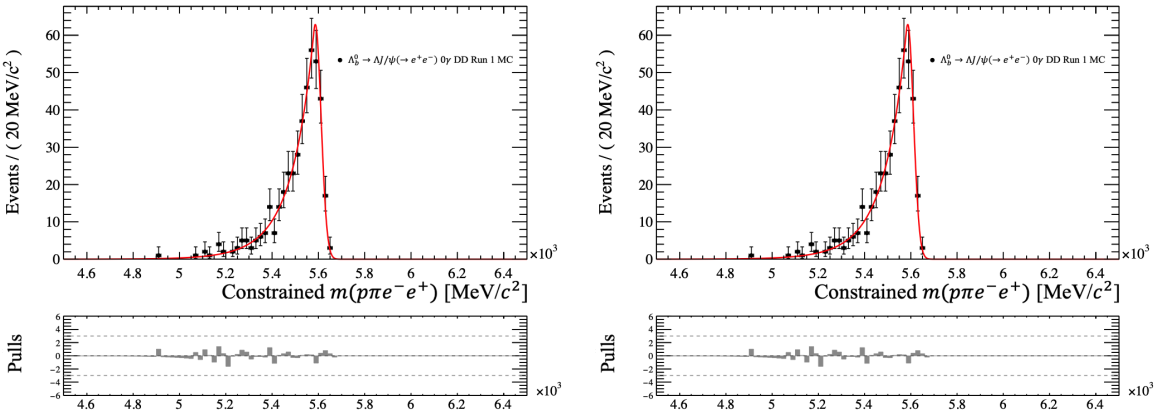
For the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ mode, the fits are similar to those shown in Chapter 8.2. However, the J/ψ mass constraint's removal means that the same PID cut on the proton and pion used in the $\Lambda_b^0 \to \Lambda J/\psi (\to e^- e^+)$ case had to be included to align with that fit. This was found to not significantly affect the peak resolution. Figure C.5 shows the fits to the $\Lambda_b^0 \to \Lambda J/\psi (\to \mu^- \mu^+)$ simulated candidates in Run 1 and Run 2 to determine the MC resolutions and fix the fit shape except for the signal peak's mean and width. Figure C.6 shows the same fits but to the data to get the data resolution values.

From these $\Lambda_b^0 \to \Lambda J/\psi (\to \ell^- \ell^+)$ fit results the peak width parameter, *i.e.* the resolution, can be extracted to calculate the $\mathcal{C}$ factors in each fit category using Equation C.0.3. Table C.1 summarises the results of these values, which showed large uncertainties. The main source of the uncertainty is the limited sample size for the $\Lambda_b^0 \to \Lambda J/\psi (\to e^- e^+)$ fits (both MC and data), particularly in Run 1. Therefore, as Run 1 and Run 2 are compatible and Run 2 has a larger sample size contributing to the fit, such that the fit resolution is likely to be better represented in the Run 2 case, the Run 2 values were proposed to be used for both runs.

However, these $\mathcal{C}$ factors were also all compatible with one, which would imply that no correction was necessary. Additionally, the estimated upper limits, as outlined in Chapter 9, were evaluated for the case where these $\mathcal{C}$ factors were included and also varied by up to 40%. The results were identical, which implies that the estimated upper limits are not sensitive to the inclusion of the $\mathcal{C}$ factor. Therefore, instead of introducing an additional free parameter that had a large uncertainty to the signal fit, it was decided to fix the signal peak resolution to the value obtained from fitting the MC samples.

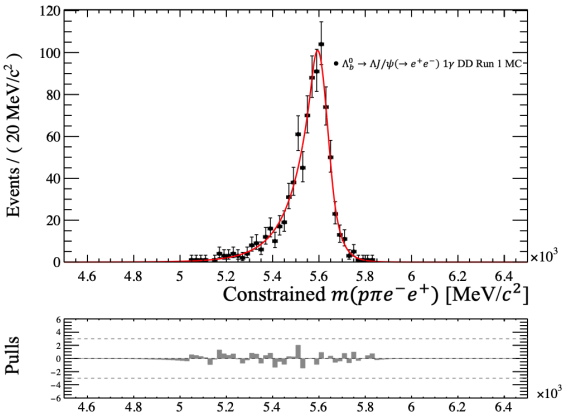
Table C.1: Results for the evaluation of the 4-body $\mathcal{C}$ factor in different categories.

$\mathcal{C}$ factor	DD Run 1	LL Run 1	DD Run 2	LL Run 2
0γ $\mathcal{C}_4$	1.213 ± 0.152	0.856 ± 0.096	1.118 ± 0.055	1.043 ± 0.058
1γ $\mathcal{C}_4$	1.157 ± 0.274	1.448 ± 0.457	1.275 ± 0.124	1.149 ± 0.135

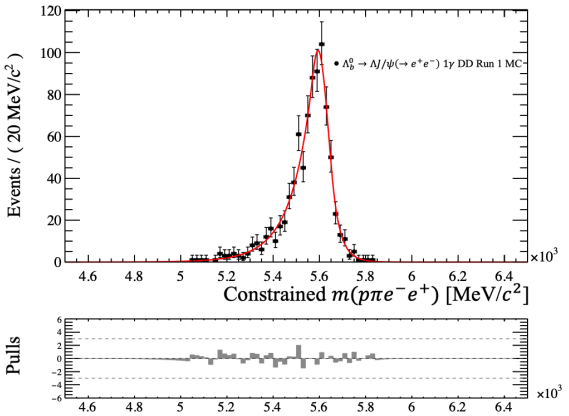


(a) 0γ-DD

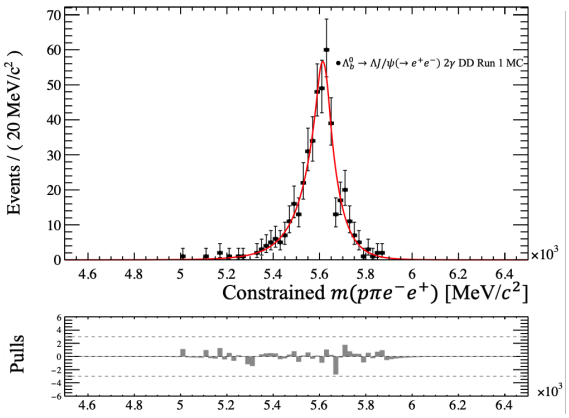
(b) 0γ-LL



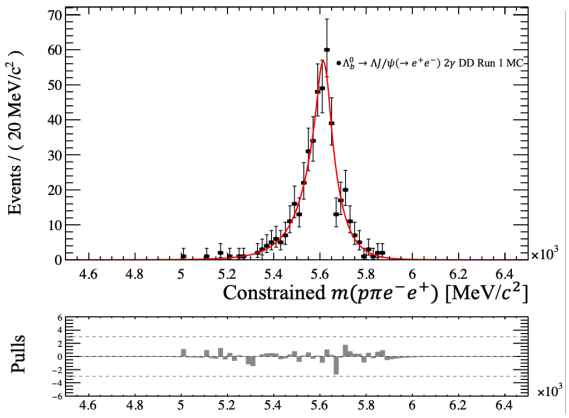
(c) 1γ-DD



(d) 1γ-LL

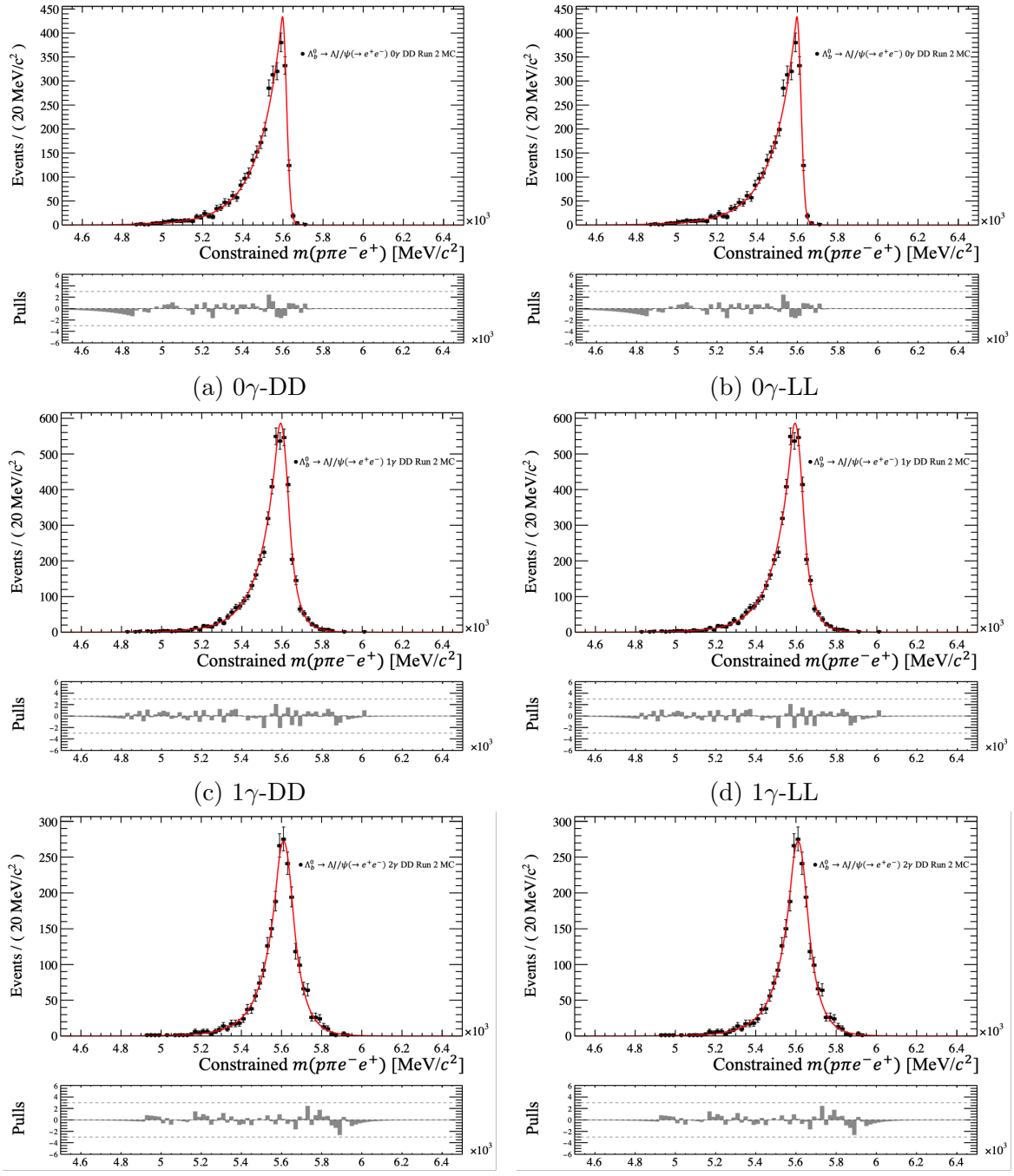


(e) 2γ-DD



(f) 2γ-LL

Figure C.1: Fits to the invariant mass for $\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$ simulated candidates for mass calibration in Run 1 with DD (LL) tracks on the left (right).

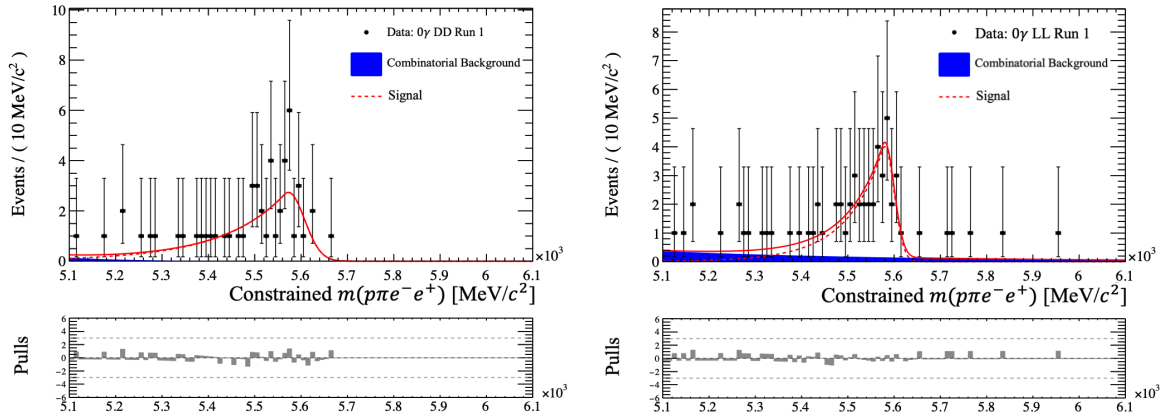


(a) 0γ-DD (b) 0γ-LL

(c) 1γ-DD (d) 1γ-LL

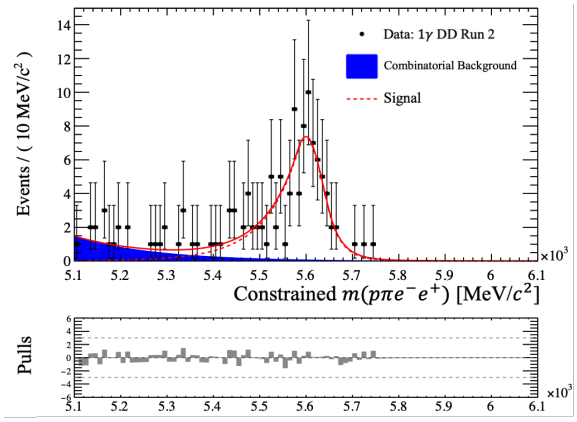
(e) 2γ-DD (f) 2γ-LL

Figure C.2: Fits to the invariant mass for $\Lambda_b^0 \to \Lambda J/\psi\,(\to e^-e^+)$ simulated candidates for mass calibration in Run 2 with DD (LL) tracks on the left (right).

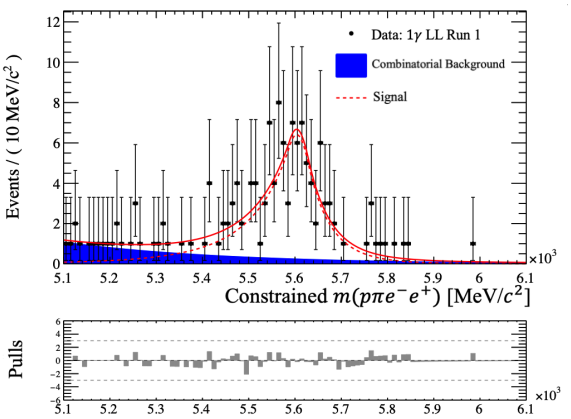


(a) 0γ-DD

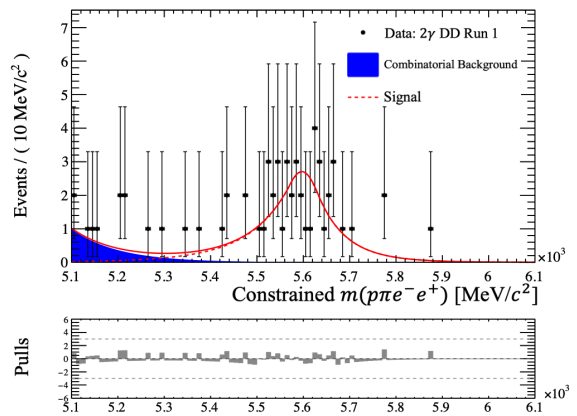
(b) 0γ-LL



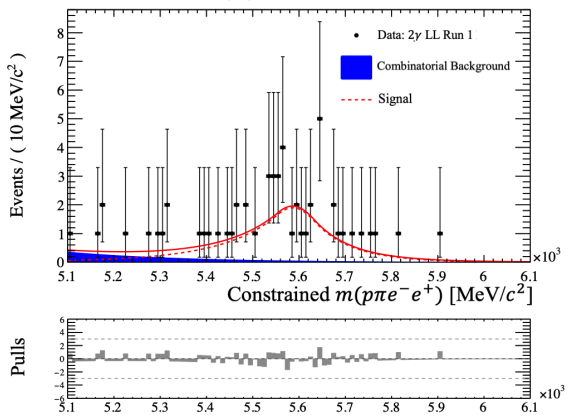
(c) 1γ-DD



(d) 1γ-LL

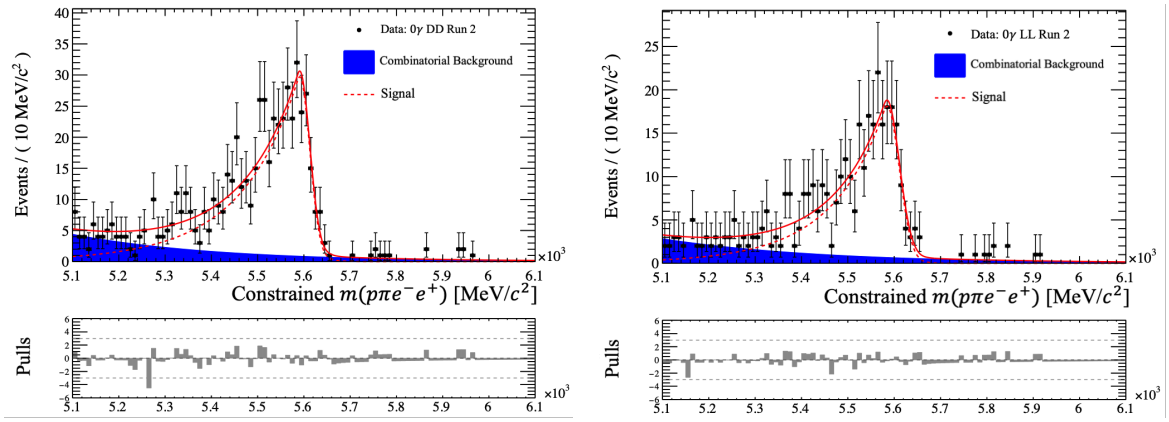


(e) 2γ-DD



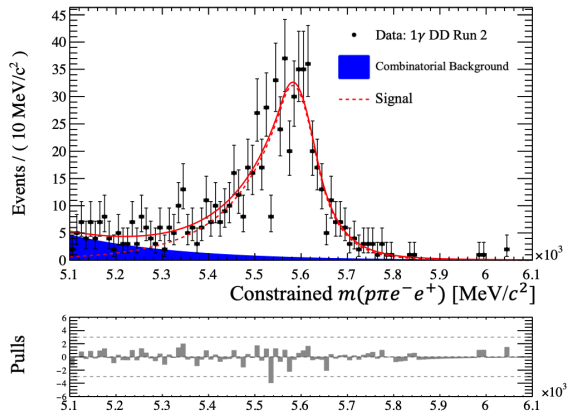
(f) 2γ-LL

Figure C.3: Fits to the invariant mass for $\Lambda_b^0 \to \Lambda J/\psi(\to e^-e^+)$ data candidates for mass calibration in Run 1 with DD (LL) on the left (right).

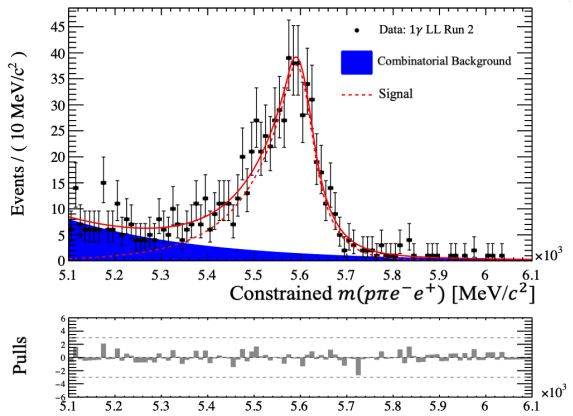


(a) 0γ-DD

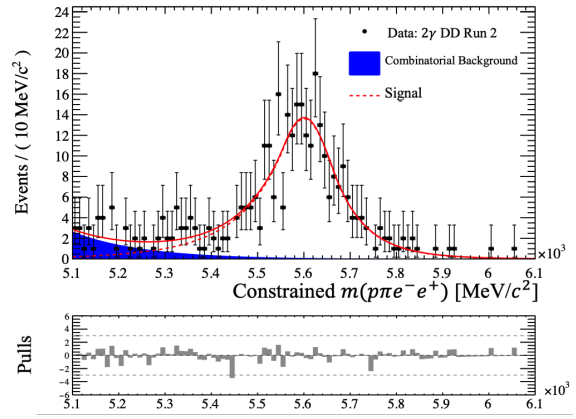
(b) 0γ-LL



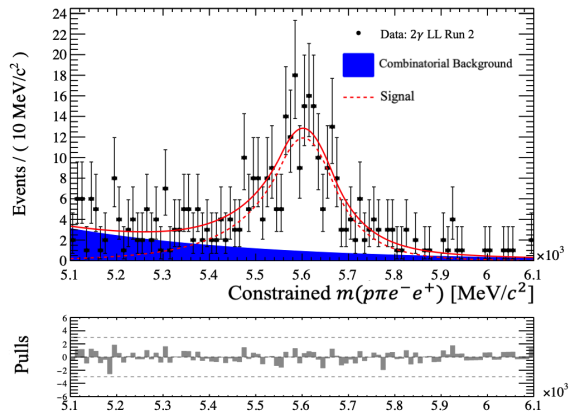
(c) 1γ-DD



(d) 1γ-LL

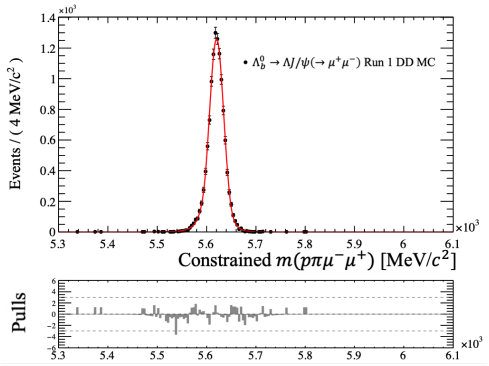


(e) 2γ-DD

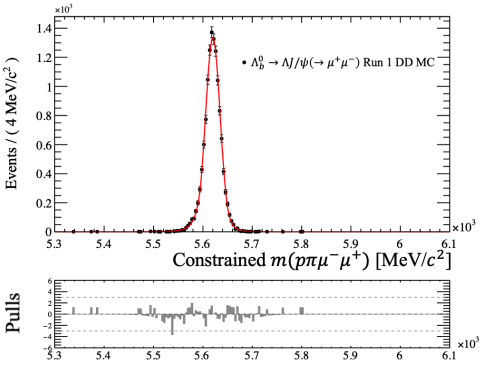


(f) 2γ-LL

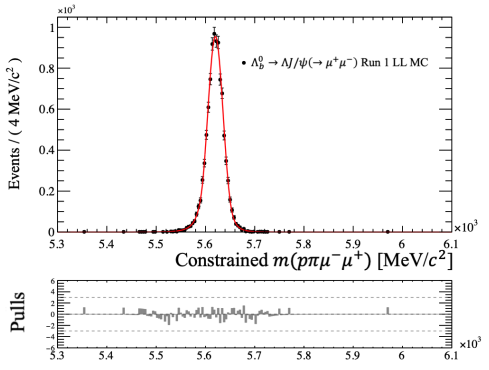
Figure C.4: Fits to the invariant mass for $\Lambda_b^0 \to \Lambda J/\psi (\to e^- e^+)$ data candidates for mass calibration in Run 2 with DD (LL) on the left (right).



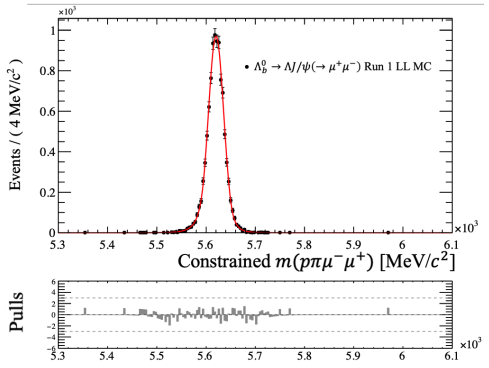
(a) 0γ GBDT selection-DD



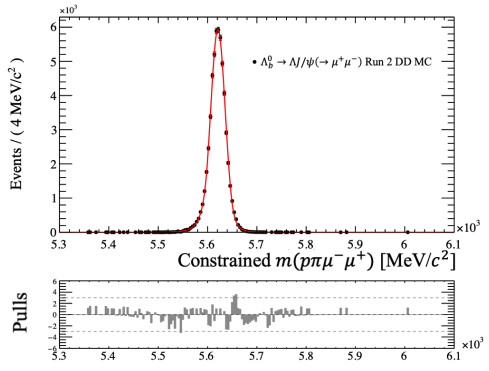
(b) 1γ GBDT selection-DD



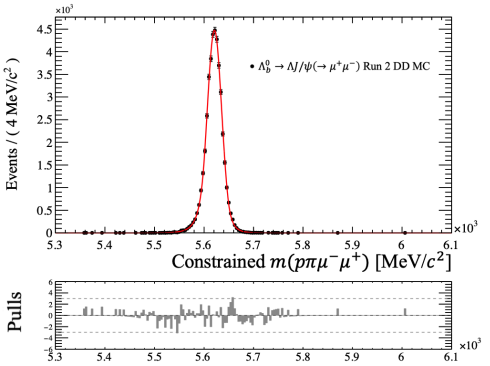
(c) 0γ GBDT selection-LL



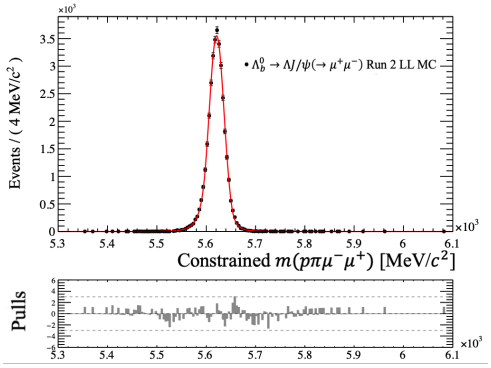
(d) 1γ GBDT selection-LL



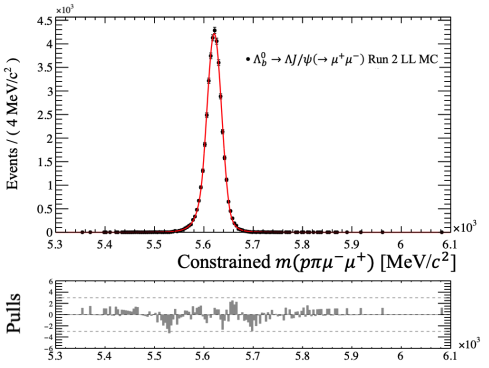
(e) 0γ GBDT selection-DD



(f) 1γ GBDT selection-DD

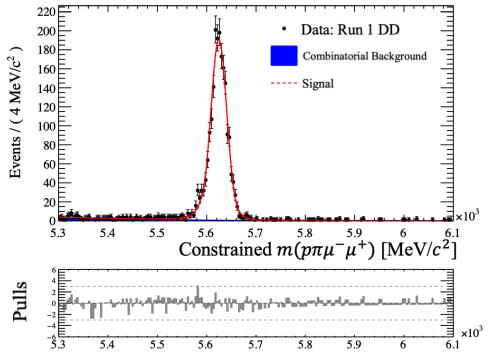


(g) 0γ GBDT selection-LL

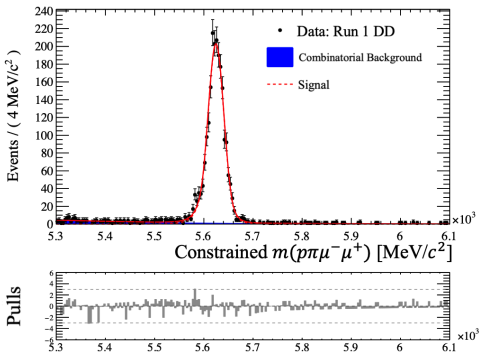


(h) 1γ GBDT selection-LL

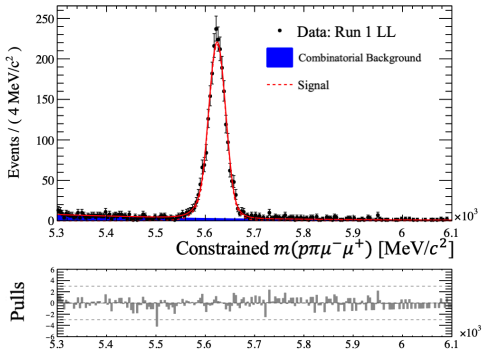
Figure C.5: Fits to the invariant mass for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ simulated candidates for mass calibration in Run 1 and Run 2.



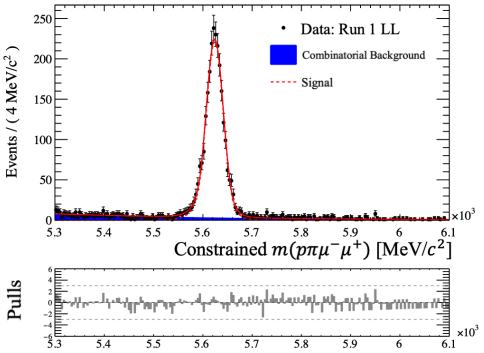
(a) 0γ GBDT selection-DD



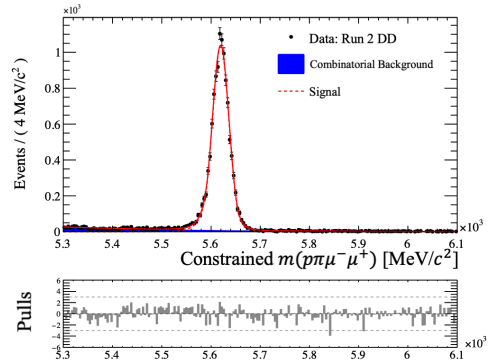
(b) 1γ GBDT selection-DD



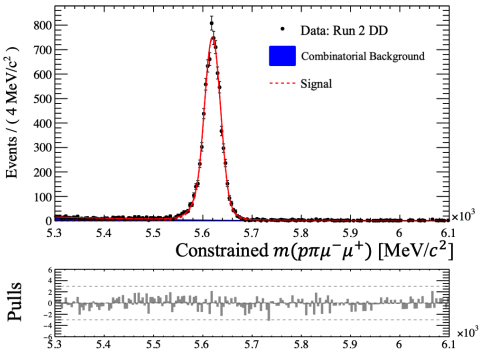
(c) 0γ GBDT selection-LL



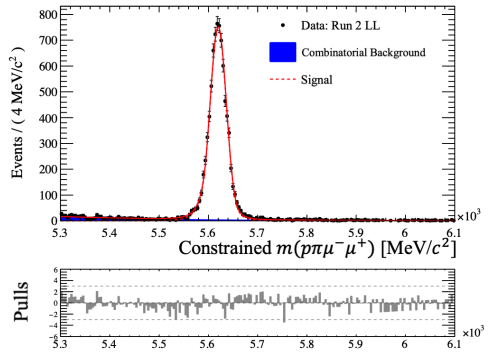
(d) 1γ GBDT selection-LL



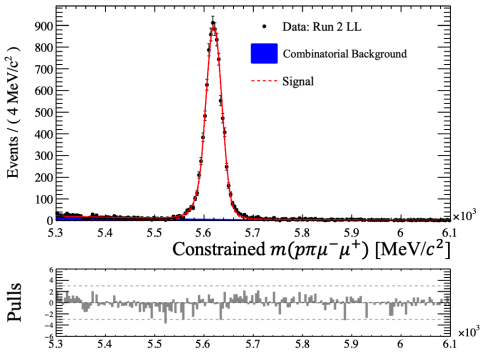
(e) 0γ GBDT selection-DD



(f) 1γ GBDT selection-DD



(g) 0γ GBDT selection-LL



(h) 1γ GBDT selection-LL

Figure C.6: Fits to the invariant mass for $\Lambda_b^0 \to \Lambda J/\psi(\to \mu^-\mu^+)$ data candidates for mass calibration in Run 1 and Run 2.

Appendix D

Estimated Limits Using the Profile Likelihood Method

As described in Chapter 9.3, in the GammaCombo implementation of the CL_s method there are two main ways of obtaining an upper limit: with the profile likelihood (Prob) method or through a pseudo-experiment based (Plugin) method. Of these two methods, the Plugin case was the nominal result quoted throughout Chapter 9.4. This was selected because it, by construction, has better statistical coverage than the profile likelihood method and does not make any assumptions about the underlying distribution of the test statistic [169]. However, there were some additional limitations to the profile likelihood method that arose due to the fact that a BG-only proxy dataset was used to derive the estimated upper limits.

It was observed that in the standard fit model without the Λ_c^+ veto, if the seed used to generate the BG-only proxy dataset was varied then the likelihood scan could vary dramatically in shape with upper limits that could range from $[0.5\text{–}3] \times 10^{-7}$. This variation can be explained by the large uncertainties within the exclusive background PDFs for this fit model, leading to a large variation of the BG-only proxy dataset. This is because it is generated from the background-only PDF, which includes the exclusive background PDFs as a major feature. This variation is mostly present when considering $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ as the parameter of interest, and is more suppressed for the case involving r as it removes the source of uncertainty due to $\mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$. However, this also means that when considering the compatibility cross-check of the results for $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and $r \cdot \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ there could be quite strong inconsistencies. For the results shown in Chapter 9.4 and in this appendix the same seed was used to generate the BG-only proxy dataset to allow easier cross-checking in this way. But, there are two different datasets being evaluated by GammaCombo as they are each generated for the background PDFs for either $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ or r as the parameter of interest, which are quite different by design. Therefore, if, for the particular seed chosen for examination, one of the datasets has fluctuated significantly more so than the other then inconsistencies can arise, which is likely as the r case appears to be more stable. It should also be noted that this is a limitation inherent both to this particular fit model without the Λ_c^+ veto due to its relatively large uncertainties, and to the estimated limit derivation generally because

the unblinded dataset, in reality, will be identical for both parameters of interest and decoupled from the fit model. Thus this variation would not affect the upper limits obtained or the compatibility of the results from the two parameters of interest.

Additionally, the limitations described in the previous paragraph do not affect the results of the pseudo-experiment method as the variations are explicitly handled within its procedure as pseudo-experiments are generated based on the combined signal and background, and background-only PDFs with different seeds. The likelihood scan only provides starting values within these. The ensemble of the toys as a whole is then considered, with the potential variation informing the upper limit and the error bands on both the observed and expected results. Therefore, for clarity, only the Plugin results were presented in Chapter 9.4. However, the Prob scan results still provide an approximate upper limit that should give an order-of-magnitude result similar to the Plugin case and can give insight into whether there is a problem within the model before evaluating a large set of pseudo-experiments.

Figure D.1 shows the profile likelihood scan results compared to the results from pseudo-experiments for both the considered parameters of interest, *i.e.* $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and r. This demonstrates that although the $\mathrm{CL_s}$ upper limit from the likelihood scan is of the same order of magnitude as the Plugin results, the limits can differ significantly. The $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ upper limit is ~ 1.4 times larger than the Plugin equivalent, but for r the limit is approximately half that of the Plugin version. This also leads to a significant incompatibility between $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and $r \cdot \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ of 15.8σ, implying an inconsistency between these two cases. This is likely due to a strong over-fluctuation in the case of $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$, which in previous minor iterations of the fit model had a profile likelihood scan shape more similar to r and a limit $\mathcal{O}(1 \times 10^{-7})$.

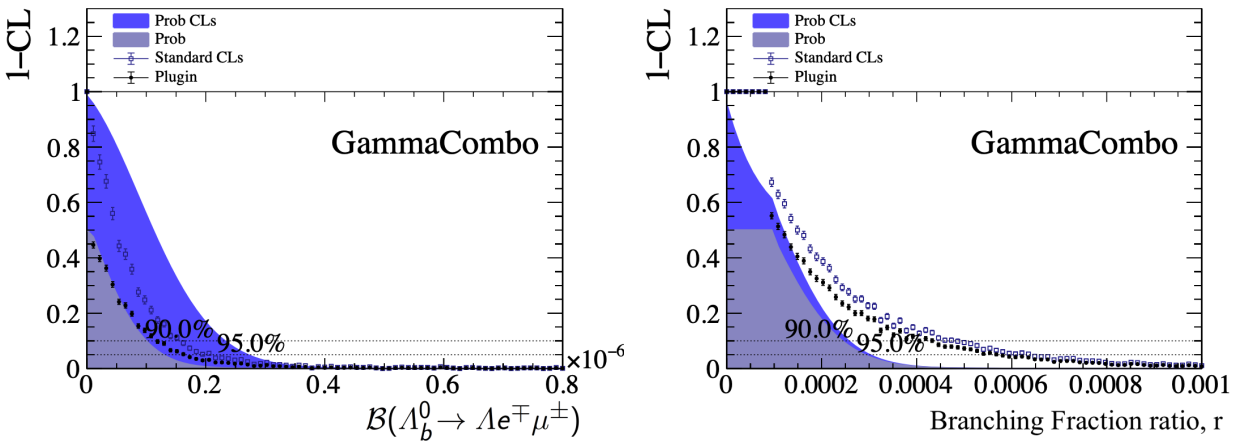


Figure D.1: $\mathrm{CL_s}$ limit profile likelihood scan with the results of 2000 $\mathrm{CL_s}$ pseudo-experiment runs overlaid. For the parameters of interest (Left) $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and (Right) the ratio, r, of the branching fraction of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ over the branching fraction of $\Lambda_b^0 \to \Lambda J/\psi$.

Chapter 9.6 describes the effect of an additional Λ_c^+ veto included after all other selections. One of its major impacts was to eliminate the exclusive Λ_c^+ backgrounds considered in the standard analysis procedure, and thus simplify the background fit model to an exponential to describe the combinatorial background. This also removes the large uncertainty of the fit model's shape, although the statistical uncertainty of the dataset itself is higher. As a result, the inconsistencies that can arise between the parameters of interest appear to be strongly suppressed, with Figure D.2 showing very similar shapes for both parameters of interest. The compatibility between $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and $r \cdot \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ is 0.04σ, and so does not appear to show the same problematic behaviour as the case without the Λ_c^+ veto included. This provides an additional reason to adopt the Λ_c^+ veto fully throughout the analysis.

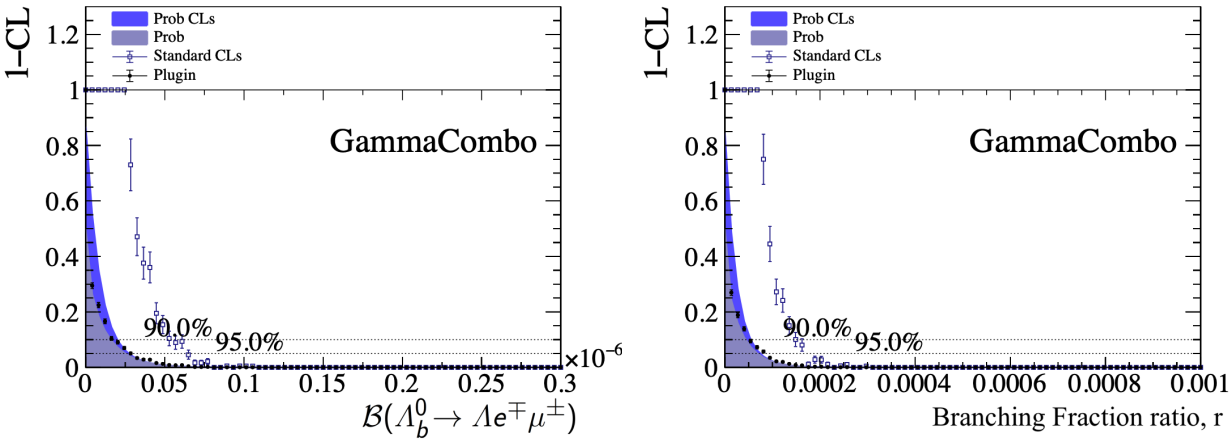


Figure D.2: CL_s limit profile likelihood scan with the results of 2000 CL_s pseudo-experiment runs overlaid with the Λ_c^+ veto included. For the parameters of interest (Left) $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and (Right) the ratio, r, of the branching fraction of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ over the branching fraction of $\Lambda_b^0 \to \Lambda J/\psi$.

Table D.1 shows a summary of the upper limits obtained with the profile likelihood method for each of the parameters of interest, both with and without the inclusion of the Λ_c^+ veto.

Table D.1: The estimated upper limits at a 95% confidence level obtained to the signal mode branching fraction with the profile likelihood method with and without a Λ_c^+ veto applied.

Parameter of Interest	**Profile Likelihood CL_s Method**	
(@ 95% CL)	No Λ_c^+ veto	With Λ_c^+ veto
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$	$\lesssim 2.70 \times 10^{-7}$	$\lesssim 2.73 \times 10^{-8}$
r	$\lesssim 3.01 \times 10^{-4}$	$\lesssim 7.49 \times 10^{-5}$
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	$\lesssim (1.09 \pm 0.10) \times 10^{-7}$	$\lesssim (2.72 \pm 0.25) \times 10^{-8}$

In Chapter 9.5.2 an approximate total uncertainty on the upper limit is derived to highlight the potential variation from the large uncertainties within the fit model. Although, this has its own limitations due to the fixed nature of the parameters to their initial central values, as discussed in Chapter 9.5.2. Figures D.3 and D.4 show the profile likelihood scan plots for the case without and with the additional Λ_c^+ veto included respectively for both considered parameters of interest, with the associated results shown in Table D.2. These show similar results to those in Figures D.1 and D.2 and Table D.1. Except for the $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ plot without the Λ_c^+ veto, which shows a profile likelihood CL_s shape that is not as strongly over-estimated compared to the Plugin case and more comparable to the equivalent shape in r. However, the compatibility cross-check has $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and $r \cdot \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$ differ by 7.26σ and 0.02σ for the upper limit without and with the Λ_c^+ veto, respectively. This highlights that alongside the large uncertainties within the fit model, the initial central values of the exclusive background PDFs are probably not fully representative of the data and require some freedom within the fit to accurately model those backgrounds. The associated approximate total relative uncertainties on the limits are given in Table D.3, and show a similar trend to the pseudo-experiment method results where the Λ_c^+ veto relative uncertainty is much smaller.

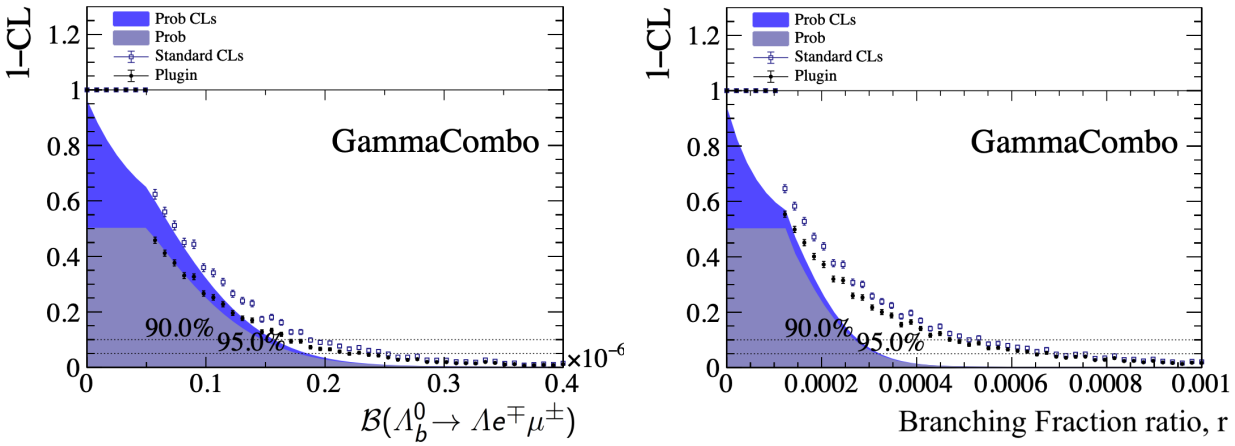


Figure D.3: CL_s limit profile likelihood scan with constrained variables fixed to their nominal value and the results of 2000 CL_s pseudo-experiment runs overlaid. For the parameters of interest (Left) $\mathcal{B}(\Lambda_b^0 \to \Lambda e^\mp \mu^\pm)$ and (Right) the ratio, r, of the branching fraction of $\Lambda_b^0 \to \Lambda e^\mp \mu^\pm$ over the branching fraction of $\Lambda_b^0 \to \Lambda J/\psi$.

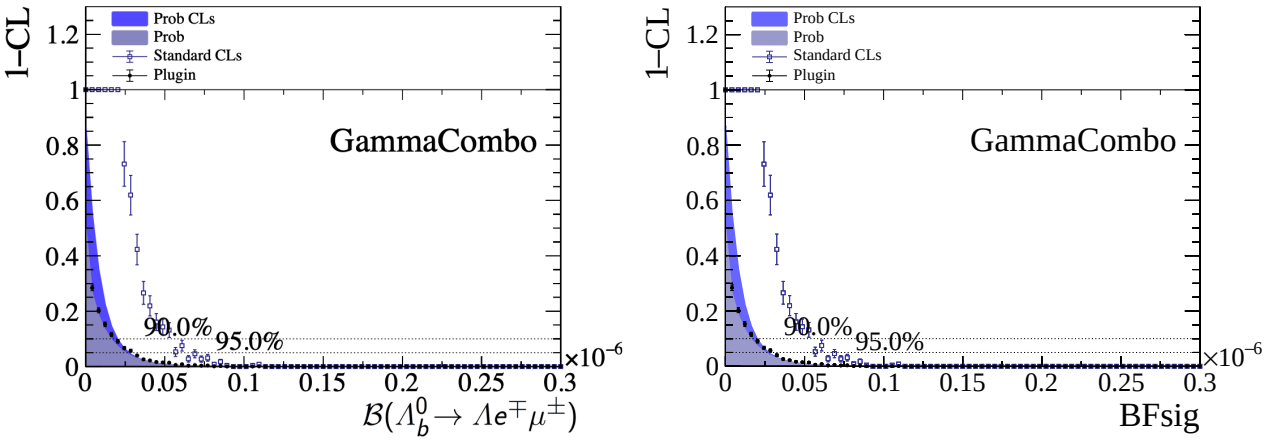


Figure D.4: CL_s limit profile likelihood scan with constrained variables fixed to their nominal value, the results of 2000 CL_s pseudo-experiment runs overlaid, and the Λ_c^+ veto applied. For the parameters of interest (Left) $\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$ and (Right) the ratio, r, of the branching fraction of $\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm}$ over the branching fraction of $\Lambda_b^0 \to \Lambda J/\psi$.

Table D.2: The estimated upper limits at a 95% confidence level obtained to the signal mode branching fraction with the profile likelihood method with and without a Λ_c^+ veto applied and the constrained variables in the fit model fixed to their nominal value.

Parameter of Interest	**Profile Likelihood CL_s Method**	
(@ 95% CL)	No Λ_c^+ veto	With Λ_c^+ veto
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$	$\lesssim 1.83 \times 10^{-7}$	$\lesssim 2.73 \times 10^{-8}$
r	$\lesssim 3.0 \times 10^{-4}$	$\lesssim 7.51 \times 10^{-5}$
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	$\lesssim (1.09 \pm 0.10) \times 10^{-7}$	$\lesssim (2.73 \pm 0.25) \times 10^{-8}$

Table D.3: The uncertainties derived for the estimated upper limits at a 95% confidence level obtained to the signal mode branching fraction with the profile likelihood method, with and without the inclusion of the Λ_c^+ veto.

Parameter of Interest	**Profile Likelihood CL_s Method**	
(@ 95% CL)	No Λ_c^+ veto	With Λ_c^+ veto
$\mathcal{B}(\Lambda_b^0 \to \Lambda e^{\mp}\mu^{\pm})$	32.2%	0.0%
r	6.3%	0.3%
$r \times \mathcal{B}(\Lambda_b^0 \to \Lambda J/\psi)$	10.3%	0.4%

Appendix E

Additional Λ_c^+ Veto Selection Plots

This appendix shows supplementary plots for all of the considered analysis categories for various mass variables, *e.g.* $m(p\pi e\mu)$, $m(p\pi\ell)$, for different data and simulated samples. The Brem-DD-Run 2 category for these plots are discussed in Section 9.6.

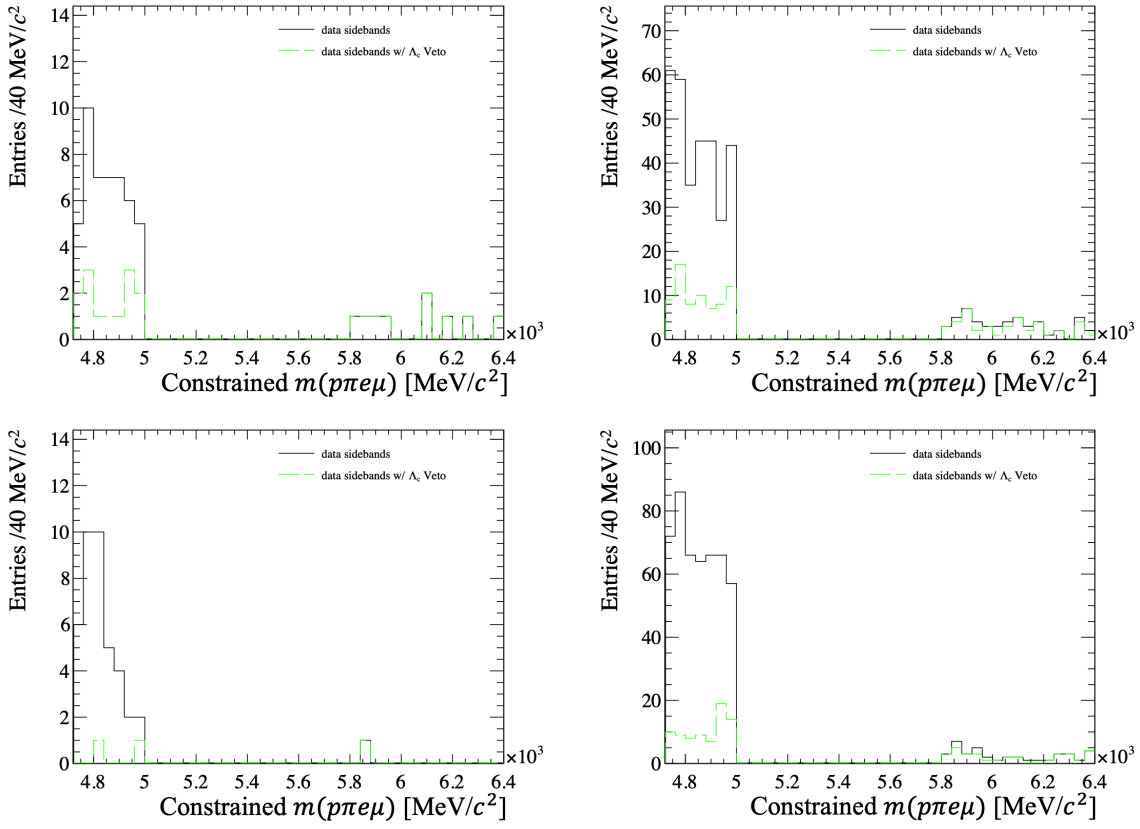


Figure E.1: Distribution of the constrained Λ_b^0 mass for the data sidebands with and without the Λ_c^+ veto applied for categories with bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

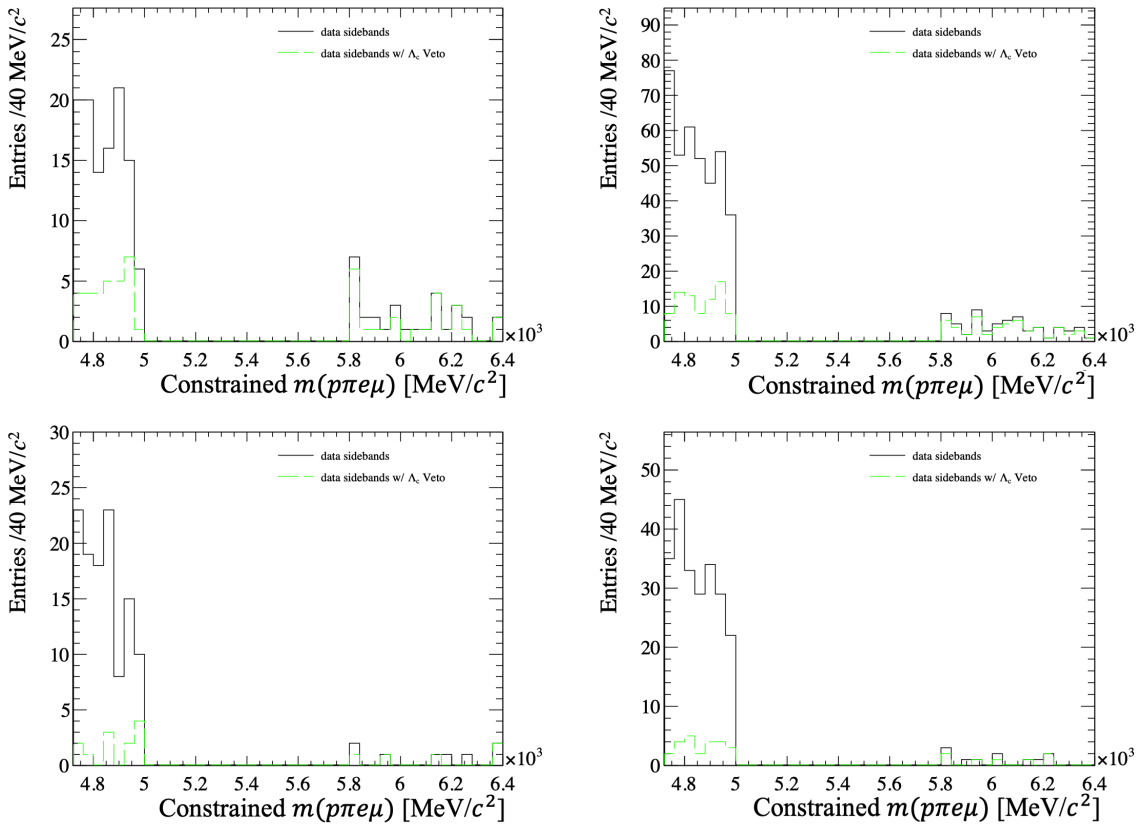


Figure E.2: Distribution of the constrained Λ_b^0 mass for the data sidebands with and without the Λ_c^+ veto applied for categories without bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

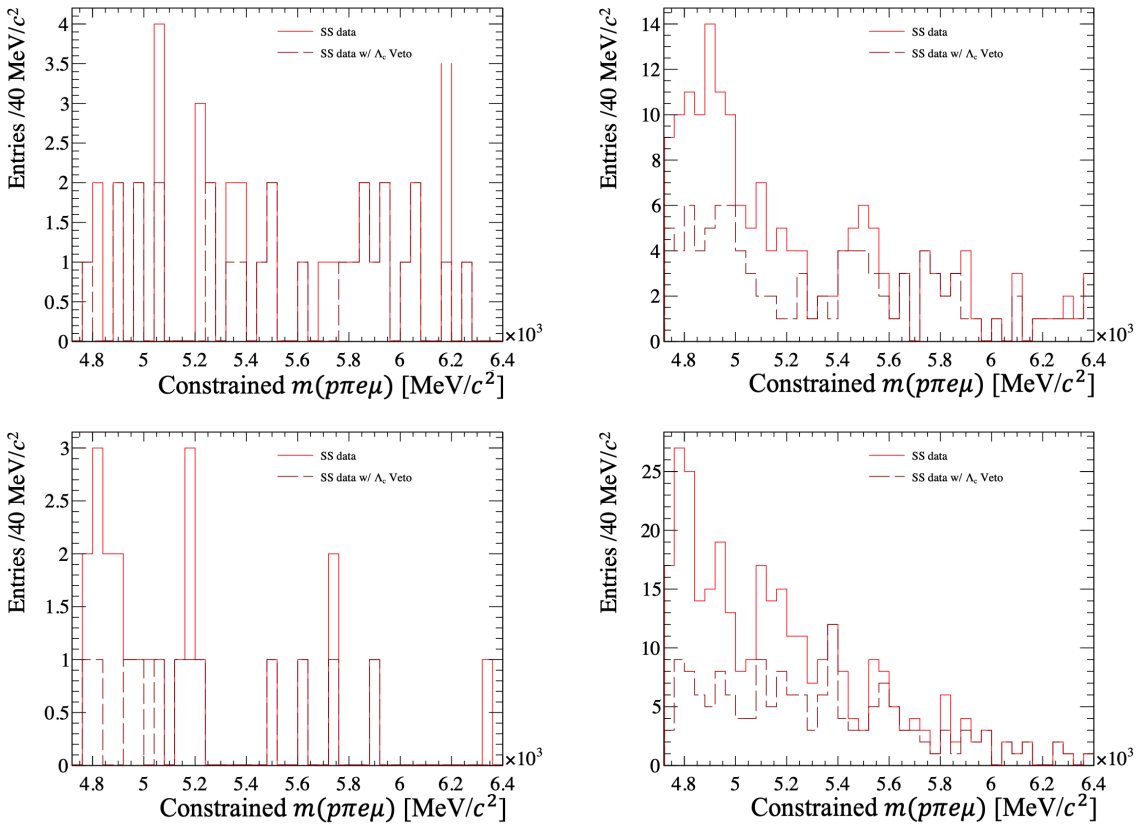


Figure E.3: Distribution of the constrained Λ_b^0 mass for same-sign data with and without the Λ_c^+ veto applied for categories with bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

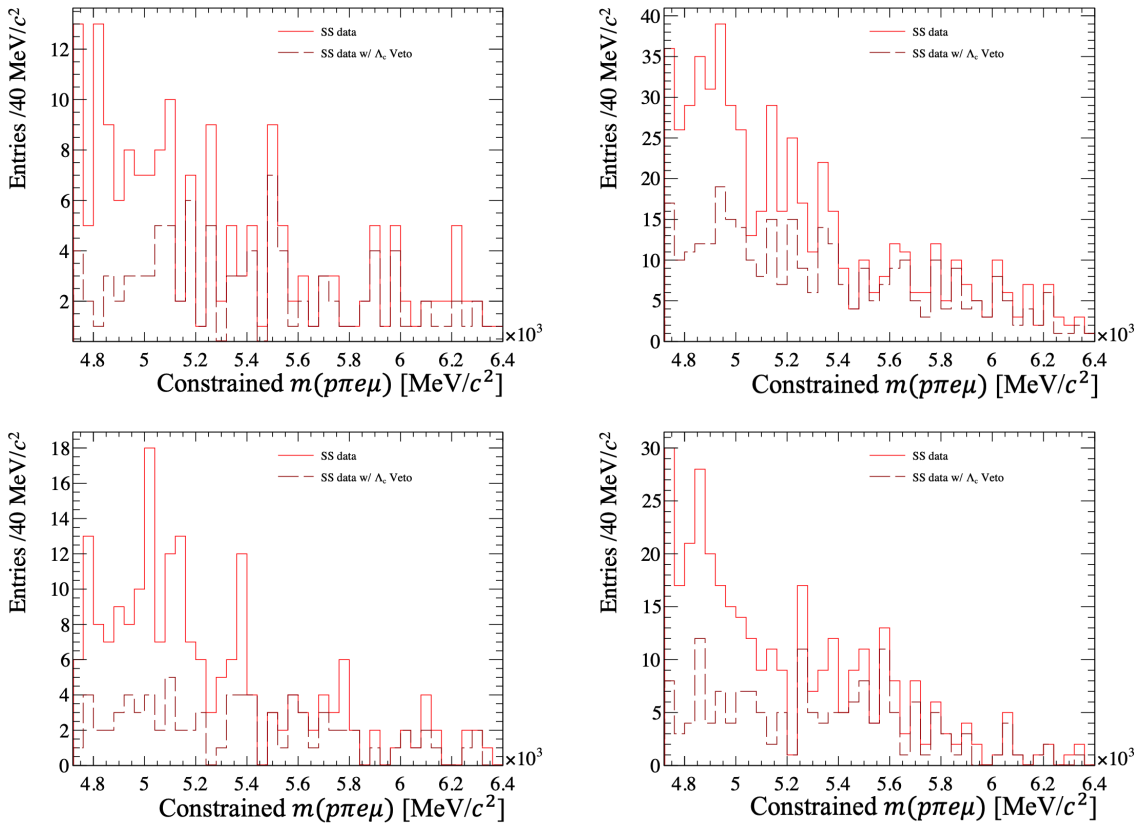


Figure E.4: Distribution of the constrained Λ_b^0 mass for same-sign data with and without the Λ_c^+ veto applied for categories without bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

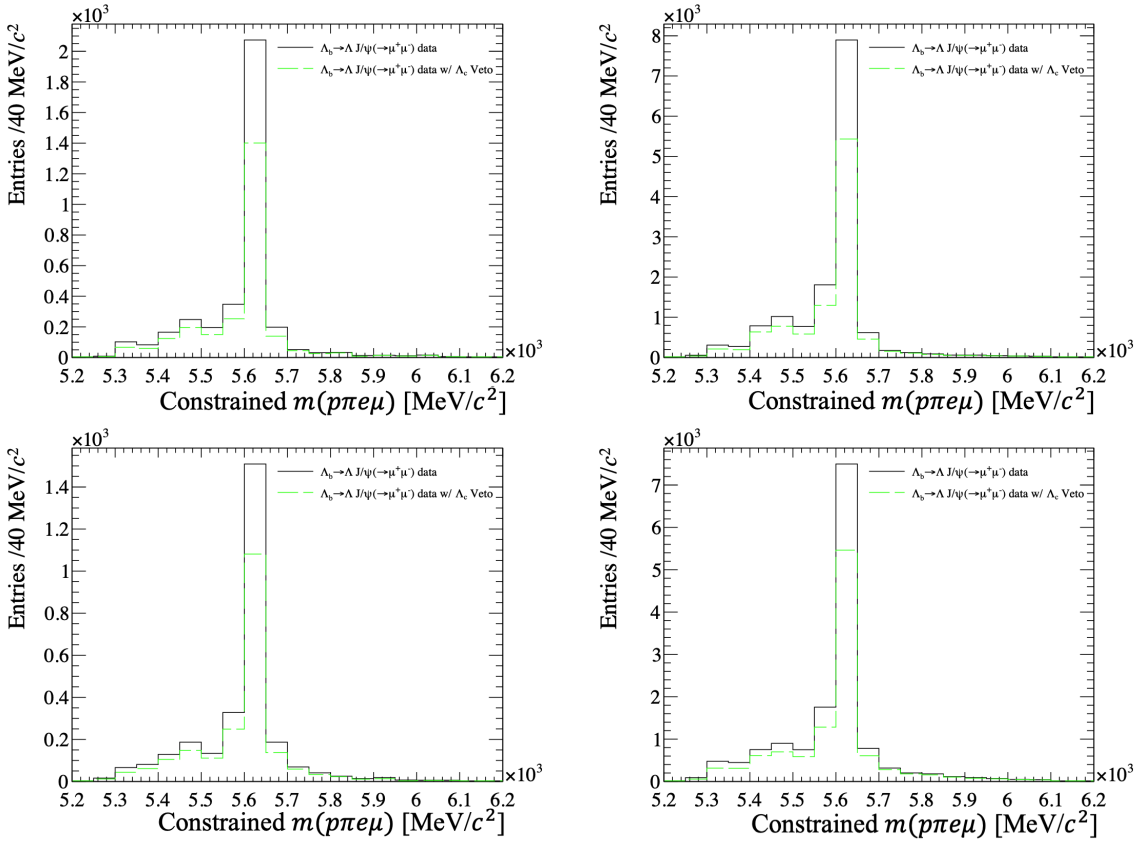


Figure E.5: Distribution of the constrained Λ_b^0 mass for the data used for the normalisation mode fits with and without the Λ_c^+ veto applied for Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

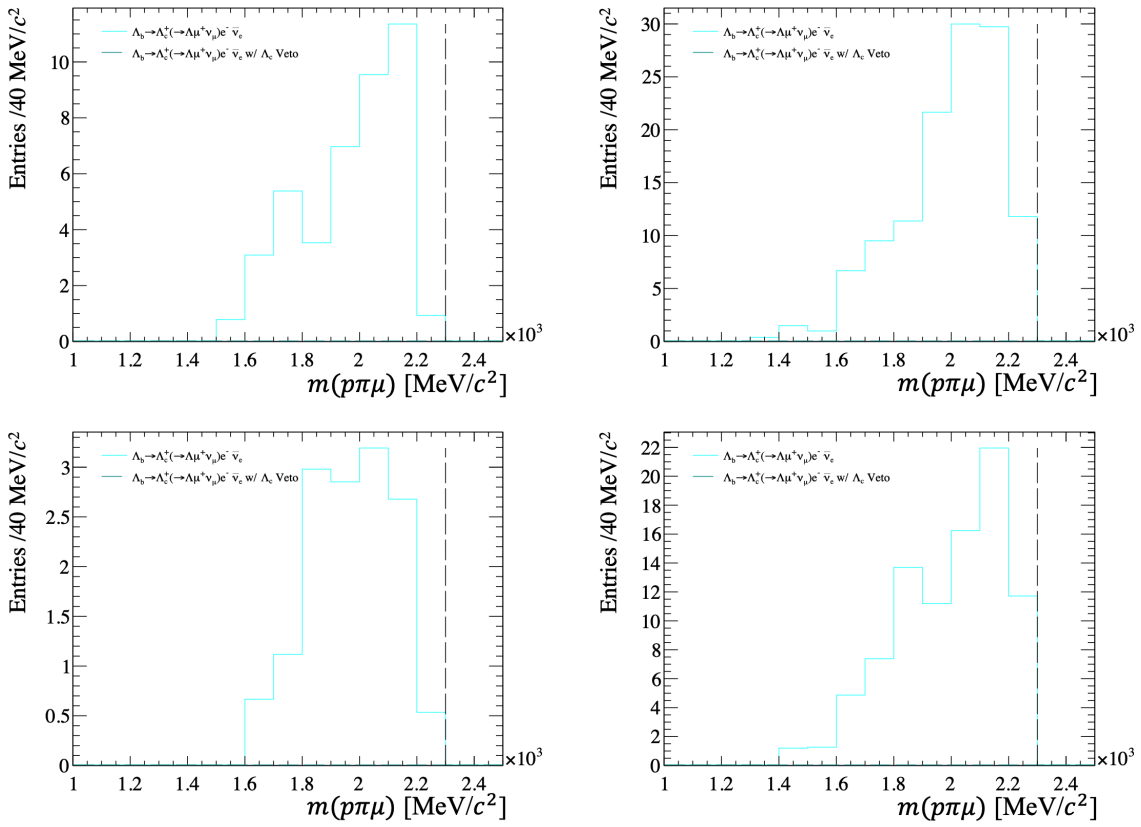


Figure E.6: Distribution of $m(p\pi\mu)$ for the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ MC with and without the Λ_c^+ veto applied for categories with bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

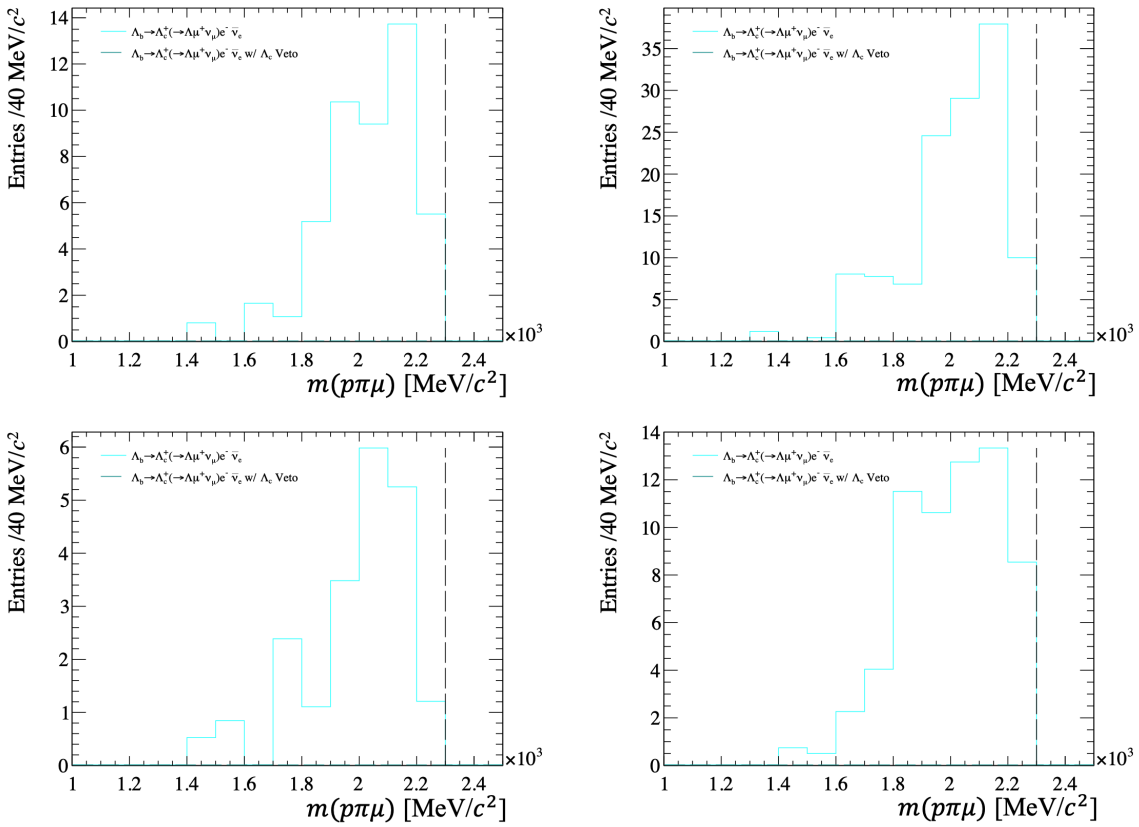


Figure E.7: Distribution of $m(p\pi\mu)$ for the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda\mu^+\nu_\mu)e^-\overline{\nu}_e$ MC with and without the Λ_c^+ veto applied for categories without bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

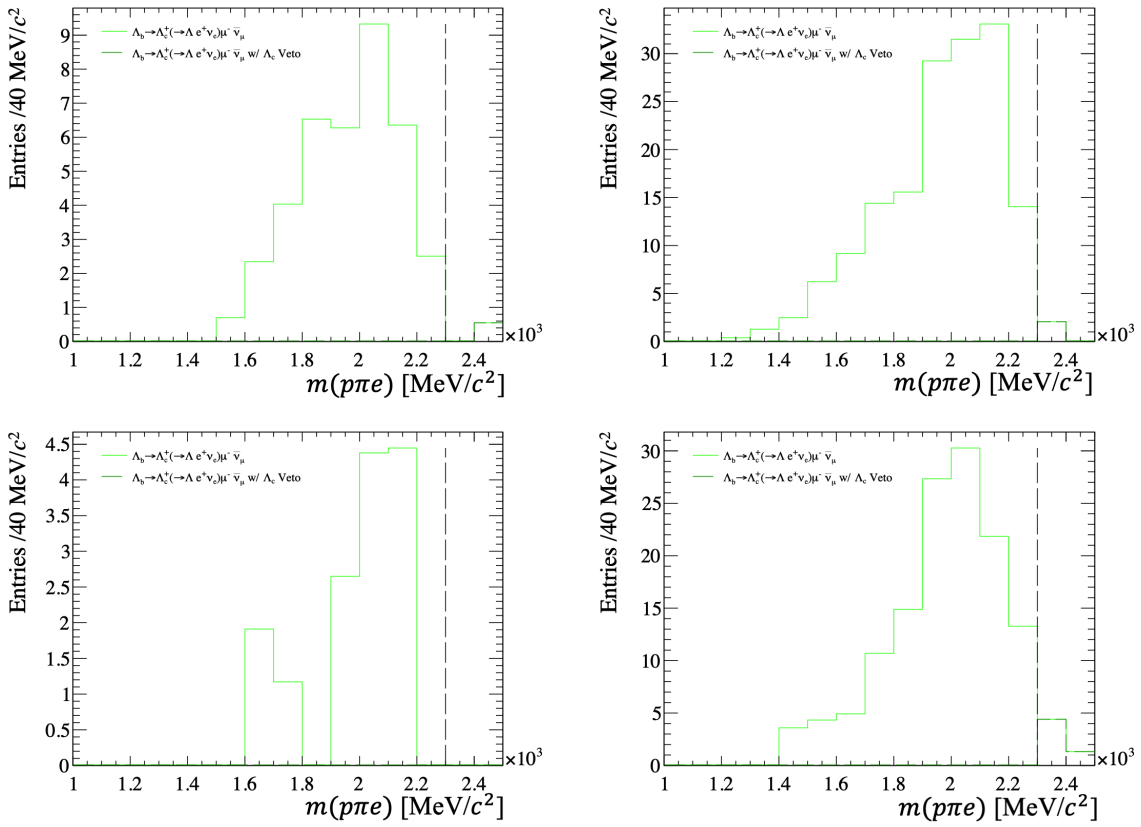


Figure E.8: Distribution of $m(p\pi e)$ for the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ MC with and without the Λ_c^+ veto applied for categories with bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).

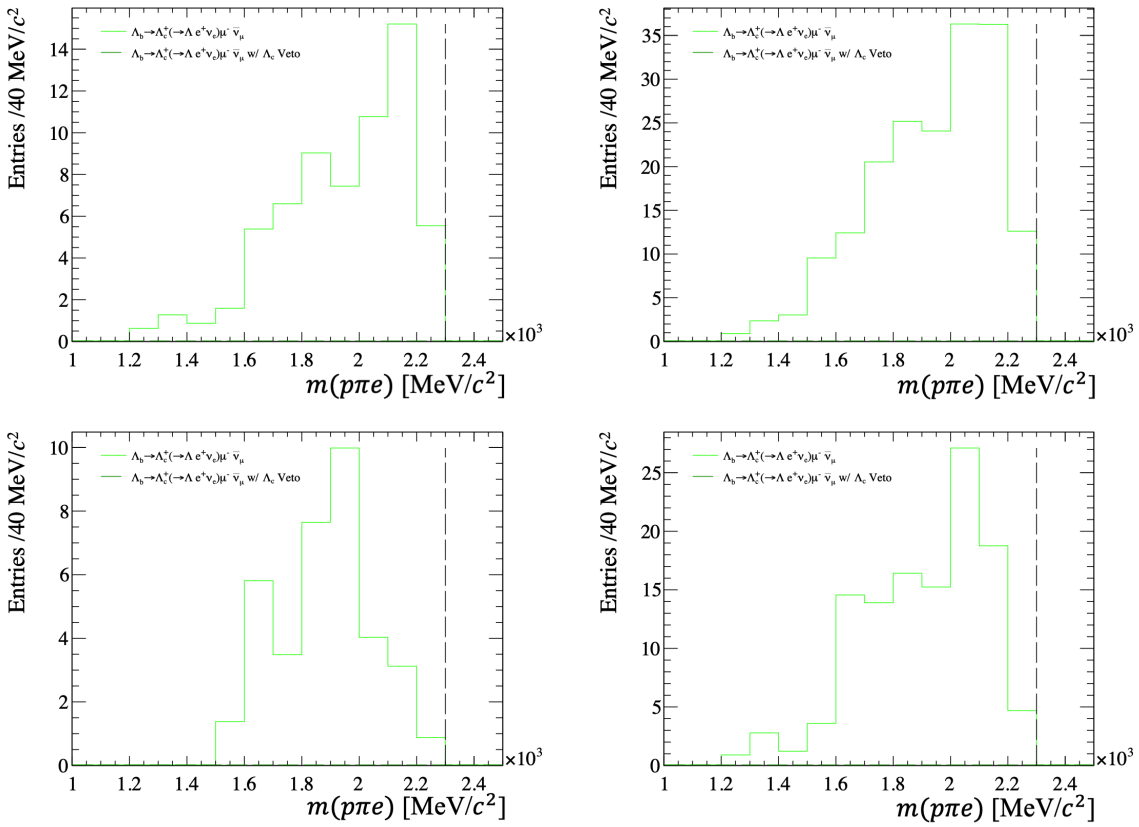


Figure E.9: Distribution of $m(p\pi e)$ for the $\Lambda_b^0 \to \Lambda_c^+(\to \Lambda e^+\nu_e)\mu^-\overline{\nu}_\mu$ MC with and without the Λ_c^+ veto applied for categories without bremsstrahlung added and Run 1 (Run 2) plots on the left (right) and DD (LL) on the top (bottom).