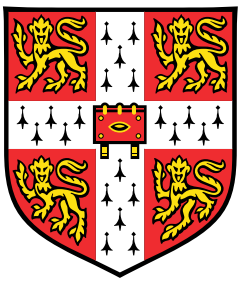
Tests of lepton universality in $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays at the LHCb experiment



John Gordon Smeaton

Supervisor: Prof. Valerie Gibson

Department of Physics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

Clare Hall

March 2022

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

John Gordon Smeaton
March 2022

Abstract

Tests of lepton universality in $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays at the LHCb experiment

John Gordon Smeaton

The decays $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$- where $\ell^+ \ell^-$ represents an electron-positron or muon-antimuon pair - are examples of $b \to s \ell^+ \ell^-$ processes. Numerous analyses of such decays have found tensions with predictions under the Standard Model (SM) of particle physics. These tensions form a coherent pattern, which could be caused by a novel physical effect.

Ratios of branching fractions for muon-mode and electron-mode $b \to s \ell^+ \ell^-$ processes can be predicted with $\mathcal{O}(1\%)$ precision under the SM, and can provide stringent experimental tests for novel physical processes which violate lepton universality. This dissertation presents new measurements of such ratios using $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays, denoted $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$. These measurements used data from pp collisions collected using the LHCb detector between 2011 and 2018, with a total sample size of 9 fb^{-1}. The results are:

$$R^{-1}_{K^0_S} = 1.51\,^{+0.40}_{-0.35}\,(\text{stat.})\,^{+0.09}_{-0.04}\,(\text{syst.})\,, \quad R^{-1}_{K^{*+}} = 1.44\,^{+0.32}_{-0.29}\,(\text{stat.})\,^{+0.09}_{-0.06}\,(\text{syst.})$$

Here, the first uncertainties are statistical, and the second systematic. Both $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ are consistent with SM expectations, at significances of 1.5σ and 1.4σ, respectively. In addition, differential branching fractions for the decays $B^0 \to K^0 e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ are measured in bins of the dilepton invariant mass squared. The first ever observations of these decays are reported.

Acknowledgements

The work outlined in this dissertation would not have been possible without the contributions and support of innumerable other people; too many to thank, though I have made an effort below.

My thanks first to Valerie Gibson, for the invitation to study at the University of Cambridge, and to work on the LHCb experiment. I thank her for being a greatly supportive supervisor, and for her many words of useful advice and guidance throughout my PhD studies.

I owe a huge thanks to Harry Cliff, without whom there is no chance I would have been able to complete my PhD studies. I don't think I could have hoped for a better collaborator and mentor, who has passed me a great breadth of knowledge, and steered me back on the road when I've been in danger of getting lost in the weeds.

The LHCb collaboration has around 1400 members, and is supported by huge numbers of people running the surrounding organisations and infrastructures. Without the efforts of them and their forebears to build, run, and maintain the LHCb detector, this work would have been very materially impossible. I am particularly grateful to all those who have offered comment on the work documented in this dissertation, at every stage of the analysis and publication process. Thank you especially to the members and convenors of the Rare Decays and Electroweak Penguins working-groups, the analysis review committee, the editorial board, and the LHCb physics coordination team. Thanks also to the members of the LHCb RICH upgrade team, for offering me a brief sojourn in the domain of hardware.

I am grateful to the Science and Technology Facilities Council (STFC) for funding my studies, including an illuminating and enjoyable year in Geneva. The STFC also funded a three-month internship at the Parliamentary Office of Science and Technology (POST) in late-2020, where I helped to produce a policy briefing note [1]. A big thanks to the members of POST for offering me this opportunity, and for being greatly welcoming and accommodating hosts at a very strange time. In particular, my thanks to Lorna Christie, for being a superb supervisor, and for offering me a great deal of insight into the world of policy.

Thank you to all the members of the Cambridge HEP and LHCb groups, who have provided a very special and pleasant working environment. In particular, a thanks to John Hill and Steve

Wotton for their tireless work maintaining the HEP computing cluster (and for tolerating my occasionally "sub-optimal" uses of the storage and network resources!).

A special thanks to the other members of the "Gang of Four"; Blaise Delaney, George Lovell, and Ifan Williams. Through thick and thin, they have made the last four years so much more interesting and enjoyable. To use a slightly clichéd analogy, a PhD is a long journey, with many twists and turns to negotiate. It has been an honour to have them as travelling companions. My thanks also to Fionn Bishop and Richard Williams, for being excellent office-mates during some of the most stressful phases of my PhD studies. I only hope that I have transferred to them more of my late-PhD knowledge than of my late-PhD weariness!

Thank you to friends and family in Ely, Cambridge, London and beyond, for keeping me sane, sharing a laugh, and showing an interest in my work (even if my explanations haven't done it justice!). Finally, but certainly not least, thank you to my Mother, for feeding, clothing, and sheltering me, and for her infinite love and support.

"Nevertheless, ultimately many experimental particle physics measurements reduce to counting events... Of course, this is not always quite as easy as it sounds."

MARK THOMSON, Modern Particle Physics [2, Sec. 1.4]

For my Father

Preface

This dissertation outlines work undertaken by the author to test the Standard Model (SM) of particle physics, by measuring whether $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays obey lepton universality. These are examples of $b \to s \ell^+ \ell^-$ decays, which can be used for a range of sensitive tests of the SM. The results of this analysis were made public in a pair of conference and seminar talks on the 19th of October 2021 [3, 4]. A preprint was also released, which has been submitted to an academic journal for publication [5].

In Chapter 1, the SM is briefly summarised, with a focus on those aspects relevant to $b \to s \ell^+ \ell^-$ decays. The phenomenology of these decays is then examined in further depth, and recent experimental results are reviewed. All data used in this dissertation was gathered using the LHCb detector at CERN, the apparatus of which is described in Chapter 2. General aspects of the software and methods used to analyse LHCb data are also outlined here.

Following this, the methodology and results of the new analysis are outlined in Chapters 3 to 7. This analysis was carried out in close collaboration with Dr Harry Cliff, who initiated it before the author commenced his PhD work. Components of the analysis which were carried out mainly or solely by the author include the "particle identification" and "residual" calibrations applied to simulation in Sections 4.2.1 and 4.2.6; the selection efficiency calculations in Section 4.3; most of the background studies in Section 4.4; the maximum likelihood fits to reconstructed B-meson masses, used to estimate the number of decays reconstructed in data, in Section 4.5; the systematic uncertainty estimates in Chapter 5; the cross-checks in Chapter 6; and the estimates of the measurement sensitivities, estimates of residual biases, and final results reported in Chapter 7.

Finally, this dissertation concludes by considering the implications of these new measurements, and reviewing future prospects in the field, in Chapter 8.

Note that charge conjugation is implied throughout this dissertation, unless otherwise specified. (For example, $B^+ \to K^{*+} \ell^+ \ell^-$ refers both to $B^+ \to K^{*+} \ell^+ \ell^-$ and $B^- \to K^{*-} \ell^- \ell^+$ decays.)

Glossary

All abbreviations and acronyms are defined in the text of this dissertation. Definitions for some of those used in multiple sections are re-iterated in this glossary, as an *aide-mémoire*.

$\mathrm{BDT}_{Comb.}$	Output of multivariate classifier trained to separate signal decays from combinatorial background.
$\chi^2_{\mathrm{DTF}}/\mathrm{nDoF}$	Goodness-of-fit of `DecayTreeFitter` fit.
χ^2_{FD}	Significance of a particle candidate's distance-of-flight.
χ^2_{IP}	Significance of a track's IP.
`DecayTreeFitter`	Method to determine kinematics for all particles in a decay process, incorporating various physical constraints.
DLL_{xy}	Difference between log of likelihood for particle hypotheses x and y, based on particle identification sub-detector outputs.
DownDown	K^0_{S} candidate built from a pair of downstream tracks. Downstream tracks are constructed from hits in all tracking subsystems, except for the Vertex Locator (VELO).
η	Pseudorapidity.
IP	*(Impact Parameter)*. Distance of closest approach between a track and the PV.
FSR	*(Final-State Radiation)*. Emission of a photon by a final-state particle in a decay process.
`isMuon`	Boolean variable indicating whether muon track matches hits in muon sub-detectors.
$\mathrm{ISO}_{\mathrm{Trk.}}$	Variable describing how isolated a given track is from other tracks in an event.

$\mathrm{ISO}_{\mathrm{Vtx.}}$	Variable describing how isolated a given vertex is from other tracks in an event.
LongLong	K^0_S candidate built from a pair of long tracks. Long tracks are constructed from hits in all tracking sub-systems.
$m(B)$	Reconstructed mass of a reconstructed B-meson candidate.
MC	*(Monte Carlo)*. Simulation of particle collisions and decays.
NP	*(New Physics)*. A physical theory which makes predictions beyond those of the Standard Model of particle physics.
`nSPDHits`	Number of Scintillating Pad Detector (SPD) hits recorded in an event.
`nTracks`	Number of tracks reconstructed in an event.
`ProbNN`	Outputs of neural networks trained to separate different particle species, based on particle identification and tracking sub-detector outputs.
p_T	Momentum component transverse to the beamline (*a.k.a.* "transverse momentum").
PV	*(Primary Vertex)*. Location of inelastic collision between two protons.
q^2	Dilepton invariant mass squared (equivalent to the square of the momentum transferred to the two leptons in the decay of a B-hadron).
τ	Lifetime of particle.

Table of contents

Part I

Introduction

Chapter 1

Theoretical overview and recent experimental results

1.1 The Standard Model

The Standard Model (SM) of particle physics describes the interactions of sub-atomic particles, and represents humanity's best understanding of the fundamental constituents of matter. Having reached its current form in the 1970s, it describes three of the fundamental forces; the electromagnetic force, the weak force, and the strong force. The SM is often cited as one of the most successful scientific theories ever [6], having predicted numerous physical phenomena with great precision, across a wide range of energies. For example, the magnetic moment of the electron can be predicted using Quantum Electrodynamics (QED)*, yielding a value which agrees with experimental measurements to within 8 ppb [7]. In this chapter, some of the mathematics governing the SM is briefly outlined†. Then, various experimental and theoretical challenges to the SM which suggest the presence of New Physics (NP) effects are explained. Finally, this chapter explains how the SM can be further tested by measuring $b \to s\ell^+\ell^-$ decay processes. Recent experimental results, and their potential implications, are reviewed.

1.1.1 Organisation of the Standard Model

There are 25 fundamental particles in the SM, which are classified according to various properties summarised in Fig. 1.1. Firstly, particles are split according to their spins; the 13

*The strong and weak forces also make small contributions.

†Much of the information and derivations in this chapter are adapted from [2, 8–11]. More specific references are given where necessary.

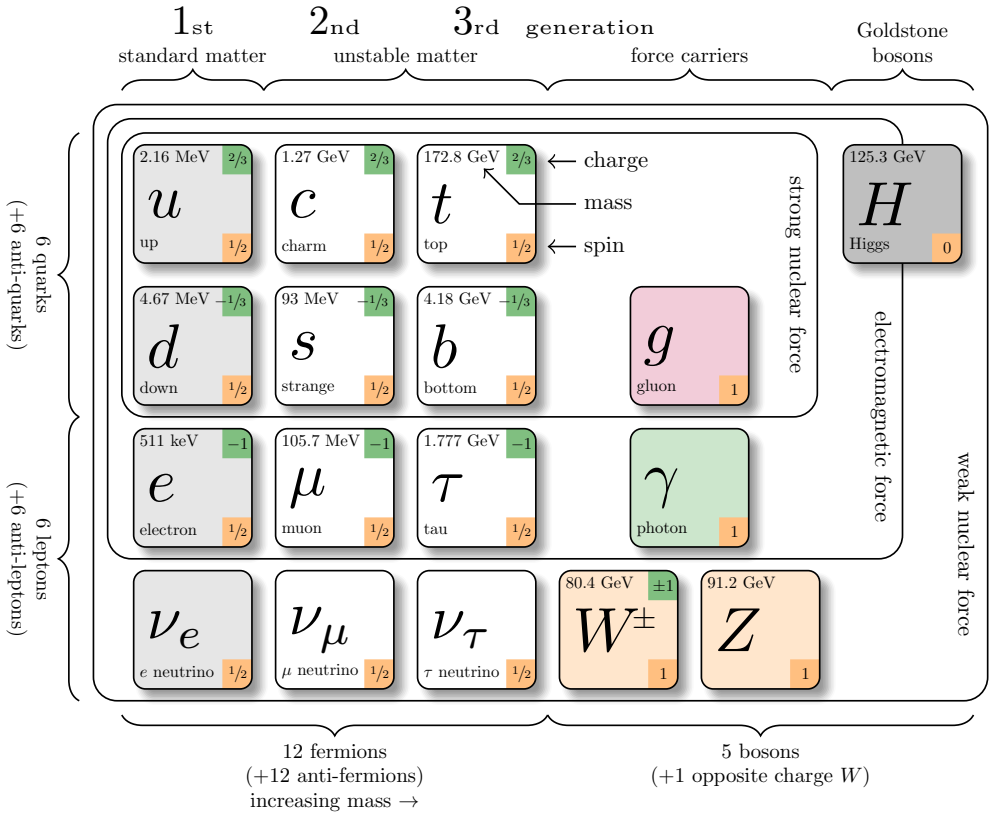


Fig. 1.1: Particles of the Standard Model (adapted from [12], with masses taken from [13]).

particles with integer spins are known as "bosons", while the 12 particles with half-integer spins are known as "fermions"*.

1.1.1.1 The Fermions

The fermions are further divided according to whether they carry "colour charge", and so can interact via the strong nuclear force. Fermions which do not carry colour charge are known as "leptons". Some leptons carry electric charges of -1, so are able to interact via the electromagnetic force. Others, known as "neutrinos" are electrically neutral, so cannot interact via the electromagnetic force. The leptons and neutrinos are found in three "generations" of increasing mass.

*The full 25-particle enumeration is as follows; six quarks, six leptons, one photon, one Z^0 boson, two $W^\pm$ bosons, eight gluons, and one Higgs boson. Other totals can be reached, depending on how one chooses to count the gluons, $W^\pm$ bosons, antimatter *etc.*

Fermions which can interact via the strong nuclear force are known as "quarks". As for the leptons, there are two different types with differing electrical charges; those with charges of $+2/3$ are known as "up-type" quarks, while those with charges of $-1/3$ are known as "down-type" quarks. The up-type and down-type quarks are also classified into three generations of increasing mass.

Due to an effect of the strong force known as "colour confinement", bound states of multiple quarks are formed, known as "hadrons". Most quarks cannot be observed outside such bound states[*]. "Baryons" are a type of hadron typically made up of three quarks (for example, the proton). "Mesons" are typically made up of one quark and one antiquark (for example the pion or B^0). Mesons and baryons comprising larger numbers of quarks, known as "tetraquarks" and "pentaquarks", have also been observed [14].

1.1.1.2 The Bosons

The Standard Model is a Quantum Field Theory with a Lagrangian which is symmetric under local Gauge transformations with the groups [15]:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.1}$$

The $SU(3)_C$ group corresponds to the strong nuclear force, with C denoting the colour charge carried by quarks. The $SU(2)_L \otimes U(1)_Y$ group corresponds to the electroweak force, with the L and Y denoting "weak isospin" and "weak hypercharge" (more on these later). However, these symmetries cannot be respected by a Lagrangian containing only the fermions. They can only be respected if a set of spin-1 "gauge bosons" are introduced to the Lagrangian, which correspond to the generators of each symmetry group. Terms allowing each gauge boson to interact with fermions carrying the relevant charge are also introduced. The eight generators of $SU(3)_C$ correspond to eight massless "gluons"[†]. The four generators of $SU(2)_L \otimes U(1)_Y$ mix via the mechanisms outlined in the following sections, to form the massless "photon" (γ), and the massive W^+, W^-, and Z^0 bosons.

1.1.1.3 Electroweak Unification

As mentioned above, the $SU(2)_L$ gauge bosons only couple to particles possessing a non-zero value of weak isospin (I_W). The left-handed chiral components of the fermions form isospin

[*]The top quark is the only exception, as its lifetime is short enough for it to decay before it forms a hadron.

[†]These generators arise from nine linearly-independent 3×3 matrices which obey unitarity. One of these is absorbed into the quark fields as a complex phase, so does not correspond to a colour transformation [2, Sec. 9.6].

doublets with $\left(I_W, I_W^3\right) = (1/2, \pm 1/2)$, while the right-handed chiral components form isospin singlets with $\left(I_W, I_W^3\right) = (0,0)$:

$$\psi_L \in \left\{ \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \quad \begin{pmatrix} u \\ d' \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s' \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b' \end{pmatrix}_L \right\} \tag{1.2}$$

$$\psi_R \in \left\{ e_R^-, \quad \mu_R^-, \quad \tau_R^-, \quad u_R, \quad c_R, \quad t_R, \quad d_R, \quad s_R, \quad b_R \right\} \tag{1.3}$$

Expressed in terms of the Pauli spin matrices $\sigma_{1,2,3}$, the three generators of the $SU(2)_L$ group give rise to three fields $W^{(1,2,3)}$, plus three "current terms" with the left-handed fermions:

$$j_i^\mu = \frac{g_W}{2} \bar{\psi}_L \gamma^\mu \sigma_i \psi_L \tag{1.4}$$

Here, g_W denotes the weak coupling strength. One finds that the physical W-bosons are given by linear combinations of $W^{(1)}$ and $W^{(2)}$:

$$W^\pm = \frac{1}{\sqrt{2}} \left(W^{(1)} \mp W^{(2)} \right) \tag{1.5}$$

The current terms for fermions coupling to the W^+ and W^- are hence as follows (with e_L and ν_L substituted as appropriate):

$$j_+^\mu = \frac{g_W}{\sqrt{2}} \bar{\nu}_L \gamma^\mu e_L, \quad j_-^\mu = \frac{g_W}{\sqrt{2}} \bar{e}_L \gamma^\mu \nu_L \tag{1.6}$$

One also finds a neutral current term for fermions coupling to the $W^{(3)}$:

$$j_3^\mu = \frac{g_w}{2} \left(\bar{\nu_L} \gamma^\mu \nu_L - \bar{e_L} \gamma^\mu e_L \right) \tag{1.7}$$

Note that the $U(1)_Y$ symmetry gives rise to a similar neutral-current interaction term, for particles with non-zero values of "weak hypercharge" Y. However, this does not correspond directly to a physical field. Instead, the $U(1)_Y$ gauge boson B mixes with the neutral $W^{(3)}$ gauge boson from $SU(2)_L$, giving the physical photon and Z^0 bosons [16–18]. The level of mixing is dictated by the Weinberg Angle θ_W, giving the physical fermion currents:

$$j_{EM}^\mu = j_Y^\mu \cos\theta_W + j_3^\mu \sin\theta_W \tag{1.8}$$

$$j_Z^\mu = -j_Y^\mu \sin\theta_W + j_3^\mu \cos\theta_W \tag{1.9}$$

Using the fact that this theory must give the measured physical values for the electromagnetic coupling strengths, and that $U(1)_Y$ transformations must not break the $SU(2)_L$ symmetry

of the SM Lagrangian, one can calculate the weak hypercharges for each left-handed and right-handed fermion. After doing this, one finds that the Z^0 boson couples to both left-handed and right-handed fermions*, with the current given by:

$$j_Z^\mu = g_Z \left(c_L \bar{f}_L \gamma^\mu f_L + c_R \bar{f}_R \gamma^\mu f_R \right) \tag{1.10}$$

where:

$$g_Z = \frac{g_W}{\cos\theta_W} \tag{1.11}$$

$$c_L = I_W^{(3)} - Q_f^{EM} \sin^2\theta_W \tag{1.12}$$

$$c_R = -Q_f^{EM} \sin^2\theta_W \tag{1.13}$$

Here, Q_f^{EM} indicates electromagnetic charge (in natural units). However, one issue remains. The $W^\pm$ and Z^0 bosons have masses of 80 GeV, and 91 GeV respectively [13]. If one naively adds mass terms for any gauge boson to the Lagrangian, local gauge invariance is broken. Therefore, some mechanism is needed which will account for the masses of the $W^\pm$ and Z^0 bosons, while retaining the local gauge symmetry of the Lagrangian.

1.1.1.4 The Higgs mechanism

The Higgs mechanism allows the $W^\pm$ and Z bosons to acquire such a mass, and is constructed by inserting two complex scalar fields into the Lagrangian [19–24]. One is charged, and the other neutral, and they form a weak isospin doublet. The neutral scalar is given a non-zero vacuum expectation value $\langle 0|\phi^0|0\rangle = v$, meaning perturbations about this minimum can be expanded:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.14}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.15}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + \eta + i\phi_4 \end{pmatrix} \tag{1.16}$$

*With the exception of right-handed neutrinos, which have both zero electromagnetic charge, and zero $I_W^{(3)}$.

Using the gauge freedom of the SM Lagrangian, one can choose to write this in the "Unitary Gauge" where the scalar fields are entirely real, finding:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H^0 \end{pmatrix} \tag{1.17}$$

By considering the effect of the $SU(2)_L \otimes U(1)_Y$ covariant derivative on this "Higgs doublet", a Lagrangian is recovered where the $W^{(1)}$, $W^{(2)}$, $W^{(3)}$, and B fields mix as asserted in the previous section, to give a massless photon, plus massive $W^{\pm}$ and Z^0 bosons, while retaining gauge invariance. In addition, the Lagrangian includes a single, massive, neutral scalar field H^0, which can couple to the gauge bosons with vertices of the form VVH^0H^0 and VVH^0, where $VV \in \{W^+W^-, Z^0Z^0\}$. Under this model, the masses of the $W^{\pm}$, Z^0, and H^0 fields, the mixing angle between the fields θ_W, and the coupling strengths of the VVH^0H^0 and VVH^0 vertices are given in terms of just four parameters: the coupling strengths of the original $SU(2)_L$ and $U(1)_Y$ fields, and two free parameters of the Higgs potential. By experimentally measuring these particle masses, plus the $W^{\pm}$ and Z decay widths and other observables sensitive to these parameters, one can test whether the model is self-consistent. To date, such observables show good compatibility with SM expectations [25]. The scalar "Higgs boson" H^0 was first observed in 2012 by the ATLAS and CMS experiments, with a mass of 125 GeV [26, 27]. With its discovery, all component particles of the SM have now been observed.

1.1.2 Flavour in the Standard Model

Under the SM, the Higgs mechanism also gives rise to the fermion masses*. The SM Lagrangian contains terms where the fermion and Higgs fields are coupled, known as the Yukawa interactions. These terms also allow the mixing of quark flavours in weak interaction decays. One first needs to construct a "conjugate" Higgs doublet, given by:

$$\phi^c = i\sigma_2\phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H^0 \\ 0 \end{pmatrix} \tag{1.18}$$

Then, the quark-sector Yukawa terms are given by:

*Note that this is not fundamentally necessary for the self-consistency of the theory. It would be possible *a priori* for there to be two separate scalar fields, with one generating the fermion masses, and the other generating the electroweak boson masses [28, 29]. However, all measurements to date have been compatible with SM expectations [13, Chapter 11, Status of Higgs Boson Physics].

$$\mathcal{L}_{Yukawa} = -\bar{Q}_{Li}\phi^{c}\hat{Y}^{u}_{i,j}u_{Rj} - \bar{u}_{Ri}\phi^{c\dagger}\hat{Y}^{u\dagger}_{j,i}Q_{Lj} - \bar{Q}_{Li}\phi\hat{Y}^{d}_{i,j}d_{Rj} - \bar{d}_{Ri}\phi^{\dagger}\hat{Y}^{d\dagger}_{j,i}Q_{Lj} \tag{1.19}$$

This expression uses the flavour eigenspace, where $u_{L,R}$ are the left-handed and right-handed up-type quarks (*i.e.*, u, c, or t), $d_{L,R}$ are the down-type quarks (*i.e.*, d, s, or b), and Q_L are $SU(2)_L$ doublets of the left-handed quarks (*i.e.* the last three terms of eq. 1.2). The indices i and j run across the three quark generations, and $\hat{Y}^{u,d}_{i,j}$ are non-diagonal matrices of coupling strengths, known as the Yukawa matrices. After the Higgs field acquires a vacuum expectation value v, the "mass-terms" of $\mathcal{L}_{Yukawa}$ are given by:

$$\mathcal{L}_{Yukawa,Mass} = -\bar{u}_L\hat{M}_1u_R - \bar{u}_R\hat{M}^{\dagger}_1u_L - \bar{d}_L\hat{M}_2d_R - \bar{d}_R\hat{M}^{\dagger}_2d_L \tag{1.20}$$

Here, the mass matrices $\hat{M}_{1,2}$ are given in terms of the Yukawa matrices and the Higgs vacuum expectation value:

$$\hat{M}_1 = \frac{v}{\sqrt{2}}\hat{Y}^{u}, \quad \hat{M}_2 = \frac{v}{\sqrt{2}}\hat{Y}^{d} \tag{1.21}$$

However, the quarks do not propagate in their flavour eigenstates - they propagate in mass eigenstates, corresponding to the basis where the mass matrices are diagonal. As for any matrix, the mass matrices can be diagonalised using a set of unitary transformations, such that the quark fields are transformed from flavour to mass eigenstates by:

$$u_{L,R} \to \hat{U}_1u_{L,R}, \quad d_{L,R} \to \hat{U}_2d_{L,R} \tag{1.22}$$

1.1.2.1 The CKM Matrix

Note that the charged-current terms for fermions coupled to the W^+ and W^- bosons given in eq. 1.6 were given in the flavour eigenspace, rather than the mass eigenspace. Transforming into the mass eigenspace, the W^+ and W^- currents for a pair of quarks are given by:

$$j^{\mu}_{+} = \frac{g_W}{\sqrt{2}}\bar{d}_{Lj}\hat{U}^{\dagger}_{2j,i}\gamma^{\mu}\hat{U}_{1i,j}u_{Li} \tag{1.23}$$

$$j^{\mu}_{-} = \frac{g_W}{\sqrt{2}}\bar{u}_{Lj}\hat{U}^{\dagger}_{1j,i}\gamma^{\mu}\hat{U}_{2i,j}d_{Li} \tag{1.24}$$

Recall that the indices i and j denote any of the quark generations - the Yukawa interaction terms have allowed for coupling between an up-type and down-type quark of different

generations! The strengths of these couplings are modified by the appropriate element of the Cabibbo-Kobayashi-Maskawa (CKM) Matrix, given by [30, 31]:

$$V_{CKM} = U_1^\dagger U_2 = U_2^\dagger U_1 \tag{1.25}$$

$$= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.26}$$

The values of the different CKM matrix elements are not given *a priori* from the Standard Model, and must be determined using experimental measurements of weak decay processes. The current global average values are [32]:

$$|V_{CKM}| = \begin{pmatrix} 0.974390^{+0.000014}_{-0.000058} & 0.224834^{+0.000252}_{-0.000059} & 0.003683^{+0.000075}_{-0.000061} \\ 0.224701^{+0.000254}_{-0.000058} & 0.973539^{+0.000038}_{-0.000060} & 0.04162^{+0.00026}_{-0.00080} \\ 0.008545^{+0.000075}_{-0.000157} & 0.04090^{+0.00026}_{-0.00076} & 0.999127^{+0.000032}_{-0.000012} \end{pmatrix} \tag{1.27}$$

One can see that this matrix exhibits a clear hierarchy, where couplings within the same generation are strong, and couplings between different generations are much weaker. These patterns can be seen by expanding the CKM matrix in the $d \leftrightarrow s$ mixing strength, denoted λ. This is known as the Wolfenstein Parameterisation, and approximates the CKM matrix according to four real parameters [33]:

$$V_{CKM} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right) \tag{1.28}$$

The measured values of these four parameters are [32]:

$$\lambda = 0.224837^{+0.000251}_{-0.000060} \tag{1.29}$$

$$A = 0.8235^{+0.0056}_{-0.0145} \tag{1.30}$$

$$\rho = 0.1569^{+0.0102}_{-0.0061} \tag{1.31}$$

$$\eta = 0.3499^{+0.0079}_{-0.0065} \tag{1.32}$$

Processes involving the electromagnetic and strong nuclear forces are unchanged if one swaps all particles with antiparticles (a "charge flip"), whilst simultaneously inverting all spatial coordinates (a "parity flip"). This effect is known as charge-parity (CP) symmetry. Weak decays of quarks are the only process in the Standard Model where CP symmetry is violated, meaning the strengths of such processes are modified by a CP transformation*. The non-zero value of η shows that there is a small, irreducible imaginary phase in the CKM matrix, allowing for this CP violation.

1.1.2.2 Flavour-changing neutral currents

But what happens when the weak neutral-current given in eq. 1.8 is transformed to the mass basis? Inserting the unitary transformation matrices, one finds that the quark interaction currents are given by:

$$j_Z^\mu = g_Z\big(c_L \bar{u}_{Lj} \hat{U}^\dagger_{1j,i} \gamma^\mu \hat{U}_{1i,j} u_{Li} + c_R \bar{u}_{Rj} \hat{U}^\dagger_{1j,i} \gamma^\mu \hat{U}_{1i,j} u_{Ri} + c_L \bar{d}_{Lj} \hat{U}^\dagger_{2j,i} \gamma^\mu \hat{U}_{2i,j} d_{Li} + c_R \bar{d}_{Rj} \hat{U}^\dagger_{2j,i} \gamma^\mu \hat{U}_{2i,j} d_{Ri}\big) \quad (1.33)$$

Due to the unitarity of the transformation matrices, one finds that $U^\dagger_{1j,i} U_{1i,j} = U^\dagger_{2j,i} U_{2i,j} = \delta_{i,j}$. Therefore, the couplings between quarks of different generations are set to zero. Hence, any Flavour-Changing Neutral Current (FCNC) processes, such as mixing between B^0_s and $\bar{B}^0_s$ mesons, or decays involving $b \to s\ell^+\ell^-$ processes, cannot proceed at tree-level under the SM. Instead, such processes can only occur at loop-level, via Feynman diagrams involving multiple $W^\pm$ vertices, and are thus highly suppressed. This effect is sometimes known as the Glashow–Iliopoulos–Maiani (GIM) mechanism [34].

1.1.2.3 Lepton flavour conservation and universality

In contrast to quark flavour, lepton flavour is conserved within the SM. This principle can be tested by searching for decays where such conservation does not hold. For example, one can search for muons decaying to electrons via $\mu^- \to e^-\gamma$ or $\mu^- \to e^-e^+e^-$ [35, 36], on-shell decays of bosons to two different leptons such as $Z^0 \to e^\pm\mu^\mp$ or $H^0 \to e^\pm\tau^\mp$ [37, 38], or decays of heavy hadrons with different leptons in the final state such as $K^0_L \to e^\pm\mu^\mp$ or $B^0_s \to K^+\mu^-\tau^+$ [39, 40]. No such lepton flavour violating decays have been observed.

*Note that weak decays also violate parity symmetry. The electromagnetic and strong forces each obey both charge and parity symmetry, and hence also obey charge-parity symmetry.

In addition, the coupling strengths of the gauge bosons are identical for the different leptons, in an effect known as "lepton universality". Measurements of branching fractions for on-shell decays of $W^- \to \ell^- \overline{\nu}_\ell$ and $Z^0 \to \ell^+ \ell^-$ have shown good agreement with lepton universality [41–43]. Lepton flavour universality can also be tested in decays of hadrons, for example by comparing the branching fractions for $\pi^+ \to e^+ \nu_e$ against those for $\pi^+ \to \mu^+ \nu_\mu$, which are seen to agree with SM expectations [44]. However, hints of lepton flavour universality violation have been seen in some decays of B-hadrons. These measurements, and their possible implications, will be discussed later in this chapter.

1.1.3 Problems with the Standard Model

The SM is able to explain a huge number of phenomena involving subatomic particles, predicting various quantities to great precision. However, there are some effects which the SM does not explain, and some theoretical issues which suggest the SM could be a small part of another, more fundamental, theory of physics. These issues include:

- **Dark Matter** - Measurements of the rotational velocities of spiral galaxies, temperature fluctuations in the cosmic microwave background, gravitational lensing by galaxy clusters, and the redshifts of Type Ia supernovae all suggest that SM particles only account for $\sim 5\%$ of the universe's matter-energy content [45]. A further $\sim 69\%$ is made up of "dark energy" responsible for the acceleration in the universe's expansion, while the remaining $\sim 26\%$ is made up of non-luminous, gravitationally-interacting "dark matter" [46]. No SM particle provides a suitable candidate for this. Therefore, if it is assumed to be a form of particle, it must interact little with SM matter*. Numerous searches for NP dark matter candidates have taken place, with attempts to produce dark matter at particle colliders [48], to directly observe interactions of dark matter particles with atomic nuclei in high-mass targets such as large tanks of xenon [49], and to indirectly observe the annihilation of dark matter into SM particles using astroparticle detectors [50]. No suitable candidate for dark matter has yet been observed.

- **Asymmetry between matter and antimatter** - At the Big Bang, matter and antimatter are expected to have been produced in equal quantities. Naively, one would expect that they would completely annihilate to photons as time progresses. Instead, there is now an excess of matter over antimatter, quantified by the baryon to photon ratio $5.8 \times 10^{-10} < (N_B - N_{\bar{B}})/N_\gamma < 6.5 \times 10^{-10}$ (95% C.L.) [13]. Under the assumption that $N_B = N_{\bar{B}}$ at the Big Bang, this asymmetry can only occur if the Sakharov Conditions are

*Dark matter need not necessarily be a particle. For example, primordial black holes have been proposed as a dark matter candidate [47].

satisfied [51]. These are the presence of baryon-number violating processes, a departure from thermal equilibrium, and presence of CP violating processes. Baryon number violation can occur in the SM via non-perturbative processes [52, 53]. While electroweak symmetry breaking can provide a phase-transition which breaks thermal equilibrium, this is only the case for Higgs masses of $m_H < 75\,\text{GeV}$ [54]. In addition, the amount of CP violation present in quark flavour mixing is around 10 orders of magnitude smaller than required [55]. This suggests that some extension to the SM is required, where departure from thermal equilibrium and large CP violating effects are present.

- **The strong CP problem** - In principle, the QCD Lagrangian also contains a term where CP violation is possible [56]. However, this term is heavily suppressed by a parameter $\theta \leq 10^{-10}$, experimentally bound mostly by measurements of the neutron's electric dipole moment [57, 58]. The small size of this parameter may be said to be "unnatural" or "fine-tuned". It could be explained by BSM models such as the Peccei-Quinn mechanism [59, 60], where θ is promoted to a field with a vacuum expectation value of zero. The "axion" particles which would correspond to this field have not been observed [61].

- **Neutrino masses** - Within the SM, neutrinos are massless. Contrary to this expectation, measurements of neutrinos originating from the Sun, from cosmic ray interactions with the atmosphere, and from nuclear reactors show that neutrinos oscillate between different lepton flavours [62]. This is only possible if neutrinos have a mass. Such a mass can be introduced by adding right-handed neutrinos to the SM Lagrangian, with the same "Dirac" mass terms as for other SM particles. However, neutrino masses are small compared to other SM particles, with $(m(\nu_e) + m(\nu_\mu) + m(\nu_\tau)) < 0.12\,\text{eV}$ (95% C.L.) [13]. This could suggest they acquire masses by an alternative means known as the "seesaw mechanism" [2, Sec. 17.8]. As right-handed neutrinos would not carry any conserved charges, they might act as their own antiparticles, meaning they could acquire additional "Majorana" mass terms. The Dirac and Majorana terms in the Lagrangian can be written in terms of a "mass matrix", the eigenvalues of which correspond to the physical neutrino masses. If there is sufficient difference between the Majorana mass and the Dirac mass, one of these physical states has a very low mass, while the other has a very high mass. With a Dirac mass of $\mathcal{O}(1\,\text{GeV})$ (comparable to the masses of the other fermions), and a Majorana mass of $\mathcal{O}(10^{11}\,\text{GeV})$, the light physical state has a mass of $\mathcal{O}(0.1\,\text{eV})$, as experimentally observed in left-handed neutrinos. Right-handed neutrinos with Majorana masses of $\mathcal{O}(10^{11}\,\text{GeV})$ can arise in some "grand unified theories". Such theories attempt to unify the strong, electromagnetic, and weak forces into a single force

at high energies, similar to unification of the latter two forces under the electroweak mechanism [63].

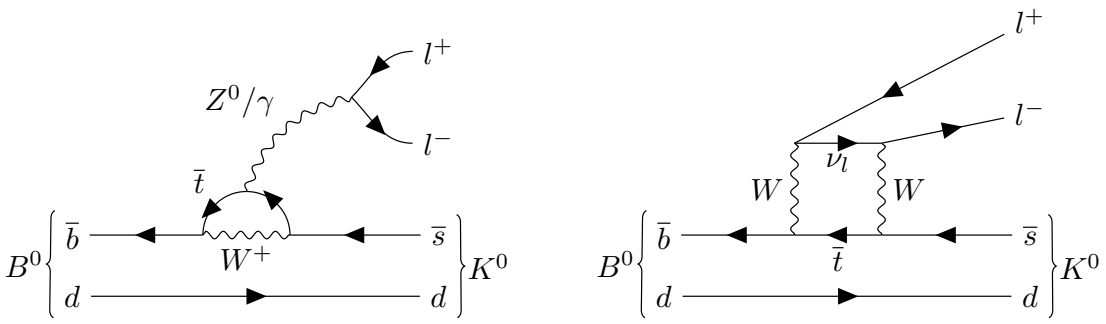
- **Gravity** - The SM describes three of the four fundamental forces, and gravity is not described. Its quantum effects are expected to be dominant at a the Planck scale of $\mathcal{O}(10^{19}\,\text{GeV})$, meaning it can be neglected for energies accessible at current particle colliders. Constructing a quantum theory of gravity which is non-divergent, and from which the SM and General Relativity can be recovered at the appropriate energy and length scales, is extremely theoretically challenging. At least 16 major approaches have been proposed [64].

- **The hierarchy problem** - As the Higgs boson is a scalar particle, its observed mass acquires corrections from virtual particle loops. The momenta of such loops are unconstrained, meaning they must be integrated up to the Planck scale of $m_P \sim \mathcal{O}(10^{19}\,\text{GeV})$. For example, a loop containing a top quark (with $m_t \sim 100\,\text{GeV}$) introduces a correction to the Higgs mass of $m_P^2/m_t^2 \sim 10^{34}$. In order to recover the measured Higgs mass of 125 GeV, its bare mass would need to be fine-tuned to cancel such virtual contributions [65–68]. Alternatively, theories such as Supersymmetry (SUSY) aim to recover the measured Higgs mass without such fine-tuning [69]. Under SUSY, each particle has a "super-partner", with its spin differing by $-1/2$. The loop-level contributions to the Higgs mass from each particle would then be cancelled by its super-partner, keeping the measured mass at $\mathcal{O}(100\,\text{GeV})$ without fine-tuning of the bare mass. No SUSY candidates have been observed, with experimental limits set at the LHC suggesting their masses would be above the $\mathcal{O}(100\,\text{GeV} - 1\,\text{TeV})$ level [70]. This suggests that if SUSY is true it is a broken symmetry, meaning some fine-tuning of the Higgs mass is still required.

It should be noted that there is some debate as to whether fine-tuned or unnatural parameters with values far from $\mathcal{O}(1)$ represent real problems with the SM [71]. For example, Hossenfelder argues that notions of fine-tuning are ill-defined and driven more by aesthetic than scientific concerns, as one cannot say *a priori* what ranges of a parameter would be theoretically acceptable [72]. In contrast, Williams argues that natural parameters are necessary when constructing a QFT where phenomena at widely separated energy scales are decoupled [73].

1.2 Rare decays of B mesons

There are two essential methods by which one may search for signs of NP at a particle collider. The first is by attempting to produce new particles on-shell, then examining them via their



(a): Penguin diagram ***(b):*** Box diagram

Fig. 1.2: Feynman diagrams showing the leading-order contributions to the $b \to s\ell^+\ell^-$ process $B^0 \to K^0\ell^+\ell^-$, under the SM.

decay products. If a particle is discovered via such a "direct search", one can then precisely determine its properties, for example its mass, spin, and coupling strengths with other particles. However, one can only produce new particles with masses at or below the centre-of-mass energy a particle collider is operating at (for example 13 TeV at the LHC prior to 2022).

Alternatively, one can attempt to search for NP effects by precisely measuring the properties of SM processes. Virtual NP particles may contribute to the amplitudes for these processes, causing them to deviate from SM expectations. With this method, one sacrifices some of the sensitivity to exact particle properties found in direct searches. However, it becomes possible to probe for NP with mass scales far above those which can be directly produced at current particle colliders. Many particles within the SM were suggested via such "indirect" methods before being observed directly, including the charm quark and neutrino, though with the notable exception of the muon [34, 74–78]. In order to carry out indirect searches successfully, one needs to define a set of observables which can be calculated precisely under the SM, can be measured precisely at a detector, and could be changed significantly by virtual NP particle contributions. One set of processes which satisfy these requirements are $b \to s\ell^+\ell^-$ decays.

As mentioned previously, $b \to s\ell^+\ell^-$ decays are an example of a FCNC process. Under the SM, these are only able to proceed via heavily-suppressed loop-level processes. The leading-order contributions to these are from the "box" and "penguin" diagrams shown in Fig. 1.2. Typically, these processes have branching fractions of $\mathcal{O}(10^{-6})$ or below; hence, they are referred to as "rare decays".

A large collection of measurable quantities can be defined for these decays, such as differential branching fractions, CP asymmetries, and observables describing the angular distributions of decay products. In general, these can be computed with reasonable precision under the SM. Many NP models allow for FCNC processes at tree-level, so could give sizeable

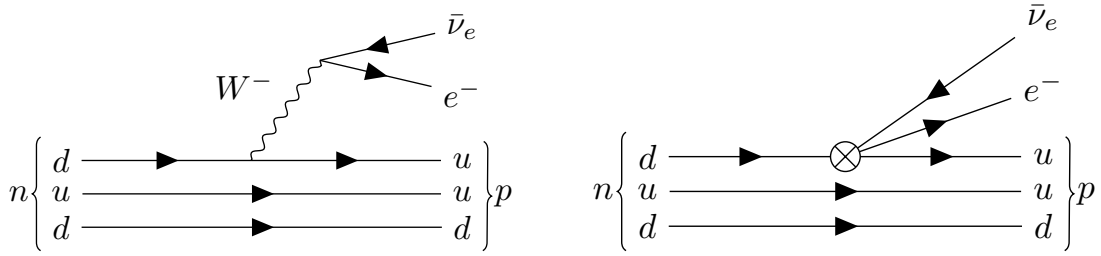


Fig. 1.3: Feynman diagrams showing the leading-order contribution to neutron beta-decay under the SM (left), and under an EFT for $q^2 \ll m_W^2$ (right).

amplitudes for $b \to s\ell^+\ell^-$ processes. Via interference with the SM amplitudes, these in turn could cause large deviations in the measured observables.

1.2.1 Effective Hamiltonian

Decays involving the $b \to s\ell^+\ell^-$ process can be modelled using an Effective Field Theory (EFT). To illustrate this, first consider weak decays of muons via $\mu^- \to e^-\nu_\mu\overline{\nu}_e$. To leading-order, this is mediated at tree-level by a W^- boson. The Feynman diagram for this process includes two W^--fermion vertices and a W^- propagator, giving a matrix element of the form [2, Sec. 11.5.1]:

$$\mathcal{M}_{fi} = -\frac{g_W^2}{8}\left[\bar{u}(\nu_\mu)\gamma^\mu\left(1-\gamma^5\right)u(\mu^-)\right]\left[\frac{g_{\mu\nu}-q_\mu q_\nu/m_W^2}{q^2-m_W^2}\right]\left[\bar{u}(e^-)\gamma^\nu\left(1-\gamma^5\right)v(\overline{\nu}_e)\right] \tag{1.34}$$

Here, g_W refers to the weak coupling constant, m_W refers to the mass of the W^- boson, and q refers to the momentum carried by the W^- boson. The spinors and adjoint spinors of particles and anti-particles are denoted by $\bar{u}$ and $\bar{v}$. However, with $m_\mu - m_e \approx 100\,\text{MeV}$, the momentum transfer q^2 will be negligible compared to m_W^2. Under the limit $q^2/m_W^2 \to 0$, the matrix element reduces to:

$$\mathcal{M}_{fi} = \frac{g_W^2}{8m_W^2}g_{\mu\nu}\left[\bar{u}(\nu_\mu)\gamma^\mu\left(1-\gamma^5\right)u(\mu^-)\right]\left[\bar{u}(e^-)\gamma^\nu\left(1-\gamma^5\right)v(\overline{\nu}_e)\right] \tag{1.35}$$

In essence, the high-energy $W^\pm$ boson propagator has been "integrated away", such that its degrees of freedom are no longer considered. At a scale of $q^2 \sim 1\,\text{GeV}^2$, we are left with a four-point contact interaction between the muon, electron, neutrino, and anti-neutrino. The $W^\pm$ boson is absorbed into an "effective coupling strength" for the point-interaction, defined by the

Fermi constant $G_F = g_W^2 / \left(4\sqrt{2} \cdot m_W^2\right)$. The same is true in neutron beta-decay (see Fig. 1.3), for which this point-interaction theory was first devised by Enrico Fermi [75–77]. It was used to model weak interactions before electroweak theory was devised.

Similarly, $b \to s\ell^+\ell^-$ processes are governed by physical interactions with a wide range of energy scales. Firstly, there is the decay of the b-quark, with a mass of $m_b \approx 4.2\,\text{GeV}$. Then, there are the t-quark, $W^\pm$ bosons and Z bosons in the box and penguin loops, with masses at the electroweak scale of $\Lambda_{EW} \sim 100\,\text{GeV}$. Finally, the initial-state B-hadron and the final-state K-hadron are bound together by the strong nuclear force, with an energy scale of $\Lambda_{QCD} \sim 200\,\text{MeV}$. Because of this separation of scales, an EFT can be constructed, where the theory is separated into the "short-distance" effects with $\Lambda >> m_b$, and "long-distance" effects with $\Lambda << m_b$. The Operator Product Expansion (OPE) technique is then used to define an effective Hamiltonian for these decays [79–82]:

$$\mathcal{H}_{Eff.} = -4\frac{G_F}{\sqrt{2}} V_{tb} V_{ts}^* \frac{e^2}{16\pi^2} \sum_i \left(\mathcal{C}_i(\mu) \cdot \mathcal{O}_i\right) \tag{1.36}$$

Here, G_F stands for the Fermi constant, $V_{ij}^{(*)}$ denote CKM matrix elements, and e stands for the electron charge. The long-distance physics are described by a set of point-operators $\mathcal{O}_i$, which correspond to different particle topologies and spin configurations. Each operator has a corresponding dimensionless Wilson coefficient $\mathcal{C}_i$, which absorbs all the high-energy effects and gives the operator's effective coupling strength at a renormalisation energy μ. In general, these Wilson coefficients are calculated using a three-step process:

1. The effective Hamiltonian $\mathcal{H}_{Eff.}$ is written in terms of the various operators and Wilson coefficients for different final states

2. The same Hamiltonian is written in terms of the full SM theory at a high energy scale (typically m_W), to a chosen order. Then, the terms found in the full theory are matched to the terms in the effective theory.

3. A set of renormalisation group equations are calculated for the Wilson coefficients. They are evolved down to the chosen renormalisation energy scale (typically $\mu = m_b$), and their numerical values are calculated.

The operators which give dominant contributions to $b \to s\ell^+\ell^-$ decays are:

$$\mathcal{O}_7 = \frac{m_b}{e}\left(\bar{s}\sigma_{\mu\nu}P_R b\right)\cdot F^{\mu\nu} \tag{1.37}$$

$$\mathcal{O}_9 = \left(\bar{s}\gamma_\mu P_L b\right)\cdot\left(\bar{\ell}\gamma^\mu \ell\right) \tag{1.38}$$

$$\mathcal{O}_{10} = \left(\bar{s}\gamma_\mu P_L b\right)\cdot\left(\bar{\ell}\gamma^\mu \gamma_5 \ell\right) \tag{1.39}$$

Here, b and $\bar{s}$ denote the quark spinors and adjoint-spinors, $\bar{\ell}$ and ℓ denote leptons of a given flavour, and $P_{L/R} = (1 \mp \gamma_5)/2$ project out left-handed and right-handed chirality components. The "electromagnetic" operator $\mathcal{O}_7$ corresponds to the decays with a virtual photon. The "semileptonic" operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$ give couplings to the leptons with vector and axial-vector spin configurations. Note that there are separate semileptonic operators for electrons, muons, and taus. Under the SM, their coupling strengths are identical due to lepton universality. The results of next-to-next-to-leading order (NNLO) computations of these Wilson coefficients under the SM, at the renormalisation scale $\mu = 4.2\,\text{GeV}$, are [83–85]:

$$\mathcal{C}_7^{\text{SM}} = -0.30, \quad \mathcal{C}_9^{\text{SM}} = 4.1, \quad \mathcal{C}_{10}^{\text{SM}} = -4.2 \tag{1.40}$$

Any NP contributions to $b \to s\ell^+\ell^-$ decays would appear at high energy-scales, above m_b. Therefore, their effects would manifest via shifts to the measured Wilson coefficients, giving $\mathcal{C}_i = \mathcal{C}_i^{\text{SM}} + \mathcal{C}_i^{\text{NP}}$. It is this property which makes the effective field theory treatment such a useful methodology. Different models of NP will cause changes to different Wilson coefficients. Likewise, different $b \to s\ell^+\ell^-$ decay processes and observables are sensitive to different Wilson coefficients. Thus, one can use the effective Hamiltonian to easily interpret $b \to s\ell^+\ell^-$ decay results through the lens of many different NP models.

In addition, there are a set of operators with flipped quark chiralities, found by swapping $P_L \leftrightarrow P_R$:

$$\mathcal{O}'_7 = \frac{m_b}{e}\left(\bar{s}\sigma_{\mu\nu}P_L b\right)\cdot F^{\mu\nu} \tag{1.41}$$

$$\mathcal{O}'_9 = \left(\bar{s}\gamma_\mu P_R b\right)\cdot\left(\bar{\ell}\gamma^\mu \ell\right) \tag{1.42}$$

$$\mathcal{O}'_{10} = \left(\bar{s}\gamma_\mu P_R b\right)\cdot\left(\bar{\ell}\gamma^\mu \gamma_5 \ell\right) \tag{1.43}$$

These are heavily suppressed under the SM, by a factor of m_s/m_b. In addition, there are a set of operators for scalar ($\mathcal{O}_S^{(\prime)}$), pseudoscalar ($\mathcal{O}_P^{(\prime)}$), and tensor ($\mathcal{O}_T$, $\mathcal{O}_{T5}$) couplings between quarks and leptons, which are also negligible under the SM. Finally, there is the chromomagnetic operator $\mathcal{O}_8$, and a set of four-quark operators $\mathcal{O}_{1-6}^{(\prime)}$. These only contribute to

$b \to s\ell^+\ell^-$ decays at loop-level or beyond, and are expected to be relatively insensitive to NP contributions. However, they can be a source of theoretical uncertainty when computing some observables.

1.2.2 Phenomenology and theoretical uncertainties

In semileptonic $b \to s\ell^+\ell^-$ decays[*], the relative contributions of different Wilson coefficients to the decay amplitudes, and other theoretical considerations, vary as a function of q^2. This is defined as the square of the momentum transferred to the two leptons[†]. Cartoons of the differential decay-rates as a function of q^2 are shown in Fig. 1.4. Some of the key features seen as q^2 increases from zero are highlighted in the following paragraphs.

At very low values of q^2, there is an enhancement in the differential decay rate if the final-state meson is a vector with $J = 1$ (for example, a ϕ or K^*). This is because the decaying B-hadrons are pseudoscalars with $J = 0$, and the two leptons are fermions with $J = \frac{1}{2}$. Therefore, these decays can couple to virtual photons with $J = 1$, the propagators of which have a pole at $q^2 \to 0$. In the effective Hamiltonian, this corresponds to the electromagnetic operator $\mathcal{O}_7$ (and its chirality-flipped counterpart $\mathcal{O}'_7$). Such an enhancement is not present when the final-state meson is a pseudoscalar (for example, a K^+ or K^0_S), as angular momentum would no longer be conserved in a coupling with the photon. Contributions from $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are present across the full range of q^2, with both pseudoscalar and vector mesons in the final-state.

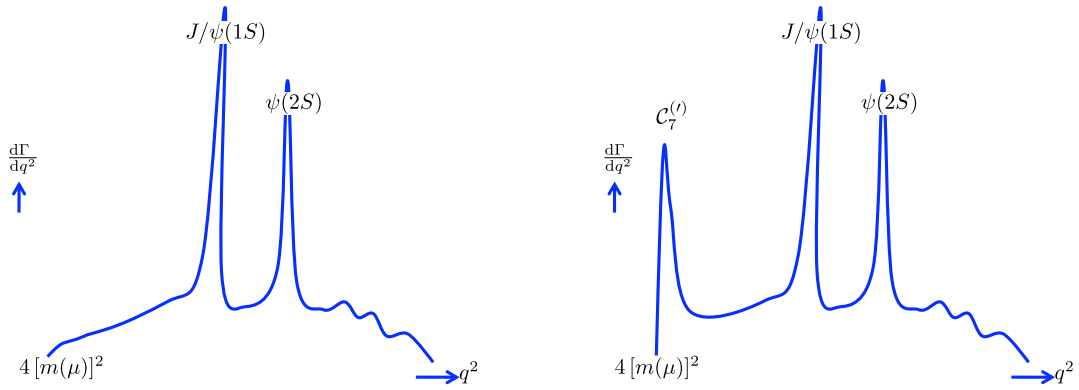
As q^2 increases, several narrow peaks can be seen in the differential decay rate. These correspond to on-shell decays of $c\bar{c}$ charmonium resonances, most notably the J/ψ and $\psi(2S)$ at $q^2 \approx 9.6\,\text{GeV}^2$ and $q^2 \approx 13.6\,\text{GeV}^2$, respectively. As shown in Fig. 1.5, these occur at leading-order via tree-level $W^\pm$ decays. Therefore their branching fractions are much higher than those of the rare FCNC $b \to s\ell^+\ell^-$ decays, with [13]:

$$\mathcal{B}(B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}) \cdot \mathcal{B}(J/\psi \to \ell^+\ell^-) \approx 10^{-3} \cdot \left(5 \times 10^{-2}\right) \tag{1.44}$$

As these resonances are not sensitive to the $b \to s\ell^+\ell^-$ operators $\mathcal{O}_i$, they are often vetoed in experimental measurements using requirements on q^2. In addition, there are a number of broad charmonium resonances in the region $q^2 > m(\psi(2S))^2$. Finally, the decay's kinematic limit is reached at $q^2 \gtrapprox 20\,\text{GeV}^2$, and the differential decay rate falls to zero.

[*]These are decays where there is a hadron and a pair of leptons in the final-state

[†]This is also equal to the invariant mass of the two leptons - the two definitions may be used interchangeably throughout this dissertation.



(a): $B \to P\ell^+\ell^-$ ***(b):*** $B \to V\ell^+\ell^-$

Fig. 1.4: Cartoons of the differential decay rates for $B \to P\ell^+\ell^-$ (left) and $B \to V\ell^+\ell^-$ (right) decays, as a function of q^2, where P is a pseudoscalar meson, and V is a vector meson. Adapted from [86].

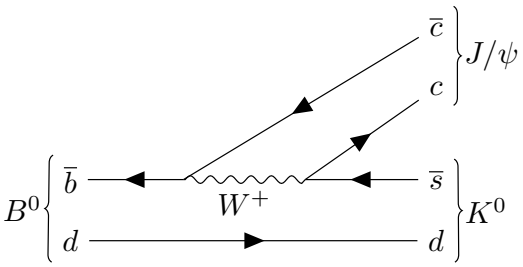


Fig. 1.5: Feynman diagram showing the tree-level decay of $B^0 \to J/\psi K^0$, which causes a large resonance in the q^2 spectrum for $B^0 \to K^0\ell^+\ell^-$ decays.

1.2.2.1 Hadronic effects

In order to calculate an observable for a given $b \to s\ell^+\ell^-$ process, one must first calculate its amplitude using:

$$\mathcal{A}(I \to F) = \langle I|\mathcal{H}_{Eff.}|F\rangle \propto \sum_i \mathcal{C}_i \cdot \langle I|\mathcal{O}_i|F\rangle \tag{1.45}$$

Note here that the Wilson coefficients $\mathcal{C}_i$ are independent of the specific initial and final states I and F. In contrast, the expectation values of the operators $\langle I|\mathcal{O}_i|F\rangle$ are dependent on I and F. The transitions between the different hadrons in I and F are parameterised using sets of *hadronic form factors*. At the energy scale where hadronisation occurs, the strong coupling constant α_S is too large to use perturbative QCD methods. Therefore, the form-factors must be calculated using various non-perturbative techniques, which can carry large theoretical uncertainties.

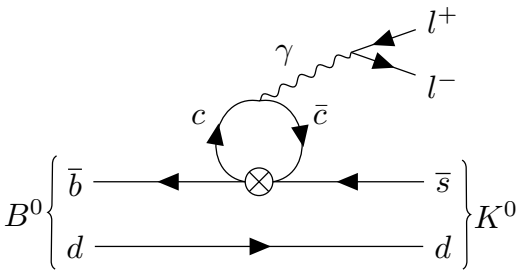


Fig. 1.6: Contribution to $B^0 \to K^0_S \ell^+ \ell^-$ from virtual charm-quark loops.

Different non-perturbative methods may be used for different levels of recoil between the initial and final-state hadrons (and hence to different regions of q^2). At high values of q^2 above the $\psi(2S)$ resonance region, *lattice QCD* (LQCD) may be used [87–91]. In this technique, a configuration of quarks is simulated across a discretised lattice of space-time points. The resulting QCD potential is then numerically integrated across this lattice. The main theoretical uncertainties relate to the chosen lattice spacing a, with computations becoming more intensive as a is decreased. At lower values of q^2, the hadronic recoil increases to become comparable to the upper momentum cut-off $1/a$ imposed by currently-practical lattice spacings, and the technique breaks down. In such a region, alternative non-perturbative techniques such as *light-cone sum rules* (LCSR) can be used instead [92–95]. One can also carry out fits to the form-factor values found using LCSR and LQCD in different regions of q^2, in order to interpolate between them and reduce theoretical uncertainties [94, 96, 97].

The other major source of theoretical uncertainty arises from loop-level contributions of the four-quark operators $\mathcal{O}_{1,2}$. As shown in Fig. 1.6, a virtual charm-quark loop is produced, which couples to a lepton-antilepton pair via a virtual photon [98, 99]. Such a contribution could mimic a lepton-universal shift to $\mathcal{C}_9$ if not properly modelled. The size of this "charm loop effect" can be difficult to quantify, particularly at high values of q^2. Its value and uncertainty are subject to debate within the theory community.

1.2.3 Observables and experimental results

A large range of observables for $b \to s\ell^+\ell^-$ processes have been measured by LHCb and other experiments. The SM predictions for some of these have large theoretical uncertainties from the charm loop effect and from hadronic form-factors. Others have been designed such that these effects cancel, at least partially. Some of these observables, and their recent experimental results, are reviewed here.

1.2.3.1 Differential branching fractions

One of the most conceptually simple observables is the differential branching fraction for a decay, measured in bins of q^2. The theoretical uncertainties for these observables, arising from hadronic form-factors and charm loops, are at the $\mathcal{O}(10-30\%)$ level, making interpretation of these measurements difficult. Nonetheless, the LHCb experiment has measured differential branching fractions for several semileptonic muon-mode decays*, including $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^{*0}\mu^+\mu^-$, $B_s^0 \to \phi\mu^+\mu^-$, and $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ [100–103]. In addition, the differential branching fractions for two of the decays examined in this dissertation, $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$, have been measured by LHCb [100]. In regions of q^2 below the J/ψ resonance, many of these differential branching fractions are seen to be slightly lower than theoretical expectations, as shown in Fig 1.7. The Belle, BaBar, CDF, and CMS experiments have also measured some of these decays [104–109]. Their results are consistent with LHCb measurements, albeit with larger experimental uncertainties (meaning their results also show good consistency with SM expectations).

*Such analyses have focussed on muon-mode decays because the LHCb detector can reconstruct them with higher efficiencies and improved resolutions, compared to the equivalent electron-mode decays

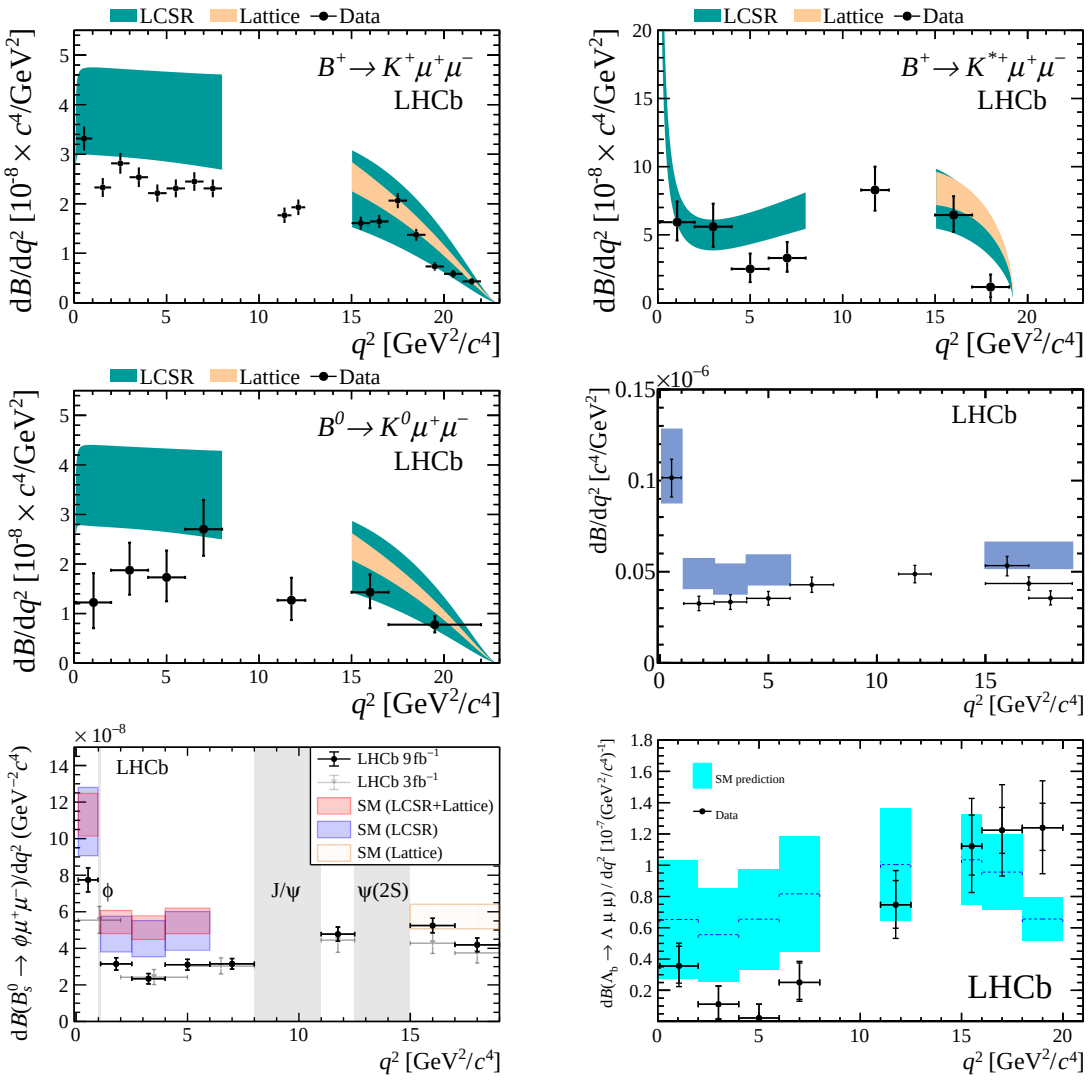


Fig. 1.7: Differential branching fractions measured by LHCb for $B^+ \to K^+\mu^+\mu^-$ (top left), $B^+ \to K^{*+}\mu^+\mu^-$ (top right), $B^0 \to K^0\mu^+\mu^-$ (middle left), $B^0 \to K^{*0}\mu^+\mu^-$ (middle right), $B_s^0 \to \phi\mu^+\mu^-$ (bottom left), and $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ (bottom right) decays, compared to theoretical expectations [100–103].

1.2.3.2 Angular analyses

The angular distributions of final-state particles in $b \to s\ell^+\ell^-$ decays are also sensitive to NP contributions. For example, the decay rate Γ for $B^+ \to K^+\ell^+\ell^-$ varies as a function of θ_ℓ, defined as angle between the ℓ^- and K^+ in the B^+ rest frame*, with the form [110]:

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta_\ell} = \frac{3}{4}(1-F_H)\left(1-\cos^2\theta_\ell\right) + \frac{1}{2}F_H + A_{FB}\cos\theta_\ell \tag{1.46}$$

By examining how θ_ℓ is distributed in data, the "forward-backward asymmetry" of the dilepton system A_{FB} and "flatness parameter" F_H can be measured. Typically, these measurements are made in bins of q^2. The LHCb experiment has carried out such measurements using $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^0_S\mu^+\mu^-$ decays, finding values of A_{FB} and F_H which are consistent with SM expectations across a range of q^2 regions [111]. Similar results have been seen by the CMS, CDF, BaBar, and Belle experiments [112, 113, 106, 104].

The angular structures of decays with vector mesons in the final state are much more complicated. For example, the $B^0 \to K^{*0}\ell^+\ell^-$ decay, followed by the p-wave decay $K^{*0} \to K^+\pi^-$, is described using three angles; the angle θ_ℓ describes the relative angles of the leptons, θ_K describes the relative angles of the K^+ and π^-, and ϕ describes the relative angles of the K^+-π^- and dilepton planes. The full differential decay-rate is then given in terms of a set of eleven spherical harmonics $f_i(\theta_\ell, \theta_K, \phi)$ and associated q^2-dependent observables $I_i\left(q^2\right)$ [114]:

$$\frac{d^4\Gamma}{dq^2 d\cos\theta_\ell d\cos\theta_K d\phi} = \frac{9}{32\pi}\sum_i \left(I_i\left(q^2\right)\cdot f_i(\theta_\ell, \theta_K, \phi)\right) \tag{1.47}$$

The angular observables can then be combined with their counterparts for the CP-conjugate decay $\bar{B}^0 \to \bar{K}^{*0}\ell^+\ell^-$, to obtain a set of CP averages S_i, and CP asymmetries A_i:

$$S_i = (I_i + \bar{I}_i) \Big/ \frac{d\left(\Gamma+\bar{\Gamma}\right)}{dq^2} \tag{1.48}$$

$$A_i = (I_i - \bar{I}_i) \Big/ \frac{d\left(\Gamma+\bar{\Gamma}\right)}{dq^2} \tag{1.49}$$

Some or all of these observables may be measured by examining the distributions of θ_ℓ, θ_K, and ϕ in data. However, the theoretical uncertainties on S_i arising from hadronic form-factors can be quite large. Instead, an optimised basis of observables $P_i^{(\prime)}$ can be used, where hadronic

*For the CP conjugate decay, θ_ℓ is defined using the ℓ^+ and K^- instead.

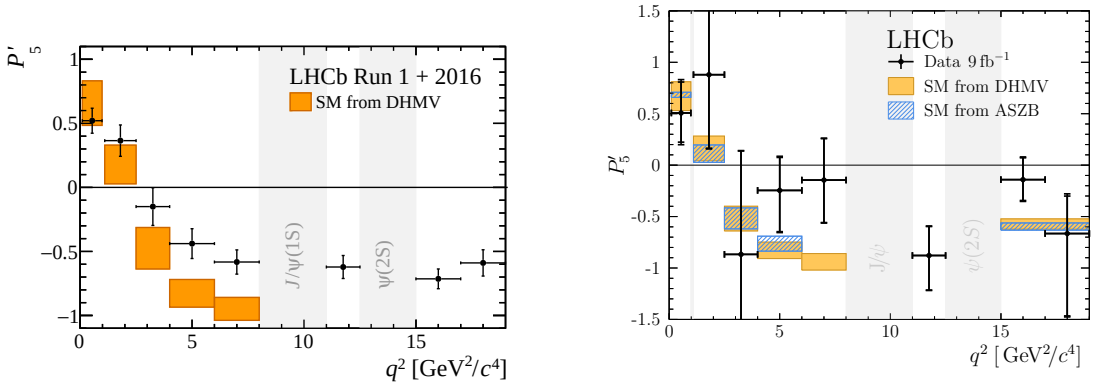


Fig. 1.8: Measurements of the angular observable P'_5 by the LHCb experiment using $B^0 \to K^{*0}\mu^+\mu^-$ decays (left) [116], and $B^+ \to K^{*+}\mu^+\mu^-$ decays (right) [117].

uncertainties partially cancel [115]. These are found by normalising S_i observables using the longitudinal polarisation fraction of the K^{*0} F_L (itself related to $S^{s,c}_{1,2}$). One such parameter is:

$$P'_5 = \frac{S_5}{2\sqrt{F_L(1-F_L)}} \tag{1.50}$$

The LHCb experiment has found that some angular observables for $B^0 \to K^{*0}\mu^+\mu^-$ deviate from SM expectations. As seen in Fig 1.8, the observable P'_5 differs from SM expectations by $2.5-2.9\sigma$ in some bins of q^2 below the J/ψ resonance [116]. Results from the Belle, CMS and ATLAS experiments are compatible with those of LHCb (although they are also compatible with SM expectations due to larger experimental uncertainties) [118–120].

These angular observables are particularly sensitive to NP in $\mathcal{C}_{9\mu}$. For example, the values of all the $B^0 \to K^{*0}\mu^+\mu^-$ angular observables measured by LHCb are found to favour a change in $\mathcal{C}_{9\mu}$ from SM expectations with $2.7-3.3\sigma$ significance, depending on the q^2 bins which are considered. The LHCb experiments have also measured angular observables for $B^+ \to K^{*+}\mu^+\mu^-$ (Fig. 1.8) and $B^0_s \to \phi\mu^+\mu^-$ decays. These are found to favour similar changes in $\mathcal{C}_{9\mu}$, with significances of 3.1σ and 1.9σ, respectively [117, 121].

The $\mathcal{C}^{(\prime)}_7$ photon pole means that the yields of $B^0 \to K^{*0}e^+e^-$ decays are greatly enhanced as $q^2 \to 0$. Using $B^0 \to K^{*0}e^+e^-$ data at very low q^2, the LHCb experiment measured a set of angular observables related to the polarisation of the virtual photon [122]. These observables were used to constrain the ratio $\mathcal{C}'_7/\mathcal{C}_7$, with results consistent with the SM expectation of $\mathcal{C}'_7 \sim 0$.

1.2.3.3 Isospin asymmetries

The mesons B^0 and B^+ are isospin partners, as the former contains a d-quark, while the latter contains a u-quark. In $b \to s\ell^+\ell^-$ decays, these "spectator quarks" do not participate in the hard interaction, and hence these decays are expected to be approximately symmetric between B^0 and B^+ decays. This assumption can be tested by measuring the isospin asymmetry:

$$A_I = \frac{\Gamma\left(B^0 \to K^{(*)0}\ell^+\ell^-\right) - \Gamma\left(B^+ \to K^{(*)+}\ell^+\ell^-\right)}{\Gamma\left(B^0 \to K^{(*)0}\ell^+\ell^-\right) + \Gamma\left(B^+ \to K^{(*)+}\ell^+\ell^-\right)} \tag{1.51}$$

Under the SM, isospin symmetry is broken by effects such as photon radiation by the spectator quark and weak annihilation of the B^+ mesons. This causes A_I to deviate from zero at the $\mathcal{O}(1\%)$ level for q^2 below the J/ψ resonance, increasing to $\mathcal{O}(10\%)$ as $q^2 \to 0$ [123–126]. NP effects such as flavour-changing neutral gauge bosons [127], or some SUSY models [128], can also affect A_I. The Belle and BaBar experiments have both measured isospin asymmetries for various $B \to K^{(*)}\ell^+\ell^-$ decays which are slightly below, though still consistent with, SM expectations [105, 106]. However, measurements for $B \to K^{(*)}\mu^+\mu^-$ decays by the LHCb experiment show good consistency with SM expectations [100].

1.2.3.4 CP asymmetries

One can also measure the asymmetry between CP-conjugate decays of $\bar{B}$ and B hadrons. In $b \to s\ell^+\ell^-$ decays, such CP violation is highly suppressed due to the small size of the CKM element V_{ts}. Under the SM, CP asymmetries are at the level of $\mathcal{O}(10^{-3})$, but can be enhanced to $\mathcal{O}(10\%)$ under some NP scenarios [114, 129]. The LHCb experiment has measured the direct CP asymmetries in the decay widths of $B^\pm \to K^\pm\mu^+\mu^-$, $\overline{(}B^0\overline{)} \to \overline{(}K^{*0}\overline{)}\mu^+\mu^-$, and $\overline{(}\Lambda_b^0\overline{)} \to \overline{(}p\overline{)}K^\mp\mu^+\mu^-$ decays, finding results consistent with CP conservation [130, 131]. Similar results have been found by the Belle and BaBar experiments [104, 106]. In addition, the LHCb experiment has measured some of the CP asymmetries in the angular observables for $B^0 \to K^{*0}\mu^+\mu^-$ and $B_s^0 \to \phi\mu^+\mu^-$, again finding results consistent with CP conservation [132, 121]. CP violation is incorporated into the effective field theory framework with imaginary components to the Wilson coefficients ($\mathcal{I}m(\mathcal{C}_i)$). The constraints on such $\mathcal{I}m(\mathcal{C}_i)$ are much more loose than those on $\mathcal{R}e(\mathcal{C}_i)$, with current data [133].

1.2.3.5 $B^0_{(s)} \to \mu^+\mu^-$ measurements

The decay $B_s^0 \to \mu^+\mu^-$ is another example of a $b \to s\ell^+\ell^-$ process, where the $\bar{b}$-quark and s-quark annihilate to produce a pair of muons. In addition to the usual suppressions applying to $b \to s\ell^+\ell^-$ processes, $B_s^0 \to \mu^+\mu^-$ decays are helicity-suppressed because the B_s^0 is a

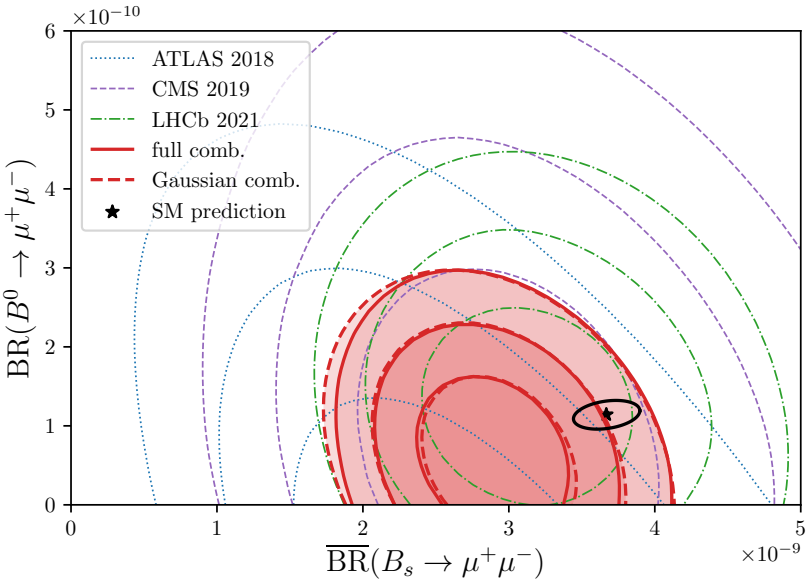


Fig. 1.9: Averages of branching fraction measurements for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays by the LHCb, CMS, and ATLAS experiments, compared to SM expectations [139]. The three contours for each experimental measurement (and their combinations) have coverage levels of 68%, 95%, and 99.7%. The contour for the SM expectations has a coverage level of 68%.

pseudoscalar meson. Therefore, it has a branching fraction of $\sim 4 \times 10^{-9}$ under the SM. The discovery of this decay, together with the CMS experiment, represented a triumph of the early LHCb physics programme [134]. Independent measurements of its branching fraction have since been carried out by the LHCb, CMS, and ATLAS experiments [135–138]. In addition, these experiments have searched for the $b \to d\mu^+\mu^-$ decay $B^0 \to \mu^+\mu^-$. Due to additional CKM suppression, the branching fraction of this decay is $\sim 1 \times 10^{-10}$ under the SM. While this decay has not yet been discovered, limits have been set on its branching fraction.

The branching fractions for $B^0_{(s)} \to \mu^+\mu^-$ are sensitive to the Wilson coefficients $\mathcal{C}_{10}^{(\prime)}$, $\mathcal{C}_S^{(\prime)}$, and $\mathcal{C}_P^{(\prime)}$ (with all except $\mathcal{C}_{10}$ negligible under the SM). Theoretical predictions for these decays are extremely clean. As they have a purely leptonic final-state, the hadronisations of the $B^0_{(s)}$ mesons are described using decay constants, which have been calculated to $\sim 0.6 - 0.7\%$ precision using lattice QCD [140]. Uncertainties on the branching fraction predictions are dominated by the experimental measurements of the relevant CKM elements. As shown in Fig. 1.9, an average of the most recent branching fraction measurements LHCb, CMS, and ATLAS is found to agree with SM predictions at the 1.8σ level [135–139]. The LHCb experiment has also measured the effective lifetime of $B_s^0 \to \mu^+\mu^-$ decays, which can disentangle NP contributions

to $\mathcal{C}_S^{(\prime)}$ and $\mathcal{C}_P^{(\prime)}$, providing complementary information to the branching fraction measurement [135, 136, 141]. The measured effective lifetime is consistent with SM expectations.

1.2.3.6 Lepton Universality Tests

Ratios of muon-mode and electron-mode branching fractions provide a stringent test of the SM. Typically, such "lepton universality ratios" are measured within a bin of $q^2 \in [q^2_{\text{min}}, q^2_{\text{max}}]$:

$$R_H \equiv \frac{\int_{q^2_{\text{min}}}^{q^2_{\text{max}}} \frac{\mathrm{d}\mathcal{B}\left(B \to H\mu^+\mu^-\right)}{\mathrm{d}q^2}\mathrm{d}q^2}{\int_{q^2_{\text{min}}}^{q^2_{\text{max}}} \frac{\mathrm{d}\mathcal{B}\left(B \to He^+e^-\right)}{\mathrm{d}q^2}\mathrm{d}q^2}, \tag{1.52}$$

Here, B represents a given B-hadron, and H a given strange hadron or set of hadrons (such as a K, K^*, or Λ). Due to lepton universality, these ratios are predicted to be close to unity under the SM. The small differences in branching fractions due to the differing muon and electron masses only become relevant at very low values of q^2, approaching the muon kinematic threshold of $q^2 = 4m_\mu^2 \approx 0.045\,\text{GeV}^2$. In contrast to the individual differential branching fractions, these ratios are extremely theoretically clean under the SM. Uncertainties due to hadronic form-factors cancel to the $\mathcal{O}(10^{-4})$ level, with the largest remaining uncertainties of $\sim 1\%$ relating to soft QED effects [142–144].

Lepton universality ratios are sensitive to NP effects which couple to muons and electrons with different strengths. Furthermore, ratios with different hadrons in the final-state can provide complementary information on potential NP. For example, considering only the leading terms from interference between SM and NP amplitudes, one can find the ratios with K and K^* mesons are given by [145]:

$$R_K = 1 + \Delta_+ \tag{1.53}$$

$$R_{K^*} = 1 + p\left(\Delta_- - \Delta_+\right) + \Delta_+ \tag{1.54}$$

Here, p relates to the polarisation of the K^*, and is measured to be close to unity [145]. The interference between the SM and NP contributions is described by $\Delta_\pm$:

$$\Delta_\pm = \frac{2}{|\mathcal{C}_9^{SM}|^2 + |\mathcal{C}_{10}^{SM}|^2}\Big[\mathcal{R}e\left(\mathcal{C}_9^{SM} \cdot \left(\mathcal{C}_9^{NP\mu} \pm \mathcal{C}_9^{\prime\mu}\right)^*\right) + \mathcal{R}e\left(\mathcal{C}_{10}^{SM} \cdot \left(\mathcal{C}_{10}^{NP\mu} \pm \mathcal{C}_{10}^{\prime\mu}\right)^*\right) - \left(\mu \to e\right)\Big] \tag{1.55}$$

So, any measured differences between R_K and R_{K^*} could give information on the right-handed operators $\mathcal{C}'_9$ and $\mathcal{C}'_{10}$.

Measurements of such ratios are currently dominated by the LHCb experiment. Owing to difficulties modelling the experimental backgrounds at high values of q^2, all LHCb measurements have considered regions of q^2 below the J/ψ resonance. Since the first such measurement in 2014 [146], the ratio R_{K^+} has been of particular interest. The most recent measurement, made using all 9 fb^{-1} of data recorded by LHCb to date, found [147]:

$$R_{K^+} = 0.846^{+0.042}_{-0.039}(Stat.)^{+0.013}_{-0.012}(Syst.) \quad \text{for } 1.1 < q^2 < 6.0\,\text{GeV}^2 \tag{1.56}$$

This 3.1σ tension with the SM expectation raised cautious excitement within the theoretical and experimental physics communities, and attracted some media interest [148–150].

The LHCb experiment has also measured the ratio $R_{K^{*0}}$, using 3 fb^{-1} of data recorded in 2011 and 2012 (referred to as "run-1") [151]. Two bins of q^2 were used in this measurement - one the same as used for R_{K^+}, and another at low values of q^2 where universality-breaking effects from the differing muon and electron masses are more significant. This measurement found:

$$R_{K^{*0}} = \begin{cases} 0.66^{+0.11}_{-0.07}(Stat.) \pm 0.03(Syst.) & \text{for } 0.045 < q^2 < 1.1\,\text{GeV}^2 \\ 0.69^{+0.11}_{-0.07}(Stat.) \pm 0.05(Syst.) & \text{for } 1.1 < q^2 < 6.0\,\text{GeV}^2 \end{cases} \tag{1.57}$$

Depending on the exact theoretical assumptions, these two results are compatible with SM expectations at the $2.1-2.3\sigma$ and $2.4-2.5\sigma$ levels, respectively. In Fig. 1.10, the LHCb measurements of R_{K^+} and $R_{K^{*0}}$ are compared to theoretical expectations under the SM and in a scenario with NP in $\mathcal{C}^{\mu}_9$.

Finally, the LHCb experiment has measured lepton universality in $\Lambda^0_b \to pK^-\ell^+\ell^-$ decays, representing the first test of lepton universality using baryons [152]. This measurement used 4.7 fb^{-1} of data recorded in run-1 and 2016, finding:

$$R_{pK} = 0.86^{+0.14}_{-0.11}(Stat.) \pm 0.05(Syst.) \quad \text{for } 0.1 < q^2 < 6.0\,\text{GeV} \tag{1.58}$$

While this result is consistent with lepton universality within 1σ, it nonetheless shows the same pattern as in other ratios, with a deficit of muon-mode decays relative to electron-mode decays. Due to a rich resonance structure for the final-state hadrons, interpreting this result in the effective field theory framework is difficult, though possible under certain theoretical assumptions [153].

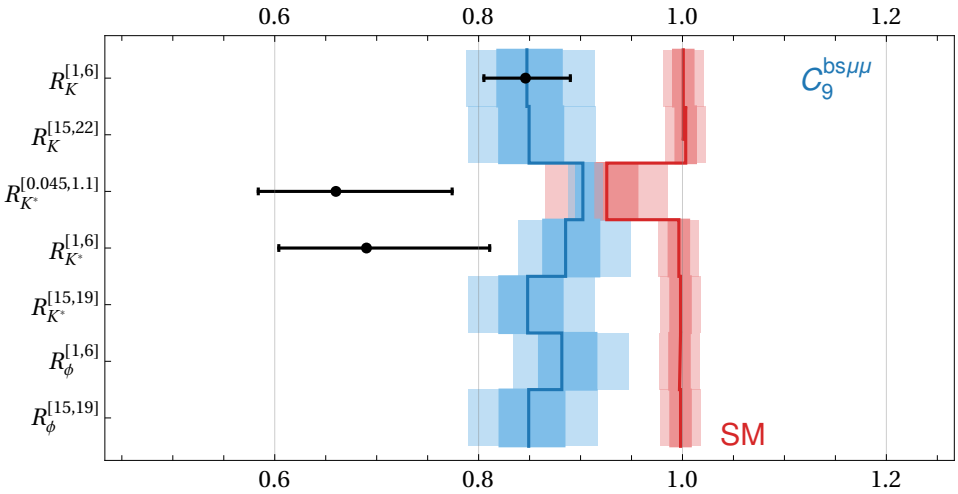


Fig. 1.10: Predicted values for various R_H ratios in different regions of q^2 under the SM (red), and under a model with NP in $\mathcal{C}_9^\mu$ (blue), compared to values measured by LHCb (black) [139]. Note that the NP predictions have larger theoretical uncertainties, as the uncertainties related to hadronic form-factors do not cancel as much as in the SM. In addition, note that the SM prediction for R_{K^*} in $0.045 < q^2 < 1.1\,\text{GeV}^2$ deviates from unity, as universality is broken by the differing masses of the electron and muon in this region.

$R_{K_S^0}$ and $R_{K^{*+}}$ The Belle experiment has measured R_{K^+} and $R_{K^{*0}}$ in several bins of q^2, though with markedly higher statistical uncertainties than the LHCb results [154, 105]. In the same papers, Belle also reported lepton universality ratios for the isospin partners of these decays: $R_{K_S^0}$ and $R_{K^{*+}}$. Prior to the work outlined in this dissertation, these were the only measurements available for $R_{K_S^0}$ and $R_{K^{*+}}$. The K_S^0 meson was reconstructed from the decay $K_S^0 \to \pi^+\pi^-$, and the K^{*+} from $K^{*+} \to K_S^0\pi^+$. Such decays can be experimentally challenging to reconstruct, as the K_S^0 decay position is often displaced from the B-hadron decay position due to its relatively long lifetime ($c\tau(K_S^0) = 2.7\,\text{cm}$, and $c\tau(B^0) = 560\,\mu\text{m}$) [13]. The values found in regions of q^2 below the J/ψ resonance were:

$$R_{K_S^0} = 0.55^{+0.46}_{-0.34}(Stat.) \pm 0.01(Syst.) \quad \text{for } 1.0 < q^2 < 6.0\,\text{GeV}^2 \tag{1.59}$$

$$R_{K^{*+}} = \begin{cases} 0.62^{+0.60}_{-0.36}(Stat.) \pm 0.07(Syst.) & \text{for } 0.045 < q^2 < 1.1\,\text{GeV}^2 \\ 0.72^{+0.99}_{-0.44}(Stat.) \pm 0.14(Syst.) & \text{for } 1.1 < q^2 < 6.0\,\text{GeV}^2 \end{cases} \tag{1.60}$$

All these measurements are compatible with unity within 1σ. The BaBar experiment has also carried out lepton universality ratio measurements using $B^0 \to K_S^0\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays [106]. However, they only report R_K and R_{K^*}, where these decays have been combined with their isospin partners $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$. It should

be noted that neither Belle, BaBar, nor any other previous experiments have reported an unambiguous observation of the electron-mode decays $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$.

Lepton universality in angular observables The angular distributions of electron-mode and muon-mode decays are also expected to be almost identical, except for small differences due to the differing lepton masses. One can test lepton universality by comparing a set of angular observables in these decays, then calculating, for example, $Q_i \equiv P_i'^{\mu} - P_i'^{e}$ [155]. As for lepton universality ratios, the theoretical uncertainties on the SM expectations of $P_i'^{\mu}$ and $P_i'^{e}$ arising from hadronic form-factors and the charm-loop effect largely cancel, meaning Q_i have very small theoretical uncertainties. Testing lepton universality using angular observables is currently a nascent area of study. The Belle experiment has reported the only results, measuring Q_4 and Q_5 in several bins of q^2 [118]. All results are compatible with SM expectations of $Q_i = 0$.

1.2.4 Global averages and theory interpretation

Tensions with SM expectations can now be seen in a host of different $b \to s\ell^+\ell^-$ observables, but are they pointing to a coherent picture? If so, what sort of NP effect could cause these tensions? These questions will be considered in this sub-section.

Global fits of $b \to s\ell^+\ell^-$ observables can be used to examine the structure of potential NP in a model-independent manner. Using the effective field theory outlined in Sec. 1.2.1, one first calculates a set of observables, plus their associated theoretical uncertainties, in terms of a set of Wilson coefficients $\mathcal{C}_i^{(\prime)}$. One then defines a test statistic which measures the compatibility between the predicted values and the measured values for this set of observables, at a given set of $\mathcal{C}_i^{(\prime)}$ values. For example, this test statistic may be a χ^2 or likelihood-ratio. Then, by scanning across different values of $\mathcal{C}_i^{(\prime)}$, one can find the point in parameter-space which gives the best fit between the theoretical and measured observable values, plus associated confidence intervals across $\mathcal{C}_i^{(\prime)}$.

Numerous such global fits have been carried out by different groups of physicists since the flavour anomalies started becoming apparent* [139, 157–159]. The exact methodologies of these global fits differ. For example, different sets of Wilson coefficients are considered, different calculations may be used for form-factors and charm-loop effects, and some observables or

*The references here are for the latest analyses from groups of physicists colloquially known as "the big four" of $b \to s\ell^+\ell^-$ global fitting. Numerous other groups have also carried out global fits, and a more comprehensive review is given in [156]. The reader is also invited to examine the citation lists for high-profile $b \to s\ell^+\ell^-$ results on https://inspirehep.net for a sense of the vast number of global fits and phenomenological interpretations which have been offered to explain these anomalies.

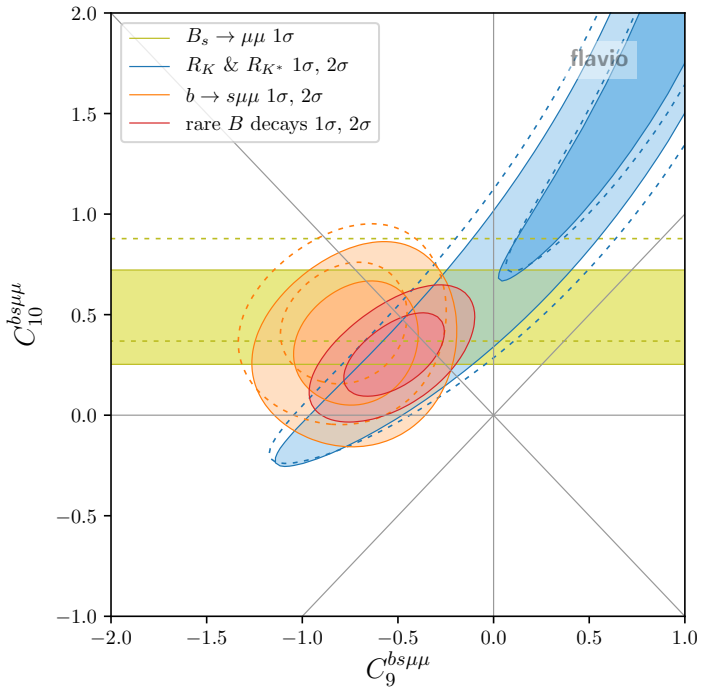


Fig. 1.11: Example of a global $b \to s\ell^+\ell^-$ fit for the Wilson coefficients $\mathcal{C}_{9\mu}^{NP}$ and $\mathcal{C}_{10\mu}^{NP}$ [139]. Note that the fits to $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions (yellow), lepton-universality ratios (blue), and $b \to s\mu^+\mu^-$ differential branching fractions and angular observables (orange) are in good agreement. Note also that the fit to all $b \to s\ell^+\ell^-$ observables (red) lies close to the line defined by $\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{10\mu}^{NP}$. The solid contours for the different observable combinations have coverage levels of 68% and 95%. The dotted contours show the fit results prior to the most recent LHCb measurements (as fully detailed in the source paper).

regions of q^2 may be excluded. Despite these differences, many of these global fits give very similar results.

In most global analyses, separate global fits are carried out using theoretically "clean" and theoretically "messy" observables, to check whether consistent results are obtained. Generally, the clean observables consist of lepton universality ratios R_H, and the $B_{(s)}^0 \to \mu^+\mu^-$ branching fractions, where hadronic effects are well understood. The messy observables consist mainly of differential branching fractions and angular observables, where uncertainties from form-factors and the charm-loop effect are much larger. Generally, the clean and messy fits favour consistent values of $\mathcal{C}_i^{(\prime)NP}$, which suggests there is a consistent NP cause for the deviations, rather than some misunderstood hadronic effect in the predictions of the differential branching fractions and angular observables. The good agreement between global fits by different groups, particularly when clean observables are used, suggests that global fits are fairly robust to different theoretical assumptions and statistical approaches [160, 161].

Scans across one or two different Wilson coefficients are usually performed at a time, on the assumption that large NP effects will only be seen in a small number of operators. Two single-parameter scenarios are strongly favoured, indicating NP coupling to the muons alone: $\mathcal{C}_{9\mu}^{NP} \approx [-1,-0.8]$ and $\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{10\mu}^{NP} \approx [-0.4,-0.5]$. These scenarios are favoured over SM values of $\mathcal{C}_{9\mu}^{NP}$ and $\mathcal{C}_{10\mu}^{NP}$ at the $\sim 4-7\sigma$ level. Some of the favoured two-dimensional scenarios have shifts to $[\mathcal{C}_{9\mu}^{NP},\mathcal{C}_{10\mu}^{NP}]$, or a mix of right-handed and left-handed operators $[\mathcal{C}_{9\mu}^{NP},\mathcal{C}_{9\mu}^{\prime NP} = -\mathcal{C}_{10\mu}^{\prime NP}]$. One example of a fit to $[\mathcal{C}_{9\mu}^{NP},\mathcal{C}_{10\mu}^{NP}]$ is shown in Fig. 1.11. A scenario with a new operator $\mathcal{C}_{i,\mu-\text{only}}^{NP}$ which couples only to muons, on top of an operator $\mathcal{C}_{i,e,\mu}^{NP}$ which couples to both muons and electrons, is also favoured, with $[\mathcal{C}_{9,\mu-\text{only}}^{NP} = \mathcal{C}_{10,\mu-\text{only}}^{NP},\mathcal{C}_{9,e,\mu}^{NP}]$.

The significances for the different models, showing how much they are favoured over the SM, are very useful for model-building purposes. However, they do not give an accurate metric for the "global" significance, *i.e.*, how much NP is favoured *in general* over the SM. This is because the basis of Wilson coefficients floated in the fit is picked according to how well it describes the data. Because of this, there is a significant "look-elsewhere effect", increasing the measured significance. To account for this, one can instead float a more complete basis of Wilson coefficients, rather than picking a basis which best describes the data. One analysis which did this, using pseudoexperiments to account for the look-elsewhere effect and with conservative assumptions for the charm-loop effect, estimated the global significance of NP over the SM to be 4.2σ [153, 162].

1.2.4.1 Simplified models

With information on the effects of NP at the m_b energy scale, one can attempt to infer what kinds of new high-mass particles could be responsible. Generally, simplified models are used for this, whereby a small number of new fields are simply added into the Lagrangian, rather than attempting to construct a full theory which is valid up to arbitrarily-high energies (or "UV-complete"). Such simplified models could mediate $b \to s\ell^+\ell^-$ decays at tree-level or at loop-level*. Assuming the coupling strengths of the new particles are $\mathcal{O}(1)$, one can find their masses must be of the scale [156]:

$$\Lambda_i^{Tree} \sim \frac{35\,\text{TeV}}{|\mathcal{C}_i^{NP}|^{\frac{1}{2}}}, \quad \Lambda_i^{Loop} \sim \frac{3\,\text{TeV}}{|\mathcal{C}_i^{NP}|^{\frac{1}{2}}} \tag{1.61}$$

So, a shift of $\mathcal{C}_{9\mu}^{NP} = -1$ could be explained by a new particle with a mass of ~ 35 TeV - far higher than the $\mathcal{O}(1\,\text{TeV})$ masses accessible via direct production at the LHC. Two popular

*Again, the references in this section are far from complete - more comprehensive discussions may be found in [156, 163].

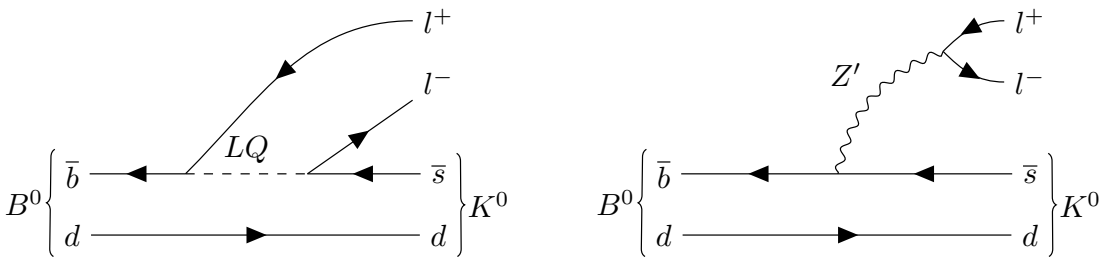


Fig. 1.12: Feynman diagrams showing potential NP tree-level contributions to $B^0 \to K^0 \ell^+ \ell^-$ decays from a leptoquark (left), and a Z' boson (right).

classes of simplified models are "Leptoquarks" and "Z' bosons". Both of these could contribute to $b \to s\ell^+\ell^-$ processes at tree-level, according to the Feynman diagrams shown in Fig. 1.12.

Z' bosons are massive, neutral, colourless gauge bosons. They may have different coupling strengths for quarks of different generations, or non-zero coupling strengths between quarks or leptons of different generations. For example, a Z' with a large $\bar{b}s$ vertex coupling strength, and different coupling strengths for the $\mu^+\mu^-$ and e^+e^- vertices, could be able to explain the $b \to s\ell^+\ell^-$ anomalies [164–170]. However, such a Z' boson would also give rise to four-quark and four-lepton vertices, modifying the amount of B_s^0-$\bar{B}_s^0$ mixing, and contributing to the neutrino trident production process ($\nu_\mu N \to \nu_\mu N \mu^+\mu^-$, where N is some nucleon) [171, 172]. The experimental results for these processes place constraints on the allowed masses and coupling-strengths of the Z' bosons, as do direct searches for resonances in $\mu^+\mu^-$ and e^+e^- invariant mass spectra at the LHC [173, 174].

Leptoquarks are bosons which carry both lepton number and baryon number, meaning they provide vertices where leptons and quarks directly couple to one another. Only ten different configurations of spin and other quantum numbers provide interaction terms which are invariant under the SM gauge group [175]. Of these, the scalar triplet (under $SU(2)_L$), vector singlet, and vector triplet leptoquarks could all contribute to $\mathcal{O}_9$ and $\mathcal{O}_{10}$ at tree-level, so are promising candidates to explain the $b \to s\ell^+\ell^-$ anomalies [176–184]. At lowest order, leptoquarks can only mediate B_s^0-$\bar{B}_s^0$ mixing via one-loop box diagrams, meaning data from this process is less constraining than for Z' bosons [171]. Searches for various decay signatures at the LHC set further limits, as leptoquarks can be singly-produced or pair-produced in pp collisions [13, Chapter 95, Leptoquarks].

In addition to the $b \to s\ell^+\ell^-$ anomalies, other lepton-related tensions with the SM have been observed which could have the same underlying NP cause. For example, some tensions with lepton universality have been seen in $b \to c\ell^-\overline{\nu}_\ell$ processes, mediated in the SM at tree-level by a W^- boson. For example, one can measure the lepton-universality ratio:

$$R_{D^{(*)}} \equiv \frac{\mathcal{B}\left(\bar{B} \to D^{(*)}\tau^-\overline{\nu}_\tau\right)}{\mathcal{B}\left(\bar{B} \to D^{(*)}\ell^-\overline{\nu}_\ell\right)} \tag{1.62}$$

Here, ℓ^- stands for either a muon or an electron. Combinations of measurements of R_D and R_{D^*} from the LHCb, Belle, and BaBar experiments show a 3.4σ tension with SM expectations [185]. Meanwhile, measurements and SM predictions of the anomalous magnetic moment of the muon $a_\mu = (g-2)_\mu$ show a long-standing tension. A recent measurement from Fermilab was combined with a previous result from Brookhaven National Laboratory [186, 187]. This experimental average shows a 4.2σ tension with a consensus SM prediction* [189]. Numerous models based on leptoquarks and Z' bosons aim to explain one or both of these anomalies, alongside the tensions in $b \to s\ell^+\ell^-$ [170, 176–180]. Some models also include possible candidates for Dark Matter [169, 181, 190]. Theorists may also make their models UV-complete, by having the new particle arise from some gauged symmetry group. For example, UV-complete Z' models are often constructed by gauging the difference between the muon and tau lepton numbers $L_\mu - L_\tau$ [167–169]. UV-complete leptoquark models can be constructed by gauging a Pati-Salam $SU(4)$ group, or the so-called "4321" group [178, 182–184].

1.3 Summary

The anomalies seen in $b \to s\ell^+\ell^-$ observables could be the first hints of physics from beyond the SM. However, more measurements are needed in order to say whether this is definitively the case, and to narrow down which NP models best describe these deviations. With $9\,\text{fb}^{-1}$ of data recorded, processed, and ready for analysis, the LHCb experiment is the best means of making these measurements in the near term. The LHCb detector is discussed in the next chapter.

*The uncertainty of this prediction is dominated by QCD-related effects, for which a mix of data-driven methods and Lattice QCD are used. It should be noted that a recent Lattice QCD prediction (the most precise to date) shows good agreement with the experimental average for a_μ [188].

Chapter 2

The LHCb experiment at the LHC

This chapter gives an overview of the Large Hadron Collider (LHC), and the LHCb experimental apparatus. Key features and sub-systems are highlighted, with a focus on those which are most relevant to the analysis presented in this dissertation.

2.1 The LHC

The LHC is a synchrotron particle accelerator, which straddles the Franco-Swiss border on the outskirts of Geneva [191, 192]. With a circumference of 26.7 km, and a design centre-of-mass energy of 14 TeV, it is the largest and most powerful particle accelerator ever built. Most of the time, the LHC is used to accelerate and collide protons. However, there is an additional "heavy ion" programme at the LHC, where lead and xenon nuclei are used* [193]. It was built (and is managed) by the European Organisation for Nuclear Research (CERN)†, in a system of tunnels buried $45 - 170$ m below the ground. These tunnels were originally constructed in the 1980s for the Large Electron Positron Collider (LEP), which then operated from 1989-2000. Following this, the LEP apparatus was removed, and the LHC was installed.

In this spirit of recycling, a chain of pre-existing accelerators is used to feed protons into the LHC, as illustrated in Fig. 2.1. Electrons are stripped from hydrogen molecules to produce protons, which are then accelerated to an energy of 450 GeV and injected into the LHC to circulate in opposite directions. The several hours during which beams circulate are known as a "fill", after which the beams are dumped.

*As the heavy ion programme are not directly relevant to this dissertation, and for brevity, only proton collisions are discussed from this point forward.

†This acronym is derived from the French name for its predecessor organisation - *La Conseil Européen pour la Recherche Nucléaire*. The laboratory complex near Geneva is also commonly known as CERN.

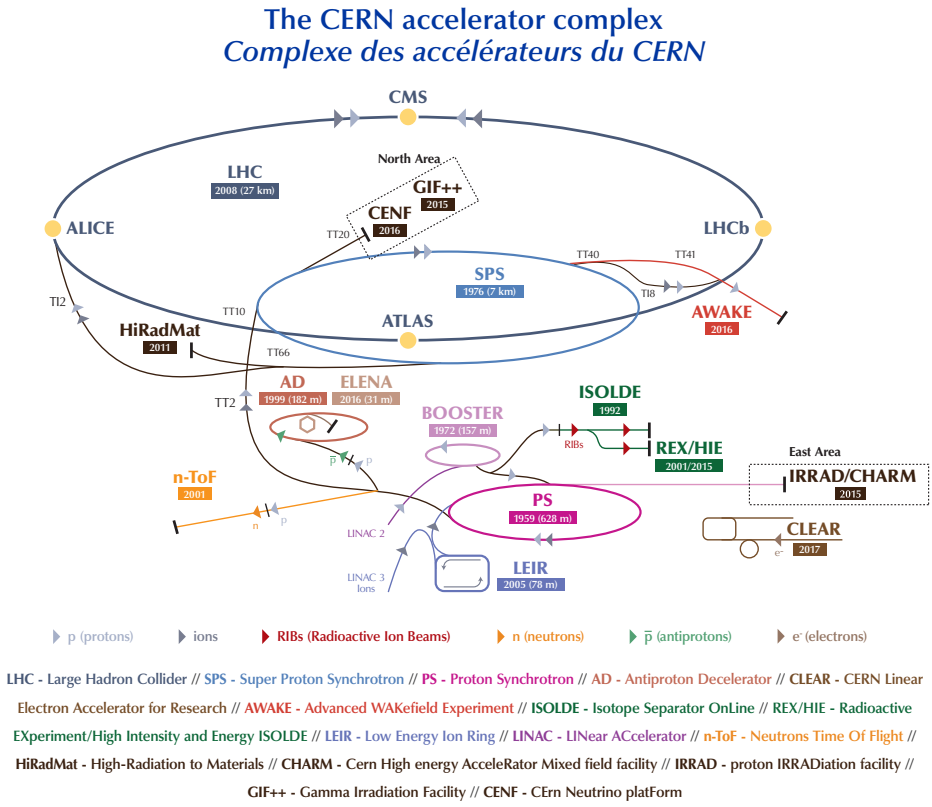


Fig. 2.1: The complex of particle accelerators at CERN, and some of the experiments which they serve [194]. Note that since this diagram was produced in 2018, Linac2 has been replaced with Linac4, which is capable of producing a more intense beam of protons [195].

Beams of protons are steered around the LHC ring using 1232 superconducting dipole magnets, each with a maximum field strength of 8.3 T. A system of 392 quadrupole magnets is used to focus the beams. The magnets are made of copper-clad niobium-titanium alloy, cooled to a temperature of 1.9 K using liquid helium. Sixteen radiofrequency (RF) cavities are used to accelerate the protons. These contain an electric field, oscillating at 400 MHz, which cause each proton's energy to increase by 16 MeV each time it orbits the LHC ring. The amount of force each proton is subjected to depends on its time of arrival in an RC cavity, due to the oscillation of the electric field. Some are subjected to no force, whereas those which arrive slightly earlier are decelerated, and those which arrive slightly later are accelerated. Because of

this, each beam is sorted into up to 2556 "bunches" containing up to $\sim 10^{11}$ protons*, separated by 25 ns time of flight in the ring† [196].

Once protons have been accelerated to the desired energy, and focussed into stable beams, the two beams are brought together at four "interaction points", where pp collisions can occur at a rate of 40 MHz. There are four major experiments situated at these interaction points, which are:

- **ATLAS and CMS** [197, 198] - Two "general-purpose" experiments, designed primarily to directly search for the Higgs boson and for particles originating from NP effects with masses at the TeV scale. Also designed for high-precision tests of electroweak physics and Higgs boson properties. Both experiments have active flavour physics programmes, and have published results using rare B-hadron decays. Each of the ATLAS and CMS collaborations contains ∼5,500 personnel‡ [199, 200].

- **ALICE** [201] - Designed to test QCD using quark-gluon plasmas produced in collisions of heavy ions. There are ∼2,000 personnel in the ALICE collaboration [202].

- **LHCb** [203] - Designed primarily to test the SM and indirectly search for NP effects by precisely measuring decays of hadrons containing b-quarks and c-quarks. This is carried out, for example, via measurements of the CKM matrix, and properties of the rare B-hadron decays outlined in the previous chapter. There are ∼1,400 personnel in the LHCb collaboration [204].

Proton beams were first circulated in the LHC in September 2008. However, within days a faulty electrical connection between two dipole magnets led to a magnet quench. This caused extensive mechanical damage to 53 dipole magnets and to the beampipe, which took over a year to repair [205, 206]. Proton beam tests resumed in November 2009, with the physics programme commencing in March 2010 [207, 208]. Collisions took place at a centre-of-mass energy of 7 TeV in 2010 and 2011, with an increase to 8 TeV in 2012. This period is referred to as "run-1" of the LHC. Following this, "long shutdown one" (LS1) took place, with collisions suspended so the detectors and accelerator complex could be upgraded [209]. Data-recording resumed for

*As a fill progresses, some protons are lost due to collisions, meaning the number of protons per bunch decreases.

†At this minimum bunch spacing, there are "slots" for 3564 bunches in the LHC beam, in principle. However, the beam is sorted into "trains" of 72 consecutive bunches, followed by several empty slots. This is due to the beam dynamics of the Proton Synchrotron and Super Proton Synchrotron, and the injection of beams into the LHC. Long sequences of empty bunch slots are also included to give sufficient time for a kicker magnet to be powered up, when the beam needs to be dumped.

‡The numbers of collaboration members quoted for the different LHC experiments are taken from public webpages. The author cannot guarantee that they were computed by consistent methods (*e.g.* whether technicians and undergraduate students are included in all these figures).

Year	Centre-of-mass energy / TeV	Integrated Luminosity / fb^{-1}		
		ATLAS	CMS	LHCb
2010	7	0.0450	0.0415	0.0377
2011	7	5.08	5.55	1.11
2012	8	21.3	21.8	2.08
2015	13	3.9	3.87	0.328
2016	13	35.6	38.3	1.67
2017	13	46.9	45.0	1.71
2018	13	60.6	63.7	2.19

Table 2.1: Centre-of-mass energies for pp collisions at the LHC in different years, and the integrated luminosities recorded by the ATLAS, CMS, and LHCb experiments [212–215].

"run-2" of the LHC, from 2015 to 2018, at a centre-of-mass energy of 13 TeV. Collisions were suspended again for further detector and accelerator upgrades in "long shutdown two" (LS2) [210]. As of March 2022, the accelerators and detectors are undergoing commissioning, with physics data-recording due to resume in mid-2022* [211].

The intensity of collisions between two particle beams is quantified using the "instantaneous luminosity", given by:

$$\mathcal{L} = \frac{fN^2}{4\pi\sigma^2} \tag{2.1}$$

Here, N refers to the number of particles in each bunch, f refers to the frequency of bunch collisions, and σ quantifies the width of the beam, perpendicular to the beam's direction of travel. The total number of particle collisions, and the amount of data recorded by a detector, is then quantified by the integrated luminosity $\mathcal{L}_{int} = \int \mathcal{L}\, dt$.

The integrated luminosities from pp collisions recorded by the ATLAS, CMS, and LHCb detectors throughout run-1 and run-2 are shown in Table 2.1. Note that the luminosities recorded by LHCb are markedly lower than those recorded by ATLAS and CMS. Because they study processes with low production cross-sections, the ATLAS and CMS experiments typically operate at the maximum possible instantaneous luminosity of $10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, with the beams colliding head-on. At each bunch collision, this leads to $\mathcal{O}(10-100)$ interactions between protons in colliding bunches, referred to as "primary vertices" (PVs).

Such a high instantaneous luminosity is not optimal at the LHCb detector, as this number of PVs would be difficult to reconstruct, and the high radiation would prematurely age sensitive

*Note that only the run-1/2 configuration of the LHCb detector is discussed in this chapter, as this was used to record the data analysed in this dissertation.

detector components. Instead, the luminosity delivered to the LHCb detector is reduced to $4 \times 10^{32}\,\text{cm}^{-2}\,\text{s}^{-1}$, by slightly defocussing the two beams to increase their widths. In addition, the two beams are slightly offset at the start of a run, so they do not collide head on. This reduces the number of PVs to ~ 1 per bunch collision, and reduces the amount of radiation the detector is subjected to. As a fill progresses, the numbers of protons in each bunch are depleted due to collisions, meaning the maximum possible instantaneous luminosity falls. Therefore, the offset between the two beams is adjusted as data-taking progresses, allowing the instantaneous luminosity to be held at a relatively constant level [216, 217]. This way, operating conditions remain stable throughout data-taking*.

2.2 The LHCb experiment

Prior to the commencement of the LHCb physics programme, the most precise measurements of B-hadron properties had predominantly been carried out by the two "B Factories"; the Belle experiment situated at KEK in Japan [220], and the BaBar experiment situated at SLAC National Accelerator Laboratory in the US [221]. These experiments collected data from 1999 to 2010 and 1999 to 2008, respectively. At the B Factories electron-positron collisions are carried out at a centre-of-mass energy equivalent to the mass of a $b\bar{b}$ resonance; typically the $\Upsilon(4S)$. The resonance then decays to a pair of B-mesons, whose properties can then be studied using a detector with full angular coverage. This approach offers a very clean environment, with little background from other collisions of the electron and positron. However, the cross-section for $\Upsilon(4S)$ production is relatively low, at $\sigma(e^+e^- \to \Upsilon(4S)) \approx 1.1$ nb [221].

By contrast, the production cross-section for $b\bar{b}$ pairs in pp collisions is $\mathcal{O}(10^5)$ higher, at $\sigma(pp \to b\bar{b}X) = 284 \pm 54\,\mu\text{b}$ for $\sqrt{s} = 7$ TeV [222]. This cross-section increases approximately linearly with $\sqrt{s}$. Therefore, even though Belle and BaBar operated with $\mathcal{O}(10-100)$ greater luminosity than the LHC [221, 223], far more $b\bar{b}$ pairs can be produced in pp collisions than at the B Factories. In addition, the energies of $b\bar{b}$ pairs at the LHC are high enough for them to hadronise into particles which cannot be produced from a $\Upsilon(4S)$ resonance, such as Λ_b^0-baryons and B_c^+-mesons. However, due to their composite nature, protons "shatter" into many particles upon collision, meaning collision-related backgrounds are much higher at the LHC than at the B Factories.

The ATLAS, CMS, and the B Factory detectors all use cylindrical designs with full angular coverage, so they can reconstruct particles flying in any direction. However, as shown in Fig.

*Such "luminosity levelling" using adjustable beam offsets was used at the LHCb and ALICE detectors throughout run-1/2. Luminosity levelling was also introduced at the ATLAS and CMS detectors for part of run-2, using a mixture of adjustable beam offsets, crossing angles, and widths [218, 219].

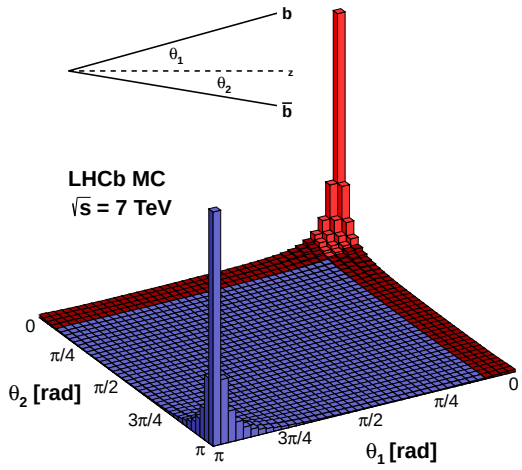


Fig. 2.2: The angular distributions of b and $\overline{b}$ quarks produced in pp collisions at the LHC, relative to the beamline. Collisions were simulated using PYTHIA, with $\sqrt{s} = 7$ TeV. The angular acceptance of the LHCb detector is shown in red [224].

2.2, $b\overline{b}$ quark pairs are highly boosted parallel to the LHC beamline, due to the production mechanisms available in pp collisions. Because of this, the LHCb detector is designed in a conical shape along the beamline, with the apex at the beam cross-over point [203, 226–229]. The detector layout is shown in Fig. 2.3. A unified set of coordinates is used for the LHCb detector elements and data; the origin is set at the beam cross-over point, with the z-axis parallel to the beamline. The x and y-axes are perpendicular to the beamline, pointing horizontally and vertically respectively. Its instruments cover angles of $10 < \theta < 300$ mrad in the horizontal direction, and $10 < \theta < 250$ mrad in the vertical direction, relative to the beamline. In collider physics, such angles often expressed as a dimensionless "pseudorapidity", given by:

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.2}$$

The LHCb detector has an acceptance of $2 < \eta < 5$, covering $\sim 4\%$ of solid angle. However, $\sim 25\%$ of $b\overline{b}$ pairs produced at the LHC fall within this region [224]. In total, the detector is 21 m long, 11 m high, and 13 m wide. There are two principal challenges for measurements of b-hadrons at the LHC. First, one must ensure that b-hadron decays can be efficiently resolved from pp collision backgrounds. Secondly, it must be possible to tell different stable charged particles apart*. This way, b-hadron decays with identical topologies can be told apart (for example $B^0 \to K^0_S \mu^+ \mu^-$ and $B^0 \to K^0_S \pi^+ \pi^-$). These challenges are tackled at LHCb using a

i.e. charged particles which do not decay before travelling past the end of the LHCb detector.

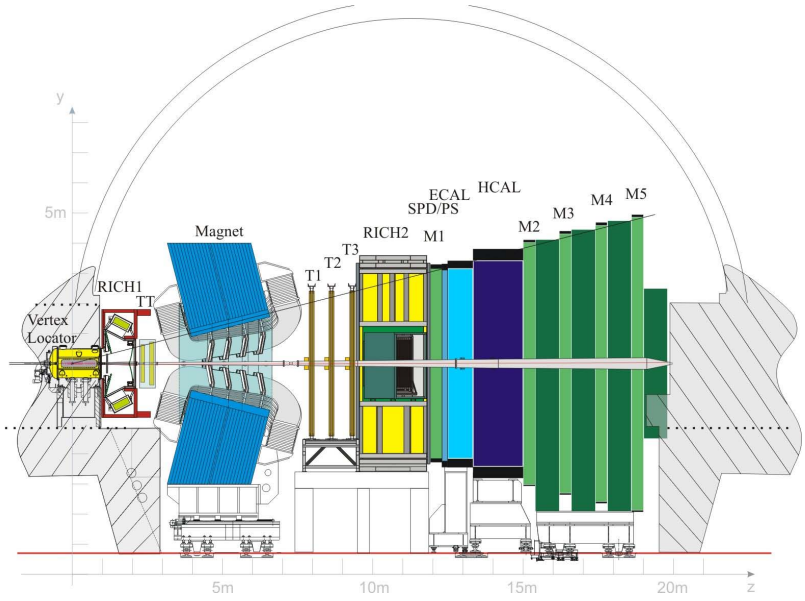


Fig. 2.3: Side-on view of the LHCb detector with various sub-systems highlighted [225].

precision tracking system, and a set of particle-identification (PID) sub-detectors, reviewed in the following sections. Features which are particularly useful for the study of $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays are highlighted.

2.2.1 Particle tracking

In a typical LHCb measurement, one wants to select pp collisions where a B-hadron was produced, and then decayed via a given topology (*i.e.* decayed to a given number of "final-state" particles in a certain order). Then, one usually measures quantities such as invariant masses, particle lifetimes, or angles between particles. In order to do this, one needs to know how far the B-hadron (and any intermediate particles) travelled before decaying, and the momenta and charges of final-state particles. For charged particles, this is accomplished using a particle tracking system.

Planes of sensitive materials are arranged perpendicular to the beamline, along the length of the LHCb detector (such as silicon semiconductors, or chambers filled with an easily-ionisable gas). As a charged particle travels through the detector, it interacts with each plane of material, causing an electrical signal (or "hit") which is then recorded. The materials are segmented, such that the locations of hits can be resolved. Often, the material can only be segmented along one dimension, so multiple planes are placed close together at different angles, allowing the three-dimensional location of each hit to be calculated. With multiple hits at different points in the detector, one can then reconstruct the path taken by the particle (known as a "track").

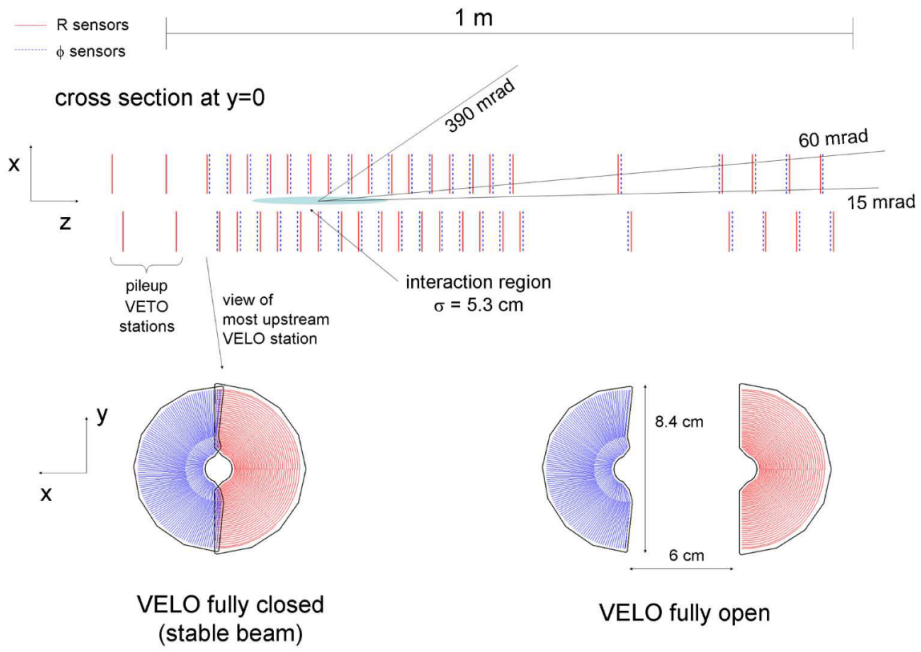


Fig. 2.4: Layout of the VELO modules along the beamline, plus the layout of a pair of modules when the VELO is fully closed and opened [203].

In addition, a dipole magnet is placed in the detector. As charged particles travel through the field from this magnet, their paths are bent. The particle's charge and momentum can then be determined from the track's direction and radius of curvature.

The LHCb tracking system is designed with relatively few tracking planes, and with light-weight materials. This reduces the "material budget" of the tracking system, in order to reduce particle scattering which would adversely affect the system's efficiencies and resolutions. A previous B-physics experiment - HERA-B, which took data in the early 2000s - suffered from poor detector performance due to high tracking material budget [230]. The components of the LHCb tracking system are outlined in the following sections, in rough order of their distances from the beam cross-over point.

2.2.1.1 Vertex Locator

With a boost typical of production at the LHC, B-hadrons only travel ~ 1 cm before decaying. Therefore, it is essential to resolve tracks originating from the primary vertex from tracks originating from the decay of the B-hadron (known as the "secondary vertex"). This separation is quantified by a track's "impact parameter" (IP), defined as its distance of closest approach to the PV. The *Vertex Locator* (VELO) sub-detector is designed for precise measurement of track IPs [231, 232].

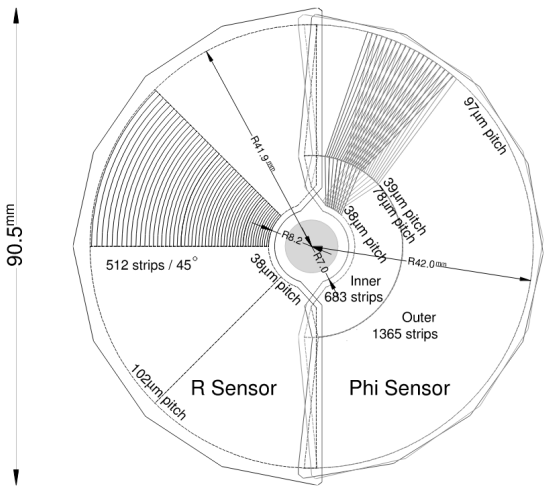
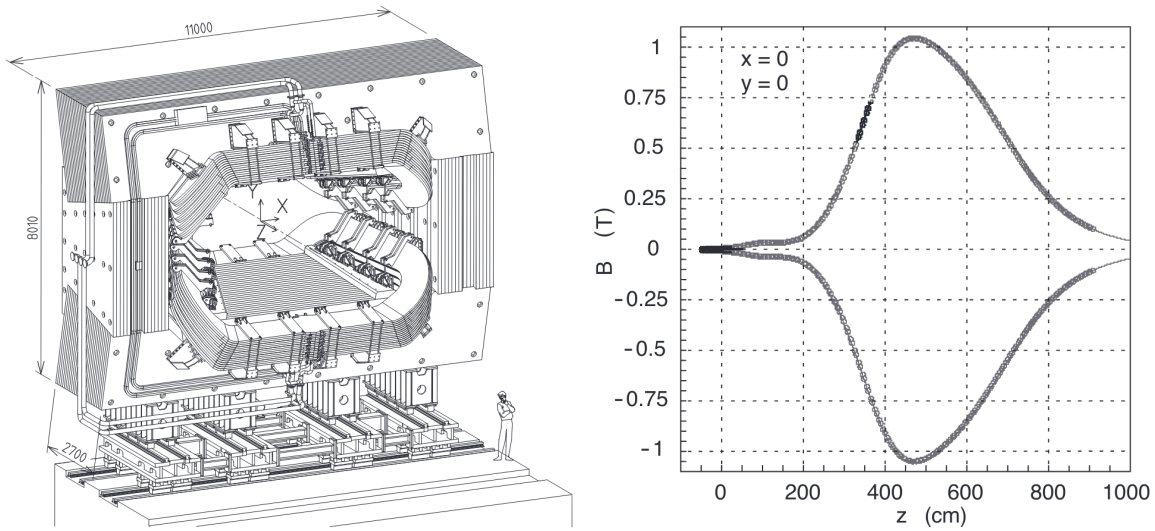


Fig. 2.5: Sketch showing the geometry of a pair VELO modules, highlighting the silicon strip layout in an R-sensor and a ϕ-sensor [203].

The VELO is a series of 21 "stations", arranged along 1 m of the beamline, and surrounding the beam cross-over point. This arrangement is shown in Fig. 2.4. Each station is made up of two semi-circular "modules", composed of two silicon strip sensors. As shown in Fig. 2.5, one strip sensor (the "R-sensor") is segmented in the radial direction, while the other (the "ϕ-sensor") is segmented in the azimuthal direction. This allows the two-dimensional locations of hits in each module to be determined. The VELO stations can be retracted along the x direction. While an LHC beam is initially being injected, they are withdrawn to a distance of 3 cm from the unfocussed beam to reduce radiation damage. Once the beam is stable and focussed, they are pushed in, sitting within 8 mm of the beam at their closest. The two sides of the VELO are offset by 1.5 cm along the z direction, allowing the modules at each station to overlap slightly, providing full azimuthal coverage.

To reduce material budget, the VELO sensors are held in a vacuum of $\sim 10^{-4}$ mbar. However, an ultra-high vacuum of $< 10^{-8}$ mbar is used by the LHC to prevent beam scattering. To prevent contamination of the LHC vacuum, an "RF box" made of aluminium with a thickness of ≤ 0.5 mm is used to separate the two vacua. The inner face of the box, known as the "RF foil", also shields the VELO electronics from electromagnetic waves induced by the beam.

Excellent tracking performance is achieved with the VELO [232]. Tracks can be reconstructed with efficiencies above 98%. A PV with 25 tracks originating from it can be located with resolutions of 13 μm in the $x-y$ plane and 71 μm along the z-axis, with some deterioration as the number of tracks decreases. For particles with momentum transverse to the beamline (known as "transverse momentum" or p_T) ≥ 1 GeV, the track IP can be measured with 35 μm resolution in the $x-y$ plane, with some deterioration as p_T decreases. These resolutions are far



(a): Magnet layout [203] ***(b):*** Magnet field strength [203]

Fig. 2.6: Left - sketch of the magnet looking towards the interaction point, showing its dimensions. Right - measured magnetic field strength at different positions along z, in the *MagUp* and *MagDown* polarities.

smaller than the ~ 1 cm typically travelled by a B-hadron, showing the VELO can effectively distinguish between particles originating from the PV, and particles originating from B-hadron decays.

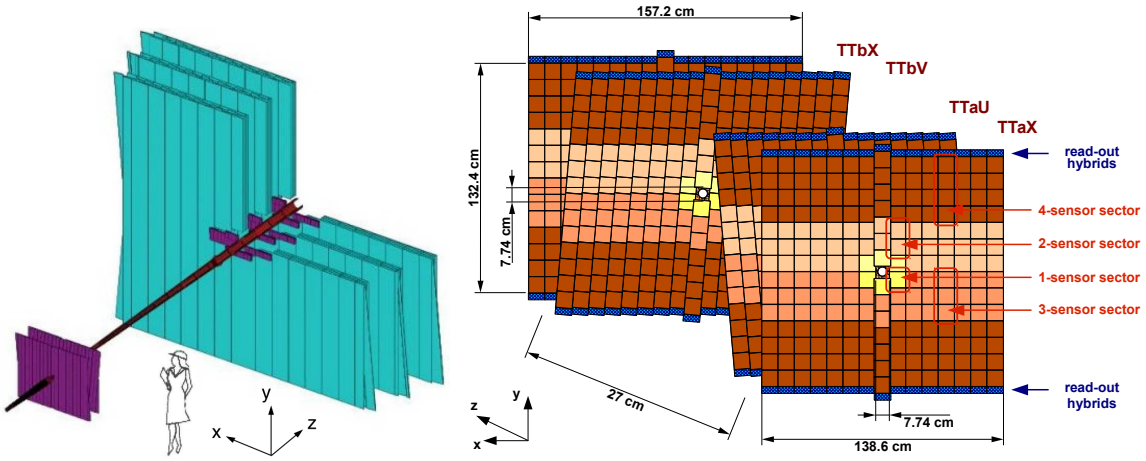
2.2.1.2 Magnet

A warm* dipole magnet is used to bend particle tracks so that their charges and momenta can be measured [233]. Shown in Fig. 2.6a, the magnet is placed ~ 5 m from the beam cross-over point. It is made up of a 1500 tonne iron yoke, and a pair of 27 tonne aluminium coils. Tracks of 10 m in length are subject to an integrated field strength of 4 T m.

The direction of the magnetic field is flipped approximately every two weeks, such that roughly equal amounts of data are recorded with "*MagUp*" and "*MagDown*" magnet polarities [234]. This way, any systematic effects from detector asymmetries are partially cancelled, which is particularly useful for studies of CP violation.

The field strength must be precisely mapped in order to accurately extrapolate tracks through the detector, and achieve good momentum resolution. A system of Hall probes was used to measure the magnetic field direction and strength across a grid of $8 \times 8 \times 10$ cm, to a relative precision of 4×10^{-4} [229]. The field strengths found for each polarity are shown in Fig. 2.6b.

i.e. non-superconducting.



(a): All tracking sub-detectors [203] *(b):* Layers of TT [235]

Fig. 2.7: Left - Arrangement of the LHCb tracking sub-detectors, showing the TT (purple, bottom-left), IT (purple, top-right), and OT (cyan). Right - Staggered view of the four layers in the TT. Similar arrangements of angled layers are also used in the IT and OT.

2.2.1.3 Tracker Turicensis and Tracking Stations

A set of four tracking detectors are situated upstream and downstream of the magnet, as shown in 2.7a. These measure the curvatures of tracks traversing the magnetic field [236, 237].

The Tracker Turicensis (TT) is located upstream of the magnet, and is particularly useful for reconstructing tracks which are later swept out of the LHCb acceptance by the magnetic field. As shown in Fig. 2.7b, it is $\sim 170 \times 130\,\text{cm}$ in size, and is made up of four layers of silicon microstrip detectors, spanning 32 cm along the z-axis. The middle layers are angled at $+5^\circ$ and -5° from the vertical, allowing two-dimensional hit locations to be resolved. Using readout strip pitches of $\lesssim 200\,\mu\text{m}$, a hit resolution of $50\,\mu\text{m}$ is achieved in the TT.

Three tracking stations (T1-3) are located downstream of the magnet, and cover areas of $\sim 600 \times 490\,\text{cm}$. However, the tracking stations use two different technologies. The region surrounding the beampipe is covered by the Inner Tracker (IT), with a cross-shaped configuration $\sim 120 \times 40\,\text{cm}$ in size. While this only represents $\sim 1\%$ of the tracking station area, $\sim 20\%$ of tracks produced close to the beam cross-over point pass through the IT, meaning good spacial resolution is essential to separate different particle hits [236]. The IT uses the same silicon strip technology as the TT, also with four layers of staggered material and strip pitches of $\lesssim 200\,\mu\text{m}$ to achieve a hit resolution of $50\,\mu\text{m}$.

Outside this region hit densities are lower, meaning such a high spacial resolution is not necessary. Because of this, and due to the high monetary cost of silicon detectors, gas straw tube detectors are used for the Outer Tracker (OT). As for the IT, each OT station is comprised

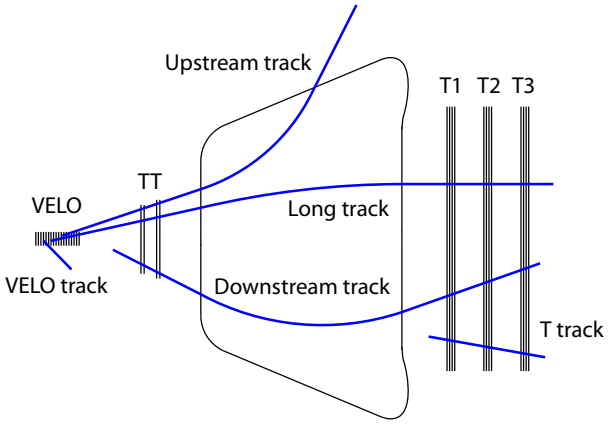


Fig. 2.8: Definitions of different types of tracks reconstructed using the LHCb detector [240].

of four layers, the inner two of which are angled at $\pm 5^{\circ}$. Each layer is made of two staggered arrays of straw tubes, allowing hit positions to be inferred from the drift times in different straws. Hit resolutions of 200 µm were found in run-1, with improved alignment and calibration giving hit resolutions of 170 µm in run-2 [238, 239].

2.2.1.4 Track reconstruction

Once hit positions have been calculated in the different tracking sub-detectors, one can reconstruct the trajectories and momenta of particle tracks. As shown in Fig. 2.8, there are various different "track types" at LHCb, defined according to which tracking elements a particle has passed through. Of particular interest for this analysis are "long" and "downstream" tracks.

- **Long tracks** pass through the full tracking system. They have hits recorded in the VELO and T-stations, and may have hits in the TT. These are reconstructed first by finding a set of straight tracks in the VELO. Each VELO track is then combined with information from the T-stations, to construct a track spanning the full length of the detector. The track is then extrapolated through the TT, and any hits which are consistent with it are included. Such tracks are reconstructed with efficiencies above 95%, with some dependence on momentum and detector occupancy* [241].

*The "business" of an event in the detector is referred to as the "occupancy". Proxies for the event occupancy include the number of PVs, number of tracks, and number of hits in the calorimeters. In general, the detector performance degrades as occupancy increases.

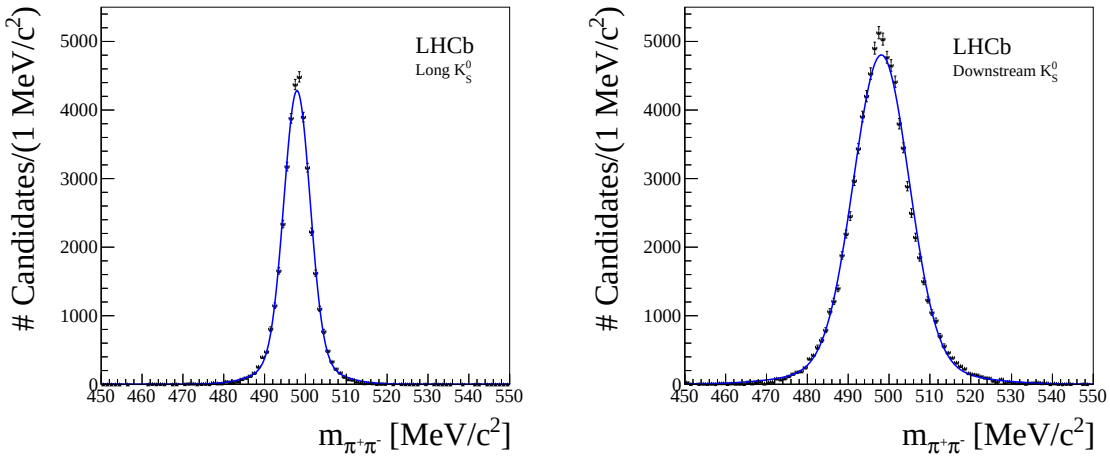


Fig. 2.9: Invariant masses of $K^0_S \to \pi^+\pi^-$ candidates reconstructed using pairs of long tracks (left), and pairs of downstream tracks (right) [229].

- **Downstream tracks** only pass through the TT and T-stations. Track segments are first constructed through the T-stations. Then, these segments are extrapolated back to the TT, and compatible hits are sought. Downstream tracks typically originate from decays of long-lived particles such as K^0_S-meson or Λ-baryon, which travel outside the VELO acceptance before decaying. Downstream tracks have reconstruction efficiencies of only $\sim 76\%$ [242]. All the same, two thirds of K^0_S tracks originating from B-hadron decays are reconstructed from pairs of downstream tracks [229], meaning they give an essential boost in statistics when studying decay processes with long-lived particles.

Various pattern recognition algorithms are used to find hits which could originate from a single particle, using simplified track parameterisations. Then, each track is fitted using a Kalman Filter algorithm [243]. This estimates the track's trajectory and momentum, accounting for scattering and energy loss as the particle traverses the detector [244–246]. The algorithm also provides a χ^2 metric which indicates the quality of the fit, allowing fake tracks to be rejected. Momenta for long tracks are measured with a resolution of 0.5% for particles with $p < 20\,\text{GeV}$, increasing to 0.8% for particles with $p \sim 100\,\text{GeV}$ [229]. Due to the missing VELO information, the momentum resolution is worse for downstream tracks. For instance, the invariant masses of $K^0_S \to \pi^+\pi^-$ candidates are measured with resolutions of 3.5 MeV using pairs of long tracks (or "LongLong" candidates), and 7 MeV using downstream tracks (or "DownDown" candidates), as shown in Fig. 2.9.

2.2.2 Particle Identification

Often, different decay processes leave similar signatures in the tracking sub-detectors, making them difficult to separate using information from these systems alone. For example, the decays $B^0 \to K^0_S \mu^+ \mu^-$ and $B^0 \to K^0_S \pi^+ \pi^-$ both have the same topology, and would peak in similar regions of reconstructed B^0 mass, owing to the similar masses of the muon and pion. Therefore, a set of dedicated PID sub-detectors are used to distinguish between different stable charged particles produced at LHCb; namely pions, kaons, protons, muons, electrons, and deuterons. These are the Ring Imaging Cherenkov (RICH) detectors, the calorimeters, and the muon chambers. These systems are reviewed in the following sections, as are the methods used to combine information from the different sub-detectors and quantitatively discriminate between the different particles.

2.2.2.1 Ring Imaging Cherenkov Detectors

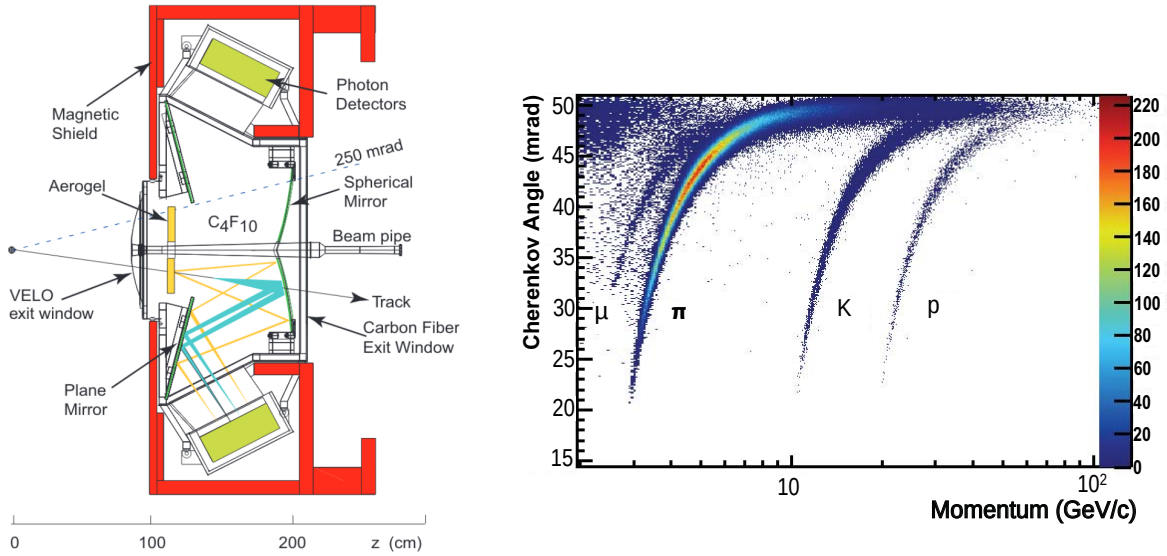
A pair of RICH detectors - RICH1 and RICH2 - are used chiefly to discriminate between different charged hadrons [247–249]. These operate by capturing Cherenkov radiation emitted by charged particles as they pass through the sub-detectors [2, pp. 17-18]. When a charged particle moves through a medium with refractive index n at a velocity $v > c/n$, polarisation and depolarisation of molecules in the medium causes photons to be emitted in a coherent wavefront*. The opening angle θ of this cone-shaped wavefront of Cherenkov radiation, relative to the particle trajectory, varies according to the velocity $v = \beta c$ of the charged particle:

$$\theta = \cos^{-1}\left(\frac{1}{n\beta}\right) \tag{2.3}$$

Therefore, by measuring θ for a particle one can calculate its velocity. By combining this with the momentum found using the tracking system, one can calculate the particle's mass, allowing one to discriminate between different charged particles. Note that this function is only defined for values of $n\beta < 1$, meaning there is a lower momentum threshold before each charged particle emits Cherenkov radiation. At high momentum, θ tends towards a fixed value for all particles, meaning they can no longer be told apart. Therefore, the choice of Cherenkov medium (or "radiator") in a RICH detector is informed by which particles one wishes to discriminate between, and what momentum range one wishes to be sensitive across.

The two RICH detectors at LHCb have similar constructions and layouts. Both consist of a gas radiator covering some portion of the LHCb acceptance. Cherenkov radiation is then reflected and focussed by a system of spherical and planar mirrors, onto an array of photon

*Here, c/n is the phase velocity of light in that medium



(a): RICH1 layout [203]

(b): RICH1 Cherenkov angle data [229]

Fig. 2.10: Left - Side-view of the RICH1 detector as found in run-1, including typical paths taken by Cherenkov photons emitted in the aerogel and C_4F_{10} radiators. Right - Distribution of reconstructed Cherenkov angles against track momentum, for photons produced in the C_4F_{10} radiator in RICH1, as recorded in data. The "band" structures arising from different particles are highlighted.

pixel detectors placed outside the RICH acceptance. Hybrid Photon Detectors (HPDs) are used for this purpose. These detectors are very sensitive to magnetic fields, so they are shielded using a mix of iron and mu-metal. In the RICH1 HPD plane, this reduces the stray magnetic field from $\sim 60\,\text{mT}$ to below $2.4\,\text{mT}$ [203].

Cherenkov photons from a given particle appear with a ring shape on the HPD arrays. Pattern recognition algorithms are used to reconstruct these rings, match them to tracks, and calculate emission angles [250, 251]. This can be challenging, as each Cherenkov ring may only consist of a few photons, and rings from different tracks may overlap (particularly when the detector occupancy is high).

The RICH1 detector is placed upstream of the magnet, and is designed to provide PID information from low-momentum particles (particularly those particles which are later swept out of the LHCb acceptance). The detector layout is shown in Fig. 2.10a. Initially, it was filled with a C_4F_{10} gas radiator covering a momentum range of $\sim 10-40\,\text{GeV}$, plus a slab of aerogel radiator covering a momentum range of $\sim 2-10\,\text{GeV}$. The aerogel was intended to provide separation between kaons and pions at $p \lesssim 10\,\text{GeV}$. However, the large numbers of tracks in RICH1 meant the aerogel did not give good PID information in practice, so it was removed after run-1 [249]. The Cherenkov angles recorded by RICH1 for tracks of various momenta

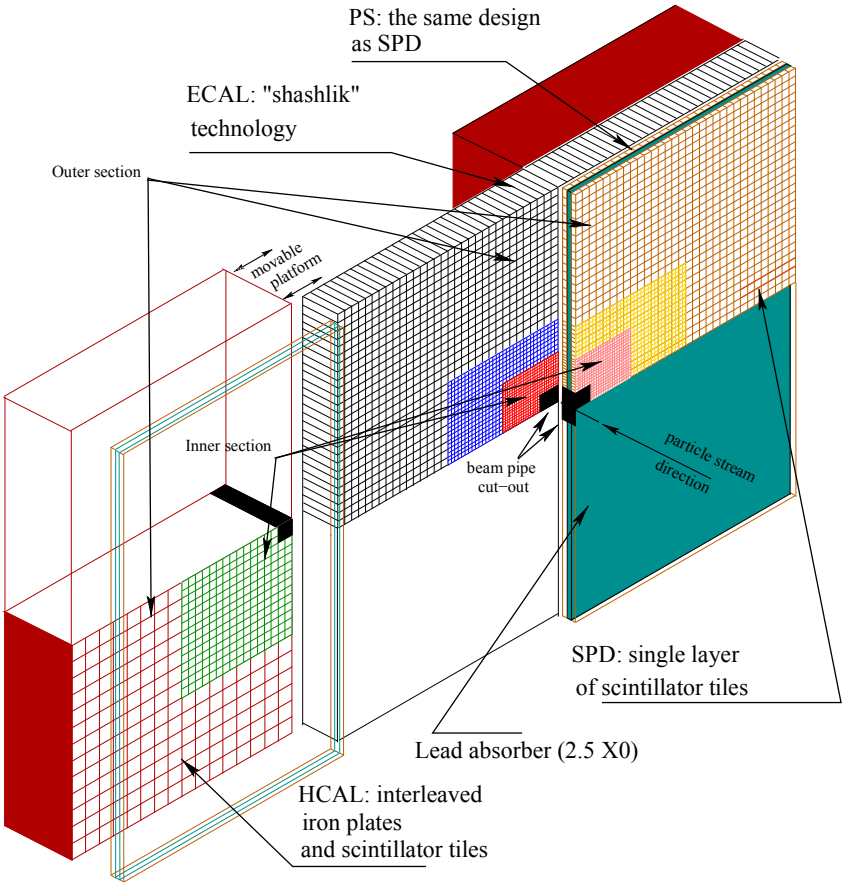


Fig. 2.11: Layout of the LHCb calorimeter system, highlighting the SPD, PS, ECAL, and HCAL, plus the segmentation of these sub-detectors [252]. Note that, as displayed on the page, the LHCb beam-crossing point would lie to the bottom-right of this diagram.

are shown in Fig. 2.10b, where one can clearly see a set of "bands" corresponding to different particles.

The RICH2 detector is placed downstream of the magnet. It is designed to provide PID information for high-momentum particles, using a CF_4 radiator with a momentum coverage of ~ 15 to ≥ 100 GeV. As high-momentum particles generally fall close to the beamline after the magnet, RICH2 has a reduced angular coverage of $15 - 120$ mrad in the horizontal plane, and $15 - 100$ mrad in the vertical plane.

2.2.2.2 Calorimeters

A system of calorimeters is used to identify hadrons, electrons, and photons, and to provide estimates of their energies* [253, 252]. These energy estimates are used to flag potentially interesting collision events for retention, in a process called "triggering". The calorimeter system is placed downstream of RICH2, with the layout shown in Fig. 2.11. Its main components are the Electromagnetic Calorimeter (ECAL), and the Hadron Calorimeter (HCAL) downstream of this. The former is designed to detect electrons and photons, and the latter to detect hadrons.

Both the ECAL and the HCAL use a "sampling" or "shashlik" design, comprised of alternating layers of a heavy material and a scintillating material along the z-axis [2, pp. 19-22]. When a high-energy electron moves through the heavy material, it may emit a photon via bremsstrahlung. If this photon is sufficiently energetic, it may pair-produce an electron and positron. These in turn may emit bremsstrahlung photons which may pair-produce, and so on. This process, known as an "electromagnetic shower" continues until all electrons have too little energy to emit bremsstrahlung, and all photons have too little energy to pair-produce. High-energy incident photons cause the same process, though starting with pair-production instead of bremsstrahlung emission. As charged particles in the electromagnetic shower move through the layers of scintillator, light is produced, which may be measured. The energy of the incident particle may be calculated from the number of charged particles in the shower, which can be measured using the light emitted as the charged particles pass through the scintillating layers. The length-scales for electromagnetic showers are characterised by the "radiation length".

The HCAL uses similar principles. However, hadrons instead interact with the heavy material via the strong force, with the resulting "hadronic shower" caused by hard collisions between the incident hadrons and nuclei in the HCAL material. In contrast to electromagnetic showers, hadronic showers contain many particle species (pions, kaons, protons *etc.*), in varying quantities. In addition, any π^0 mesons produced in these showers decay to photons and/or electrons, initiating small electromagnetic showers. This large variation in shower structure means the energy resolution is much worse in the HCAL than in the ECAL. The length-scales for hadronic showers are characterised by the "interaction length".

Another two detectors are placed just upstream of the ECAL; the Scintillating Pad Detector (SPD), and the Preshower Detector (PS). The SPD is a layer of scintillator material placed upstream of all other calorimeter material. This gives a means of discriminating between photons and charged particles, which is not possible using information from the ECAL alone. Behind the SPD, there is a 15 mm thick layer of lead, followed by another layer of scintillator.

*The calorimeter system is also used to identify and measure the energies of neutral pions, via their decays to pairs of photons.

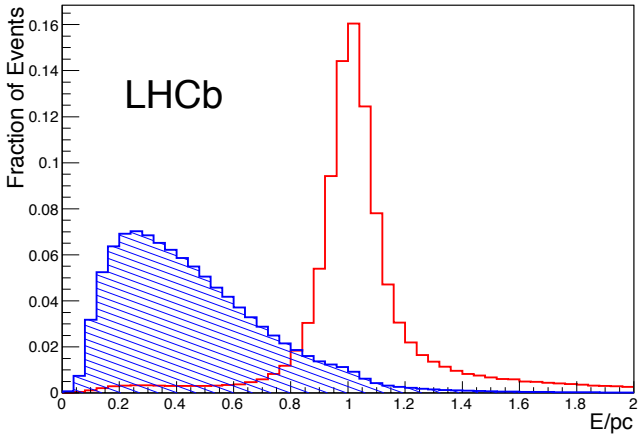


Fig. 2.12: Distributions of ECAL energies reconstructed data for electrons (red, open), and hadrons (blue, hatched), relative to the particles' track momenta [229].

The thickness of the lead corresponds to 2.5 radiation lengths, such that electrons and photons are highly likely to initiate an electromagnetic shower causing a signal in the PS scintillator, while hadrons and muons are unlikely to do so. By measuring the energy deposited in the PS, one can reject 99.6% of pions with $p = 20\,\text{GeV}$, while retaining 92% of electrons with this momentum [203].

The ECAL is made up of alternating layers of 2 mm thick lead plates, and 4 mm thick scintillator plates. These layers are repeated 66 times to give a total thickness of 42 cm, equivalent to 25 radiation lengths, and large enough to contain a full electromagnetic shower [254]. The energy resolution in each ECAL cell is given by [252]:

$$\frac{\sigma_E}{E} = \frac{(9 \pm 0.5)\%}{\sqrt{E}} \oplus (0.8 \pm 0.2)\% \oplus \frac{0.003}{E \sin\theta} \tag{2.4}$$

Here, E denotes the particle energy in GeV, and θ denotes the angle between between the z-axis, and a line drawn from the beam cross-over point to the ECAL cell. Note that charged hadrons may also cause showers within the ECAL. However, as illustrated in Fig. 2.12, hadrons generally deposit a much lower proportion of their energy than electrons, owing to their larger masses. Hence, the energy deposited in the ECAL can be used to discriminate between electrons and hadrons.

The HCAL is built from layers of 6 mm thick steel plates, and 3 mm thick scintillator plates. With a total depth of 1.65 m along z, corresponding to only 5.6 interaction lengths, the HCAL is not deep enough to capture a full hadronic shower [255]. The variation in hadronic shower structure leads to a much worse energy resolution than for the ECAL, at $\sigma_E/E = (67 \pm 5)\%/\sqrt{E} \oplus (9 \pm 2)\%$ (with E in GeV) [252]. The energy readings from the

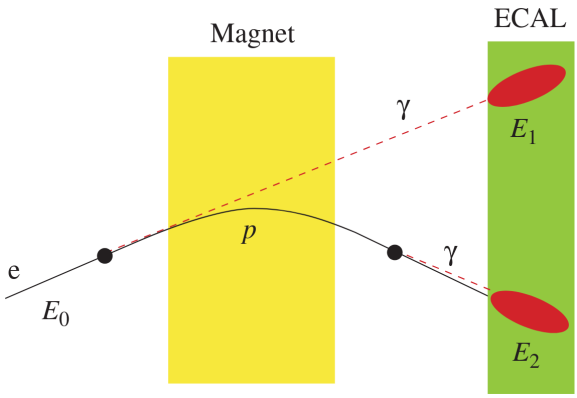


Fig. 2.13: Cartoon showing the effects of bremsstrahlung emission by electrons in different regions of the LHCb detector [228].

HCAL are chiefly used for triggering purposes, as for this one only needs to check whether the hadronic energy is above a certain threshold, rather than needing a completely accurate value.

All the calorimeter components are segmented in the x and y directions into sets of "cells", allowing energy deposits to be spacially resolved, and matched to particle tracks. The SPD, PS, and ECAL each contain 6016 cells, with lined-up angular coverages, while the HCAL only contains 1488 cells. Cells close to the beamline are smaller than those far away, to cope with the higher densities of particles hitting the calorimeters in this region. In each cell and for each sub-detector, optical fibres are used to transmit light from the scintillators to a set of photomultiplier tubes (PMTs), for electronic readout. As showers may travel through several adjacent cells, energy deposits are grouped together into "clusters" [256].

Bremsstrahlung emission and correction As a charged particle passes close to an atomic nucleus, it may emit a bremsstrahlung photon [2, p. 18]. The rate for this process in inversely proportional to the charged particle's mass, meaning the rate of emission by muons is $\sim 40,000$ times lower than by electrons. Therefore, electrons are the only stable particle examined at LHCb for which bremsstrahlung emission is a significant issue.

Due to the magnetic field, a bremsstrahlung photon will have different effects on electron reconstruction depending on where it was emitted within the LHCb detector. If the photon is emitted upstream of the magnet, the electron's momentum will be reduced accordingly before its trajectory is bent by the magnetic field. Therefore, the momentum calculated from the track curvature will be lower than the electron's true momentum when it was produced. Moreover, the emitted photon will continue along the electron's original trajectory, meaning the electron and photon will hit different calorimeter cells, as shown in Fig. 2.13. This leads to a degradation in the momentum resolution for electrons, relative to muons. In contrast, if photon emission

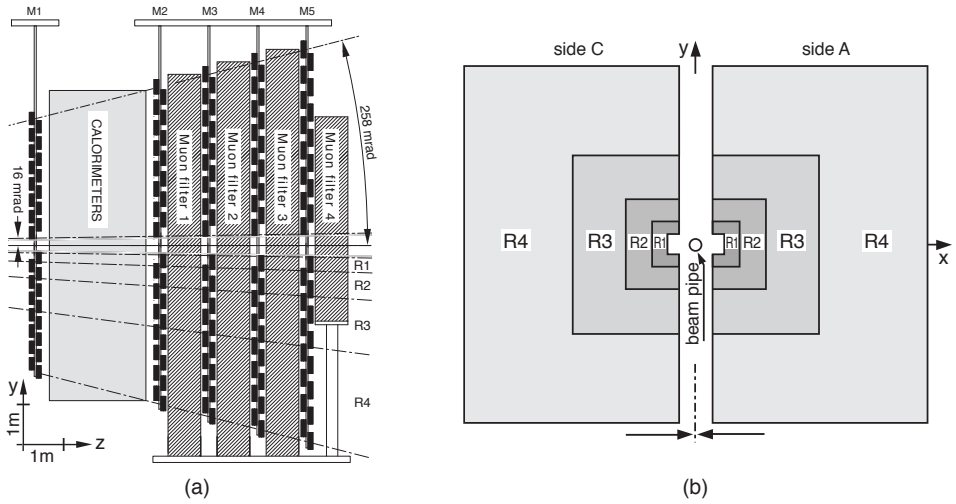


Fig. 2.14: Side-views of the muon system. Left - view of z, y plane, showing the positions of the muon stations and iron absorbers, relative to the calorimeters. Right - view of the x, y plane, showing the regions of each station where different chamber sizes are used, increasing from R1 to R4. [259].

occurs downstream of the magnet, the electron's momentum will be correctly calculated from its track curvature, and the electron and photon will hit the same calorimeter cell.

An algorithm is used to mitigate the effects of bremsstrahlung emission on measured electron momenta [257, 258]. The pre-magnet trajectory of an electron is extrapolated to the ECAL. Then, the momenta of any energy clusters in this region of the ECAL are calculated*, and added to the electron's momentum calculated from its track curvature. The region where ECAL clusters are searched for is defined using the track's trajectory at the VELO, the track's trajectory at the VELO, and the spacial resolution of the ECAL. This algorithm correctly recovers $\sim 90\%$ of bremsstrahlung photons emitted by electrons [258]. However, $\sim 20\%$ of photons recovered by the algorithm did not originate from bremsstrahlung, and are spuriously added to electrons, causing their reconstructed momenta to be too high. Overall, momentum resolutions remain poorer for electrons than for muons.

2.2.2.3 Muon system

A dedicated set of detectors is used to identify muons and measure their p_T values, mainly for triggering purposes [260, 259]. These five "muon stations" are essentially specialised tracking detectors, laid out as shown in Fig. 2.14. The stations are segmented into a total of 1,380 "chambers", further segmented into "logical pads", which allow hits to be spacially resolved.

*This momentum is calculated by assuming the cluster originates from the PV, and using the angles of the clusters relative to the beamline to get the x, y, and z components of the total cluster energy.

Finer segmentation is used closer the beamline, where particle flux is higher. Each chamber has an average hit efficiency of $\geq 99\%$, leading to a total efficiency of 95% for muon identification [229].

Because of their high masses, and the fact they do not interact via the strong force, muons are able to penetrate the detector material with little scattering. Therefore, four of the stations (M2-M5) are placed downstream of the ECAL and HCAL, which absorb the vast majority of electrons and hadrons. These stations use multi-wire proportional chambers (MWPCs) to detect muon hits. They are interspersed with 80 cm thick blocks of iron, to absorb any hadrons which have penetrated the calorimeters. Due to the presence of these absorbers, increasing momentum is required for muons to reach stations further from the beam cross-over point. For example, a momentum of $\gtrsim 3$ GeV is required to reach M1, while a momentum of $\gtrsim 6$ GeV is required to reach M5 [260, 203]. The chambers M4 and M5 are mainly designed to identify such highly-penetrating particles, so have coarser segmentation than M2 and M3, which are mainly used for p_T measurement.

The other muon chamber - M1 - is placed upstream of the calorimeter system. The addition of this detector improves the resolution of p_T measurements from 35% to 25% [259]. However, an MWPC detector would degrade too quickly in this region, owing to the high radiation levels from the large number of particles. Instead, a more radiation-hard triple gas electron multiplier (triple-GEM) detector is used for M1.

2.2.2.4 PID selection and performance

Information from the PID detectors is combined to provide several sets of variables, which analysts can use to select particles of interest. Those which are relevant to this analysis are described below.

A binary selection variable called `isMuon` can be used to select muon-like particles [261]. This is defined by extrapolating a track to the muon system, and checking whether any consistent hits are registered in each muon station. Hits are searched for within a momentum-dependent field-of-interest about the extrapolated track position. Because muons require high momenta to reach the final stations, `isMuon` uses different requirements in different momentum ranges. Tracks with $3 < p < 6$ GeV require hits only in M2 and M3. Those with $6 < p < 10$ GeV require hits in M2, M3, and then either M4 or M5. Finally, tracks with $p > 10$ GeV require hits in all stations M2-M5.

For each PID sub-detector, a set of likelihoods are calculated for each track. Each one describes the probability of observing a certain detector signature, if the track is from a certain particle species. These likelihoods are computed for all stable particles. The difference in the logarithm of the likelihood (or "log-likelihood") for the pion hypothesis, and the log-likelihood

for a different particle hypothesis x (*i.e.* a kaon, proton, muon, or electron) is computed. Such "$\text{DLL}_{x\pi}$ variables" are commonly used for particle selection. As the $\text{DLL}_{x\pi}$ variables have continuous distributions, analysts can choose a cut which provides optimal signal-efficiency and background-rejection for the decay-mode they are studying. To account for information from all PID sub-detectors, a "combined" $\text{DLL}_{x\pi}$ is used, found by summing the DLL values from different sub-detectors. The likelihoods for each sub-detector are found follows:

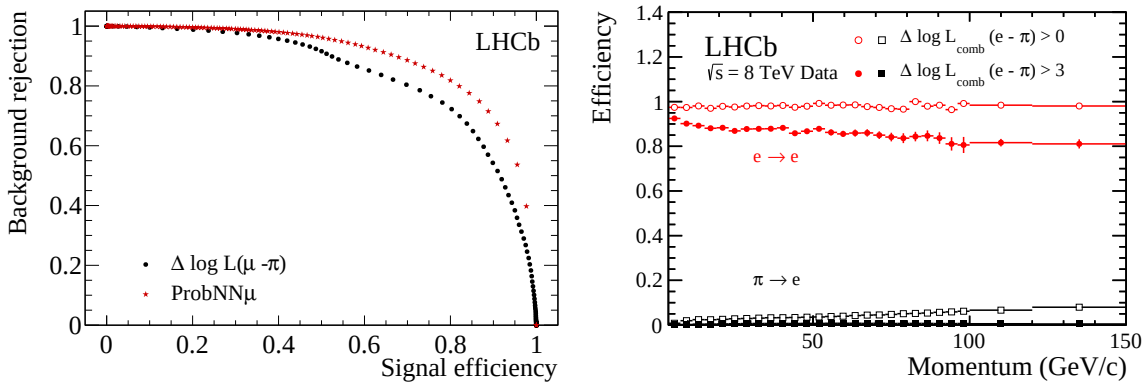
- **RICH** - A global likelihood is computed across both sub-detectors, describing how likely it is that all photons originated from all the tracks, under the assumption that each track is from a certain, hypothesised particle species. The hypothesised particle species are then varied for each track, until this global likelihood is optimised. This global approach ensures overlapping Cherenkov rings are properly accounted for. After this, the particle hypothesis can be changed for an individual track. The corresponding change in *global* likelihood is then taken as that track's *individual* RICH-DLL value, and used as an input for its combined $\text{DLL}_{x\pi}$ [248].

- **Calorimeters** - Likelihoods for electrons and hadrons are calculated using the energy deposited in the PS, ECAL, and HCAL (relative to the track momentum), plus a χ^2 metric describing the distance between an ECAL cluster and an extrapolated track position in the x, y plane [229].

- **Muon system** - The "average squared distance significance" is used, which gives a measure of the distances between muon station hits and extrapolated track positions [261].

In addition, a set of neural networks have been trained to separate particles of different species, the outputs of which are known as "`ProbNN` variables" [262, 229]. These were trained using PID information from all the sub-detectors, plus tracking information. This way, correlations between the responses of the different PID and tracking systems are accounted for, which is not the case for the summed global likelihood variables. As shown for example in Fig. 2.15a, this leads to improved separation of some particle species. The `ProbNN` variables are available in different "tunings", which provide optimal performance for different years and particle types.

Overall, the LHCb PID systems provide excellent separation of different particle types [263]. For instance, kaon can be identified with $\sim 95\%$ efficiency, with a $\sim 5\%$ probability of misidentifying a pion. Owing to the clear experimental signatures in their dedicated sub-detectors, muons can be identified with $\sim 97\%$ efficiency, at a $1-3\%$ probability of misidentifying a pion. In contrast, electrons are more difficult to identify, with efficiencies



(a): Muon PID performance [229]

(b): Electron PID performance [229]

Fig. 2.15: Left - Receiver Operating Characteristic (ROC) curve showing muon identification rates vs. proton misidentification rates obtained via cuts to $\mathrm{DLL}_{\mu\pi}$ and $\mathtt{ProbNN}(\mu)$ PID variables. Right - Electron identification efficiencies, and pion misidentification efficiencies found from two different cuts to $\mathrm{DLL}_{e\pi}$, as a function of track momentum.

of $\sim 90\%$ corresponding to $\sim 5\%$ probability of misidentifying a hadron. Of course, the efficiencies and misidentification probabilities depend on the exact PID variable cuts which are used, as illustrated in Fig. 2.15b. Efficiencies also vary with track momentum, due to the lower momentum-thresholds of the RICH, calorimeters, and muon chambers, and the fact that Cherenkov angles for all particles tend to $\cos^{-1}(1/nc)$ at high momentum.

2.2.3 Trigger

The vast majority of particle collisions at the LHC are not useful for the physics studied at LHCb. For example, at a typical operating luminosity of $4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and a centre-of-mass energy of 7 TeV, $b\bar{b}$ pairs are only produced at a rate of $\sim 100\,\mathrm{kHz}$. Only $\sim 25\%$ of these will fall within the LHCb detector acceptance, and only some decay-modes would be of physical interest. In addition, the storage and computing resources which would be needed to record the full 40 MHz of collision events are unfeasibly vast.

For these reasons, a "trigger" system is used to select events of interest for storage, so they can be used for later "offline" analysis [265–268]. This consists of three stages; first a hardware-based trigger (L0), then a two-stage software-based "High-Level Trigger" (HLT1 and HLT2). The trigger system reduced data-taking rates to 3 kHz in run-1 and 12.5 kHz in run-2, using the schemes illustrated in Fig. 2.16.

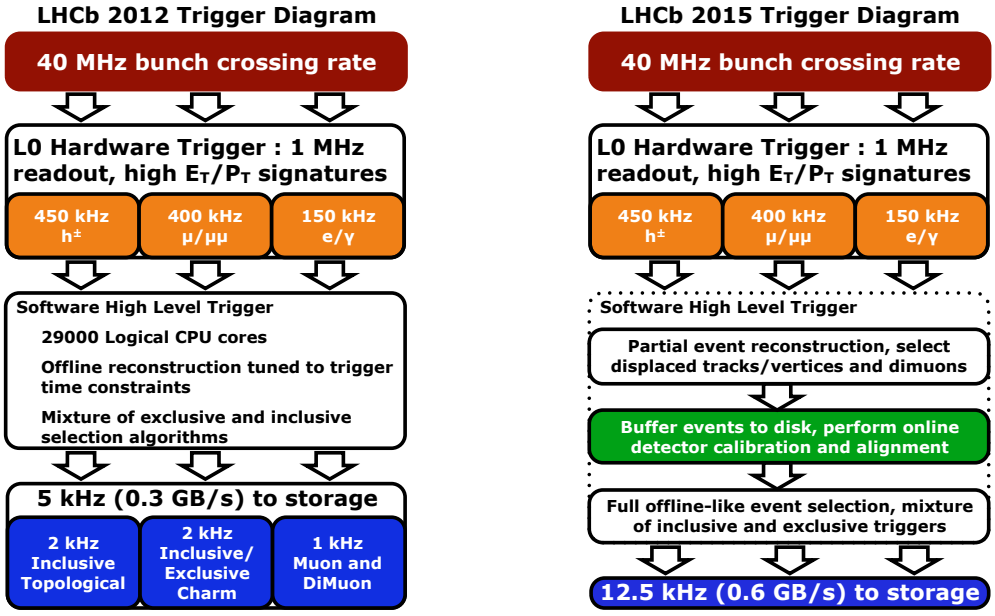


Fig. 2.16: Trigger schemes used in run-1 (left) and run-2 (right) [264].

2.2.3.1 L0 trigger

The L0 trigger reduces the rate from 40 MHz to 1 MHz, where the full detector information can then be read out. To do this, it must provide a decision within 4 µs. Only the calorimeters and muon stations can provide information suitable for triggering within this time. Using a system of Field Programmable Gate Arrays (FPGAs), the L0 trigger looks for signatures from high-energy particles, with different "trigger lines" for different signatures.

The lines of particular interest in this analysis are `L0Muon`, `L0Electron`, and `L0Hadron`. The `L0Muon` line searches for straight-line tracks through the muon stations M1-M5, with p_T above a certain threshold. The `L0Electron` and `L0Hadron` lines search for 2×2-cell clusters in the ECAL and HCAL, with a "transverse energy" above thresholds. For each cluster, this is calculated as $E_T = \sum_i (E_i \cdot sin\theta_i)$, where θ_i is the angle of a given cell from the beamline. The energy and momentum thresholds varied both within and between data-taking years, depending on detector conditions. Across run-1 and run-2 data-taking, the threshold for `L0Muon` varied between $750 < p_T < 1950$ MeV, for `L0Electron` between $1780 < E_T < 2960$ MeV, and for `L0Hadron` between $3000 < E_T < 3912$ MeV. Additionally, cuts were placed on the number of SPD cell hits (`nSPDHits`), to reject events with high detector occupancies which may take too long for the HLT to process [269]. For most lines, this cut was `nSPDHits` < 600 in run-1 and `nSPDHits` < 450 in run-2.

2.2.3.2 HLT

Events selected by the L0 trigger are then processed by HLT1 and HLT2. These are implemented using a C++ application called MOORE, running on a dedicated set of 1,700 computers (800 of which were added in run-2) which are capable of running $\sim 50,000$ single-threaded processes. As in the L0 trigger, HLT1 and HLT2 are divided into lines, designed to select processes with different detector signatures. The number of lines, and applied thresholds, varied between and within data-taking years.

In the first stage - HLT1 - a partial event reconstruction is carried out. Tracks are reconstructed using simplified algorithms, and PVs are located. In addition, muon PID information is associated to tracks, using the `isMuon` algorithm. Many of the lines run in HLT1 (including those used in this analysis) are designed to select interesting events with an "inclusive" strategy, admitting large varieties of decays. For example, they may accept events which contain at least one track which is well separated from the PV (according to its IP), and has high transverse momentum [270, 271]. This suppresses events that only contain tracks which originate from the PV, and/or are too "soft" for the detector sub-systems to give useful information. On average, events were accepted by HLT1 at rates of 70 kHz in run-1, and 150 kHz in run-2.

At HLT2, sufficient computing power is available for events to be fully reconstructed, including all tracks, calorimeter clusters and PID information. In addition to the reduction in rates from HLT1, this is because some events are temporarily buffered to disk during data-taking, allowing them to be processed between LHC fills. In run-2, further improvements were made to the trigger, with real-time alignment and calibration of sub-systems. This allowed the HLT2 reconstruction to be of the same quality as the offline reconstruction. Full event reconstruction allows for more "exclusive" selection strategies, targeting specific decays. In addition, some semi-inclusive lines exist. These include the topological lines, which search for sets of 2, 3, or 4 tracks originating from a common B-hadron vertex. Multivariate cuts are applied to the distances of closest approach between the tracks (DOCA), the significances of the track IPs (χ^2_{IP}), the flight distance of the B-hadron vertex (FD), and several kinematic variables. Input tracks are required to have low fit χ^2 values, and PID requirements may be applied to select tracks from muons or electrons.

2.2.4 Software and simulation

After data from events passing the trigger is saved to disk, it is further processed to make it suitable for use by analysts, using a series of software applications [272]. These applications (and MOORE) are all based on a C++ framework called GAUDI, providing common methods for certain processing tasks, to prevent duplication of effort in the different applications [273].

First, events are reconstructed using the BRUNEL application. This finds and fits all the tracks and PVs in the event, and calculates the associated PID information discussed in Section 2.2.2.4. For simulated events (discussed later) tracks are also associated to the simulated particles which are responsible for them, in a process known as "truth-matching". This allows the true particle species and kinematic information for a simulated track to be accessed.

Following this, events are further processed using the the DAVINCI application, which reconstructs events under the hypothesis that they contain a certain decay process. Decay vertices are fitted for tracks which might originate from a common point. Sets of tracks and vertices are then combined into "candidates" for a certain decay process, then kinematic and geometric quantities are calculated. Once these candidates have been formed, the associated information can be saved to a ROOT file, allowing analysis of the data [274].

However, the full datasets recorded by LHCb would be too large for direct use by individual analysts. Instead, various loose requirements, designed to select certain decay modes, are applied centrally in a process known as "stripping". Analysts can then access events passing a certain "stripping line", to quickly retrieve candidates for a decay they are interested in. The stripping is occasionally re-run, in case analysts want to add new stripping lines to search for decays which were not considered in previous "stripping campaigns".

2.2.4.1 Simulation

Simulated collisions and decays - known as "Monte Carlo" (MC) events - are used to model the responses of the LHCb detector to different processes, and to calculate things such as selection efficiencies and invariant mass distributions. The GAUSS application manages the first steps of this simulation [275, 276]. For brevity, only the steps used to simulate B-hadron decays are mentioned here; other steps may be used for the simulation of different physical processes.

Collisions of protons, and the hadronisation of quarks produced in these collisions, are simulated using a version of PYTHIA tuned for LHCb requirements [277–279]. Collisions are simulated until one occurs which contains the B-hadron of interest (referred to as the "signal particle"). The decay of the signal particle to a specific final-state (or set of final-states) is simulated using EVTGEN [280]. Kinematic and angular distributions of the decay can be configured using a file known as a "DECFILE". Here, cuts to generated particle angles or kinematics may also be specified, to discard events with particles which fall outside the LHCb acceptance or have very low momenta. Such events would otherwise needlessly consume computing resources in further simulation and processing. Final-state radiation (FSR) is then simulated using PHOTOS [281]. The decays of all non-signal particles are simulated by PYTHIA.

Interactions of particles with the detector material (*e.g.* the ionisation of gas in the muon stations, emission of bremsstrahlung, and showering in the calorimeters), are simulated using GEANT4. The digitised responses of the LHCb detector are then simulated using a separate LHCb software application, known as BOOLE. These digitised responses are input to MOORE, which emulates the L0 and HLT trigger responses. Note that, in contrast to real data, simulated events which do not pass the trigger are retained, enabling the study of trigger efficiencies. From this point, simulated data follows the same reconstruction and stripping steps as for real data.

Fast simulation The production of MC consumes a large proportion of the CPU resources available to LHCb. These pressures are expected to increase in the future, as increased numbers of simulated decays are required to match increased amounts of recorded data. Therefore, several "fast-simulation" methods have been adopted to reduce the amount of CPU time required to produce MC samples [282].

One such method is REDECAY, which utilises the fact that, within each event, the detector response to the signal decay is mostly uncorrelated with the detector response to other particles from the collision [283]. First, collisions containing the signal particle of interest are simulated using PYTHIA. All other particles in the event are known as the "rest of the event". The passage of such particles is simulated using GEANT4, and the results are set aside for later use. Then, the decay of the signal decay is simulated with EVTGEN, and detector interactions modelled with GEANT4. This decay process is repeated $\sim 50-100$ times, producing an ensemble of detector responses for decays of the signal. These decay responses are each merged with the detector response for the rest of the event. Using this technique, MC samples can be simulated $10-20$ times faster than with the usual LHCb simulation methods. However, re-use of the rest of the event can lead to small correlations between events, particularly for variables related to the detector occupancy.

The RAPIDSIM software package models the kinematic properties of decays reconstructed at LHCb (and other detectors), without modelling other aspects such as the RICH or trigger responses [284]. First, the signal particle's momentum and pseudorapidity are randomly chosen according to theoretical expectations. Then, its decay is simulated using multi-body kinematic phase-space calculations [285], or a more physically-motivated model implemented in EVTGEN [280]. The particle momenta are then smeared, to mimic the kinematic signatures seen in the LHCb detector. Using this approach, millions of events can be simulated within seconds. However, as the full detector response is not known, such samples cannot be used to calculate reconstruction or selection efficiencies.

Part II

Measurements of $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$

Chapter 3

Analysis Strategy

This chapter, and the following chapters, outline tests of lepton universality using the decays $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$, in data recorded with the LHCb detector. They also outline measurements of the differential branching fractions of the electron-modes of these decays. This chapter outlines the general strategies used for these measurements. The following chapter details the core methodology of the analysis. Then, various systematic uncertainties and cross-checks are discussed, followed by a chapter giving the final results of the measurements.

Lepton universality is tested in the decays $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ with measurements of the ratios $R_{K^0_S}$ and $R_{K^{*+}}$ within a bin of q^2 (see Section 1.2.3.6). In order to measure these quantities, one needs to select a sample of data enriched with muon-mode or electron-mode decays, and the number $N_{Sel.}$ (or "yield") of such decays in this sample. Then, one estimates the efficiencies ε with which such decays are reconstructed in the detector and selected, using this to calculate the "true" number of decays which took place. However, as outlined in the previous chapter, electrons and muons leave very different signatures in the LHCb apparatus. Therefore, the ratios of (efficiency-adjusted) muon-mode and electron-mode yields could be subject to large systematic biases, if either signature is inadequately understood.

To mitigate such biases, the tree-level decays $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ are used as control modes. The branching fractions for $J/\psi \to \mu^+\mu^-$ and $J/\psi \to e^+e^-$ are consistent within $\sim 1\%$ [13]. Therefore, one can measure "double-ratios" of efficiency-adjusted yields, where systematic biases partially cancel between the signal and control modes*:

*Note that this is the inverse of the conventional R_X ratio, where the electron-mode signal yield is in the denominator. Because this yield is so small, it can cause instabilities when attempting to directly measure R_X with a maximum likelihood fit.

$$R_X^{-1} = \frac{N_{Sel}\left(B \to X e^+ e^-\right)}{N_{Sel}\left(B \to X J/\psi\left(e^+ e^-\right)\right)} \frac{N_{Sel}\left(B \to X J/\psi\left(\mu^+ \mu^-\right)\right)}{N_{Sel}\left(B \to X \mu^+ \mu^-\right)} \times \frac{\varepsilon\left(B \to X J/\psi\left(e^+ e^-\right)\right)}{\varepsilon\left(B \to X e^+ e^-\right)} \frac{\varepsilon\left(B \to X \mu^+ \mu^-\right)}{\varepsilon\left(B \to X J/\psi\left(\mu^+ \mu^-\right)\right)} \quad (3.1)$$

In addition, the differential branching fractions* are measured for the electron-mode decays within a bin of $q^2 \in [q^2_{Min.}, q^2_{Max.}]$, again using J/ψ decays as a control mode:

$$\frac{d\mathcal{B}}{dq^2}\left(B \to X \ell^+ \ell^-\right) = \frac{1}{q^2_{Max.} - q^2_{Min.}} \times \frac{N_{Sel}\left(B \to X \ell^+ \ell^-\right)}{N_{Sel}\left(B \to X J/\psi\left(\ell^+ \ell^-\right)\right)} \times \frac{\varepsilon\left(B \to X J/\psi\left(\ell^+ \ell^-\right)\right)}{\varepsilon\left(B \to X \ell^+ \ell^-\right)} \times \mathcal{B}\left(B \to X J/\psi\left(\ell^+ \ell^-\right)\right) \quad (3.2)$$

Note that these measurements are normalised by the branching fractions for the control-mode decays, calculated from [13]:

$$\mathcal{B}\left(B^0 \to J/\psi K^0\right) = (8.91 \pm 0.21) \times 10^{-4}, \quad \mathcal{B}\left(B^+ \to J/\psi K^{*+}\right) = (1.43 \pm 0.08) \times 10^{-3}$$
$$\mathcal{B}\left(J/\psi \to \mu^+ \mu^-\right) = (5.971 \pm 0.032) \times 10^{-2}, \quad \mathcal{B}\left(J/\psi \to e^+ e^-\right) = (5.961 \pm 0.033) \times 10^{-2}$$

The statistical significances of the measured numbers of $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ decays are also evaluated. All the measurements outlined above, and any associated uncertainties, are measured using the statistical procedures documented in Section 7.1, with the results given in Section 7.2. Systematic uncertainties on these measurements are evaluated in Chapter 5. A variety of cross-checks are carried out to ensure that efficiencies and yields are well understood, as documented in Chapter 6. These chiefly use the $B \to J/\psi\left(\ell^+ \ell^-\right) K^{(*)}$ control modes, as well as another lepton-universal, tree-level process; $B \to \psi(2S)\left(\ell^+ \ell^-\right) K^{(*)}$.

All the signal and control-mode decays are selected with cuts to q^2, using the scheme outlined in Table 3.1, which was originally devised for use in the run-1 LHCb measurement of $R_{K^{*0}}$ [151]. The selection requirements in this analysis are designed to optimise the electron-mode signal significances, as these have the greatest effect on the precision of the R_X^{-1} measurements. Signal $B^0 \to K^0_S \ell^+ \ell^-$ decays are selected in $q^2_{central}$, as studies indicated this would give the

*Use of the term "differential branching fraction" is perhaps contentious given these measurements are only carried out in one bin of q^2 for each mode, rather than multiple bins. However, this term is used to stress that absolute branching fractions are *not* reported. Instead, the reported values are normalised by the ranges of q^2 used in the measurements, making them "differential".

q^2 bin	Muon Modes		Electron Modes	
	Min. / GeV^2	Max. / GeV^2	Min. / GeV^2	Max. / GeV^2
low	0.045	1.1	0.045	1.1
central	1.1	6.0	1.1	6.0
J/ψ	8.98	10.21	6.0	11.0
$\psi(2S)$	12.86	14.33	11.0	15.0

Table 3.1: Dilepton invariant mass squared (q^2) bins used to select signal and control-mode decays in this analysis. Note that the J/ψ and $\psi(2S)$ muon-mode bins are found from $|q - m(J/\psi)| = 100\,\text{MeV}$ and $|q - m(\psi(2S))| = 100\,\text{MeV}$, using measured values of $m(J/\psi)$ and $m(\psi(2S))$ [13]. The equivalent electron-mode bins are much wider, to account for degraded resolution from bremsstrahlung emission and recovery.

optimal $B^0 \to K^0_S e^+ e^-$ significance (see Section 4.1.5.2). By contrast, $B^+ \to K^{*+}\ell^+\ell^-$ decays are selected in q^2_{low} and $q^2_{central}$, as studies indicated a range $q^2 \in [0.045, 6.0\,\text{GeV}^2]$ would give the optimal $B^+ \to K^{*+}e^+e^-$ significance. This is because rates of $b \to s\ell^+\ell^-$ decays with vector mesons in the final state are greatly enhanced at low values of q^2, due to the photon pole (see Section 1.2.2). However, there is a trade-off. The photon pole is lepton-universal, meaning the inclusion of q^2_{low} could potentially dilute the sensitivity of the $R^{-1}_{K^{*+}}$ measurement to new physics. Such dilution must be accounted for when comparing the measured $R^{-1}_{K^{*+}}$ value to theoretical expectations.

3.1 Data samples

Data recorded by LHCb during all of run-1 and run-2 is considered in this analysis. However, data from 2010 and 2015 is discarded, as relatively small amounts of data were recorded in these years, with very different trigger conditions to other years, making it difficult to calculate well-calibrated selection efficiencies.

Candidates of interest are selected using the requirements outlined in Section 4.1. Candidates for K^0_S mesons are reconstructed from pairs of charged pions, and K^{*+} candidates are reconstructed from the decay $K^{*+} \to K^0_S \pi^+$. The pion from the K^{*+} is known as the "soft pion" (π^+_{Soft}). In general, electron-mode decays are selected with $\sim 20\%$ the efficiency of muon-mode decays, with increased background contamination due to poorer resolutions and PID performance. Therefore, the precisions of R^{-1}_X measurements are dictated by $B \to K^{(*)}e^+e^-$ yields, so selection requirements are designed to maximise these. Muon-mode selection requirements are kept similar to electron-mode requirements, to minimise the potential for systematic biases.

Therefore, the muon-mode yields are lower than they could be in a dedicated analysis, so the muon-mode branching fractions are not measured in this analysis except as a cross-check outlined in Section 6.3.

Simulated samples of signal and control-mode decays, produced using the procedures in Section 2.2.4.1, are used throughout this analysis. Simulated samples are also used to estimate contamination rates from various physical background sources, and devise selection requirements which reject them (see Section 4.4). Various samples are produced which simulate conditions in different years. When samples for different years are combined, their relative proportions are weighted to match the recorded luminosities and $b\bar{b}$ cross-sections in data [286]. Similarly, samples are simulated for the *MagUp* and *MagDown* magnet polarities, and their relative proportions are reweighted to match data*.

These samples contain "reconstruction-level" information about the signatures which simulated events leave in the detector. They also contain information on the "true" kinematic and geometric behaviour of the particles in a decay process, known as "generator-level" information. At reconstruction-level, q^2 is calculated with $q^2 = |p_{\ell^+} + p_{\ell^-}|^2$. However, at generator-level $p_{\ell^\pm}$ is defined after emission of FSR. Instead, the definition $q^2 = |p_B - p_{K^{(*)}}|^2$ is used, to recover the value before emission of FSR by leptons.

3.2 Calculating efficiencies

Selection efficiencies in each year are calculated using MC samples, following the methodology in Section 4.3. However, some aspects of the detector response are difficult to simulate accurately. To account for such mismodelling, a set of data-driven corrections are applied, as outlined in Section 4.2. These mostly use weights w_i associated to each candidate, which transform overall variable distributions or describe the efficiency of a certain selection requirement. Correcting the MC samples in this way ensures that estimates of selection efficiencies are accurate.

Most of the time, the uncertainty on an estimate of the efficiency ε for a selection requirement can be calculated using the "normal approximation". Here, the efficiency estimate is assumed to be Gaussian-distributed, with mean $\varepsilon = k/N$. The associated uncertainty is given by [287, Sec. 8.5.2]:

$$\sigma_\varepsilon^2 = \frac{1}{N^2}\left(k_2\left(1 - 2\frac{k}{N}\right) + N_2\left(\frac{k}{N}\right)^2\right) \tag{3.3}$$

*Note that approximately equal amounts of data were recorded with each magnet polarity. To match this, approximately equal numbers of candidates were simulated for each magnet polarity. The reweighting of residual discrepancies in magnet polarity proportions hence amounts to a very minor correction.

$$k = \sum_{i}^{Passed} w_i, \quad N = \sum_{i}^{Tot} w_i, \quad k_2 = \sum_{i}^{Passed} w_i^2, \quad N_2 = \sum_{i}^{Tot} w_i^2 \tag{3.4}$$

This approximation gives good coverage for large k and N. However, it can give under-covering or unphysical uncertainties for $k \to N$ or $k \to 0$. This can occur when estimating efficiencies for background sources, where few candidates in MC pass full selection requirements. In such cases, a Bayesian method can be used to estimate the efficiency [288]. Here, the binomial distribution is taken as a likelihood function $P(k|\varepsilon,N)$. Then, Bayes' theorem is used to compute a posterior distribution $P(\varepsilon|k,N)$, with a uniform prior $P(\varepsilon)$. As this posterior distribution is asymmetric and non-Gaussian, standard uncertainty propagation cannot be used. Instead, a toy-based method is used, utilising the MCERP module in PYTHON* [290].

3.3 Calculating yields

Unbinned extended maximum likelihood fits are used to estimate the yields of signal and control-mode decays, while distinguishing these from various physical and detector-related backgrounds [291, pp. 83-57]. The likelihoods are constructed using the invariant masses of reconstructed B-meson candidates ($m(B)$), following the procedures outlined in Section 4.5. They can be denoted $\mathcal{L}(\vec{x}|\vec{\theta})$, where $\vec{x}$ represents a set of mass-values observed in data, and $\vec{\theta}$ represents the fit-model parameters (such as signal and background yields). The negative of the log of each likelihood (or "NLL" for short) is minimised using the MINUIT software library [292], to find a set of optimal parameter values $\vec{\hat{\theta}}$ with corresponding "maximum likelihood" $\mathcal{L}(\vec{x}|\vec{\hat{\theta}})$. The uncertainties on each optimal parameter estimate $\hat{\theta}_i$ can be found via two methods [287, pp. 233-238]:

1. **HESSE** - The NLL is assumed to be parabolic about $\hat{\theta}_i$. This can be shown to be true in the "asymptotic limit" $N \to \infty$, where N is the number of observations in $\vec{x}$. Uncertainties are found by numerically calculating the second derivatives of the NLL with respect to each θ_i. A matrix of second derivatives is then inverted, to find a covariance matrix, indicating the uncertainties on each $\hat{\theta}_i$, and correlations between different parameter estimates. Most fitted parameter uncertainties in this analysis are computed using HESSE.

2. **MINOS** - A parameter is held at a certain value $\theta_i = c$, and the likelihood is maximised with respect to all other parameters, yielding the "profile likelihood" $\mathcal{L}(\vec{x}|\theta_i = c, \vec{\hat{\theta}}_{j\neq i})$.

*Note that both the normal and Bayesian efficiency calculations are implemented in ROOT [289]. These implementations were ported to PYTHON, for use in this analysis.

This is calculated for various values of $\theta_i = c$, to construct a "profile likelihood scan". In the asymptotic limit, one can show that $\Delta_{PLL} = \ln\mathcal{L}(\vec{x}|\theta_i = c, \vec{\hat{\theta}}_{j\neq i}) - \ln\mathcal{L}(\vec{x}|\vec{\hat{\theta}})$ is distributed as a χ^2 distribution with one degree of freedom [293]. Therefore, the regions of $\theta_i = c$ where $\Delta_{PLL} < 2$ represent a 68.3% confidence interval for $\hat{\theta}_i$, providing a pair of asymmetric uncertainties on $\hat{\theta}_i$. MINOS is used to calculate uncertainties on the final results presented in Section 7.2, where the likelihoods are highly non-Gaussian, and symmetric uncertainties from HESSE would not be well-motivated.

In order to improve the resolutions of the B-candidate masses, they are computed using the `DecayTreeFitter` method [294]. Here, all particles in a decay process (or "decay tree") are parameterised in terms of their momenta, vertex positions, and decay times. These parameters are then fit to a set of tracking detector hits and ECAL clusters, subject to constraints such as conservation of momentum at vertices, and the masses of final-state particles matching their physical values. A Kalman Filter algorithm is used for this, which provides a χ^2 value indicating goodness-of-fit. Additional constraints can be imposed in the fit, at an analyst's discretion. Unless otherwise specified, all B-candidate masses in this analysis are calculated with the masses of K^0_S candidates constrained to their measured value [13], and with the B candidate required to originate from the PV. In addition, the dilepton masses are constrained to the measured J/ψ and $\psi(2S)$ masses in some fits for control-mode yields [13]. This constraint is particularly useful in the electron control-modes, where it greatly mitigates the effects of bremsstrahlung emission.

In many other R_X^{-1} analyses, simultaneous fits are used to extract yields for data recorded in each year, and with different L0 trigger lines. However, the numbers of candidates are comparatively low in this analysis, meaning background components could be poorly constrained in simultaneous fits. Therefore, the yield for each given decay mode is found with a single fit, where data from different years and L0 lines is combined.

3.4 Blinding

This analysis was carried out "blind". The values of R_X^{-1} and $\mathcal{B}(B \to K^{(*)}e^+e^-)$ in data were only examined after the full analysis methodology was devised, and the cross-checks outlined in Chapter 6 showed satisfactory results. In addition, $B^0 \to K^0_S e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ candidates in the region $m(B) \in [4900, 5400\,\text{MeV}]$ were not examined, as neither decay had been previously examined at LHCb (or indeed observed at any other experiment). The process of "unblinding" the final results was agreed with, and overseen by, a "review committee" of three LHCb personnel.

Chapter 4

Core Methodology

This chapter outlines the core methodology by which candidates are selected, the number of candidates in data estimated using maximum likelihood fits, and the efficiency of their detection estimated using MC. Studies of physical background contamination are also presented.

4.1 Selection

A set of selection requirements are used to retrieve samples of data enriched with $B \to K^{(*)} e^+ e^-$ and $B \to K^{(*)} \mu^+ \mu^-$ decays, and with minimal background contamination. Candidates are required to pass a specific set of trigger lines (see Section 2.2.3). Then, a central stripping selection is applied (see Section 2.2.4), followed by a loose "preselection". This includes requirements which suppress background from physical sources. Finally, a multivariate classifier is used to suppress "combinatorial background", where tracks from multiple different sources happen to overlap, in a way which looks like a $B \to K^{(*)} \ell^+ \ell^-$ candidate. These requirements are outlined in the following section.

4.1.1 Trigger

To ensure that the trigger performance can be easily calibrated to match data (see Section 4.2.5), each candidate is required to pass a specific set of trigger lines. The majority of the time, each line is required to be "triggered on signal" (TOS), where only a particle (or particles) associated to the candidate caused the trigger to fire [295]. This is opposed to being "triggered independent of signal" (TIS), where only particles unassociated to the candidate were responsible.

The muon-mode data is selected using only the `L0Muon` trigger line. However, due the high occupancy of the ECAL, the thresholds for the `L0Electron` line are set comparatively high to suppress backgrounds, making it comparatively inefficient for signal. Therefore, electron-mode

Lepton	L0 Category	Definition	Prop. Data
Muon	L0M	L0Muon-TOS (μ^+) | L0Muon-TOS (μ^-)	100%
Electron	L0E	L0Electron-TOS (e^+) | L0Electron-TOS (e^-)	$\sim 80-82\%$
	L0H	[L0Hadron-TOS (π^+) | L0Hadron-TOS (π^-) | L0Hadron-TOS (π^+_{Soft})] & !L0E	$\sim 3-4\%$
	L0I	L0Global-TIS & !(L0E|L0H)	$\sim 14-16\%$

Table 4.1: Definitions of different L0 categories used to select muon-mode and electron-mode data, together with the approximate proportions of fully-selected data in each category. Note that π^+_{Soft} refers to the pion originating from the decay $K^{*+} \to K^0_S \pi^+$, in $B^+ \to K^{*+} \ell^+ \ell^-$ candidates.

candidates which pass L0Electron, L0Hadron, or with any line being TIS (L0Global-TIS) are used. The definitions of these "L0 categories" are given in Table 4.1.

At HLT1, candidates are required to contain at least one track with good fit-quality, and large p_T and IP values. Various topological trigger lines are then used at HLT2. These lines, summarised in Table 4.2, are kept similar between the muon and electron modes, to ensure any systematic biases in their efficiencies are partially cancelled.

4.1.1.1 Fiducial requirements

The L0 thresholds used to collect data in each year varied depending on detector conditions. However, each sample of MC used in this analysis was only simulated using one set of thresholds, for each given year. Therefore, the applied thresholds do not match between data and MC. This could cause systematic biases.

To mitigate such biases, additional "fiducial" cuts are imposed on the E_T or p_T values of particles which fired a given L0 line. These are summarised in Table 4.3. Note that in 2017, MC was produced with a lower L0Muon p_T threshold than used in data. Therefore, MC candidates were randomly assigned fiducial cuts between $1450-1950$ MeV, to match the proportions of data recorded with different thresholds. In addition, the E_T value measured by L0Electron and L0Hadron cannot be matched with an offline track candidate, for technical reasons. Therefore, tightened cuts are applied to offline E_T measurements, to account for differing resolutions between the offline and L0 measurements. In addition, cuts of nSPDHits $<$ 600 and nSPDHits $<$ 450 are imposed in run-1 and run-2, respectively. These are equivalent to the tightest cuts imposed by an L0 line in each period.

Stage	Run-1	Run-2
HLT1	Hlt1TrackAllL0	Hlt1TrackMVA
HLT2	Hlt2Topo2BodyBBDT	Hlt2Topo2Body
	Hlt3Topo2BodyBBDT	Hlt2Topo3Body
	Hlt2Topo{Mu,E}2BodyBBDT	Hlt2Topo{Mu,E}2Body
	Hlt2Topo{Mu,E}3BodyBBDT	Hlt2Topo{Mu,E}3Body
		Hlt2Topo{MuMu,EE}2Body
		Hlt2Topo{MuMu,EE}3Body

Table 4.2: HLT lines used to select candidates in run-1 and run-2. Note that the terms in angular brackets denote lines used in the muon modes and electron modes, respectively. *i.e.*, Hlt2Topo{MuMu,EE}2Body denotes Hlt2TopoMuMu2Body in the muon modes, and Hlt2TopoEE2Body in the electron modes.

L0 Category	Year				
	2011	2012	2016	2017	2018
`L0M` (p_T / MeV)	-		1850	1450-1950	-
`L0E` (E_T / MeV)	3000		2700	2955	3150
`L0H` (E_T / MeV)	3500				

Table 4.3: Fiducial cuts imposed on particles triggering the `L0Muon`, `L0Electron`, and `L0Hadron` lines. Note that the `L0M` cuts are imposed using the p_T values measured by the L0 trigger, while the `L0E` and `L0H` cuts use values of E_T measured offline.

Fiducial cuts are also imposed for the `Hlt1TrackMVA` line. This line imposes a multivariate cut using the significance of a track's IP (χ^2_{IP}), and its p_T (in GeV) [296]:

$$\log \chi^2_{\text{IP}} > \frac{1}{(p_T - 1)^2} + \frac{b}{25}(25 - p_T) + \log 7.4 \tag{4.1}$$

For most of run-2, a threshold of $b = 1.1$ was imposed, and this is used in all MC samples. However, it was tightened to $b = 2.3$ for $\sim 30\%$ of 2016 data-taking. This tight threshold is imposed on a randomly chosen set of 2016 MC events, to match the proportions of data with the different thresholds.

Particle	Requirement
B^0/ B^+	$\chi^2_{\text{FD}}(\text{PV}) > 100$, $\cos\theta_{\text{DIRA}} > 0.9995$, $\chi^2_{\text{IP}} < 25$, $\chi^2_{\text{Vtx.}} < 9$, $\lvert m - m_{B^+}\rvert < 1500\,\text{MeV}$
$\ell^+\ell^-$ (dilepton)	$\chi^2_{\text{FD}}(\text{PV}) > 16$, $\chi^2_{\text{Vtx.}} < 9$, $m < 5500\,\text{MeV}$
$\mu^\pm$	$\chi^2_{\text{IP}} > 9$, $p_T > 300\,\text{MeV}$, `isMuon`
$e^\pm$	$\chi^2_{\text{IP}} > 9$, $p_T > 300\,\text{MeV}$, $\text{DLL}_{e\pi} > 0$
K^{*+}	$p_T > 400\,\text{MeV}$, $\lvert m - m_{K^{*+}}\rvert < 300\,\text{MeV}$, $\chi^2_{\text{DOCA}}\left(K^0_S\pi^+\right) < 30$, $\chi^2_{VTX} < 25$
K^0_S	$p_T > 400\,\text{MeV}$, $m < 2600\,\text{MeV}$, $\min\left(\chi^2_{\text{IP}}(\pi^+), \chi^2_{\text{IP}}(\pi^-)\right) > 9$
π^+_{Soft}	$p_T > 400\,\text{MeV}$, $\chi^2_{\text{IP}} > 9$

Table 4.4: Stripping selections imposed on $B^0 \to K^0_S\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ candidates.

4.1.2 Stripping

As outlined in Section 2.2.4, a set of "stripping" requirements are used to produce samples of manageable size for use by analysts. The requirements used for this analysis are given in Table 4.4. Here, m_{B^+} and $m_{K^{*+}}$ refer to the world-average measured masses of these particles [13]. Other variables used here include the significance of a candidate's distance-of-flight relative to the PV ($\chi^2_{\text{FD}}(\text{PV})$), the goodness-of-fit of a candidate's decay vertex ($\chi^2_{\text{Vtx.}}$), the significance of the distance of closest approach between a set of particles (χ^2_{DOCA}), and a candidate's "direction angle" (θ_{DIRA}). This is the angle between a line drawn from a candidate's production vertex to its decay vertex, and the momentum vector of its decay products. Other variables were defined previously.

To increase sample sizes, both DownDown and LongLong K^0_S candidates are used (see Section 2.2.1.4). For the same reason, a comparatively wide $m(K^0_S, \pi^+_{Soft})$ window is used for $B^+ \to K^{*+}\ell^+\ell^-$ candidates. In this window, there is a $\mathcal{O}(20\%)$ contribution from non-resonant "s-wave" $B^+ \to K^0_S\pi^+\ell^+\ell^-$ decays [101]. No effort is made to unfold these from the resonant "p-wave" $B^+ \to K^{*+}\ell^+\ell^-$ decays which dominate this region.

4.1.3 Preselection

An offline "preselection" is used to suppress misidentified particles and fake tracks, using PID information and track-quality indicators. These include the track fit χ^2 normalised by the

Particle	Requirement
All tracks	$\chi^2_{\text{Trk.}}/\text{nDoF} < 3$, `hasRich`, $\texttt{GhostProb}_{\text{Trk.}} < 0.4$ (see caption)
$\mu^\pm$	$\texttt{ProbNN}(\mu) > 0.2$, $p_T > 800\,\text{MeV}$
$e^\pm$	$\texttt{ProbNN}(e) > 0.2$, $p_T > 500\,\text{MeV}$, `hasCalo` $!(\lvert x_{ECAL}\rvert < 363.3\,\text{mm}\ \&\ \lvert y_{ECAL}\rvert < 282.6\,\text{mm})$
Long $\pi^\pm$, π^+_{Soft}	$\texttt{ProbNN}(\pi) \times (1 - \texttt{ProbNN}(K)) \times (1 - \texttt{ProbNN}(p)) > 0.1$
π^+_{Soft}	$\texttt{ProbNN}(e) < 0.1$ (electron mode only)
K^0_S	$m \in [470, 530\,\text{MeV}]$

Table 4.5: Offline preselection requirements. The `ProbNN` variables are computed using the `MC12TuneV2` and `MC12TuneV3` tunings for hadrons and leptons respectively in run-1, and with the `MC15TuneV1` tuning for all particles in run-2. In addition, the $\texttt{GhostProb}_{\text{Trk.}}$ cut is not imposed on downstream pions in run-1, due to low signal efficiency and poor agreement between data and MC.

fit's degrees of freedom ($\chi^2_{\text{Trk.}}/\text{nDoF}$), and the output of a neural network designed to classify "ghost tracks" constructed from incorrectly-associated tracker hits ($\texttt{GhostProb}_{\text{Trk.}}$) [297]. To ensure PID information is meaningful, tracks are required to have high p_T, plus associated RICH and calorimeter signatures (`hasRich`, `hasCalo`) with energy clusters falling outside the innermost region of the calorimeter. In addition, the K^0_S mass window is tightened to reject combinatorial candidates. These requirements are outlined in Table 4.5.

4.1.4 Physical background vetoes

Various requirements are used to suppress background from different physical sources, as summarised in Tables 4.6 and 4.7. The motivations for, and meanings of, these cuts are discussed in Section 4.4. Note that while the yields for many of these backgrounds are too low to cause significant contamination in the J/ψ and $\psi(2S)$ control modes, the physical background vetoes are still applied to these channels. This ensures that any potential systematic biases in the estimated efficiencies for the vetoes should partially cancel.

Back. Type	Electron-mode cut	Muon-mode cut
$B^0 \to (D^- \to K^0_S\pi^-)X$	$\lvert m(K^0_S\pi^\pm) - m(D^-)\rvert > 40\,\text{MeV}$	
$B^0 \to (D^- \to K^0_S\ell^-\bar{\nu})X$	$m(K^0_S\ell^\pm) > m(D^-)$	
$\Lambda^0_b \to \Lambda^0\ell^+\ell^-$	$!(m(\pi^+_{\pi\to p}\pi^-) \in [1100, 1135\,\text{MeV}]$ & $m(\pi^+_{\pi\to p}\pi^- e^+ e^-)$ $\in [4800, 5800\,\text{MeV}])$	$!(m(\pi^+_{\pi\to p}\pi^-) \in [1100, 1135\,\text{MeV}]$ & $m(\pi^+_{\pi\to p}\pi^- e^+ e^-)$ $\in [5500, 5675\,\text{MeV}])$
Prompt Decays	$\tau(K^0_S) > 0.0005\,\text{ns}$	

Table 4.6: Summary of physical background vetoes applied to B^0-mode data. The sources of background, and motivations for these cuts, are discussed in Section 4.4.

Back. Type	Electron-mode cut	Muon-mode cut
$B^0 \to (D^- \to K^0_S\pi^-)X$	$\lvert m(K^0_S\pi^\pm) - m(D^-)\rvert > 40\,\text{MeV}$	
$B^0 \to (D^- \to K^0_S\ell^-\bar{\nu})X$	$m(K^0_S\ell^\pm) > m(D^-)$	
$\Lambda^0_b \to \Lambda^0\ell^+\ell^-$	$m(\pi^+_{\pi\to p}\pi^-) \notin [1100, 1135\,\text{MeV}]$	
Prompt Decays	$\tau(K^0_S) > 0.0005\,\text{ns}$	
Over-reconstructed B^0	$m(K^0_S\ell^+\ell^-) \notin [5150, 5350\,\text{MeV}]$	
$\pi^+_{Soft} \leftrightarrow \ell^+$ Swaps	$\texttt{ProbNN}(e)(\pi^+_{Soft}) < 0.1$	$m(\pi^+_{Soft.,\,\pi\to\mu},\,\mu^-) \notin$ $[3050, 3150\,\text{MeV}], [3650, 3750\,\text{MeV}]$

Table 4.7: Summary of physical background vetoes applied to B^+-mode data. The sources of background, and motivations for these cuts, are discussed in Section 4.4.

4.1.5 Multivariate classifier to suppress combinatorial background

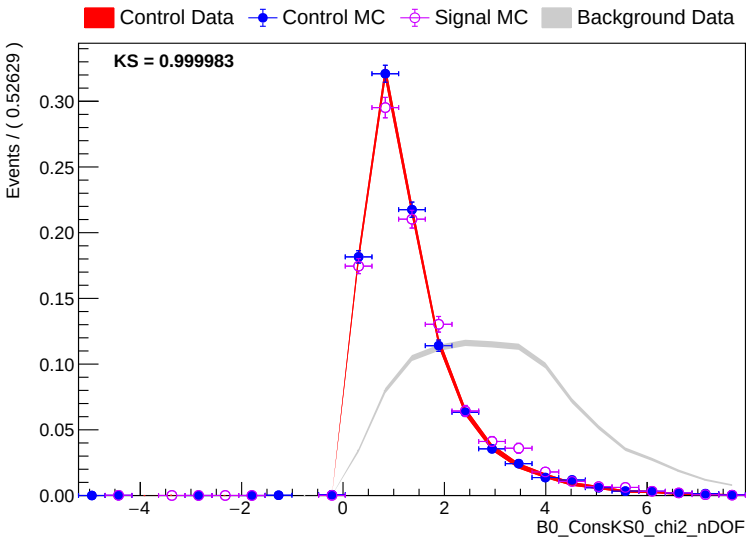
Significant contamination from combinatorial background remains after the trigger, stripping, and preselection requirements are imposed. No single variable discriminates strongly between signal and combinatorial background candidates. Therefore, multivariate classifiers are used, which combine information from multiple variables and take correlations between them into account, providing high discrimination power [298]. The "training variables" given in Table 4.8 are used as inputs to the classifiers. These include the quality of the `DecayTreeFitter` fit normalised by this fit's degrees of freedom ($\chi^2_{\text{DTF}}/\text{nDoF}$), the reconstructed lifetimes of intermediate particles (τ), and a pair of "isolation variables", which indicate whether additional tracks in an event could have actually originated from the B-meson candidate. The first of these ($\text{ISO}_{\text{Vtx.}}$) refers to the change in $\chi^2_{\text{Vtx.}}$ when an additional track is added to a candidate vertex. The second ($\text{ISO}_{\text{Trk.}}$) is found by counting the number of additional tracks which pass close to a given candidate track [299, Sec. 6.1] [300, 301]. To be counted, the angle between

Particle	Variable
B^0/ B^+	$\chi^2_{\text{DTF}}/\text{nDoF}$, $\log\chi^2_{\text{IP}}$, $\log\chi^2_{\text{FD}}\,(\text{PV})$, p_T, τ, $\text{ISO}_{\text{Vtx.}}$
$\ell^+\ell^-$ (dilepton)	τ, $\text{ISO}_{\text{Trk.}}(\ell^+) + \text{ISO}_{\text{Trk.}}(\ell^-)$
K^0_{S}	p_T
$\pi^\pm$ (from K^0_{S})	$\text{ISO}_{\text{Trk.}}(\pi^+) + \text{ISO}_{\text{Trk.}}(\pi^-)$
$\pi^\pm$ (from LongLong K^0_{S} only)	min. $\left(\log\chi^2_{\text{IP}}(\pi^+), \log\chi^2_{\text{IP}}(\pi^-)\right)$, max. $\left(\log\chi^2_{\text{IP}}(\pi^+), \log\chi^2_{\text{IP}}(\pi^-)\right)$

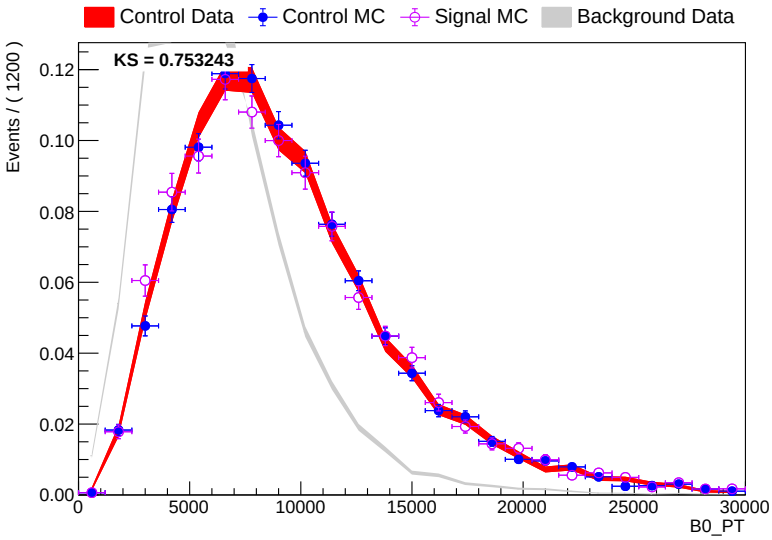
Table 4.8: Variables used to discriminate between signal and combinatorial background candidates.

an additional track and the candidate track must be small, and the two tracks must form a good-quality vertex. The contamination from combinatorial K^0_{S} candidates is much larger with the LongLong reconstruction than with the DownDown reconstruction. Therefore, additional χ^2_{IP} information is utilised in the classifiers for LongLong K^0_{S} candidates, providing increased discrimination power.

The detector performance varies between run-1 and run-2, $B^0 \to K^0_{\text{S}}\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ decays, muon and electron modes, and DownDown and LongLong K^0_{S} candidates. Therefore, different classifiers are trained for each run, decay mode, lepton, and K^0_{S} category, giving 16 classifiers in total. The classifiers are trained to discriminate between "signal" samples of $B \to K^{(*)}\ell^+\ell^-$ MC, and "background" samples of data in the "upper-mass sideband" of $m_B > 5500\,\text{MeV}$, which is dominated by combinatorial candidates. The background sample is required to fall outside $q^2_{J/\psi}$ and $q^2_{\psi(2S)}$, to reject real $c\bar{c}$ candidates. Both the signal and background samples are required to pass the trigger, stripping, and preselection requirements, as well as the physical background vetoes outlined in Section 4.4. Example comparisons of the training variables in control-mode data and MC are shown in Fig. 4.1, as well as signal-mode MC, and data in the upper-mass sideband. Good separation can be seen between signal and background, as well as good agreement between control-mode MC and data.



(a): B^0- $\chi^2_{\mathrm{DTF}}/\mathrm{nDoF}$



(b): B^0- p_T

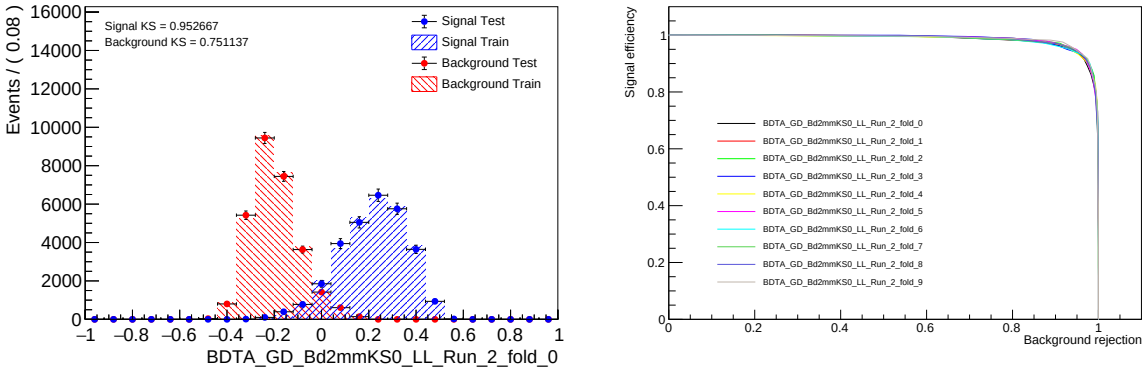
Fig. 4.1: Distributions of some of the multivariate classifier input variables found in $B^0 \to K^0_S e^+ e^-$ signal MC (magenta), $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ data (red), $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ MC (blue) and data from the upper-mass sideband (grey) for 2018. The corrections outlined in Section 4.2 are applied to the MC samples, to improve agreement with data. Background in $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ data is statistically subtracted using the *sPlot* method (see Section 4.2.1.1). Kolmogorov-Smirnov test p-values are shown, indicating the agreement between $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ MC and data.

4.1.5.1 Choice and training of classifier

A classifier known as a "boosted decision tree" (BDT) is trained to discriminate between signal and background [302, 303]. A series of decision trees are constructed which use sets of rectangular cuts to separate the signal and background samples. The construction of each decision tree is terminated under certain conditions, such as the tree reaching a certain depth, imposing a certain number of cuts, or the number of events in the sample passing a given path along the tree falling below a certain threshold. Therefore, no single decision tree will be able to correctly classify all entries in the training sample. Any entry which is misclassified by a given decision tree is up-weighted, such that the next tree to be trained is more likely to classify that entry correctly. This process is known as "boosting". After a certain number of trees have been trained, their results are aggregated, and each sample entry is assigned an output score showing how "signal-like" or "background-like" it is (denoted $\mathrm{BDT}_{Comb.}$). This method can give good separation between signal and background, while remaining robust against "overtraining", where a classifier learns the differences based on statistical fluctuations in the training sample. An overtrained classifier will show degraded performance when applied to an independent sample. In addition, if the performance of the classifier is assessed using the training sample, it will appear inflated. This could, for example, lead to a systematic bias if the training sample is subsequently used to estimate the efficiency of a cut on $\mathrm{BDT}_{Comb.}$.

Different algorithms are available which follow this boosting methodology. In this analysis, the `AdaBoost` method is used [304], as implemented in the TMVA software library [305, 306]. Before training each classifier, the ranges of the training variables are transformed such that the training samples follow Gaussian distributions. Then, a matrix transformation is applied which ensures the variables are linearly decorrelated. These transformations were found to improve classifier performance. In addition, k-fold cross-validation, with ten folds, is used to increase training sample statistics [307]. Here, 90% of the training sample is used to train a classifier, whose output score is then assigned to the remaining 10% of the data (known as a "hold-out sample"). This process is repeated ten times with different hold-out samples, such that each entry in the full sample is assigned a score from a classifier which it was not used to train. This way, efficiencies of cuts on $\mathrm{BDT}_{Comb.}$ can be estimated without bias, even if classifiers are overtrained. Candidates which were not used in any training sample (for example MC samples for $B \to \psi(2S)\,(\ell^+\ell^-)\,K^{(*)}$) are randomly assigned a score from one of the ten classifiers.

After training, each classifier, for each k-fold, can be checked for over-training by comparing the $\mathrm{BDT}_{Comb.}$ distributions found for training and hold-out samples. The statistical agreement between these distributions is assessed using a Kolmogorov-Smirnov (KS) test [308]. As shown for example in Fig. 4.2a, the training and hold-out distributions are in good agreement for both signal and background, suggesting negligible over-training. In addition, the "receiver operating



(a): BDT output ***(b):*** ROC curves

Fig. 4.2: Comparisons of BDT output score ($\text{BDT}_{Comb.}$) distributions in the training and hold-out (shown as "test") samples of k-fold 0, for signal and background, with KS test p-values shown in the legend (left). A comparison of the ROC curves for the hold-out samples of different k-folds (right). These plots correspond to classifiers for $B^0 \to K^0_S \mu^+ \mu^-$ decays in run-2, with LongLong K^0_S candidates.

characteristic" (ROC) curves - showing the values of signal efficiency and background rejection when cutting on different values of $\text{BDT}_{Comb.}$ - are examined for different k-folds. As shown for example in Fig. 4.2b, these are generally in good agreement, again suggesting a lack of over-training.

4.1.5.2 Classifier cut optimisation

After all BDTs have been trained, and applied to data and MC samples, cuts are placed on $\text{BDT}_{Comb.}$ to suppress combinatorial background. The working-points for these cuts were chosen to maximise the figure-of-merit $\sigma = S/\sqrt{S+B}$ for each rare-mode channel, which provides an estimate of signal significance in the presence of background [309]. Here, S and B are yields for rare-mode signal and background candidates passing the full selection requirements, in "signal mass windows" of $m(B) \in [5240, 5320\,\text{MeV}]$ and $m(B) \in [5100, 5350\,\text{MeV}]$ for muon and electron modes, respectively. The signal yield at a given working-point was calculated with eq. 3.2, using the world-average branching fractions of $\mathcal{B}\left(B^0 \to K^0 \ell^+ \ell^-\right) = \left(3.1^{+0.8}_{-0.7}\right) \times 10^{-7}$ and $\mathcal{B}\left(B^+ \to K^{*+} \ell^+ \ell^-\right) = (1.01 \pm 0.11) \times 10^{-6}$, and the equivalent control modes [13]. Signal and control-mode efficiencies were estimated from MC, and the control-mode yields were calculated using the mass-fits documented in Section 4.5.2. The background yield was calculated by fitting an exponential PDF to $m(B)$ in data, in regions above and below the signal mass windows. Then, these PDFs were extrapolated into the signal mass window. Examples of such fits are shown in Fig. 4.3.

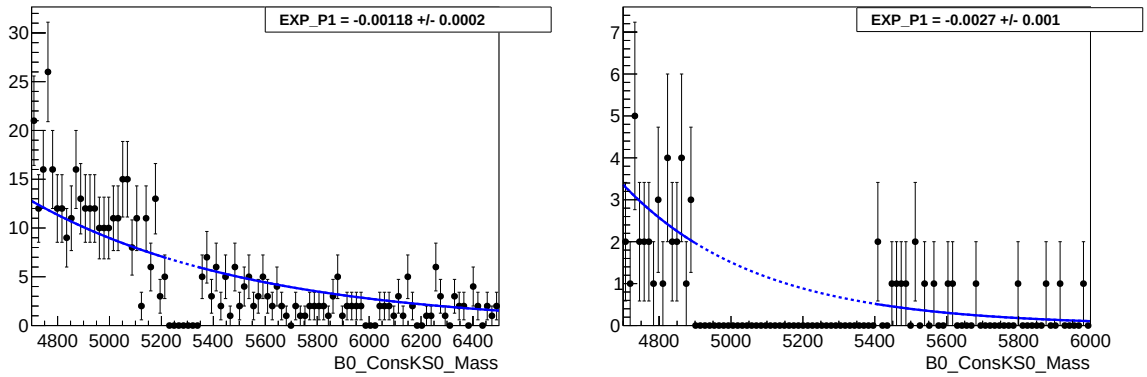


Fig. 4.3: Examples of fits to $m(B)$ used to estimate the background yields in the signal mass regions for $B^0 \to K^0_S \mu^+ \mu^-$ (left), and $B^0 \to K^0_S e^+ e^-$ (right). These signal regions are blinded. The optimal BDT$_{Comb.}$ cuts were applied when carrying out these fits.

Data from run-1 and run-2 was combined, to increase data sample statistics when fitting for B. Scans were carried out across different working-points. Optimal values of σ were found with cuts of BDT$_{Comb.} > 0.15$ for $B^0 \to K^0_S \mu^+ \mu^-$ and $B^+ \to K^{*+} \mu^+ \mu^-$, BDT$_{Comb.} > 0.225$ for $B^0 \to K^0_S e^+ e^-$, and BDT$_{Comb.} > 0.250$ for $B^+ \to K^{*+} e^+ e^-$. These cuts were used to select data for the final calculations of R_X^{-1}, and were also applied to the equivalent control modes, to mitigate potential systematic biases from BDT mismodelling. This study also informed the q^2 ranges used for each channel. When q^2_{low} was included, σ was seen to increase for $B^+ \to K^{*+} e^+ e^-$, but decrease for $B^0 \to K^0_S e^+ e^-$.

4.1.5.3 Check for mass sculpting

If a classifier is trained using variables which are highly dependent on the kinematics of a B-hadron decay, it may be able to reconstruct the value of $m(B)$ in each candidate. Such a classifier could learn that signal decays lie close to the nominal value of $m(B) \approx 5280\,\text{MeV}$, while background decays lie far from this value. If this happens, a cut on the classifier output would "sculpt" the combinatorial background, causing its distribution in $m(B)$ to form a peak at $\sim 5280\,\text{MeV}$, similar to signal. This would artificially increase the measured signal yield in a fit to $m(B)$.

To check for this effect, samples of $B^0 \to K^0_S e^\pm \mu^\mp$ and $B^+ \to K^{*+} e^\pm \mu^\mp$ candidates were examined in 2018 data, as these can be assumed to be composed almost exclusively of combinatorial background. The samples were required to pass the trigger, stripping, and preselection requirements outlined previously. Output values for the muon-mode and electron-mode BDTs were calculated for each sample (*i.e.* $B^0 \to K^0_S \mu^+ \mu^-$ and $B^0 \to K^0_S e^+ e^-$ BDT$_{Comb.}$ values were added to the $B^0 \to K^0_S e^\pm \mu^\mp$ sample, and likewise for the $B^+ \to K^{*+} e^\pm \mu^\mp$ sample). Then, the

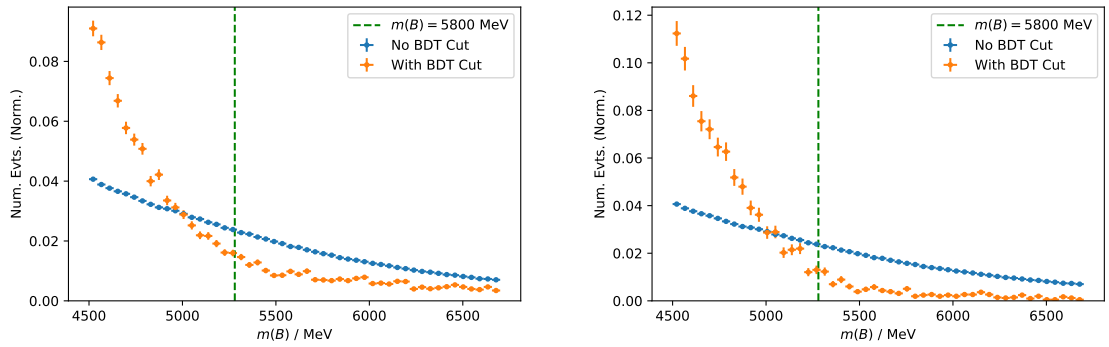


Fig. 4.4: Distributions of $m(B)$ found for $B^0 \to K^0_S e^{\pm}\mu^{\mp}$ candidates in 2018 data, before a BDT$_{Comb.}$ cut is applied (blue), and after a BDT cut is applied (orange). The nominal mass of a B-meson is shown by the vertical green lines. These plots show the effect of the muon-mode BDT (left), and the electron-mode BDT (right).

Decay mode	Electron mode		Muon mode	
	Lower / MeV	Upper / MeV	Lower / MeV	Upper / MeV
$B \to K^{(*)}\ell^+\ell^-$	4700	6000	5175	5700
$B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$	5175			
$B \to \psi(2S)\,(\ell^+\ell^-)\,K^{(*)}$				

Table 4.9: Lower and upper limits of mass fits used to extract signal and control-mode yields.

distributions of $m(B)$ were examined in these samples before and after BDT$_{Comb.}$ cuts were applied. As seen for example in Fig. 4.4, these have smooth distributions both with and without BDT$_{Comb.}$ cuts, with no signs of peaking structures at $m(B) \approx 5280\,\text{MeV}$. This suggests the BDTs to not sculpt the $m(B)$ distributions in combinatorial background.

4.1.6 Invariant b-meson mass regions

The ranges of $m(B)$ used in fits to extract rare-mode and control-mode yields, as outlined in Section 4.5, are summarised in Table 4.9. The lower limits are informed by the need to discard sources of physical background with low values of $m(B)$, while the upper limits ensure that sufficient combinatorial background is retained for its shape and yield to be constrained by the fit. As mentioned in Section 3.3, the control-mode masses are calculated using `DecayTreeFitter`, with constraints to the dilepton masses. Such constraints cannot be used for the rare-mode

Particle	Decay
π^+, K^+	$D^0 \to K^-\pi^+$, from $D^{*+} \to D^0\pi^+$
p	$\Lambda \to p\pi^-$
μ^+	$J/\psi \to \mu^+\mu^-$
e^+	$J/\psi \to e^+e^-$, from $B^+ \to K^+J/\psi$

Table 4.10: Decay modes used to select PID calibration samples for different particles [262, 310, 311].

masses. Due to bremsstrahlung emission, a wide mass range is therefore needed for the rare electron modes, to ensure signal is not cut away.

4.2 Corrections to simulation

Numerous data-driven corrections are applied to the MC samples used in this analysis, to ensure the estimates of selection efficiencies are accurate. These corrections are documented in the following pages. Many of them are found using samples of data and MC dedicated for use in calibration. Others are calculated using the $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ data and MC samples selected in this analysis. The corrections are found in a series of sequential steps. When deriving a set of corrections using the $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ samples, the results of all previous corrections are applied to the MC samples. This ensures there is no "double-counting" of any corrections.

4.2.1 Particle identification efficiencies

Due to imperfect modelling of the RICH and calorimeter systems, the distributions of PID variables found in data are not properly reproduced in MC. Hence, it is not possible to evaluate PID efficiencies using the variables in MC, so dedicated calibration data samples are used instead [262, 310, 311]. These samples are constructed using kinematically well-constrained decays, where backgrounds are suppressed by applying tight requirements to a sub-set of the final-state particles, known as the "tag" particles. Minimal selection requirements are imposed on the remaining "probe" particle, meaning the PID distributions associated to it are left unbiased, and can be used for calibration purposes. The decays used to construct calibration samples for different particle types are shown in Table 4.10.

Distributions of PID variables associated to a track are dependent on the track's kinematics, and on the detector occupancy for the associated event. As the distributions of the final-state track kinematics and detector occupancy vary between the calibration sample and an analyst's

MC sample of interest (also referred to as the "target sample"), the PID responses will also vary between them. To account for this, a set of "PID efficiency weights" are calculated using the calibration samples, with their computation managed by the PIDCALIB software package from LHCb [262, 310, 311].

To calculate the weights, each calibration sample is divided into bins according to the probe track's p and p_T, and the number of tracks reconstructed in the associated event (`nTracks`). Within each "kinematic bin", the efficiency of a PID cut is calculated by comparing the number of candidates in the calibration sample before and after this cut is applied to the probe track. These efficiencies are then folded into the MC target samples, such that each track within each candidate is assigned an efficiency weight ($w^{PID\ Trk.}$). The PID efficiencies for different tracks in the target sample are assumed to factorise, meaning the total efficiency for all PID cuts imposed on the target sample can be found by taking the product of $w^{PID\ Trk.}$ within each candidate, then averaging across all candidates in the sample:

$$\varepsilon_{PID} = \frac{1}{N_{Cands.}} \sum_{i}^{Cands.} w_i^{PID} = \frac{1}{N_{Cands.}} \sum_{i}^{Cands.} \left[\prod_{j}^{Tracks} \left[w_{i,j}^{PID\ Trk.} \right] \right] \tag{4.2}$$

Here, each candidate is indexed by i, and the tracks within each candidate by j. The total number of candidates is given by $N_{Cands.}$, and the total "per-candidate" PID weight by w_j^{PID}. Such weights are used to compute the efficiencies of the PID cuts imposed in the stripping and preselection stages for this analysis, for the signal and control modes. Weights are also calculated for misidentified particles, which are used in some of the background studies documented in Section 4.4.

A set of fiducial requirements are imposed on the calibration samples, to ensure the probe track's selection requirements are well aligned with the target sample's selection requirements. These consist of cuts to `nSPDHits`, and requirements that probe tracks have associated RICH, ECAL, and muon chamber signatures where necessary. Separate sets of PIDCALIB weights are calculated and used for each magnet polarity, and track charge. Electron samples are further split according to whether the probe track has an associated bremsstrahlung photon. Weights are computed separately for each year, to account for differing PID performances.

4.2.1.1 Background subtraction

The PID calibration samples contain some background contributions. To compute an accurate set of efficiency weights, the PID variable distribution for the signal must be unfolded from the distributions for such background. To do this, the *sPlot* method is used [312]. First, a maximum likelihood fit is carried out to find the relative yields of the signal and backgrounds in data, using some "discriminating variable" x (typically a reconstructed particle mass). The results of

Particle	p, p_T Bins	nTracks Bins	Merge Tol.
π^+, K^+, p	25	10	2.5σ
e^+	10	5	1.5σ

Table 4.11: The number of initial bins used to construct binning schemes for each binning variable and particle type, plus the tolerance parameter used to decide whether adjacent bins should be merged.

this fit are used to calculate a set of "*sWeight*s", which are folded in to the data. The *sWeight*s vary as a function of x, with large positive values where the relative signal contribution is high, and small or negative values where the signal contribution is low. The *sWeight*s are then used to weight the distributions of a set of "control variables" $\vec{y}$. In this case, $\vec{y}$ are the PID variables of interest. When the *sWeight*s are applied, the background contributions to $\vec{y}$ are statistically subtracted, such that the distributions of $\vec{y}$ reflect the contributions from signal alone.

Mass-fits are used to calculate *sWeight*s for all of the PIDCALIB samples, which are used to calculate the efficiency weights in each kinematic bin. However, the *sPlot* method requires that each control variable is uncorrelated with the discriminating variable, in signal and background. Any correlations between them could cause systematic biases in the unfolded distributions. The electron calibration sample is comprised of $B^+ \to K^+ J/\psi\,(e^+e^-)$ candidates. Electron momenta and PID variables both depend on the energies of associated ECAL clusters. Therefore, the mass of a $B^+ \to K^+ J/\psi\,(e^+e^-)$ signal candidate is correlated to its PID response, meaning efficiency calculations using *sWeight*s would be biased. Instead, a dedicated "fit-and-count" method is used to find electron PID weights [313, Sec. 2.8.2]. Within each kinematic bin, mass-fits are carried out using data which passes or fails a certain PID cut. The two resulting signal yields are used to find the PID efficiency within that bin.

4.2.1.2 Binning

The binning schemes across p, p_T and nTracks should be fine enough to capture any variations in PID efficiencies, while remaining coarse enough to ensure that large numbers of events fall within each bin, reducing statistical noise. To find such schemes, a binning optimisation algorithm implemented in PIDCALIB is used for each variable, calibration sample, and PID cut. Fine binning schemes are constructed across each binning variable, with each bin equally populated by the calibration sample. PID efficiencies and associated uncertainties are calculated in each bin. If the uncertainty-normalised difference in PID efficiencies between two adjacent bins is below a certain threshold, the bins are merged. This process of efficiency calculation

and bin merging is iterated, until no more merging takes place. The number of initial bins and merging thresholds used in this analysis are shown in Table 4.11.

4.2.2 Electron tracking efficiency

The reconstruction efficiencies for long electron tracks are not correctly modelled in MC. Therefore, a set of weights is used to correct the reconstruction efficiencies for each electron, in each MC candidate. Centrally-produced weight maps were used, containing the ratios of track reconstruction efficiencies in data and MC, as a function of the track's p_T and η, plus whether it passed through the RF foil (as indicated by the azimuthal angle ϕ in the $x-y$ plane). These maps were found using samples of $B^+ \to K^+ J/\psi\,(e^+e^-)$ decays, where a "tag-and-probe" approach was used to examine the tracking efficiency for one of the final-state electrons [314].

The reconstruction efficiencies for each electron track are assumed to be uncorrelated. Therefore, the total correction weight for each candidate is found from the product of the correction weights for each electron ($w^i_{Trk.}(Tot.) = w^i_{Trk.}(e^-) \times w^i_{Trk.}(e^+)$).

4.2.3 Generated kinematics and detector occupancy

The production of B-hadrons in pp collisions involves complicated QCD processes, which are difficult to accurately model. Therefore, when simulated with PYTHIA, the kinematic distributions of B-mesons at LHCb reflect data imperfectly. Hence, the kinematic distributions of final-state particles are also modelled imperfectly. Many of the selection requirements, such as the trigger, place requirements on particle momenta, meaning efficiency estimates could be systematically biased if the B-meson kinematics are not corrected to match data. In addition, the numbers of particles produced in pp collisions are generally under-estimated by simulation, meaning the detector occupancies of MC events are lower than in data, on average. As PID and candidate reconstruction efficiencies are generally dependent on the detector occupancy, such mismodelling could also cause systematic biases.

Two sets of weights are used to correct for these effects. The first corrects the detector occupancy, as described by two "proxies" `nTracks` and `nSPDHits`, while the second corrects the generated B-meson kinematics, as described by the p_T and η values. As they arise from generator-level effects, the mismodelling of B-meson kinematics and the detector occupancy should not be affected by the decay mode of the B-meson. Therefore, the correction weights are derived using data and MC for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ and $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ decays, then applied to the MC samples for all corresponding electron and muon decay modes. The muon control modes are preferred over the electron modes because more $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ candidates survive selection, reducing statistical uncertainties on the correction weights. In

addition, the lack of bremsstrahlung emission means the reconstructed values of $p_T(B)$ and $\eta(B)$ are very close to their true values, and any resolution-related effects can be neglected. The $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ data and MC samples are required to pass full selection requirements, except for the $\text{BDT}_{Comb.}$ cuts. Separate sets of weights are calculated for the B^0 and B^+ modes, and each data-taking year, via the following method:

1. The $m(B)$ mass-fit documented in Section 4.5.2.1 is used to calculate signal and background yields, with the *sPlot* method then used to unfold the signal distributions of the four corrected variables in data (see Section 4.2.1.1).

2. A two-dimensional 50×10 binning scheme is constructed across `nTracks` and `nSPDHits`, such that bins are equally populated by $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ MC sample. This ensures a reasonable number of MC candidates fall within each bin, reducing statistical uncertainties on the weight values.

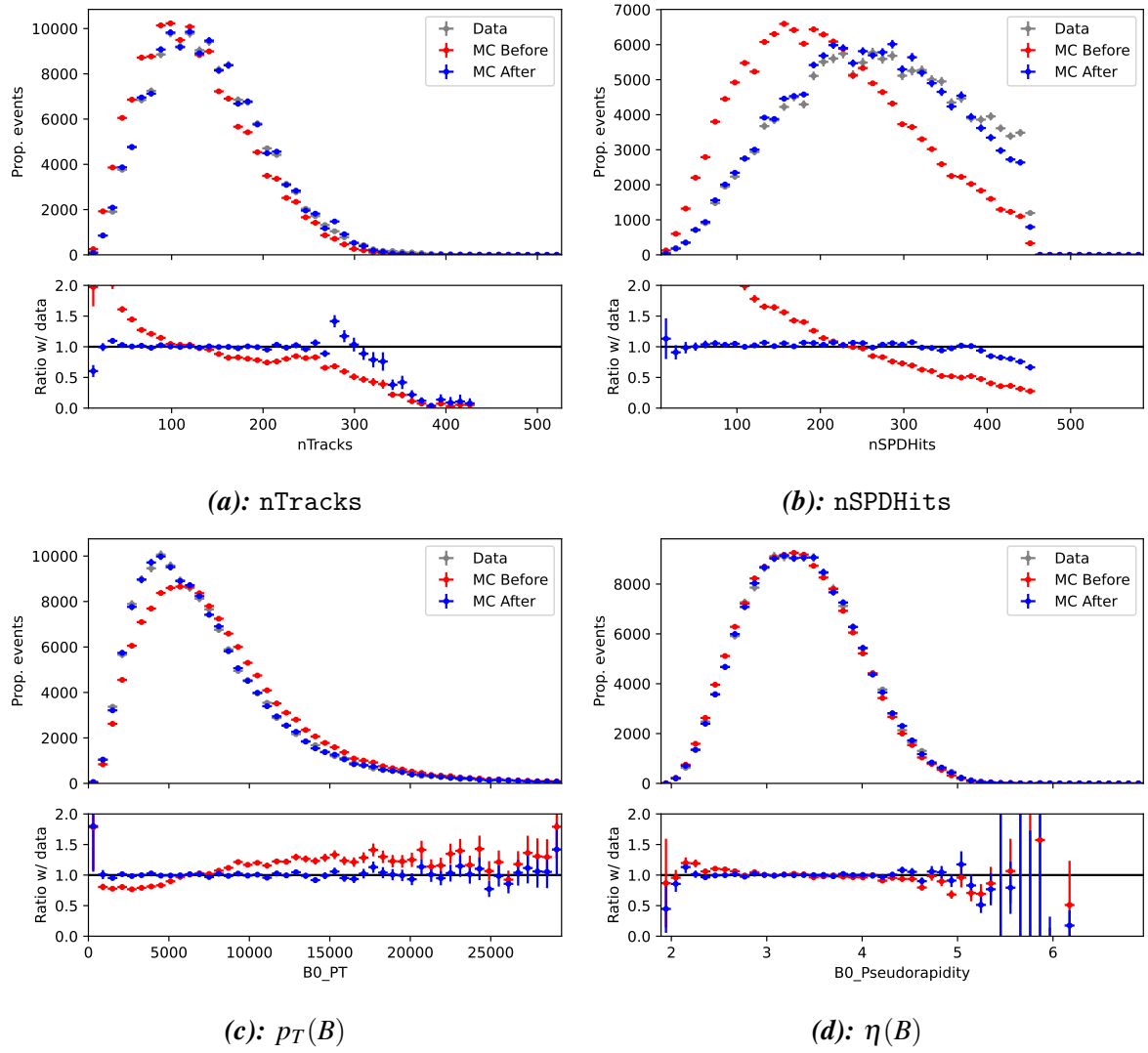
3. The $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ MC and *sWeight*ed data samples are used to populate two histograms constructed with this binning scheme. The ratios of bin contents in the data and MC histograms are taken as correction weights for the occupancy ($w_{Gen.Occ.}$).

4. A two-dimensional 25×25 binning scheme is constructed across $p_T(B)$ and $\eta(B)$. This is used to construct two histograms, occupied by MC and *sWeight*ed data. The occupancy weights are applied to the MC sample, to account for any correlations between the occupancy and B-meson kinematic mismodelling (though this is anticipated to be negligible). The ratios of data and MC bin contents are taken as correction weights for the B-meson kinematics ($w_{Gen.Kin.}$).

These weights are calculated using five k-folds, to ensure each control-mode MC candidate is assigned a weight which was calculated using an independent sample. When an MC candidate was not used to calculate the weights, it is randomly assigned weights from one of the k-folds. The weights are added both to reconstruction-level and generator-level MC samples, enabling the efficiencies of the reconstruction and stripping selection requirements to be corrected. The distributions of the four reweighted variables are compared between *sWeight*ed data and MC for $B \to J/\psi(\ell^+\ell^-)K^{(*)}$ decays in each year. As shown for example in Fig. 4.5, agreement between data and MC is improved when the correction weights were applied. In the following pages, the total weight $w_{Gen.} = w_{Gen.Occ.} \times w_{Gen.Kin.}$ is generally used.

4.2.4 K^0_S-meson reconstruction categories

The relative reconstruction efficiencies for LongLong and DownDown K^0_S-mesons are mismodelled in simulation [315], meaning the fractions of LongLong vs. DownDown K^0_S-mesons in



(a): `nTracks` *(b):* `nSPDHits`

(c): $p_T(B)$ *(d):* $\eta(B)$

Fig. 4.5: Distributions of variables used to correct the detector occupancy and generated B-meson kinematics, as found for $B^0 \to J/\psi(\mu^+\mu^-)K^0_S$ candidates *sWeight*ed data (grey), and MC before (red) and after (blue) the correction weights $w_{Gen.Occ.} \times w_{Gen.Kin.}$ are applied. The data and MC samples are required to pass full selection requirements, except for cuts to $\text{BDT}_{Comb.}$. The bottom portions of each plot show the ratios of MC and data.

$B \to J/\psi(\ell^+\ell^-)K^{(*)}$ decays do not match data. These fractions are examined in data samples of $B^0 \to J/\psi(\mu^+\mu^-)K^0_S$ and $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$ candidates passing full selection requirements (except for $\text{BDT}_{Comb.}$ cuts). The signal yields with each K^0_S-category, in each year, are extracted with separate mass-fits to data, using the model documented in Section 4.5.2.1. The fractions of each K^0_S-category are also computed in $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ MC, and weights ($w_{K^0_S}$) are calculated to make the relative LongLong and DownDown fractions match data. These weights are then applied to other MC samples in a given year.

4.2.5 Trigger performance

Particle signatures in the ECAL, HCAL, and muon chambers are imperfectly simulated, meaning the efficiencies of the L0 trigger are not described accurately. Ideally, it would be possible to obtain a sample of $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ data without imposing any trigger requirements, allowing one to directly examine trigger efficiencies $\varepsilon_{Trig.True}$ by counting the number of candidates before and after a given cut is imposed. However, this is not the case, and data is only recorded by the LHCb detector if it passes a set of trigger lines.

To account for this, a tag-and-probe approach known as the TISTOS method is used to evaluate trigger uncertainties in data [295]. For example, if examining the efficiency of the `L0Muon-TOS` requirement for a final-state muon, a sample of $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ decays would be selected using some "tag" requirement (for example `L0Muon-TIS` with respect to that muon). Then, the efficiency of `L0Muon-TOS` - the "probe" requirement - would be examined in this tagged sample, giving $\varepsilon_{Trig.Tag}$. If the tag and probe requirements are uncorrelated, this efficiency estimate would be an unbiased estimate of the *untagged* efficiency (*i.e.* $\varepsilon_{Trig.Tag} = \varepsilon_{Trig.True}$). MC can be used to check whether the requirements are indeed uncorrelated.

The efficiencies of each L0 requirement are examined in MC and data using this method. The MC and data samples are required to pass the preselection, background veto, and relevant HLT requirements. Any discrepancies between the data and MC efficiencies are parameterised as a function of some kinematic variable relevant to the trigger line, allowing differences in signal and control-mode kinematics to be accounted for when applying corrections to MC samples. These "efficiency profiles" are used to compute correction weights (w_{L0}) for each candidate. Corrections to HLT efficiencies were also considered, but efficiencies were found to be consistent between MC and data. Also, few candidates passed suitable tagging requirements, resulting in large statistical uncertainties.

4.2.5.1 `L0Muon`

To correct the `L0Muon-TOS` efficiencies, samples of $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ candidates in MC and *sWeight*ed data are used. These samples are tagged by requiring one of the final-state muons to satisfy `L0Muon-TIS`. The efficiencies of the `L0Muon-TOS` requirement, for each data-taking year, are then measured in 16 bins of the muon's offline p_T. These bins are constructed to be equally populated by the tagged muon candidates. To increase sample sizes, both positively and negatively charged muons passing `L0Muon-TIS` are considered. In addition, the $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ and $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ samples are combined, as the `L0Muon-TOS` efficiency should be independent of where the muon originated from.

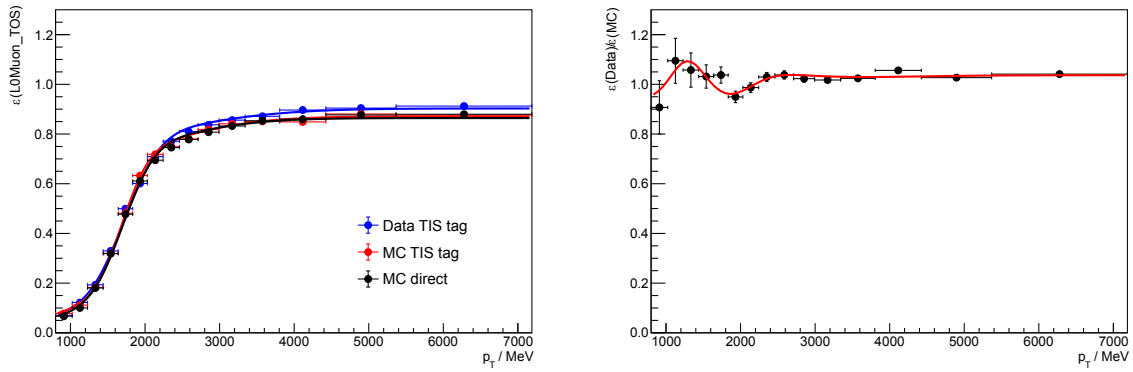


Fig. 4.6: Left - Efficiency profiles for `L0Muon-TOS` as a function of $p_T(\mu)$, in tagged data (blue), tagged MC (red), and untagged MC (black). Right - ratios of efficiencies for tagged data and tagged MC, as a function of $p_T(\mu)$. In both plots, the points show values in bins of $p_T(\mu)$ (plus associated uncertainties), while the curves show the values found from the functional parameterisations. These plots were produced using 2018 $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ samples.

In order to smooth out statistical fluctuations, the resulting efficiency profiles for each data-taking year are parameterised using a sum of two error functions [316]:

$$\varepsilon(p_T) = \frac{a}{2}\left[1 + \mathrm{Erf}\left(\frac{p_T - b}{\sqrt{2}c}\right)\right] + \frac{d}{2}\left[1 + \mathrm{Erf}\left(\frac{p_T - b}{\sqrt{2}e}\right)\right] \tag{4.3}$$

The parameters $a - e$ are all found from fits to the efficiency profiles. This parameterisation accounts for resolution effects on the applied `L0Muon` thresholds. These parameterised efficiency profiles can then be used to calculate the probability that at least one of the muons triggers `L0Muon`, for data and MC*.

$$\mathcal{P}_{\texttt{L0M}} = 1 - \left(1 - \varepsilon^{\mu TOS}(p_T[\mu^+])\right) \cdot \left(1 - \varepsilon^{\mu TOS}(p_T[\mu^-])\right) \tag{4.4}$$

Correction weights are then calculated as $w_{\texttt{L0M}} = \mathcal{P}^{Data}_{\texttt{L0M}} / \mathcal{P}^{MC}_{\texttt{L0M}}$. These weights are applied to all muon-mode MC samples for each given year. In addition, efficiency profiles are calculated using MC samples without any tagging requirements imposed, to examine the values of $\varepsilon_{\texttt{L0Muon-TOS}\,True}$ in MC. These efficiency profiles show good agreement with those found from tagged MC samples, suggesting the `L0Muon-TIS` tag causes minimal bias. Examples of the efficiency profiles, and the ratios between data and MC efficiencies, are shown in Fig. 4.6.

*Similar equations are used for the probability that at least one of the electrons triggers `L0Electron`, and that at least one of the pions triggers `L0Hadron`.

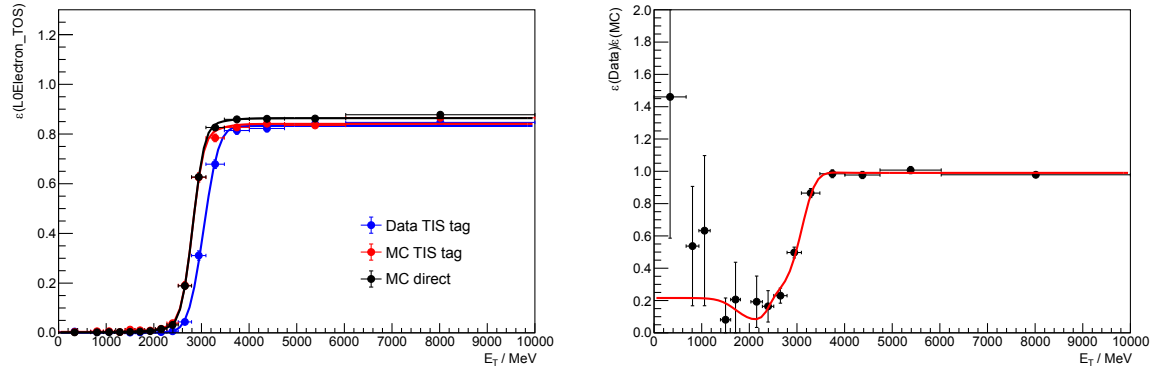


Fig. 4.7: Left - Efficiency profiles for `L0Electron-TOS` as a function of $E_T(e)$, in tagged data (blue), tagged MC (red), and untagged MC (black). Right - ratios of efficiencies for tagged data and tagged MC, as a function of $E_T(e)$. In both plots, the points show values in bins of $E_T(e)$ (plus associated uncertainties), while the curves show the values found from the functional parameterisations. These plots were produced using 2018 $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ samples.

4.2.5.2 `L0Electron`

Similarly to `L0Muon`, the `L0Electron-TOS` efficiencies are corrected using samples of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC and *sWeight*ed data, using both positively and negatively charged electrons, and combining the $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ samples to increase sample sizes. These samples are tagged using `L0Electron-TIS`. For each data-taking year, the `L0Electron-TOS` efficiencies are examined in bins of the $E_T(e^\pm)$ value measured offline, and are then fit using a sum of two error functions similar to eq. 4.3, plus an additional constant term. This term accounts for random noise in the ECAL, which can cause the `L0Electron` trigger to fire.

Such parameterised efficiency profiles are then used to compute efficiency correction weights according to $w_{\texttt{L0E}} = \mathcal{P}^{Data}_{\texttt{L0E}} / \mathcal{P}^{MC}_{\texttt{L0E}}$, which are applied to all electron-mode MC samples for a given year. Example efficiency profiles, and efficiency ratios, are shown in Fig. 4.7. As for `L0Muon-TOS`, the efficiency curves for tagged and untagged MC are in good agreement, suggesting no bias is caused by the `L0Electron-TIS` tag.

Note that the efficiency profiles are fit, rather than their ratios. If any imperfections in these parameterisations do not cancel between the data and MC, this could lead to poor agreement between the binned efficiency ratios and their parameterisations, as seen at low values of E_T in Fig. 4.7. A systematic uncertainty is assigned for such imperfect parameterisations, as documented in Section 5.7.2.

4.2.5.3 L0Hadron

The efficiencies of the L0Hadron-TOS trigger are calculated using combined samples of $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ MC and *sWeight*ed data from each data-taking year. Here, both pions from K^0_S-meson decays are considered, as well as soft pions from $K^{*+} \to K^0_S\pi^+$ decays. These samples are tagged by requiring one of the muons from the parent B-meson candidate to satisfy L0Muon-TOS, and the efficiencies of L0Hadron-TOS are then examined in 16 bins of $E_T(\pi)$. Efficiency profiles are parameterised using the same equation as for the L0Electron-TOS profiles. Then, correction weights are calculated according to the probabilities that at least one of the pions triggered L0Hadron, but none of the electrons triggered L0Electron:

$$w_{\texttt{L0H}} = \frac{\mathcal{P}^{Data}_{\texttt{L0H}} \cdot \left(1 - \mathcal{P}^{Data}_{\texttt{L0E}}\right)}{\mathcal{P}^{MC}_{\texttt{L0H}} \cdot \left(1 - \mathcal{P}^{MC}_{\texttt{L0E}}\right)} \tag{4.5}$$

As mentioned in Section 2.2.2.2, different segmentation sizes are used for different regions of the HCAL, with finer segmentation close to the beamline, and coarser segmentation far from the beamline. As shown for example in Fig. 4.8, large differences in efficiency ratios can be seen at low $E_T(\pi)$, depending on which region of the HCAL an energy cluster was recorded in. Therefore, separate sets of efficiency correction weights are calculated for each of these two HCAL regions, and applied to electron-mode MC samples for a given year.

As an additional check, the L0Hadron-TOS efficiency profiles are also measured using samples of pions from $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC and *sWeight*ed data, which are tagged with the requirement L0Electron-TOS on electrons from the parent B-meson candidate. As shown for example in Fig. 4.8, the resulting efficiency profiles show reasonable agreement with the muon-mode efficiency profiles. The same is true of the ratios of data and MC efficiencies, in each HCAL region. Therefore, only the muon-mode efficiency profiles are used to compute efficiency correction weights, as their sample sizes are larger, reducing statistical uncertainties.

4.2.5.4 L0Global-TIS

The efficiencies of the L0Global-TIS trigger are calculated using combined samples of $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ MC and *sWeight*ed data from each data-taking year. These samples are tagged by requiring one of the candidate muons to satisfy L0Muon-TOS, and are additionally required not to have any pions satisfying L0Hadron-TOS (as such pions were found to bias the L0Global-TIS efficiency). Data and MC efficiencies are calculated in bins of $p_T(B)$, as L0Global-TIS is usually fired by the other B-hadron from the parent $pp \to b\bar{b}X$, whose momentum is correlated to that of the candidate B-meson. No functions are fitted to the resulting efficiency profiles, due to a lack of any obvious, physically-motivated parameterisation choice.

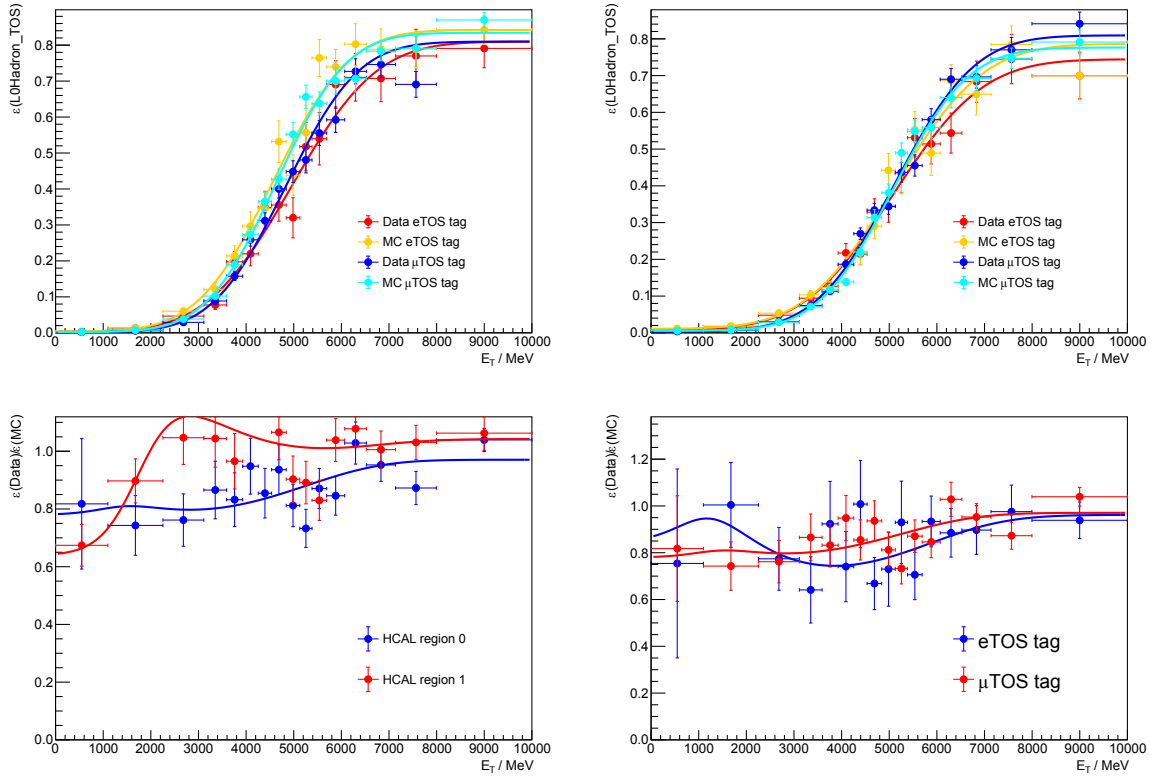


Fig. 4.8: Top - Efficiency profiles for `L0Hadron-TOS` as a function of $E_T(\pi)$, in `L0Muon`-tagged data (blue) and MC (cyan), and `L0Electron`-tagged data (red) and MC (orange), for energy clusters recorded in the inner (left) and outer (right) regions of the HCAL. Bottom left - ratios of efficiencies in `L0Muon`-tagged data and MC, for energy clusters in the inner (blue) and outer (red) regions of the HCAL. Bottom right - ratios of efficiencies in `L0Muon`-tagged (blue) and `L0Electron`-tagged (red) data and MC, in the inner HCAL region. These plots were produced using 2018 $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ and $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ samples.

Instead, correction weights are calculated using the binned efficiency values, and applied to electron-mode MC samples in a given year. These are found according to the probabilities that `L0Global-TIS` is satisfied, but that none of the electrons or pions in the candidate trigger `L0Electron` or `L0Hadron`:

$$w_{\texttt{L0I}} = \frac{\varepsilon^{TIS}_{Data}(p_T[B]) \cdot \left(1 - \mathcal{P}^{Data}_{\texttt{L0H}}\right) \cdot \left(1 - \mathcal{P}^{Data}_{\texttt{L0E}}\right)}{\varepsilon^{TIS}_{MC}(p_T[B]) \cdot \left(1 - \mathcal{P}^{MC}_{\texttt{L0H}}\right) \cdot \left(1 - \mathcal{P}^{MC}_{\texttt{L0E}}\right)} \tag{4.6}$$

The `L0Global-TIS` efficiencies are also examined using samples of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC and *sWeight*ed data, tagged by requiring `L0Electron-TOS` for one of the candidate electrons. Some discrepancies can be seen between the efficiency profiles found using the electron-

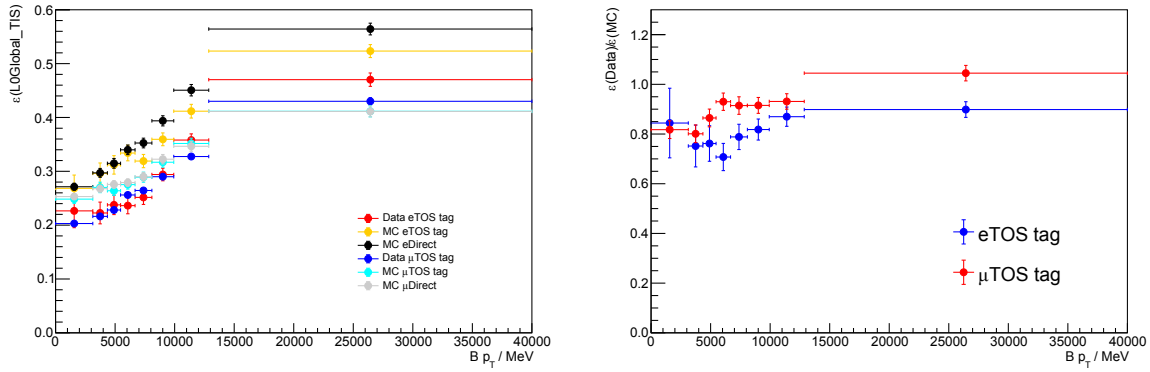


Fig. 4.9: Left - Efficiency profiles for `L0Global-TIS` as a function of $p_T(B)$, in muon control-mode tagged data (blue), tagged MC (cyan), untagged MC (grey), plus electron control-mode tagged data (red), tagged MC (orange), untagged MC (black). Right - ratios of efficiencies for muon control-mode data and MC (red), and electron control-mode data and MC (blue). These plots were produced using 2016 samples.

mode and muon-mode samples, indicating that one of the tags may bias the `L0Global-TIS` efficiency. This is also the case for ratios of efficiencies in data and MC, and particularly for data recorded in 2016 (see Fig. 4.9). Because of these discrepancies, the electron-mode efficiency profiles are used to evaluate a systematic uncertainty related to the choice of tag, as outlined in Section 5.7.3.

4.2.6 Residual discrepancies between data and simulation

Even after the previous corrections are applied, discrepancies can be seen between distributions in data and MC for some variables. This is particularly the case for some of the $\text{BDT}_{Comb.}$ input variables, which are sensitive to subtle mismodelling of the track and vertex reconstruction. Such discrepancies could cause biases in estimates of the $\text{BDT}_{Comb.}$ cut efficiencies, so the distributions in MC are corrected to mitigate these.

The simultaneous correction of so many variables presents some difficulties. One could try to compute separate correction weights for each variable, using a series of one-dimensional histograms. However, this is not possible if the variables are correlated. For example, imagine some variable a is first corrected to match data. Then, a second variable b is corrected. However, because a and b are correlated, the corrected MC distribution for a is pushed out of alignment with data when the corrections for b are also applied. One could deal with this by computing weights simultaneously using a multi-dimensional histogram, as used to compute the generator mismodelling weights in Section 4.2.3. However, this approach suffers from the "curse of dimensionality", as increasingly-large samples are needed to populate histograms

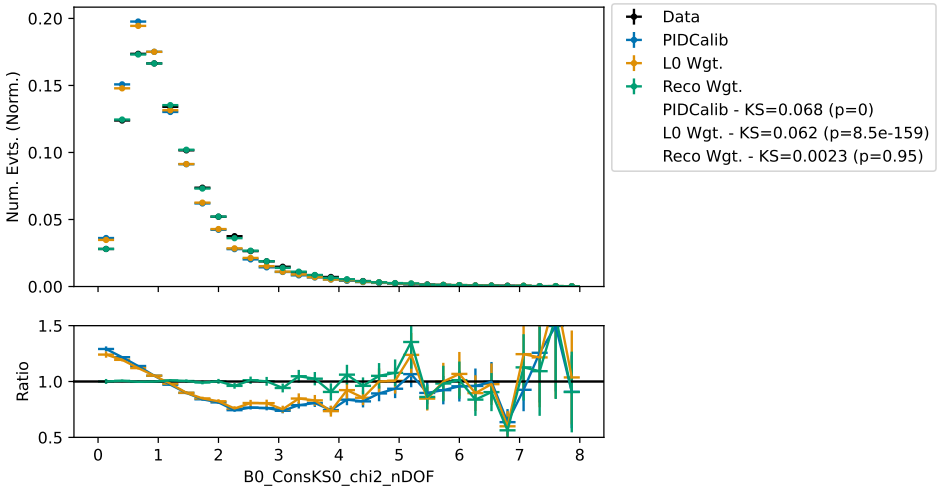
with many dimensions, controlling statistical uncertainties on the correction weights. Instead, a method based on decision trees is used. This is known as the `GBReweighter` algorithm, and is implemented in the HEPML package in PYTHON [317, 318].

Decision trees used for classification (as in Section 4.1.5) are optimised to find partitions which give the best separation between two samples. However, the decision trees here are instead optimised to find partitions which give regions with the largest discrepancies between MC and data. Once a decision tree has been constructed, the relative densities of data and MC in each region are used to calculate a set of correction weights. This process of training decision trees and calculating weights is then iterated. Importantly, when training each tree, the weights from previous trees are applied to the MC samples. Therefore, each decision tree is incentivised to correct the MC distribution in regions which were not well-corrected by the previous trees. This way, good agreement can eventually be found in all regions of variable-space.

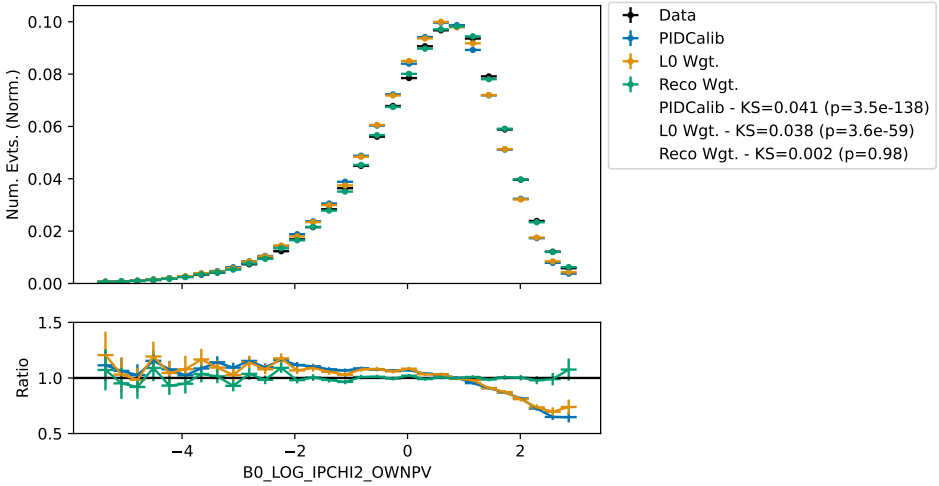
Such `GBReweighter` algorithms are trained to correct the distributions of variables where the disagreement between data and MC is most severe, namely:

- B-meson - p_T, η, $\chi^2_{\text{DTF}}/\text{nDoF}$, $\log\chi^2_{\text{IP}}$, $\log\chi^2_{\text{FD}}\,(\text{PV})$, $\text{ISO}_{\text{Vtx.}}$
- Long $\pi^\pm$ from K^0_{S}-meson decays - $\min(\log\chi^2_{\text{IP}}(\pi^\pm))$, $\max(\log\chi^2_{\text{IP}}(\pi^\pm))$.

Separate algorithms are trained for each J/ψ control-mode, data-taking year, and K^0_{S} reconstruction category, with samples of MC and data required to pass full selection requirements (except for cuts to $\text{BDT}_{Comb.}$). The data samples are *sWeight*ed using the mass-fit models in Section 4.5.2.1 and 4.5.2.3. Ten k-folds are used, to ensure the generalisation of the resulting weights is not impacted by any over-training of the `GBReweighter`. When adding weights to MC samples, any candidates which were not used to train the `GBReweighters` are assigned weights from randomly-chosen k-folds. In the following pages, these weights are denoted $w_{Resid.}$ or $w_{Reco.}$. As shown for example in Fig. 4.10, the agreement between J/ψ-mode MC and data is greatly improved when the resulting weights are applied. A set of hyper-parameters controlling the sizes and numbers of decision trees were tuned until such agreement was achieved.



(a): *B*-meson - $\chi^2_{\text{DTF}}/\text{nDoF}$



(b): *B*-meson - $\log \chi^2_{\text{IP}}$

Fig. 4.10: Comparison of some of the variables used in reconstruction-level reweighting for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ decays recorded across all years in *sWeight*ed data (black), and in MC when PID efficiency weights are applied (blue), when additional weights for generated kinematics and occupancy, K^0_S category, and L0 performance are applied (orange), and when additional weights for residual discrepancies are applied (green). The ratios of distributions found in MC and data are shown in the lower portions of the plots. Kolmogorov-Smirnov test values, and corresponding p-values, showing the compatibility between data and MC are given in the legends, for the different MC reweighting schemes.

4.2.7 q^2 resolution

Due to the finite resolutions on q^2 measurements, rare-mode candidates which fall within a given bin of q^2_{True} may end up being reconstructed in a different bin of $q^2_{Reco.}$. Such migration effects must be accounted for in the efficiency estimates (see Section 4.3.1). If momentum resolutions for leptons are not well described in simulation, the q^2 resolutions would also not be well described, causing systematic biases in efficiency calculations. Crucially, any such biases would not necessarily cancel between the rare mode and J/ψ control-mode for a given channel, as different q^2 regions are used to select them.

To correct for any such mismodelling, the "normalised q^2 resolution", given by $\left(q^2_{Reco.} - q^2_{True}\right)/q^2_{True}$, is measured in data and MC for each J/ψ control-mode. Note that for these modes q^2_{True} can be taken as the measured J/ψ mass [13]. The data and MC are required to pass full selection requirements. In addition, tight cuts are imposed on the B-meson mass computed with the dilepton mass constraint - $m(B) \in [5240, 5325\,\text{MeV}]$ for muon modes and $m(B) \in [5150, 5500\,\text{MeV}]$ for electron modes. These cuts reduce any combinatorial background surviving the $\text{BDT}_{Comb.}$ cuts to negligible levels.

For each channel, in each data-taking year*, unbinned maximum likelihood fits are carried out, using the normalised q^2 resolution in MC. This is parameterised using a sum of two Crystal Ball PDFs (see Section 4.5.1) with common mean μ_{MC} and width σ_{MC}. Such a parameterisation was chosen empirically, as it is able to describe lower- and upper-mass tails in the normalised q^2 resolution. These tails result from effects such as variation in momentum resolution between candidates, and bremsstrahlung emission and recovery. After this, a second unbinned maximum likelihood fit is carried out to the normalised q^2 resolution in data. The same PDF is used, but with the mean and width parameterised with two additional freely-floating "smearing parameters" μ_{Smear} and σ_{Smear}, such that $\mu_{Data} = \mu_{MC} + \mu_{Smear}$ and $\sigma_{Data} = \sqrt{\sigma^2_{MC} + \sigma^2_{Smear}}$. Note that no "under-smearing" of $q^2_{Reco.}$ in MC is permitted. In other words, the constraint $\sigma_{Data} > \sigma_{MC}$ is imposed, as seen in other $b \to s\ell^+\ell^-$ analyses. Other parameters controlling the shapes of the tails in the Crystal Ball PDFs are fixed to the values found in MC. Following this, the values of $q^2_{Reco.}$ in the relevant signal and control-mode MC channels are "smeared" to $q^2_{Reco.} + \Delta_{Smear} \cdot q^2_{True}$, where Δ_{Smear} is randomly sampled from a Gaussian distribution with mean μ_{Smear} and width σ_{Smear}, for each candidate.

Examples of the fits to control-mode MC and data are shown in Fig. 4.11, while the values of the smearing parameters in each channel and year are given in Table 4.12.

*Except for run-1, where 2011 and 2012 data is combined to reduce statistical uncertainties

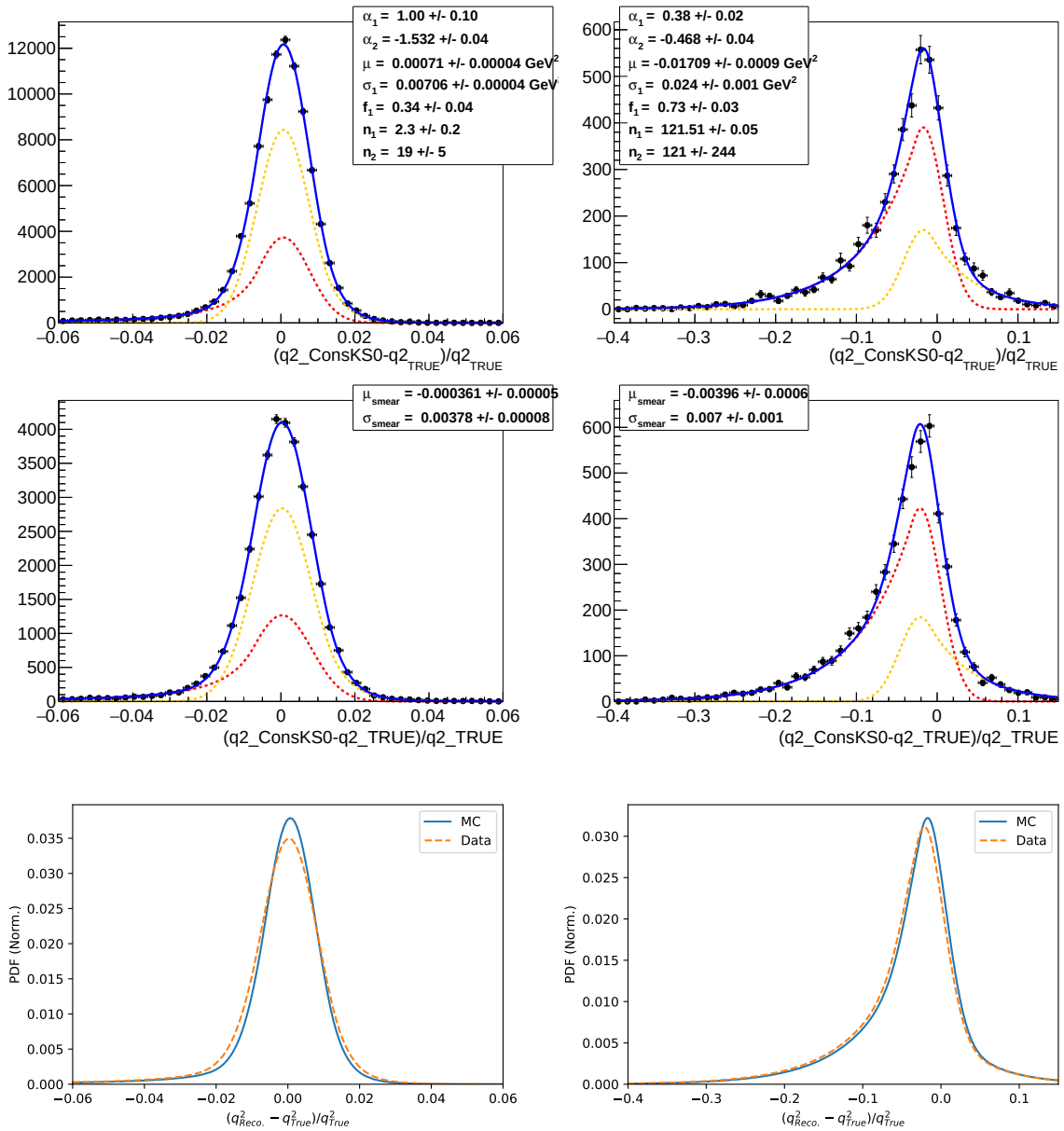


Fig. 4.11: Fits to the normalised q^2 resolution in 2018 MC (top) and data (middle) for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ decays (right). The Crystal Ball PDF parameters and smearing parameters are shown in the plot legends. Comparisons of the PDFs fitted in data and MC are also shown (bottom).

Decay	Year	μ_{Smear}	$\sigma_{Data}/\sigma_{MC}$
$B^0 \to K^0_S J/\psi(e^+e^-)$	Run-1	0.0029 ± 0.0008	1.048 ± 0.020
	2016	-0.0071 ± 0.0007	1.021 ± 0.018
	2017	-0.0057 ± 0.0007	1.0000000 ± 0.0000010
	2018	-0.0040 ± 0.0006	1.039 ± 0.016
$B^0 \to K^0_S J/\psi(\mu^+\mu^-)$	Run-1	0.00128 ± 0.00006	1.128 ± 0.007
	2016	-0.00071 ± 0.00006	1.160 ± 0.007
	2017	-0.00081 ± 0.00005	1.156 ± 0.006
	2018	-0.00036 ± 0.00005	1.134 ± 0.006
$B^+ \to K^{*+} J/\psi(e^+e^-)$	Run-1	0.0006 ± 0.0009	1.067 ± 0.024
	2016	-0.0064 ± 0.0009	1.031 ± 0.024
	2017	-0.0056 ± 0.0009	1.013 ± 0.022
	2018	-0.0059 ± 0.0008	1.00000 ± 0.00005
$B^+ \to K^{*+} J/\psi(\mu^+\mu^-)$	Run-1	0.00139 ± 0.00007	1.142 ± 0.008
	2016	-0.00065 ± 0.00007	1.153 ± 0.008
	2017	-0.00067 ± 0.00007	1.116 ± 0.008
	2018	-0.00035 ± 0.00006	1.128 ± 0.007

Table 4.12: Results of fits to J/ψ control-mode MC and data, for each channel and year. The definitions of these parameters are outlined in Section 4.2.7. Note that $\sigma_{Data}/\sigma_{MC}$ is shown here, rather than σ_{Smear} (which is actually floated in the fits to data). This is because $\sigma_{Data}/\sigma_{MC}$ is more easily interpretable. Note that some of the fits converged to $\sigma_{Smear} = 0$, corresponding to no additional smearing of $q^2_{Reco.}$ in MC, and giving $\sigma_{Data}/\sigma_{MC} = 1$.

4.3 Selection efficiencies

The efficiencies of the selection requirements for each decay are calculated using MC samples, to which the corrections outlined in Section 4.2 have been applied. Each selection efficiency is calculated using the equation:

$$\varepsilon_{Tot.} = \varepsilon_{Geo.} \times \varepsilon_{Gen.K^0_S} \times \frac{\sum^{Presel.} w_{Gen.}}{\sum^{Gen.} w_{Gen.}} \times \frac{\sum^{Norm.} \left(w_{PID} \cdot w_{Track.} \cdot w_{Gen.} \cdot w_{K^0_S} \cdot w_{L0} \right)}{\sum^{Presel.} \left(w_{K^0_S} \cdot w_{Gen.} \right)} \times \frac{\sum^{Sel.} \left(w_{PID} \cdot w_{Track.} \cdot w_{Gen.} \cdot w_{K^0_S} \cdot w_{L0} \cdot w_{Resid.} \right)}{\sum^{Norm.} \left(w_{PID} \cdot w_{Track.} \cdot w_{Gen.} \cdot w_{K^0_S} \cdot w_{L0} \cdot w_{Resid.} \right)} \quad (4.7)$$

Here, the sum $\sum^{Gen.}$ runs over all generator-level candidates, and $\sum^{Presel.}$ runs over all reconstructed candidates passing the stripping and preselection requirements (including cuts to q^2). The sum $\sum^{Sel.}$ runs over candidates passing full selection requirements, while $\sum^{Norm.}$ runs over candidates passing full selection requirements except for cuts to $\text{BDT}_{Comb.}$. In these sums, each candidate is weighted using the corrections defined in Section 4.2, with the sums in the denominators ensuring these weights are properly normalised.

The term $\varepsilon_{Geo.}$ refers to the efficiency for final-state particles to fall within the LHCb detector's angular acceptance. This cut is imposed during generation, to avoid simulating out-of-acceptance candidates, so $\varepsilon_{Geo.}$ is calculated by the LHCb simulation group when each MC sample is produced. For technical reasons, generator-level candidates where the K^0_S-meson has scattered from the detector material are not retained, and the factor $\varepsilon_{Gen.K^0_S}$ accounts for this.

Each decay's selection efficiencies are averaged across all data-taking years, with the efficiencies in each year weighted according to the corresponding $pp \to b\bar{b}X$ cross-sections and luminosities recorded in data [286]. Ratios of efficiencies are then used when extracting observables from data. All uncertainties are calculated using the normal approximation (see Section 3.2), and propagated using standard uncertainty propagation.

4.3.1 Accounting for q^2 bin migration

Note that eq. 4.7 gives efficiencies $\varepsilon_{Tot.}$ for candidates produced across a full range of q^2_{True}, relative to those reconstructed in a certain region of $q^2_{Reco.}$. So, normalising a fitted rare-mode yield $N^{Sel.}$ by this efficiency would give an estimate of the total number of decays $N^{Tot.}$ across the full range of q^2:

Decay Mode	f_{q^2}
$B^0 \to K^0_S e^+ e^-$	0.2708 ± 0.0007
$B^0 \to K^0_S \mu^+ \mu^-$	0.2707 ± 0.0007
$B^+ \to K^{*+} e^+ e^-$	0.2513 ± 0.0007
$B^+ \to K^{*+} \mu^+ \mu^-$	0.3279 ± 0.0007

Table 4.13: Values of f_{q^2} for each channel, averaged across data-taking years.

$$N^{Sel.}\left(q^2_{Reco.} \in \left[q^2_{Min.}, q^2_{Max.}\right]\right) = \varepsilon_{Tot.} \times N^{Tot.}\left(All\right) \tag{4.8}$$

However, this analysis aims to measure the *differential* electron-mode branching fractions, and ratios R_X^{-1}, within specific bins of q^2_{True}. Therefore, estimates are needed of the true numbers of decays which took place within those bins of q^2_{True} (*i.e.* $N^{Tot.}\left(q^2_{True} \in \left[q^2_{Min.}, q^2_{Max.}\right]\right)$). To do this, one can normalise the efficiency estimates by a factor f_{q^2}, defined as the fraction of decays which fall within a given bin of q^2_{True}:

$$N^{Tot.}(q^2_{True} \in \left[q^2_{Min.}, q^2_{Max.}\right]) = f_{q^2} \times N^{Tot.}(All) \tag{4.9}$$

$$\implies N^{Sel.}(q^2_{Reco.} \in \left[q^2_{Min.}, q^2_{Max.}\right]) = \varepsilon_{Tot.} \times \frac{N^{Tot.}\left(q^2_{True} \in \left[q^2_{Min.}, q^2_{Max.}\right]\right)}{f_{q^2}} \tag{4.10}$$

The values of f_{q^2} are calculated using samples of generator-level MC without any generator-level cuts applied, to avoid any biases which could result from potential correlations between such cuts and f_{q^2}. These samples were produced using the Cambridge HEP computing cluster. The resulting values of f_{q^2}, averaged across all years, are shown in Table 4.13.

4.3.2 Efficiency values

The total selection efficiencies estimated for each decay, within each year and also averaged across years, are shown in Table 4.14. Here, the rare-mode efficiencies have been normalised by the f_{q^2} factors. This table also shows the ratios between rare-mode and J/ψ control-mode efficiencies, and between the $\psi(2S)$ and J/ψ control-modes. These are used for the observable measurements, and some of the cross-checks in Chapter 6. While many of the selection efficiencies vary by up to $\mathcal{O}(50\%)$ between years for a given decay, the ratios of efficiencies

vary much less, illustrating the improved control offered by using J/ψ decays as a normalisation mode.

In addition, the efficiencies of different steps in the selection chain, averaged across years, are shown in Tables 4.15 and 4.16. Here also many of these efficiencies cancel in ratio with the J/ψ control modes, most notably those of the PID cuts, generator-level acceptance cuts, and (to an extent) the cuts to $\text{BDT}_{Comb.}$. For other parts of the selection where q^2 dependencies are greater, there is less cancellation. The trigger efficiency, for example, is strongly dependent on the momenta of the final-state leptons, leading to higher efficiencies as q^2 increases.

It should be noted that the electron-mode trigger efficiencies are $\sim 1.5-2$ times higher in run-2 than in run-1, as improved HLT bandwidth meant the L0 thresholds could be loosened in these years. However, this leads to increased background surviving the trigger cuts, meaning greater overlap between signal and background in the $\text{BDT}_{Comb.}$ distributions. Therefore, the $\text{BDT}_{Comb.}$ cuts in run-2 are ~ 0.5 times as efficient as in run-1.

Decay Mode	2011	2012	2016	2017	2018	Year Ave.
$B^0 \to K^0_S e^+ e^-$	$(3.86 \pm 0.20) \times 10^{-4}$	$(3.21 \pm 0.22) \times 10^{-4}$	$(4.21 \pm 0.16) \times 10^{-4}$	$(4.77 \pm 0.20) \times 10^{-4}$	$(4.20 \pm 0.11) \times 10^{-4}$	$(4.15 \pm 0.08) \times 10^{-4}$
$B^0 \to J/\psi\,(e^+e^-)\,K^0_S$	$(4.18 \pm 0.17) \times 10^{-4}$	$(3.69 \pm 0.14) \times 10^{-4}$	$(5.11 \pm 0.07) \times 10^{-4}$	$(5.12 \pm 0.10) \times 10^{-4}$	$(4.67 \pm 0.08) \times 10^{-4}$	$(4.69 \pm 0.05) \times 10^{-4}$
$B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$	$(4.58 \pm 0.14) \times 10^{-4}$	$(3.50 \pm 0.09) \times 10^{-4}$	$(5.07 \pm 0.13) \times 10^{-4}$	$(5.06 \pm 0.11) \times 10^{-4}$	$(4.23 \pm 0.08) \times 10^{-4}$	$(4.53 \pm 0.05) \times 10^{-4}$
$\frac{B^0 \to K^0_S e^+ e^-}{B^0 \to J/\psi(e^+e^-)K^0_S}$	0.92 ± 0.06	0.87 ± 0.07	0.823 ± 0.034	0.93 ± 0.04	0.899 ± 0.029	0.886 ± 0.019
$\frac{B^0 \to \psi(2S)(e^+e^-)K^0_S}{B^0 \to J/\psi(e^+e^-)K^0_S}$	1.10 ± 0.05	0.95 ± 0.04	0.993 ± 0.029	0.988 ± 0.028	0.905 ± 0.023	0.966 ± 0.014
$B^0 \to K^0_S \mu^+ \mu^-$	$(1.68 \pm 0.09) \times 10^{-3}$	$(1.58 \pm 0.07) \times 10^{-3}$	$(2.20 \pm 0.09) \times 10^{-3}$	$(2.39 \pm 0.06) \times 10^{-3}$	$(2.28 \pm 0.05) \times 10^{-3}$	$(2.134 \pm 0.032) \times 10^{-3}$
$B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$	$(2.043 \pm 0.015) \times 10^{-3}$	$(1.99 \pm 0.04) \times 10^{-3}$	$(2.499 \pm 0.010) \times 10^{-3}$	$(2.981 \pm 0.034) \times 10^{-3}$	$(2.771 \pm 0.014) \times 10^{-3}$	$(2.581 \pm 0.011) \times 10^{-3}$
$B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$	$(2.25 \pm 0.07) \times 10^{-3}$	$(2.00 \pm 0.05) \times 10^{-3}$	$(2.62 \pm 0.04) \times 10^{-3}$	$(2.98 \pm 0.04) \times 10^{-3}$	$(2.831 \pm 0.014) \times 10^{-3}$	$(2.642 \pm 0.017) \times 10^{-3}$
$\frac{B^0 \to K^0_S \mu^+ \mu^-}{B^0 \to J/\psi(\mu^+\mu^-)K^0_S}$	0.82 ± 0.04	0.80 ± 0.04	0.880 ± 0.035	0.803 ± 0.023	0.823 ± 0.018	0.827 ± 0.013
$\frac{B^0 \to \psi(2S)(\mu^+\mu^-)K^0_S}{B^0 \to J/\psi(\mu^+\mu^-)K^0_S}$	1.102 ± 0.034	1.006 ± 0.033	1.048 ± 0.018	0.999 ± 0.017	1.022 ± 0.007	1.024 ± 0.008
$B^+ \to K^{*+} e^+ e^-$	$(3.07 \pm 0.16) \times 10^{-4}$	$(2.19 \pm 0.15) \times 10^{-4}$	$(2.25 \pm 0.13) \times 10^{-4}$	$(2.54 \pm 0.12) \times 10^{-4}$	$(2.55 \pm 0.07) \times 10^{-4}$	$(2.46 \pm 0.05) \times 10^{-4}$
$B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$	$(3.06 \pm 0.09) \times 10^{-4}$	$(2.76 \pm 0.06) \times 10^{-4}$	$(2.85 \pm 0.07) \times 10^{-4}$	$(2.93 \pm 0.06) \times 10^{-4}$	$(2.72 \pm 0.05) \times 10^{-4}$	$(2.831 \pm 0.030) \times 10^{-4}$
$B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$	$(3.04 \pm 0.09) \times 10^{-4}$	$(2.68 \pm 0.06) \times 10^{-4}$	$(2.48 \pm 0.06) \times 10^{-4}$	$(2.63 \pm 0.06) \times 10^{-4}$	$(2.63 \pm 0.06) \times 10^{-4}$	$(2.634 \pm 0.029) \times 10^{-4}$
$\frac{B^+ \to K^{*+} e^+ e^-}{B^+ \to J/\psi(e^+e^-)K^{*+}}$	1.00 ± 0.06	0.80 ± 0.06	0.79 ± 0.05	0.87 ± 0.05	0.939 ± 0.031	0.869 ± 0.021
$\frac{B^+ \to \psi(2S)(e^+e^-)K^{*+}}{B^+ \to J/\psi(e^+e^-)K^{*+}}$	0.99 ± 0.04	0.971 ± 0.030	0.870 ± 0.030	0.896 ± 0.029	0.968 ± 0.028	0.930 ± 0.014
$B^+ \to K^{*+} \mu^+ \mu^-$	$(1.06 \pm 0.05) \times 10^{-3}$	$(8.4 \pm 0.4) \times 10^{-4}$	$(1.054 \pm 0.027) \times 10^{-3}$	$(1.310 \pm 0.029) \times 10^{-3}$	$(1.210 \pm 0.018) \times 10^{-3}$	$(1.128 \pm 0.013) \times 10^{-3}$
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$	$(1.372 \pm 0.033) \times 10^{-3}$	$(1.250 \pm 0.025) \times 10^{-3}$	$(1.347 \pm 0.029) \times 10^{-3}$	$(1.549 \pm 0.021) \times 10^{-3}$	$(1.532 \pm 0.019) \times 10^{-3}$	$(1.437 \pm 0.011) \times 10^{-3}$
$B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$	$(1.307 \pm 0.028) \times 10^{-3}$	$(1.147 \pm 0.028) \times 10^{-3}$	$(1.175 \pm 0.023) \times 10^{-3}$	$(1.448 \pm 0.017) \times 10^{-3}$	$(1.419 \pm 0.015) \times 10^{-3}$	$(1.318 \pm 0.009) \times 10^{-3}$
$\frac{B^+ \to K^{*+} \mu^+ \mu^-}{B^+ \to J/\psi(\mu^+\mu^-)K^{*+}}$	0.77 ± 0.04	0.67 ± 0.04	0.782 ± 0.027	0.845 ± 0.022	0.790 ± 0.016	0.785 ± 0.011
$\frac{B^+ \to \psi(2S)(\mu^+\mu^-)K^{*+}}{B^+ \to J/\psi(\mu^+\mu^-)K^{*+}}$	0.953 ± 0.031	0.917 ± 0.029	0.872 ± 0.026	0.934 ± 0.017	0.926 ± 0.015	0.917 ± 0.010

Table 4.14: Total efficiencies for the selection requirements outlined in Section 4.1, for different decay modes. The efficiencies are shown for each data-taking year, and averaged across these years. Ratios of efficiencies relative to the J/ψ control-modes are also shown.

Decay Mode	Generator-Level	Reco., Strip., Presel.	PID	Back. Vetoes
$B^0 \to K^0_S e^+ e^-$	0.20038 ± 0.00016	0.02622 ± 0.00016	0.8121 ± 0.0028	0.8204 ± 0.0022
$B^0 \to J/\psi\,(e^+e^-)\,K^0_S$	0.19563 ± 0.00016	0.02418 ± 0.00008	0.8110 ± 0.0017	0.7369 ± 0.0016
$B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$	0.19199 ± 0.00016	0.02189 ± 0.00008	0.8119 ± 0.0018	0.6348 ± 0.0018
$\frac{B^0 \to K^0_S e^+ e^-}{B^0 \to J/\psi(e^+e^-)K^0_S}$	1.0242 ± 0.0012	1.084 ± 0.008	1.001 ± 0.004	1.113 ± 0.004
$\frac{B^0 \to \psi(2S)\left(e^+e^-\right)K^0_S}{B^0 \to J/\psi(e^+e^-)K^0_S}$	0.9814 ± 0.0012	0.905 ± 0.005	1.0011 ± 0.0031	0.8614 ± 0.0031
$B^0 \to K^0_S \mu^+ \mu^-$	0.20139 ± 0.00016	0.03497 ± 0.00029	0.9393 ± 0.0026	0.9134 ± 0.0022
$B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$	0.19624 ± 0.00016	0.04367 ± 0.00010	0.9406 ± 0.0008	0.7970 ± 0.0010
$B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$	0.19387 ± 0.00017	0.04764 ± 0.00016	0.9431 ± 0.0011	0.6874 ± 0.0016
$\frac{B^0 \to K^0_S \mu^+ \mu^-}{B^0 \to J/\psi(\mu^+\mu^-)K^0_S}$	1.0262 ± 0.0012	0.801 ± 0.007	0.9986 ± 0.0029	1.1460 ± 0.0031
$\frac{B^0 \to \psi(2S)\left(\mu^+\mu^-\right)K^0_S}{B^0 \to J/\psi(\mu^+\mu^-)K^0_S}$	0.9879 ± 0.0012	1.091 ± 0.005	1.0027 ± 0.0015	0.8625 ± 0.0023
$B^+ \to K^{*+} e^+ e^-$	0.18067 ± 0.00014	0.01152 ± 0.00010	0.787 ± 0.004	0.8616 ± 0.0028
$B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$	0.17351 ± 0.00014	0.01045 ± 0.00005	0.7898 ± 0.0024	0.8077 ± 0.0019
$B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$	0.17519 ± 0.00014	$(7.34 \pm 0.04) \times 10^{-3}$	0.7942 ± 0.0026	0.7882 ± 0.0022
$\frac{B^+ \to K^{*+} e^+ e^-}{B^+ \to J/\psi(e^+e^-)K^{*+}}$	1.0413 ± 0.0012	1.102 ± 0.011	0.997 ± 0.006	1.067 ± 0.004
$\frac{B^+ \to \psi(2S)\left(e^+e^-\right)K^{*+}}{B^+ \to J/\psi(e^+e^-)K^{*+}}$	1.0096 ± 0.0012	0.702 ± 0.005	1.006 ± 0.005	0.9759 ± 0.0035
$B^+ \to K^{*+} \mu^+ \mu^-$	0.17751 ± 0.00014	0.01588 ± 0.00012	0.9239 ± 0.0028	0.8834 ± 0.0024
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$	0.16804 ± 0.00014	0.01941 ± 0.00010	0.9295 ± 0.0018	0.8220 ± 0.0020
$B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$	0.17593 ± 0.00015	0.01501 ± 0.00007	0.9337 ± 0.0017	0.8131 ± 0.0020
$\frac{B^+ \to K^{*+} \mu^+ \mu^-}{B^+ \to J/\psi(\mu^+\mu^-)K^{*+}}$	1.0564 ± 0.0012	0.818 ± 0.008	0.994 ± 0.004	1.075 ± 0.004
$\frac{B^+ \to \psi(2S)\left(\mu^+\mu^-\right)K^{*+}}{B^+ \to J/\psi(\mu^+\mu^-)K^{*+}}$	1.0470 ± 0.0012	0.773 ± 0.005	1.0046 ± 0.0027	0.9891 ± 0.0034

Table 4.15: Efficiencies for different steps in the selection chain, for each decay mode, averaged across years. Each column is defined as follows: "Generator-Level" refers to the generator-level angular acceptance cuts, "Reco., Strip., Presel.," refers to the reconstruction, stripping, and preselection requirements, including the q^2 cuts, but excluding the PID cuts. This term also includes the factor $\varepsilon_{Gen.K^0_S}$. "PID" refers to the PID cuts imposed in the stripping and preselection. "Back. Vetoes" refers to kinematic and geometric vetoes for physical backgrounds. Efficiencies are calculated sequentially - *i.e.* when calculating the efficiency for a given selection requirement, all previous requirements are applied. This table is continued in Table 4.16.

Decay Mode	Trigger	Mass	BDT
$B^0 \to K^0_S e^+ e^-$	0.2765 ± 0.0027	0.9289 ± 0.0030	0.499 ± 0.008
$B^0 \to J/\psi\,(e^+e^-)\,K^0_S$	0.3312 ± 0.0018	0.9823 ± 0.0010	0.548 ± 0.004
$B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$	0.3962 ± 0.0023	0.9975 ± 0.0004	0.562 ± 0.004
$\frac{B^0 \to K^0_S e^+ e^-}{B^0 \to J/\psi (e^+e^-) K^0_S}$	0.835 ± 0.009	0.9457 ± 0.0032	0.912 ± 0.015
$\frac{B^0 \to \psi(2S)\left(e^+e^-\right) K^0_S}{B^0 \to J/\psi (e^+e^-) K^0_S}$	1.196 ± 0.009	1.0155 ± 0.0011	1.026 ± 0.010
$B^0 \to K^0_S \mu^+ \mu^-$	0.576 ± 0.004	0.9579 ± 0.0023	0.646 ± 0.006
$B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$	0.6097 ± 0.0013	0.99721 ± 0.00018	0.6620 ± 0.0018
$B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$	0.6787 ± 0.0020	0.99829 ± 0.00018	0.6505 ± 0.0027
$\frac{B^0 \to K^0_S \mu^+ \mu^-}{B^0 \to J/\psi (\mu^+\mu^-) K^0_S}$	0.945 ± 0.007	0.9605 ± 0.0023	0.976 ± 0.009
$\frac{B^0 \to \psi(2S)\left(\mu^+\mu^-\right) K^0_S}{B^0 \to J/\psi (\mu^+\mu^-) K^0_S}$	1.113 ± 0.004	1.00109 ± 0.00026	0.983 ± 0.005
$B^+ \to K^{*+} e^+ e^-$	0.329 ± 0.004	0.917 ± 0.004	0.615 ± 0.008
$B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$	0.4049 ± 0.0025	0.9668 ± 0.0013	0.656 ± 0.004
$B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$	0.5211 ± 0.0029	0.9848 ± 0.0010	0.666 ± 0.004
$\frac{B^+ \to K^{*+} e^+ e^-}{B^+ \to J/\psi (e^+e^-) K^{*+}}$	0.812 ± 0.011	0.949 ± 0.005	0.937 ± 0.014
$\frac{B^+ \to \psi(2S)\left(e^+e^-\right) K^{*+}}{B^+ \to J/\psi (e^+e^-) K^{*+}}$	1.287 ± 0.011	1.0186 ± 0.0017	1.015 ± 0.008
$B^+ \to K^{*+} \mu^+ \mu^-$	0.624 ± 0.004	0.9584 ± 0.0020	0.818 ± 0.005
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$	0.6958 ± 0.0027	0.9862 ± 0.0008	0.8420 ± 0.0028
$B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$	0.7947 ± 0.0023	0.9884 ± 0.0007	0.8368 ± 0.0029
$\frac{B^+ \to K^{*+} \mu^+ \mu^-}{B^+ \to J/\psi (\mu^+\mu^-) K^{*+}}$	0.897 ± 0.007	0.9718 ± 0.0022	0.972 ± 0.006
$\frac{B^+ \to \psi(2S)\left(\mu^+\mu^-\right) K^{*+}}{B^+ \to J/\psi (\mu^+\mu^-) K^{*+}}$	1.142 ± 0.006	1.0023 ± 0.0011	0.994 ± 0.005

Table 4.16: Efficiencies for different steps in the selection chain, for each decay mode, averaged across years. Each column is defined as follows: "Trigger" refers to the L0 and HLT selections, plus fiducial trigger cuts. "Mass" refers to the $m(B)$ windows. "BDT" refers to cuts to the $\text{BDT}_{Comb.}$ output. Efficiencies are calculated sequentially - *i.e.* when calculating the efficiency for a given selection requirement, all previous requirements are applied. This table is a continuation of Table 4.15

4.3.2.1 Values with different simulation corrections

The selection efficiencies are also calculated with variations to the MC corrections. First, efficiencies are calculated without any corrections applied to MC. Then, the efficiencies are recalculated with sequential application of the corrections outlined in Section 4.2. These efficiencies are shown in Tables 4.17 and 4.18, relative to the efficiencies found with uncorrected MC. These tables also show the changes caused to measurements of the branching fractions, R_X^{-1} ratios, and the $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ cross-checks outlined in Chapter 6.

One can see that the selection efficiencies for each channel are reduced to $\sim 65-80\%$ times their uncorrected values when all corrections are applied, with the largest changes occurring when the generator-level kinematic and occupancy weights are included. However, the changes of efficiency ratios relative to the J/ψ control-modes are much smaller. Hence, the changes to measurements of rare-mode branching fractions, R_X^{-1}, and $R_{\psi(2S)}^{-1}$ cross-checks are also small. For example, the values of $R_{K_S^0}^{-1}$ and $R_{K^{*+}}^{-1}$ are within 5% of their uncorrected values, illustrating the stability offered when normalising measurements with the J/ψ control-modes.

In addition, the efficiencies are calculated with different reweighting schemes for the detector occupancy, compared to the nominal scheme in Section 4.2.3. Two such alternative schemes are examined; one with only `nTracks` reweighted as an occupancy proxy, and the other without any occupancy proxy reweighted at all. The generator kinematic weights, K_S^0 category weights, and trigger weights are all re-evaluated for each of these alternative schemes. This is particularly important for the trigger weights, as the `L0Global-TIS` efficiency is highly dependent on the detector occupancy. These efficiencies are shown in Table 4.19, relative to the nominal efficiencies where `nTracks` and `nSPDHits` are both reweighted. Once again, while the changes in individual selection efficiencies are quite large, the changes in ratio with the J/ψ control-mode are much smaller. Even with these extreme modifications to the occupancy reweighting, the $R_{K_S^0}^{-1}$ and $R_{K^{*+}}^{-1}$ ratios remain within $\sim 2-5\%$ of their nominal values. This suggests any systematic uncertainties associated to the occupancy reweighting would be far below these values.

	PID Weights	Electron Trk.	Gen. Level.	K^0_S Cat.	L0 Trig.	Resid.	q^2 Smear
$B^0 \to K^0_S \mu^+ \mu^-$	0.986	0.986	0.803	0.804	0.806	0.788	0.787
$B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$	0.988	0.988	0.809	0.810	0.813	0.797	0.797
$B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$	0.989	0.989	0.806	0.806	0.812	0.793	0.792
$B^0 \to K^0_S e^+ e^-$	0.958	0.945	0.716	0.716	0.692	0.673	0.675
$B^0 \to J/\psi\,(e^+e^-)\,K^0_S$	0.948	0.934	0.688	0.689	0.665	0.654	0.655
$B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$	0.932	0.919	0.686	0.686	0.664	0.655	0.655
$\mathcal{BF}\left(B^0 \to K^0_S e^+ e^-\right)$	0.990	0.989	0.962	0.962	0.961	0.972	0.970
$\mathcal{BF}\left(B^0 \to K^0_S \mu^+ \mu^-\right)$	1.00	1.00	1.01	1.01	1.01	1.01	1.01
$R^{-1}_{K^0_S}$	0.988	0.987	0.954	0.955	0.952	0.962	0.958
$r^{-1}_{J/\psi}\left(B^0\right)$	1.04	1.06	1.18	1.18	1.22	1.22	1.22
$R^{-1}_{\psi(2S)}\left(B^0\right)$	1.02	1.02	0.999	0.999	0.999	0.994	0.994

Table 4.17: Variation in selection efficiencies averaged across all years, and associated observables, as the MC corrections outlined in Section 4.2 are applied. All values are shown relative to those without any corrections applied. Efficiencies and ratios for the B^0-meson channels are shown in Table 4.18

	PID Weights	Electron Trk.	Gen. Level.	K^0_S Cat.	L0 Trig.	Resid.	q^2 Smear
$B^+ \to K^{*+}\mu^+\mu^-$	0.992	0.992	0.790	0.790	0.791	0.769	0.769
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$	0.992	0.992	0.826	0.825	0.829	0.810	0.810
$B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$	0.994	0.994	0.803	0.803	0.809	0.790	0.790
$B^+ \to K^{*+}e^+e^-$	0.964	0.948	0.722	0.723	0.692	0.668	0.669
$B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$	0.957	0.943	0.712	0.713	0.689	0.675	0.675
$B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$	0.951	0.938	0.708	0.709	0.687	0.676	0.674
$\mathcal{BF}\,(B^+ \to K^{*+}e^+e^-)$	0.993	0.994	0.986	0.987	0.997	1.01	1.01
$\mathcal{BF}\,(B^+ \to K^{*+}\mu^+\mu^-)$	1.00	1.00	1.05	1.05	1.05	1.05	1.05
$R^{-1}_{K^{*+}}$	0.992	0.994	0.943	0.944	0.951	0.958	0.958
$r^{-1}_{J/\psi}\,(B^+)$	1.04	1.05	1.16	1.16	1.20	1.20	1.20
$R^{-1}_{\psi(2S)}\,(B^+)$	1.01	1.01	0.978	0.978	0.98	0.973	0.976

Table 4.18: Variation in selection efficiencies averaged across all years, and associated variables, as the MC corrections outlined in Section 4.2 are applied. All values are shown relative to those without any corrections applied. Efficiencies and ratios for the B^+-meson channels are shown in Table 4.17

Decay Mode	nTracks Only / Nominal	No Occ. / Nominal
$B^0 \to K^0_S \mu^+ \mu^-$	0.916	1.099
$B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$	0.902	1.097
$B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$	0.915	1.107
$B^0 \to K^0_S e^+ e^-$	0.902	1.106
$B^0 \to J/\psi\,(e^+e^-)\,K^0_S$	0.933	1.140
$B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$	0.922	1.136
$\mathcal{BF}\left(B^0 \to K^0_S e^+ e^-\right)$	0.967	0.970
$\mathcal{BF}\left(B^0 \to K^0_S \mu^+ \mu^-\right)$	1.015	1.002
$r^{-1}_{J/\psi}\left(B^0\right)$	0.967	0.962
$R^{-1}_{\psi(2S)}\left(B^0\right)$	1.025	1.014
$R^{-1}_{K^0_S}$	1.049	1.033
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$	0.948	1.035
$B^+ \to K^{*+} \mu^+ \mu^-$	0.969	1.045
$B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$	0.978	1.068
$B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$	0.972	1.073
$B^+ \to K^{*+} e^+ e^-$	0.999	1.103
$B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$	0.973	1.078
$\mathcal{BF}\left(B^+ \to K^{*+} e^+ e^-\right)$	1.029	1.027
$\mathcal{BF}\left(B^+ \to K^{*+} \mu^+ \mu^-\right)$	1.021	1.009
$r^{-1}_{J/\psi}\left(B^+\right)$	0.976	0.964
$R^{-1}_{\psi(2S)}\left(B^+\right)$	1.030	1.027
$R^{-1}_{K^{*+}}$	0.993	0.982

Table 4.19: Variation in selection efficiencies averaged across all years, and associated observables, if the detector occupancy reweighting procedure in Section 4.2.3 is modified. Changes are shown when only `nTracks` is reweighted, and when no occupancy proxy is reweighted at all. All values are shown relative to those without any corrections applied.

4.4 Backgrounds

Physical sources of background are outlined in this section, as are the strategies used to suppress them, and estimates of contamination levels where necessary. Backgrounds to $B^0 \to K^0_S \ell^+ \ell^-$ decays are discussed first, followed by backgrounds to $B^+ \to K^{*+} \ell^+ \ell^-$.

4.4.1 Backgrounds to $B^0 \to K^0_S \ell^+ \ell^-$ decays

4.4.1.1 b-hadron decays without long-lived particles

If all final-state hadrons are identified as pions, the decays $B^0 \to \pi^+ \pi^- \ell^+ \ell^-$, $\Lambda_b^0 \to p K^- \ell^+ \ell^-$, and $B^0 \to K^{*0} \ell^+ \ell^-$ (followed by $K^{*0} \to K^+ \pi^-$) can all be misidentified as $B^0 \to K^0_S \ell^+ \ell^-$ decays. This is also the case for the equivalent J/ψ and $\psi(2S)$ control modes. Such backgrounds would peak close to the signal in $m(B)$, as all particles are reconstructed. The K^0_S candidates in such backgrounds would have negligible lifetime, so can be discarded using the cut $\tau(K^0_S) > 0.0005$ ns. This cut reduces these backgrounds to a negligible level, and is $\sim 98-99\%$ efficient on signal and control-mode MC.

4.4.1.2 $\Lambda_b^0 \to \Lambda \ell^+ \ell^-$

The decay $\Lambda_b^0 \to \Lambda \ell^+ \ell^-$, followed by $\Lambda \to p\pi^-$, can form a peaking structure in the $m(B)$ signal region if the final-state proton is misidentified as a pion. This is also the case for the equivalent J/ψ-mode and $\psi(2S)$-mode decays. Kinematic requirements can be used to suppress such decays, using the reconstructed Λ_b^0 and Λ masses. These are calculated by assuming the reconstructed "pion" with the highest value of $\texttt{ProbNN}(p)$ is really a proton, and assigning it the corresponding mass hypothesis. The Λ region is defined as $m(\Lambda) \in [1100, 1135\,\text{MeV}]$, while the Λ_b^0 region is defined as $m(\Lambda_b^0) \in [5500, 5675\,\text{MeV}]$ for muon-mode decays and $m(\Lambda_b^0) \in [4800, 5800\,\text{MeV}]$ for electron-mode decays. Candidates falling simultaneously within the Λ and Λ_b^0 regions are discarded. These cuts are $\sim 97-99\%$ efficient on signal and control-mode MC, and reduce Λ_b^0 contamination to a negligible level.

4.4.1.3 Partially reconstructed kaon resonances

The decays $B^0 \to K^{*0} \ell^+ \ell^-$ followed by $K^{*0} \to K^0_S \pi^0$, and $B^+ \to K^{*+} \ell^+ \ell^-$ followed by $K^{*+} \to K^0_S \pi^+$, can appear as $B^0 \to K^0_S \ell^+ \ell^-$ candidates if the pion is missed during reconstruction. This is also the case for the equivalent control-mode decays. These "partially-reconstructed" (PR) backgrounds form "shoulder" shapes in $m(B^0)$ below the signal region, due to missing momentum carried by the pion, as shown in Fig. 4.12.

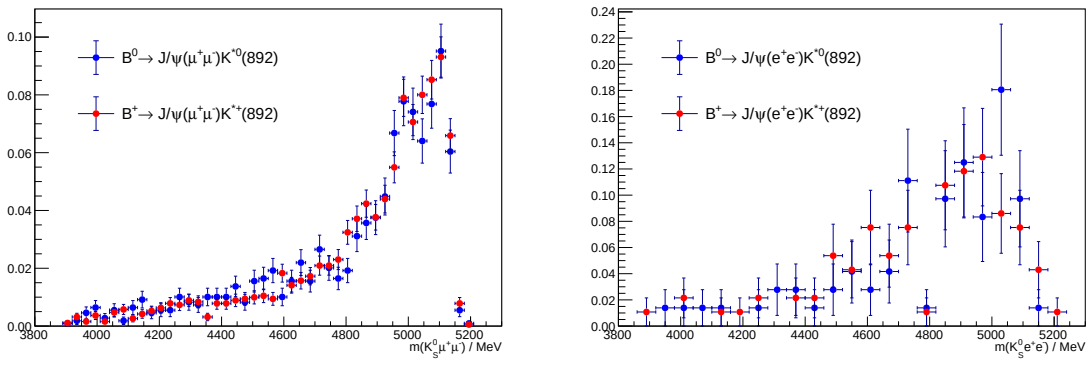


Fig. 4.12: Distributions of $m(B^0)$ found for partially-reconstructed $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^{*0}$ (blue) and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ decays (red), in the muon mode (left) and electron mode (right). These plots were produced using 2012 MC samples, passing full selection requirements (except for the $m(B)$ requirements). Note that both the B^0 and B^+ meson decays have similar mass distributions owing to their similar masses, and the similar masses of the missing π^+ and π^0 mesons.

As mentioned in Section 3.3, the $m(B^0)$ resolution for control-mode decays is improved greatly from the use of dilepton mass constraints in `DecayTreeFitter`. Therefore, the lower limit of $m(B^0)$ used in the control-mode mass-fits is high enough to discard PR backgrounds. This is also the case in the rare muon-mode mass-fits, as the $m(B^0)$ resolution remains good, even without a dilepton mass constraint. However, there is a large overlap in $m(B^0)$ between signal and PR background in the rare electron-mode channels, owing to bremsstrahlung emission. Therefore, the PR backgrounds are included as components in the mass-fits outlined in Section 4.5.4.2, with freely-varying yields.

Decays with higher-order K^* resonances may also form partially-reconstructed backgrounds. However, these lie below the signal $m(B)$ regions due to the high momenta carried by unreconstructed pions, so are neglected.

4.4.1.4 Open charm cascade decays

There are numerous B-meson decays involving intermediate open charm mesons (referred to as "cascade" decay processes) which could contaminate the signal $m(B^0)$ region. The processes with the highest potential contamination rates involve the semileptonic and hadronic decays $B^0 \to D^-\ell^+\nu$ and $B^0 \to D^-\pi^+$, followed by either $D^- \to K^0_S\ell^-\overline{\nu}$ or $D^- \to K^0_S\pi^-$. The branching fractions of these processes are as follows [13]:

$$\mathcal{B}\left(B^0 \to D^- \ell^+ \nu\right) = (2.20 \pm 0.10)\,\%, \quad \mathcal{B}\left(B^0 \to D^- \pi^+\right) = (2.52 \pm 0.13) \times 10^{-3}$$
$$\mathcal{B}\left(D^- \to K^0 e^- \overline{\nu}_e\right) = (8.73 \pm 0.10)\,\%, \quad \mathcal{B}\left(D^- \to K^0_S \pi^-\right) = (1.47 \pm 0.08)\,\%$$

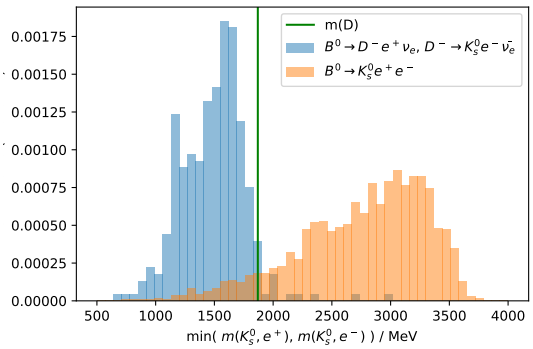
The four combinations of these decays have total branching fractions of $\mathcal{O}(10^{-3} - 10^{-4})$, far higher than the signal branching fraction* of $\mathcal{B}\left(B^0 \to K^0 \mu^+ \mu^-\right) = (3.39 \pm 0.34) \times 10^{-7}$ [13]. However, the processes involving $B^0 \to D^- \pi^+$ and $D^- \to K^0_S \pi^-$ include one or more $\pi^\pm \to \ell^\pm$ misidentification, so are suppressed by PID requirements. In addition, as the intermediate D-mesons in all cascade processes fly a short distance, the reconstructed electrons form poor dilepton and B^0 vertices, so they will be suppressed by the vertex-quality requirements imposed by the stripping and BDT selections. Finally, processes involving $B^0 \to D^- \ell^+ \nu$ or $D^- \to K^0_S \ell^- \overline{\nu}$ have low reconstructed B-meson masses, due to the momentum carried by unreconstructed neutrinos.

Kinematic requirements are used to further suppress cascade processes. For processes with $D^- \to K^0_S \pi^-$, vetoes are applied on the invariant masses of the K^0_S and each "lepton", where the lepton is assigned a pion mass. In the electron mode, the lepton momentum found without bremsstrahlung recovery is used, as it would not have emitted bremsstrahlung if it is really a pion. The veto applied is $|m\left(K^0_S, \pi^\pm\right) - m\left(D^-\right)| > 40\,\text{MeV}$, where $m\left(D^-\right)$ is the global average measured mass of 1870 MeV. This cut is 97% and 99% efficient, respectively in $B^0 \to K^0_S e^+ e^-$ and $B^0 \to K^0_S \mu^+ \mu^-$ MC.

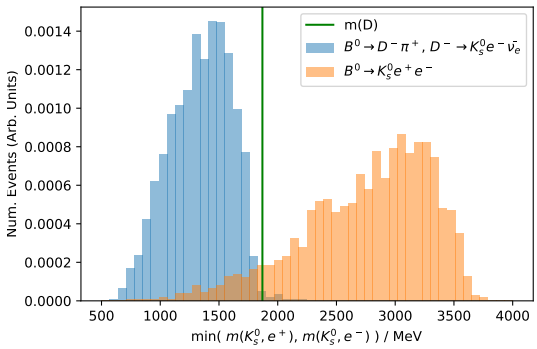
For processes with $D^- \to K^0_S \ell^- \overline{\nu}$, similar vetoes are applied. Due to the unreconstructed neutrinos, the invariant mass of the K^0_S and one of the leptons will lie below $m\left(D^-\right)$. Therefore, the veto $m\left(K^0_S, \ell^\pm\right) > m\left(D^-\right)$ is applied, using both leptons. This requirement is 91% and 94% efficient in $B^0 \to K^0_S e^+ e^-$ and $B^0 \to K^0_S \mu^+ \mu^-$ MC, respectively.

Expected contamination Some cascade background candidates may survive these vetoes, and the expected contamination rates must be assessed. Only the electron-mode contamination is considered, as if an ECAL cluster is falsely associated to an electron candidate, the reconstructed value of $m(K^0_S, e^\pm)$ will be larger than its true value, potentially allowing it to survive the kinematic vetoes. The improved resolution and narrower $m(B)$ ranges in the muon modes mean cascade processes are less likely to cause significant contamination there.

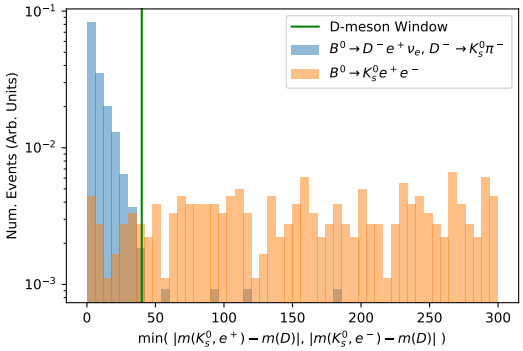
*Due to poorly-constrained experimental values for the rare electron-mode branching fractions, the muon-mode branching fractions were used to calculate background contamination rates, relative to both the muon and electron mode signals, throughout this section. The assumption of lepton-universality underlying this approach is a conservative one, as muon-mode branching fractions are seen to be lower than electron-mode branching fractions in other $b \to s\ell^+\ell^-$ processes, meaning physical values of B/S in the electron mode were likely smaller than estimated here.



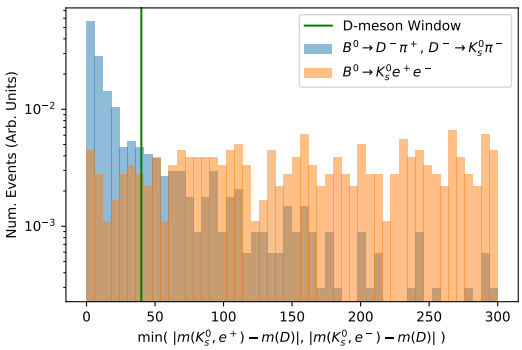
(a): $B^0 \to D^- e^+ \nu_e, D^- \to K^0_S e^- \overline{\nu}_e$



(b): $B^0 \to D^- \pi^+, D^- \to K^0_S e^- \overline{\nu}_e$



(c): $B^0 \to D^- e^+ \nu_e, D^- \to K^0_S \pi^-$



(d): $B^0 \to D^- \pi^+, D^- \to K^0_S \pi^-$

Fig. 4.13: Distributions of $B^0 \to K^0_S e^+ e^-$ decays and cascade processes in $min(m(K^0_S, e^-), m(K^0_S, e^+))$, and $min(|m(K^0_S, e^+) - m(D)|, |m(K^0_S, e^-) - m(D)|)$. Cuts imposed on these variables, which veto cascade background processes, are highlighted. Signal events are required to pass full selection requirements (except cascade vetoes). $D^- \to K^0_S e^- \overline{\nu}_e$ events are required to pass preselection and $m(B^0)$ requirements. $D^- \to K^0_S \pi^-$ events are only required to pass preselection.

The expected contamination rates for the four cascade processes are calculated using MC samples, simulated under 2012 conditions. Prior to this analysis, MC was only available for $B^0 \to D^- \pi^+$ followed by $D^- \to K^0_S \pi^-$. Therefore, the other three samples were produced using the Cambridge HEP group computing system. New DecFiles were also prepared for these decay processes. To avoid unnecessary use of computing resources, generator-level requirements were used to discard candidates which would not pass full selection requirements. These included cuts to the invariant mass of the final-state particles except for the neutrinos (*i.e.* the "visible mass" for reconstructed B^0 candidates), and cuts of $p_T > 200\,\text{MeV}$ on these particles. In addition, ReDecay was used to speed up the generation of events (see Section 2.2.4.1 for details). The distributions of $m\left(K^0_S, e^\pm\right)$ seen in MC samples for the four cascade

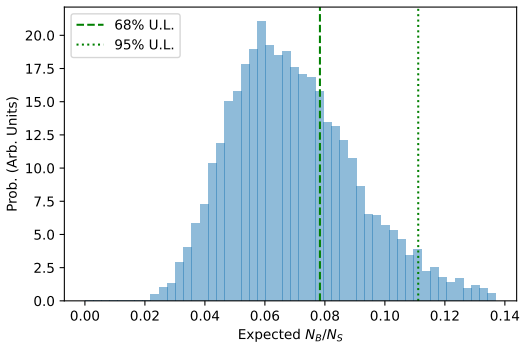
First decay	Second decay	No Vetoes	With Vetoes
$B^0 \to D^- e^+ \nu_e$	$D^- \to K^0_S e^- \overline{\nu}_e$	1.8	0.078
	$D^- \to K^0_S \pi^-$	0.18	0.008
$B^0 \to D^- \pi^+$	$D^- \to K^0_S e^- \overline{\nu}_e$	1.1	0.011
	$D^- \to K^0_S \pi^-$	0.0034	2.8×10^{-5}

Table 4.20: Upper-limits (from 68% quantiles) for the expected background / signal ratio from different cascade-background processes, relative to $B^0 \to K^0_S e^+ e^-$, with and without the application of kinematic vetoes for these backgrounds.

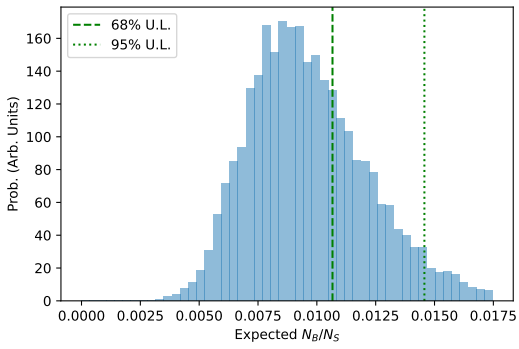
processes, as compared to $B^0 \to K^0_S e^+ e^-$, are shown in Fig. 4.13. One can see that a small number of cascade candidates fall outside the vetoed regions.

The selection efficiencies for the different cascade processes are estimated both with and without application of the kinematic vetoes. The efficiencies of PID requirements are corrected using a set of data-driven weights as detailed in Section 4.2.1. This includes the efficiencies for misidentified pions. Very small numbers of MC candidates pass the full selection requirements, meaning the normal approximation used for evaluation of signal and control-mode efficiencies in Section 4.3 cannot be applied. Instead, the Bayesian method detailed in Section 3.2 is used to calculate efficiencies of the selection steps after the preselection (namely the trigger, BDT, $m(B^0)$ window, q^2 region, and vetoes for other backgrounds). Furthermore, the efficiencies of the cascade vetoes are assumed to factorise from all selection requirements (except for the $m(B^0)$ window, as the efficiencies of the $D^- \to K^0_S e^- \overline{\nu}_e$ vetoes were seen to be higher for candidates falling within this region). Therefore, the efficiencies of the cascade vetoes are calculated using MC candidates passing the preselection and $m(B^0)$ requirements.

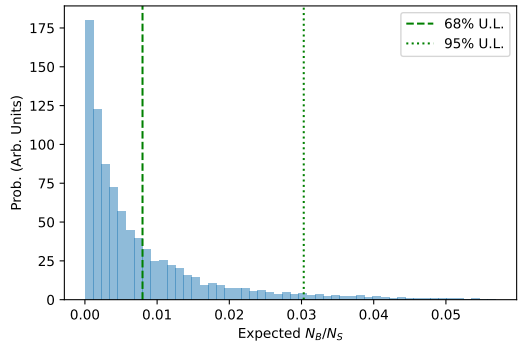
After this, the total selection efficiency for $B^0 \to K^0_S e^+ e^-$ decays is calculated using the nominal method in Section 4.3. The expected ratios of background yields to signal yields (B/S) are then calculated for each cascade process, using the efficiencies and branching fractions for signal and background [13]. In addition to the uncertainties on selection efficiencies discussed above, uncertainties on these branching fraction measurements are also propagated. Note that one could instead simply calculate the total number of expected background candidates. However, by presenting B/S instead, one can easily see which sources of background may be



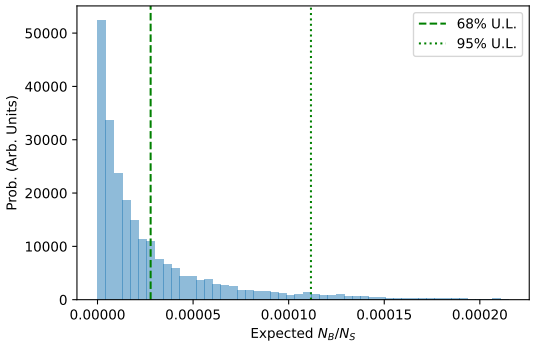
(a): $B^0 \to D^- e^+ \nu_e$, $D^- \to K^0_S e^- \overline{\nu}_e$



(b): $B^0 \to D^- \pi^+$, $D^- \to K^0_S e^- \overline{\nu}_e$



(c): $B^0 \to D^- e^+ \nu_e$, $D^- \to K^0_S \pi^-$



(d): $B^0 \to D^- \pi^+$, $D^- \to K^0_S \pi^-$

Fig. 4.14: Probability distributions for the expected background / signal ratio from different cascade-background processes, relative to $B^0 \to K^0_S e^+ e^-$. The 68% and 95% quantiles of these distributions are highlighted.

significant, requiring further study or incorporation into a mass-fit model. Throughout this section, any sources of background with $S/B \gtrsim 1\%$ are studied further*†.

The resulting probability distributions of B/S for each process are shown in Fig. 4.14. Note that they are very asymmetric, owing to the Bayesian method used for the background efficiency estimates. Upper-limits for the expected contamination rates are taken from the 68% upper-quantiles of these distributions, as summarised in Table 4.20. This is particularly large for $B^0 \to D^- e^+ \nu_e$ followed by $D^- \to K^0_S e^- \overline{\nu}_e$, with a contamination rate of $B/S = 7.8\%$. Such

*This represents a fairly conservative threshold, as it falls well below the statistical uncertainties of the final measurements ($\sim 20-25\%$, see Section 7.2), and the statistical uncertainties anticipated from toy studies ($\sim 30\%$, see Section 7.1)

†When carrying out such further study or incorporation into a mass-fit model, the total number of expected background candidates is instead calculated relative to the relevant $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ control mode. This provides a more precise estimate for the number of candidates, which is not affected by the large uncertainties on the $B \to K^{(*)}\ell^+\ell^-$ signal branching fractions [13].

First decay	Second decay	$\frac{N(B^0_s\to\ldots)}{N(B^0\to\ldots)}$	Upper-limits	
			B^0-mode	B^0_s-mode
$B^0_{(s)} \to D^-_{(s)} e^+ \nu_e$	$D^-_{(s)} \to K^0_S e^- \bar{\nu}_e$	0.036	1.8	0.065
	$D^-_{(s)} \to K^0_S \pi^-$	0.075	0.18	0.014
$B^0_{(s)} \to D^-_{(s)} \pi^+$	$D^-_{(s)} \to K^0_S e^- \bar{\nu}_e$	0.012	1.1	0.012
	$D^-_{(s)} \to K^0_S \pi^-$	0.024	0.0034	8.2×10^{-5}

Table 4.21: Estimates of the upper-limits on background / signal ratios from $D^\pm_s$-mode cascade processes, prior to the application of cascade background vetoes. These are found from the ratios of $D^\pm_s$-mode and $D^\pm$-mode yields, and the previously estimated upper-limits for $D^\pm$-mode cascade backgrounds (Table 4.20).

contamination could bias the measured yield for $B^0 \to K^0_S e^+ e^-$, and this effect is considered in Section 5.13.

Cascade processes with D^-_s mesons Cascade processes of the form $B^0_s \to D^-_s (K^0_S X^-) Y^+$ could also pollute the $B^0 \to K^0_S e^+ e^-$ signal region. Given their identical topologies, the expected contamination rates for $B^0 \to D^- (K^0_S X^-) Y^+$ decays can be used to estimate the contamination from these analogous D^-_s-mode processes.

The ratios of expected yields from a D^-_s-mode process, relative to the equivalent D^--mode process, can be calculated using the relevant branching fractions and the fragmentation-fraction $f_s/f_d \approx 0.25$, which describes the relative numbers of B^0_s and B^0 mesons produced in pp collisions at LHCb [13, 319]. Then, these ratios can be combined with the estimated upper-limits on contamination from a given D^--mode process, to find the upper-limit on contamination from the equivalent D^-_s-mode process. This makes the assumption that selection efficiencies are the same for these decays, which should be approximately correct for all selection requirements except for the cascade background vetoes. The increased mass of the D^-_s means more candidates are likely to survive these vetoes. Ratios of contamination from D^--mode and D^-_s-mode processes are shown in Table 4.21, together with the upper-limits on the contamination rates without the application of cascade background vetoes.

In order to get an idea of the kinematics of the D^-_s-mode processes, MC samples were produced using RapidSim (see Section 2.2.4.1). A custom smearing technique is used for electron momenta, to mimic the effects of bremsstrahlung emission and recovery (see Section 5.13). The kinematic stripping and preselection requirements were applied to these samples. The equivalent D^--mode processes and the $B^0 \to K^0_S e^+ e^-$ signal were also simulated using

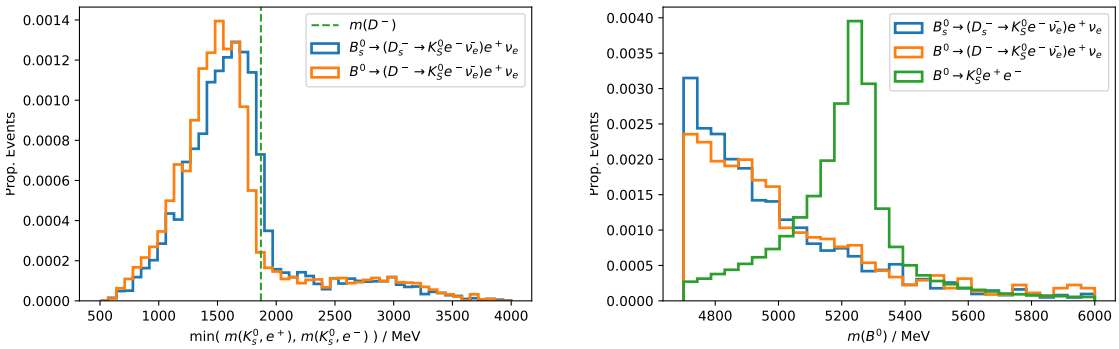


Fig. 4.15: Distributions of kinematic distributions for RapidSim samples of decays of $B_s^0 \to D_s^- e^+ \nu_e$ followed by $D_s^- \to K_S^0 e^- \overline{\nu}_e$, compared to the equivalent D^--meson cascade process and $B^0 \to K_S^0 e^+ e^-$ signal. Left - Distributions of $min(m(K_S^0, e^-), m(K_S^0, e^+))$ for candidates falling within $4700 < m(B^0) < 6000\,\text{GeV}$. The cut of $m(D^-)$ used to veto cascade backgrounds is indicated by the vertical dashed line. Right - Distributions of $m(B^0)$ for candidates passing kinematic vetoes for $m(D^-)$ cascade processes.

RapidSim, for comparison. Note that estimates of veto efficiencies using these samples would probably not be particularly accurate. However, the samples should give a good indication of the kinematics for these decays, relative to the equivalent D^--meson cascade processes.

Decays of $B_s^0 \to D_s^- e^+ \nu_e$ followed by $D_s^- \to K_S^0 e^- \overline{\nu}_e$ cause the most contamination before the imposition of cascade vetoes, at $\sim 5\%$ relative to the signal yield. However, due to the loss of energy from the D_s^+-daughter neutrino, these decays have similar distributions of $min(m(K_S^0, e^-), m(K_S^0, e^+))$ to the equivalent D^--mode process, meaning they should be vetoed efficiently by the veto for these decays (see Fig. 4.15). In addition, these decays have a smoothly-falling distribution in $m(B^0)$, so should be reasonably well parameterised as part of the combinatorial background. Decays of $B_s^0 \to D_s^- \pi^+$ followed by $D_s^- \to K_S^0 e^- \overline{\nu}_e$ have a pre-veto contamination rate of $\sim 1\%$, and should also be efficiently vetoed by the cuts to $m(K_S^0, e^\pm)$, so can be neglected.

The next-most problematic process, of $B_s^0 \to D_s^- e^+ \nu_e$ followed by $D_s^- \to K_S^0 \pi^-$, has a background-to-signal ratio of $\sim 1\%$. Note that in this case, these events will not be vetoed by the equivalent D^--mode veto of $|m(K_S^0, e^\pm) - m(D^-)| > 40\,\text{MeV}$. However, due to the energy carried by the final-state neutrino, these decays lie at low reconstructed B-mass, with a smoothly-falling distribution (see Fig. 4.16). Hence, these decays are not anticipated to bias the fitted signal yield at an appreciable level, and are neglected. The process of $B_s^0 \to D_s^- \pi^+$ followed by $D_s^- \to K_S^0 \pi^-$ gives a pre-veto contamination of $\mathcal{O}(0.01\%)$ relative to the signal, so is neglected.

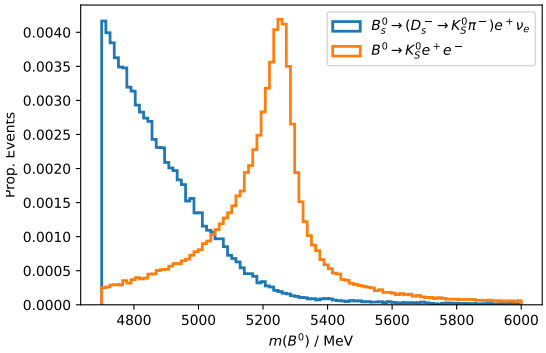


Fig. 4.16: Distribution of $m(B^0)$ for decays of $B_s^0 \to D_s^- e^+ \nu_e$ followed by $D_s^- \to K_S^0 \pi^-$, compared to $B^0 \to K_S^0 e^+ e^-$ signal decays. These distributions are taken from RAPIDSIM candidates which pass kinematic vetoes for D^- cascade processes.

Cascade processes involving decays to D_s^{*+} mesons may also contaminate the signal region. However, these are are suppressed by a factor of ~ 0.66 compared to D_s^+-mode decays, due to their lower branching fractions. In addition, they will lie at lower reconstructed B^0-mass, due loss of energy in the $D_s^{*+} \to D_s^+ \pi^+$ transition. Hence, they are neglected.

4.4.1.5 Fully hadronic decays

If both pions are misidentified as leptons, $B^0 \to K_S^0 \pi^+ \pi^-$ decays could pollute both the electron and muon channels. These $B^0 \to K^0 \pi^+ \pi^-$ decays have a total branching fraction of $(4.97 \pm 0.18) \times 10^{-5}$, approximately 150 times that of $B^0 \to K^0 \ell^+ \ell^-$ [13], meaning they could present a significant background. With all final-state particles reconstructed, and no emission of bremsstrahlung by the pions, the reconstructed B-masses of such decays form a peak very close to the signal. Hence, the contamination rates from such decays should be quantified. In contrast to the cascade processes outlined previously, $B^0 \to K_S^0 \pi^+ \pi^-$ decays lack any intermediate particles with finite lifetime. Therefore, such decays will form a good-quality B^0 vertex, so are unlikely to be suppressed by stripping or BDT requirements. Instead, they are suppressed by the trigger and PID requirements.

In addition to non-resonant $B^0 \to K_S^0 \pi^+ \pi^-$ decays, many resonances are present in the $\pi^+ - \pi^-$ and $K_S^0 - \pi^\pm$ systems. Some of these are shown in Table 4.22. As these resonances may affect the reconstructed q^2 distribution of $B^0 \to K_S^0 \pi^+ \pi^-$ decays, they must be accounted for in estimates of the contamination rates.

Simulation This background is studied using MC samples of $B^0 \to K_S^0 \pi^+ \pi^-$ decays in all years. Two different decay models were used to generate them: The PHSP model generates non-

Resonance Struct.	Branching fraction
Non-Resonant	$(1.39^{+0.26}_{-0.18}) \times 10^{-5}$
$K^0\rho^0$	$(3.4 \pm 1.1) \times 10^{-6}$
$K^*(892)^+\ \pi^-$	$(7.5 \pm 0.4) \times 10^{-6}$
$K_0^*(1430)^+\ \pi^-$	$(3.3 \pm 0.7) \times 10^{-6}$
$K^0\ f_0(980)$	$(8.1 \pm 0.8) \times 10^{-6}$
$K^0\ f_2(1270)$	$(2.7^{+1.3}_{-1.2}) \times 10^{-6}$
Total	$(4.97 \pm 0.18) \times 10^{-5}$

Table 4.22: The total branching fraction of $B^0 \to K^0\pi^+\pi^-$ decays, along with the branching fractions of the non-resonant component, and a (non-exhaustive) selection of contributing resonances [13].

resonant decays, according to the kinematic phase-space available to the final-state particles [285]. The `FLATSQDALITZ` model generates events which are uniformly distributed in the square Dalitz plane*; a two-dimensional co-ordinate system defined according to the mass of the π^+-π^- pair (m'), and the helicity angle between the π^+ and K^0_S, in the rest-frame of the π^+-π^- pair (θ'), each normalised to a range of 0-1 [320].

In order to use these samples, their decay kinematics must be reweighted to reflect those seen in physical $B^0 \to K^0_S\pi^+\pi^-$ decays. This should account for the branching fractions of the contributing resonances, and preferably any interference between them. Luckily, an amplitude analysis of $B^0 \to K^0_S\pi^+\pi^-$ decays, studying CP violation, has been published by LHCb [321]. The corresponding analysis note includes a full specification of the different components included in the analysis, including their line-shapes, fitted magnitudes, and relative phases. By generating toys according to this model, one can obtain the physical distribution of $B^0 \to K^0_S\pi^+\pi^-$ decays in (m', θ'). By comparing these to the distributions for the `PHSP` and `FLATSQDALITZ` models, one can obtain weights which map between the three different decay models.

To do this, the LHCb amplitude model was re-implemented in LAURA++; a software package used for three-body amplitude fits using the Dalitz plot formalism [322]. Then, 1,000,000 toy MC events were generated from this model, and used to produce a two-dimensional histogram in 50 bins each of m' and θ'. Similarly, toys were generated from the `PHSP` and `FLATSQDALITZ` models, using RAPIDSIM's interface with EVTGEN, and corresponding histograms were produced. By taking the ratios of histogram bin-contents from two different

*In amplitude analyses, this coordinate scheme is often used to simplify the description of efficiency profiles.

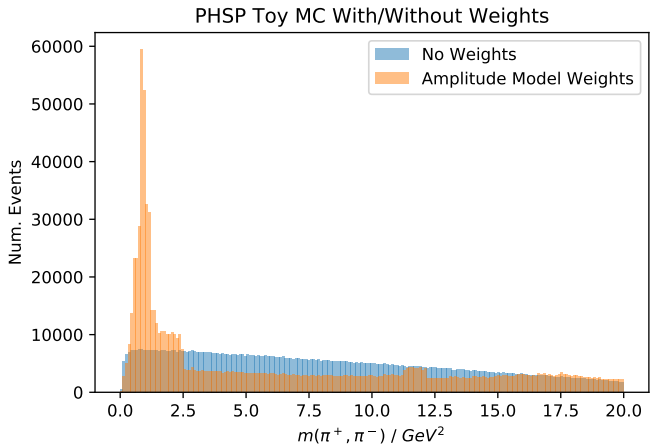


Fig. 4.17: Distribution of q^2_{True} seen in a RapidSim sample of non-resonant $B^0 \to K^0_S\pi^+\pi^-$ decays, and with the decay kinematics reweighted to match the amplitude model given in [321].

models, sets of weights were obtained. The distribution of q^2 found in a RapidSim sample of $B^0 \to K^0_S\pi^+\pi^-$ events, generated with the PHSP model, is shown in Fig. 4.17. This also shows the distribution found when the decay kinematics are reweighted to match those given by the amplitude model.

Expected contamination As noted above, the L0 trigger and PID requirements are the main aspects of the selection which suppress $B^0 \to K^0_S\pi^+\pi^-$ decays. Therefore, the total efficiencies ($\varepsilon(PID) \times \varepsilon(Trig.)$) must be calculated accurately when estimating the corresponding contamination rates. Note that both misidentified pions must pass the leptonic PID requirements, while only one pion needs to pass the L0 trigger in the `L0M` and `L0E` trigger categories. Therefore, in order to estimate $\varepsilon(PID) \times \varepsilon(Trig.)$, one needs to estimate the values of $\varepsilon(PID\ \&\ L0)$ and $\varepsilon(PID\ \&\ !L0)$ found for misidentified pions.

This presents some difficulties, as the efficiencies of PID requirements for a misidentified pion are highly dependent on whether it also fired the `L0Muon` or `L0Electron` line (and *vice-versa*). As shown in Table 4.23, the PID efficiencies are significantly higher for pions which fire a leptonic trigger (*i.e.* $\varepsilon(PID|L0) > \varepsilon(PID|!L0)$, and similarly $\varepsilon(L0|PID) > \varepsilon(L0|!PID)$). This effect is particularly large for pions reconstructed as muons, as the requirements used to define the `L0Muon` trigger and the `isMuon` PID variable are extremely similar. In addition, the trigger requirements for $B^0 \to K^0_S\mu^+\mu^-$ candidates in this analysis mean at least one pion must satisfy `L0Muon`, so the total contribution to $\varepsilon(PID) \times \varepsilon(Trig.)$ from such pions is very large. While the effect is also large for pions reconstructed as electrons, the trigger requirements for $B^0 \to K^0_S e^+e^-$ in this analysis mean many candidates are selected using

		Run-1	Run-2
Muon	μTOS PID Eff.	0.99	0.99
	$!\mu TOS$ PID Eff.	0	0
	μTOS Prop.	0.52	0.46
Electron	$eTOS$ PID Eff.	0.37	0.16
	$!eTOS$ PID Eff.	0.016	0.0047
	$eTOS$ Prop.	0.043	0.058

Table 4.23: The efficiencies of leptonic PID requirements on a misidentified π^+, depending on whether the lepton fired the appropriate leptonic L0 trigger. The proportion of π^+ which fired the leptonic trigger are also shown. These quantities were calculated using run-1 and run-2 MC samples of $B^0 \to K^0_S \pi^+ \pi^-$ decays (via cuts to simulated PID variables, rather than any data-driven weights).

`L0Hadron` and `L0Global-TIS`, so the corresponding effect on $\varepsilon(PID) \times \varepsilon(Trig.)$ is lesser (though still significant). Note also that PID requirements are more efficient for misidentified pions in run-1, compared to run-2.

In general, efficiencies of PID requirements are calculated using sets of data-driven weights (see Section 4.2.1). To evaluate $\varepsilon(PID) \times \varepsilon(Trig.)$, accounting for differing efficiencies depending on whether pions have fired an L0 line, one could adopt a "PID-unbiased" method. Here, the efficiencies for misidentified pions would be calculated as $\varepsilon(PID|L0) \times \varepsilon(L0)$ and $\varepsilon(PID|!L0) \times \varepsilon(!L0)$, with $\varepsilon(L0)$ and $\varepsilon(!L0)$ calculated using $B^0 \to K^0_S \pi^+ \pi^-$ MC. Then, two sets of PID efficiency weights would be calculated - one using pions which fired an L0 line (giving $\varepsilon(PID|L0)$), and the other using pions which did not fire an L0 line (giving $\varepsilon(PID|!L0)$). Pions which did not fire the L0 can be readily selected in the pion PID calibration samples, meaning $\varepsilon(PID|!L0)$ can be easily calculated. In contrast, for technical reasons, one cannot select pions which *did* fire a given L0 line, meaning $\varepsilon(PID|L0)$ cannot be calculated.

Because of this, a "trigger-unbiased" method is used to calculate the selection efficiencies of misidentified $B^0 \to K^0_S \pi^+ \pi^-$ decays. Rather than evaluating the mis-ID efficiency as $\varepsilon(PID|Trig.) \times \varepsilon(Trig.)$, with a "PID-unbiased" calculation for $\varepsilon(Trig.)$, the efficiency is calculated as $\varepsilon(Trig.|PID) \times \varepsilon(PID)$, with a "trigger-unbiased" value for $\varepsilon(PID)$. Trigger-unbiased PID weights are used to calculate the PID efficiencies for the misidentified pions (*i.e.* using $\varepsilon(PID|!L0)$). Finally, lower and upper bounds for $\varepsilon(Trig.|PID)$ are calculated using two extreme assumptions:

	Reco. Lepton	Back. Trig. Contam	Sig. Trig. Contam	Ave. Contam
Run-1	Electron	0.055 ± 0.007	0.095 ± 0.012	0.075 ± 0.015
	Muon	$(2.8 \pm 0.8) \times 10^{-5}$	0.00113 ± 0.00016	0.00058 ± 0.00033
Run-2	Electron	0.0052 ± 0.0007	0.0081 ± 0.0010	0.0067 ± 0.0012
	Muon	$(1.35 \pm 0.21) \times 10^{-5}$	0.00104 ± 0.00012	0.00053 ± 0.00031
Total	Electron	0.0155 ± 0.0018	0.0259 ± 0.0030	0.021 ± 0.004
	Muon	$(1.61 \pm 0.25) \times 10^{-5}$	0.00105 ± 0.00012	0.00053 ± 0.00031

Table 4.24: Estimated contamination from $B^0 \to K^0_S \pi^+ \pi^-$ decays compared to muon-mode and electron-mode $B^0 \to K^0_S \ell^+ \ell^-$ signal. The background efficiency is calculated using the trigger-unbiased method outlined above. Contamination rates found using the lower, upper, and nominal estimates for the trigger efficiency are shown.

1. $\varepsilon_{Back.}(Trig.)$ - MC for $B^0 \to K^0_S \pi^+ \pi^-$ is used. The increase in $\varepsilon(Trig.)$ due to the correlations between the PID and L0 for misidentified pions is not modelled. Hence, the estimated efficiency will be lower than its true value.

2. $\varepsilon_{Sig.}(Trig.)$ - MC for $B^0 \to K^0_S e^+ e^-$ is used. As this sample contains correctly-identified leptons, the estimated efficiency will be higher than the true value of $\varepsilon(Trig.)$ for $B^0 \to K^0_S \pi^+ \pi^-$ decays.

In both cases, PID weights for trigger-unbiased misidentified pions are applied to the reconstructed "leptons" when calculating $\varepsilon(Trig.)$. This ensures that any correlation between a misidentified pion's momentum and its PID performance is accounted for. The nominal value of $\varepsilon(Trig.)$ is then taken as follows, using a method recommended for an estimate of systematic uncertainty from two extreme values [323, Section 6.5.3]:

$$\varepsilon(Trig.) = \frac{\varepsilon_{Sig.}(Trig.) + \varepsilon_{Back.}(Trig.)}{2}, \quad \sigma_{\varepsilon(Trig.)} = \frac{\varepsilon_{Sig.}(Trig.) - \varepsilon_{Back.}(Trig.)}{\sqrt{12}} \tag{4.11}$$

The efficiencies of all selection requirements apart from the trigger and lepton PID are calculated in a standard manner. The total selection efficiencies for $B^0 \to K^0_S \pi^+ \pi^-$ decays are then combined with the corresponding selection efficiencies for signal decays, plus the $B^0 \to K^0 \mu^+ \mu^-$ and $B^0 \to K^0 \pi^+ \pi^-$ branching fractions [13], to calculate expected background-to-signal contamination rates. These are shown in Table 4.24. The contamination rates found

with the upper, lower, and averaged estimates of $\varepsilon(Trig.)$ are shown. In addition to the contamination rates across all years, the rates are also shown separately for run-1 and run-2.

The contamination rates in the $B^0 \to K^0_S \mu^+ \mu^-$ mode are well below 1%, so $B^0 \to K^0_S \pi^+ \pi^-$ decays are can be neglected here. In contrast, contamination of $\sim 2\%$ relative to signal can be seen in the $B^0 \to K^0_S e^+ e^-$ mode, for the combination of all years. The contamination is higher in run-1, at $\sim 7\%$, compared to $\sim 0.6\%$ in run-2. This reflects the differing PID efficiencies between the two runs noted earlier.

With this comparatively high contamination rate, $B^0 \to K^0_S \pi^+ \pi^-$ decays would ideally be suppressed with additional selection requirements. However, kinematic vetoes using $m(K^0_S, \pi^+, \pi^-)$ proved inefficient on $B^0 \to K^0_S e^+ e^-$ MC, as did tightened electron PID requirements. Therefore, a component for $B^0 \to K^0_S \pi^+ \pi^-$ decays is included in the maximum likelihood fit to data, with its yield constrained using the selection efficiencies calculated here (see Section 4.5.4.2 for details).

4.4.2 Backgrounds to $B^+ \to K^{*+} \ell^+ \ell^-$ decays

There are various sources of background to $B^+ \to K^{*+} \ell^+ \ell^-$ decays. Many of these are analogous to backgrounds to $B^0 \to K^0_S \ell^+ \ell^-$, and are dealt with in similar ways:

- Kaon resonance decays $K^{(*)} \to K^{*+} \pi$ may be partially-reconstructed, giving low values of $m(B^+)$. They must be parameterised in the $B^+ \to K^{*+} e^+ e^-$ mass-fit model (see Section 4.5.4.2).
- The cascade decay processes $B^+ \to \overline{D}{}^0 \ell^+ \nu$ and $B^+ \to \overline{D}{}^0 \pi^+$ followed by $\overline{D}{}^0 \to K^{*+} \ell^- \overline{\nu}$ or $\overline{D}{}^0 \to K^{*+} \pi^-$ are vetoed with the kinematic requirements $m\left(K^{*+}, \ell^-\right) > m\left(D^0\right)$ and $|m\left(K^{*+}, \pi^-\right) - m\left(D^0\right)| > 40\,\text{MeV}$.
- Decays without long-lived particles are vetoed with a cut to the reconstructed K^0_S lifetime.
- Decays involving Λ^0_b baryons are discarded with cuts to $m(\Lambda)$.
- Fully-hadronic decays $B^+ \to K^{*+} \pi^+ \pi^-$ are suppressed by PID and trigger requirements. The residual contamination is estimated in this section.

Other background sources are particular to $B^+ \to K^{*+} \ell^+ \ell^-$, and are outlined in the following pages.

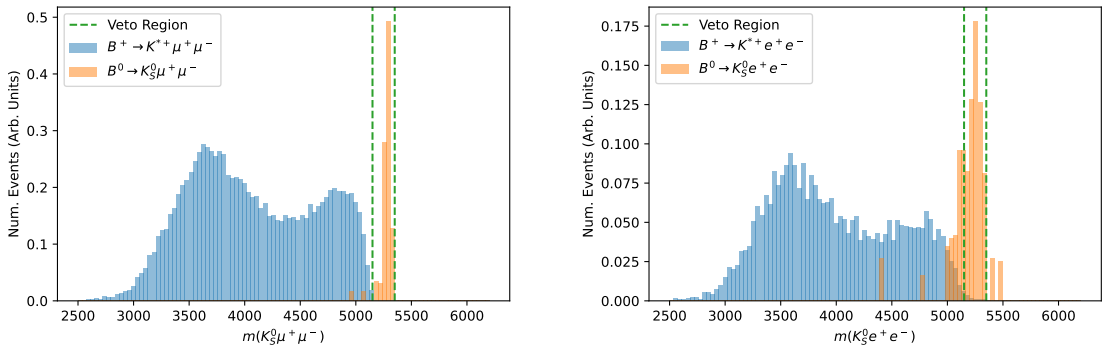


Fig. 4.18: Distributions of $m(K^0_S l^+ l^-)$ found in MC for over-reconstructed $B^0 \to K^0_S \mu^+\mu^-$ decays (left) and $B^0 \to K^0_S e^+e^-$ decays (right), compared to $B^+ \to K^{*+}\ell^+\ell^-$ signal decays. The region used to veto such background decays is indicated by the vertical dashed lines

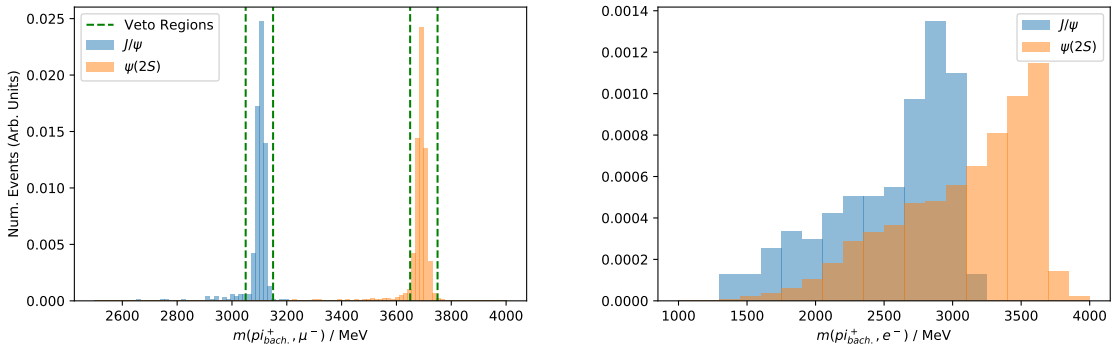
	No Veto Contam.	Veto Contam.
$B^0 \to K^0_S e^+e^-$	0.0044	0.0017
$B^0 \to K^0_S \mu^+\mu^-$	0.0031	0.00018

Table 4.25: Estimates of the upper-limits on the contamination rates from over-reconstructed $B^0 \to K^0_S \ell^+\ell^-$ decays, relative to $B^+ \to K^{*+}\ell^+\ell^-$ signal, both with and without application of the veto $m(K^0_S\ell^+\ell^-) \notin [5150, 5350\,\text{MeV}]$.

4.4.2.1 Over-reconstructed decays

If a pion (or ghost track) passes close to the secondary vertex of a $B^0 \to K^0_S\ell^+\ell^-$ decay, it may be misreconstructed as a $B^+ \to K^{*+}\ell^+\ell^-$ candidate. Such "over-reconstructed" decays can be vetoed using the invariant mass of the K^0_S and leptons, with $m(K^0_S\ell^+\ell^-) \notin [5150, 5350\,\text{MeV}]$. As shown in Fig. 4.18, over-reconstructed decays form a sharp peak in $m(K^0_S\ell^+\ell^-)$. This cut is $\sim 99\%$ efficient on signal and control-mode MC.

MC samples of over-reconstructed $B^0 \to K^0_S\ell^+\ell^-$ decays, from all years, are used to assess residual contamination rates. The selection efficiencies for such candidates cannot be calculated with the nominal method given in Section 4.3, as many of the additional "pions" are actually ghost tracks, for which no PID efficiency calibration samples are available. Instead, PID efficiencies are calculated via cuts to the simulated PID variables. Few over-reconstructed MC candidates pass the full selection requirements, so selection efficiencies are calculated using the Bayesian method (see Section 3.2). The resulting 68% upper-limits on the background contamination rates, relative to muon-mode and electron-mode signal, are given in Table 4.25.



(a): Muon Modes ***(b):*** Electron Modes

Fig. 4.19: Distribution of $m(\pi^+_{Soft,\ \pi\to\ell},\ \ell^-)$ for control-mode decays with the soft pion and same-sign lepton swapped during reconstruction, with $B^+ \to K^{*+}\mu^+\mu^-$ reconstruction (left), and $B^+ \to K^{*+}e^+e^-$ reconstruction (right). The kinematic requirements used to veto such decays in the muon mode are highlighted by the vertical lines. These plots were produced using control-mode MC candidates passing the preselection requirements.

After vetoes are applied, these contamination rates are at the $\mathcal{O}(0.1\%)$ level or below, so are neglected.

4.4.2.2 Control-mode decays with swapped tracks

The J/ψ and $\psi(2S)$ control-mode decays can form a background to $B^+ \to K^{*+}\ell^+\ell^-$ decays, if the soft pion π^+_{Soft} and same-sign lepton ℓ^+ are "swapped" in reconstruction (*i.e.* via a simultaneous $\pi^+ \to \ell^+$ and $\ell^+ \to \pi^+$ misidentification). Such decays are suppressed by the trigger and PID requirements, owing to these misidentifications. However, the q^2 distributions for such candidates pollute $q^2_{low|central}$. As all final-state particles are reconstructed, they also form a peak in $m(B^+)$, close to the signal. Therefore, they could cause significant contamination in the $B^+ \to K^{*+}\ell^+\ell^-$ channels. MC samples of $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ and $B^+ \to \psi(2S)\,(\ell^+\ell^-)\,K^{*+}$ decays, from all years, are used to study these swap background candidates.

As the soft pion and opposite-sign lepton in these decays originate from a $c\bar{c}$ resonance, the invariant masses of these two particles should peak at the resonance's mass (when the soft pion is assigned a lepton mass hypothesis). The distributions of $m(\pi^+_{Soft,\ \pi\to\ell},\ \ell^-)$ in J/ψ-mode and $\psi(2S)$-mode swap candidates, are shown in Fig. 4.19. In the muon mode, swap candidates can be seen to lie very close to $m(J/\psi)$ and $m(\psi(2S))$. Therefore, such decays are vetoed here using the requirement $m(\pi^+_{Soft,\ \pi\to\mu},\ \mu^-) \notin [3050, 3150\,\text{MeV}], [3650, 3750\,\text{MeV}]$, with a signal efficiency of $\sim 96\%$.

	No Vetoes	With Vetoes
$B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$	0.04	0.003
$B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$	0.0059	0.00062
$B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$	0.053	0.0036
$B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$	0.0071	0.00041

Table 4.26: 68% upper-quantiles for the expected background contamination rate, relative to signal, from different control-mode swap processes, with and without vetoes for these backgrounds being applied.

However, the distributions of $m(\pi^+_{Soft,\ \pi\to\ell},\ \ell^-)$ are much wider in the electron mode, as the misidentified electron can emit bremsstrahlung photons which are not subsequently recovered. Hence, any cut to $m(\pi^+_{Soft,\ \pi\to e},\ e^+)$ wide enough to veto swap candidates would have a very poor signal efficiency. Instead, these backgrounds are removed in the electron mode by requiring that the soft pion passes $\texttt{ProbNN}(e) < 0.1$, with a signal efficiency of $\sim 98\%$..

The expected contamination rates from swap backgrounds are calculated before and after the application of these veto requirements. In the electron mode, the selection efficiencies for swap candidates are calculated with the usual methodology, with uncertainties found using the normal approximation. However, few candidates survive the muon-mode veto requirements, so the efficiencies here are found using the Bayesian method. In addition, to improve precision, the efficiencies of the veto requirements are assumed to factorise from all other selection requirements except the preselection, and so are calculated using events passing the preselection.

The swap candidate selection efficiencies are then combined with the $B^+ \to K^{*+}\ell^+\ell^-$ signal selection efficiencies, plus the relevant branching fractions, to find probability distributions for the number of swap candidates relative to the number of signal candidates. The upper 68% quantiles in these distributions are taken as upper limits to the background contamination rates, as shown in Table 4.26. When vetoes are applied, the contamination rates in all decay modes are well below 1% relative to signal, meaning the swap backgrounds can be neglected.

4.4.2.3 Fully hadronic decays

Analogously to $B^0 \to K^0_S\pi^+\pi^-$ decays in the B^0 channel, fully-hadronic $B^+ \to K^{*+}\pi^+\pi^-$ decays can form a background to $B^+ \to K^{*+}\ell^+\ell^-$ decays if both pions are misidentified as leptons. As $B^+ \to K^{*+}\pi^+\pi^-$ decays have the same topology as the signal, they will peak

Resonance Struct.	Branching fraction
$K^*(892)^+\,\rho^0$	$(4.6 \pm 1.1) \times 10^{-6}$
$K^*(892)^+\,f_0(980)$	$(4.2 \pm 0.7) \times 10^{-6}$
Total	$(7.5 \pm 1.0) \times 10^{-5}$

Table 4.27: The total branching fraction of $B^+ \to K^{*+}\pi^+\pi^-$ decays, along with the branching fractions of the known contributing resonances [13].

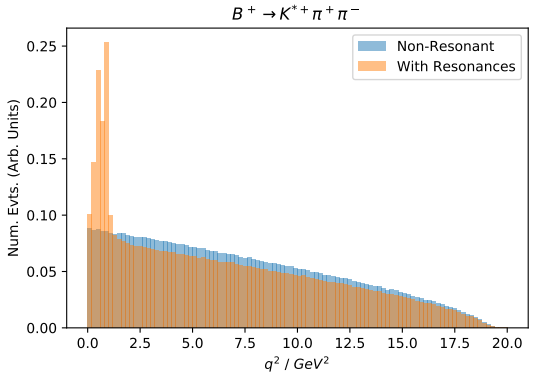


Fig. 4.20: Distribution of $B^+ \to K^{*+}\pi^+\pi^-$ events with q^2, with and without the inclusion of resonances.

close to the signal in $m(B^+)$, and will mainly be suppressed by the PID and trigger selection requirements.

In order to evaluate the contamination from these decays, MC samples of $B^+ \to K^{*+}\pi^+\pi^-$ decays from all years are used. No DecFiles were available for this decay mode prior to this analysis, so a new DecFile was prepared. To reduce computing resource requirements, generator-level requirements were imposed to discard candidates which could not be reconstructed, and which would not pass selection requirements. Such requirements included that final-state particles had $p_T > 300\,\text{MeV}$ and fell within the LHCb angular acceptance, and that K^0_S mesons had decayed upstream of the TT.

In contrast to $B^0 \to K^0_S\pi^+\pi^-$ decays, little is known about the resonant structure of $B^+ \to K^{*+}\pi^+\pi^-$ decays. As shown in Table 4.27, the *Particle Data Group (PDG) Review of Particle Physics* only lists the total branching fraction for such decays, plus branching fractions for two resonant processes [13]. In order to study the effect of the decay kinematics on contamination rates from $B^+ \to K^{*+}\pi^+\pi^-$ decays, the LHCb MC samples mentioned above were simulated with non-resonant (PHSP) kinematics for the $K^{*+} - \pi^+ - \pi^-$ system. Then,

	Reco. Lepton	Decay Model	Back. Trig. Contam	Sig. Trig. Contam	Ave. Contam
Run-1	Electron	Amp. Mod.	0.118 ± 0.026	0.19 ± 0.04	0.15 ± 0.04
		PHSP	0.137 ± 0.026	0.23 ± 0.04	0.18 ± 0.04
	Muon	Amp. Mod.	$(7.8 \pm 3.4) \times 10^{-5}$	0.0024 ± 0.0004	0.0012 ± 0.0007
		PHSP	$(5.9 \pm 1.7) \times 10^{-5}$	0.0029 ± 0.0005	0.0015 ± 0.0009
Run-2	Electron	Amp. Mod.	0.0091 ± 0.0018	0.0145 ± 0.0027	0.0118 ± 0.0027
		PHSP	0.0119 ± 0.0021	0.0199 ± 0.0035	0.016 ± 0.004
	Muon	Amp. Mod.	$(3.7 \pm 1.2) \times 10^{-5}$	0.00183 ± 0.00033	0.0009 ± 0.0005
		PHSP	$(4.7 \pm 1.1) \times 10^{-5}$	0.0022 ± 0.0004	0.0011 ± 0.0007
Total	Electron	Amp. Mod.	0.033 ± 0.007	0.053 ± 0.010	0.043 ± 0.010
		PHSP	0.039 ± 0.007	0.065 ± 0.012	0.052 ± 0.012
	Muon	Amp. Mod.	$(4.4 \pm 1.3) \times 10^{-5}$	0.00193 ± 0.00034	0.0010 ± 0.0006
		PHSP	$(5.0 \pm 1.1) \times 10^{-5}$	0.0023 ± 0.0004	0.0012 ± 0.0007

Table 4.28: Estimated contamination from $B^+ \to K^{*+}\pi^+\pi^-$ decays relative to muon-mode and electron-mode $B^+ \to K^{*+}\ell^+\ell^-$ signal. The background efficiency is calculated using the trigger-unbiased method outlined in Section 4.4.1.5. Estimated contamination rates using the lower, upper, and nominal values for the trigger efficiency are shown. The contamination is evaluated using non-resonant kinematics *(PHSP)*, and with the kinematics reweighted to include known resonances *(Amp. Mod.)*.

RapidSim was used to produce three samples; one with non-resonant kinematics, one including the ρ^0 resonance, and one including the $f_0(980)$ resonance. These RapidSim samples were then combined, in proportion according to the branching fractions for each process*. The q^2 distribution for this combined RapidSim sample, compared to the non-resonant samples alone, are shown in Fig. 4.20. As for $B^0 \to K^0_S\pi^+\pi^-$ decays (see Section 4.4.1.5), the RapidSim samples were used to produce weights in the (m', θ') plane, which were then used to reweight the decay kinematics of the full LHCb simulation samples.

Contamination rates from $B^+ \to K^{*+}\pi^+\pi^-$ decays are then calculated using the trigger-unbiased method outlined in Section 4.4.1.5, with estimates found using the upper, lower, and nominal estimates of $\varepsilon(Trig.)$ shown in Table 4.28. Here, estimates of contamination rates are shown separately in run-1, run-2, and across all years, for the muon and electron modes. In addition to contamination rates found with decay kinematics reweighted to include known resonances, they are also evaluated for non-resonant decay kinematics.

*As it is not known explicitly, the non-resonant branching fraction is calculated by subtracting the ρ^0 and $f_0(980)$ branching fractions from the total $B^+ \to K^{*+}\pi^+\pi^-$ branching fraction.

The contamination rates in the $B^+ \to K^{*+}\mu^+\mu^-$ channel are well below the $\mathcal{O}(\%)$ level relative to signal, so can be neglected. However, contamination rates across all years in the $B^+ \to K^{*+}e^+e^-$ channel are at $4-5\%$ relative to signal. Given the poor understanding of $B^+ \to K^{*+}\pi^+\pi^-$ decay kinematics, the contamination rate with non-resonant kinematics (N_{PHSP}) is taken as the nominal value. Then, a systematic uncertainty is assigned for this choice, taken as the difference in contamination rates found with MC reweighted to include known resonances, relative to the value found with non-resonant decay kinematics ($\sigma(N) = \left|N_{Amp.Mod.} - N_{PHSP}\right|$). This uncertainty is added in quadrature to the uncertainty on N_{PHSP} found from the efficiency calculations and BF values. Using this method, the nominal contamination value across all years, with uncertainty, is $(5.2 \pm 1.5)\%$. A component for $B^+ \to K^{*+}\pi^+\pi^-$ decays is included in the mass-fit for $B^+ \to K^{*+}e^+e^-$, with its yield constrained using the selection efficiencies calculated here (see Section 4.5.4.2)

4.5 Mass fitting models

As mentioned in Section 3.3, extended unbinned maximum likelihood fits to $m(B)$ are used to determined the yields of rare-mode and control-mode decays in data. These likelihoods are constructed using the RooFit software package [324]. In general, the mass-fits for each channel are carried out via a two-step process:

1. First, the probability distribution functions (PDFs) for the signal, and any physical sources of background, are determined using MC samples. Some of these "fit component" PDFs are described using an empirically-motivated functional parameterisation, where the parameters are determined from an unbinned maximum likelihood fit to MC. Others are found using non-parametric techniques. Unless otherwise specified, each fit component is determined using MC samples for that particular signal or background source, passing full selection requirements. All the corrections outlined in Section 4.2 are applied to signal and control-mode MC samples.

2. Then, a likelihood is constructed from these different fit components, and used to fit the corresponding yields in data. Some of the component PDF parameters may be permitted to float in this fit, to account for differences in mass resolutions between data and MC. In many cases, the values of some of the yields or component PDF parameters (θ) may have been determined from other sources. To incorporate this knowledge into the fit model, the likelihood is multiplied by a Gaussian term $\mathcal{G}(\theta|\mu_\theta, \sigma_\theta)$, where μ_θ and σ_θ describe the value and uncertainty of this external estimate of θ. This is referred to as a "Gaussian constraint".

In this section, the mass-fit models used in each rare-mode and control-mode channel are outlined, and the results of the fits to fully-selected data are shown. In addition, the PDFs used to model signal and physical background sources are shown, compared with the $m(B)$ distributions of the MC used to derive them. These plots are omitted in cases where the PDFs are very similar to other channels. Throughout this section, the results of fits to data passing full selection requirements, from all years combined, are shown. Other sub-sets of years and selection requirements were used when computing various cross-checks and MC corrections. The same models were used for these mass-fits, but the results are omitted for brevity.

4.5.1 Common modelling choices

Three different types of empirically-chosen PDF are used to model different signal and background processes in this section. The first is the "Crystal Ball" (CB) PDF; a piecewise function comprising a Gaussian core with mean μ and standard deviation σ, and a power-law tail with slope n [325]. The point where these components are joined is described by a threshold parameter α:

$$\mathcal{P}_{CB}\left(m(B)\,|\,\mu,\sigma,\alpha,n\right) = \begin{cases} exp\left(-\frac{\left(m\left(B^0\right)-\mu\right)^2}{2\sigma^2}\right) & for \quad \frac{m\left(B^0\right)-\mu}{\sigma} > -\alpha \\ A\cdot\left(B-\frac{m(B)}{\sigma}\right)^{-n} & for \quad \frac{m\left(B^0\right)-\mu}{\sigma} \leq -\alpha \end{cases} \tag{4.12}$$

where:

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot exp\left(-\frac{|\alpha|^2}{2}\right), \qquad B = \frac{n}{|\alpha|} - |\alpha| \tag{4.13}$$

The CB PDF is designed to model the invariant mass of a signal processes with finite mass resolution, and where energy is lost via a process such as FSR, bremsstrahlung emission, or missing a particle during reconstruction. In this analysis, "Double Crystal Ball" (DCB) PDFs are often used to parameterise signal components. These are a sum of two CB PDFs with a shared mean, and power-law tails on opposite sides of their Gaussian peaks. The relative proportions of the two PDFs are controlled by a parameter f:

$$\begin{aligned} \mathcal{P}_{DCB}\left(m(B)\,|\,\mu,\sigma_1,\sigma_2,\alpha_1,\alpha_2,n_1,n_2,f\right) = \{f\cdot\mathcal{P}_{CB}\left(m(B)\,|\,\mu,\sigma_1,\alpha_1,n_1\right) + \\ (1-f)\cdot\mathcal{P}_{CB}\left(m(B^0)\,|\,\mu,\sigma_2,-\alpha_2,n_2\right)\} \end{aligned} \tag{4.14}$$

This additional upper-mass tail accounts for effects such as the recovery of bremsstrahlung photons, variation in momentum resolution between candidates, and the imposition of `DecayTreeFitter` mass-constraints [326]. In all fits to data, the shape parameters n_i, α_i, and f are fixed from prior fits to MC candidates. To account for differences in reconstructed mass resolution between data and MC, the mean and widths of the signal peaks are often allowed to vary from their values found from fits to MC. Such variation is expressed in terms of a mean "shift" term, and a width "scale" term:

$$\mu^{data} = \mu^{MC} + \Delta_\mu, \qquad \sigma_i^{data} = s_\sigma \times \sigma_i^{MC} \tag{4.15}$$

Within each channel, the values of these parameters are shared between all signal DCBs.

The distributions of combinatorial background in each channel have smoothly decreasing densities with increasing values of $m(B)$, which are found to be well described by exponential PDFs with freely-varying "decay parameters" γ.

$$\mathcal{P}_{Exp}\left(m(B)\,|\,\gamma\right) = exp\left(\gamma \cdot m(B)\right) \tag{4.16}$$

Some sources of background cannot be empirically described with any obvious parameterisation. In such cases, a non-parametric method is used, known as "adaptive-width Gaussian Kernel Density Estimation (GKDE)" [327]. This method constructs a PDF from a sum of Gaussian "kernels", the positions of which are determined by the values of $m(B)$ in set of n MC events (denoted $\vec{t}$):

$$\mathcal{P}_{GKDE}\left(m(B)\,|\,\vec{t}\right) = \frac{1}{n}\sum_i^n \left[exp\left(-\frac{\left(m\left(B\right) - t_i\right)^2}{2\sigma_i^2} \right) \right] \tag{4.17}$$

The widths σ_i for each kernel are allowed to vary, such that narrow kernels are used in regions of $m(B)$ with high MC density, and broad kernels are used in regions of low MC density. This way, the distribution is accurately described, while any statistical fluctuations are smoothed out. The average widths of the kernels are controlled by a global "bandwidth parameter" ρ.

4.5.2 J/ψ control modes

The yields of the J/ψ control-mode decays are used throughout this analysis, namely as inputs to the rare-mode branching fraction and R_X^{-1} measurements, and the $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$ cross-checks (see Chapter 6). Fits to the $m(B)$ values found with dilepton mass-constraints in `DecayTreeFitter` are used for these purposes. However, fits to $m(B)$ found without such constraints are also carried out as a self-consistency check, and to examine the modelling

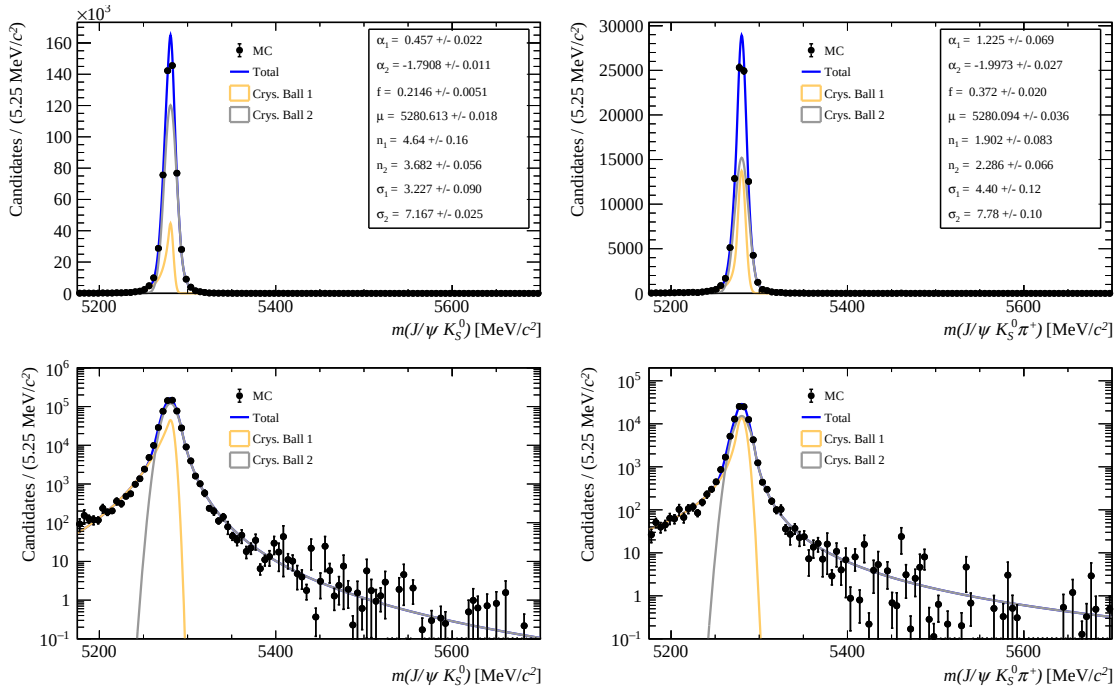


Fig. 4.21: Results of fits to $m(B)$, determining DCB fit parameters from MC for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (right) decays. These plots are shown with a linear (top) and logarithmic (bottom) scale on the y-axis. The values of $m(B)$ calculated with dilepton mass constraints are used here. Pull values between the MC and fitted PDFs are shown in the bottom portions of the plots. The fitted parameter values, and associated HESSE uncertainties, are shown in the legend.

of backgrounds at low $m(B)$ values. Resolution parameters from these fits are also used to constrain the signal distributions in the rare-mode fits.

4.5.2.1 Muon modes with dilepton constraint

Within the range $m(B) \in [5175, 5700\,\text{MeV}]$, both the $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ channels contain a signal peak (parameterised with a DCB), and combinatorial background. The PDFs found from fits to signal MC are shown in Fig. 4.21. In the $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ channel, there is an additional peak from CKM-suppressed $B^0_s \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ decays. This is parameterised using the same shape as the signal, but with the mean shifted by $m(B^0_s) - m(B^0) = 87.4\,\text{MeV}$ [13]. Its yield is left unconstrained.

The results of the fits to data are shown in Fig. 4.22. There is a small "bump" in the $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ fit, in the region $5400 \lesssim m(B^0) \lesssim 5500\,\text{MeV}$. This is believed to originate from a small number of $\Lambda^0_b \to \Lambda J/\psi\,(\mu^+\mu^-)$ candidates which survive the kinematic veto for such decays (see Section 4.4.1.2). The effect of this background was examined by inserting a

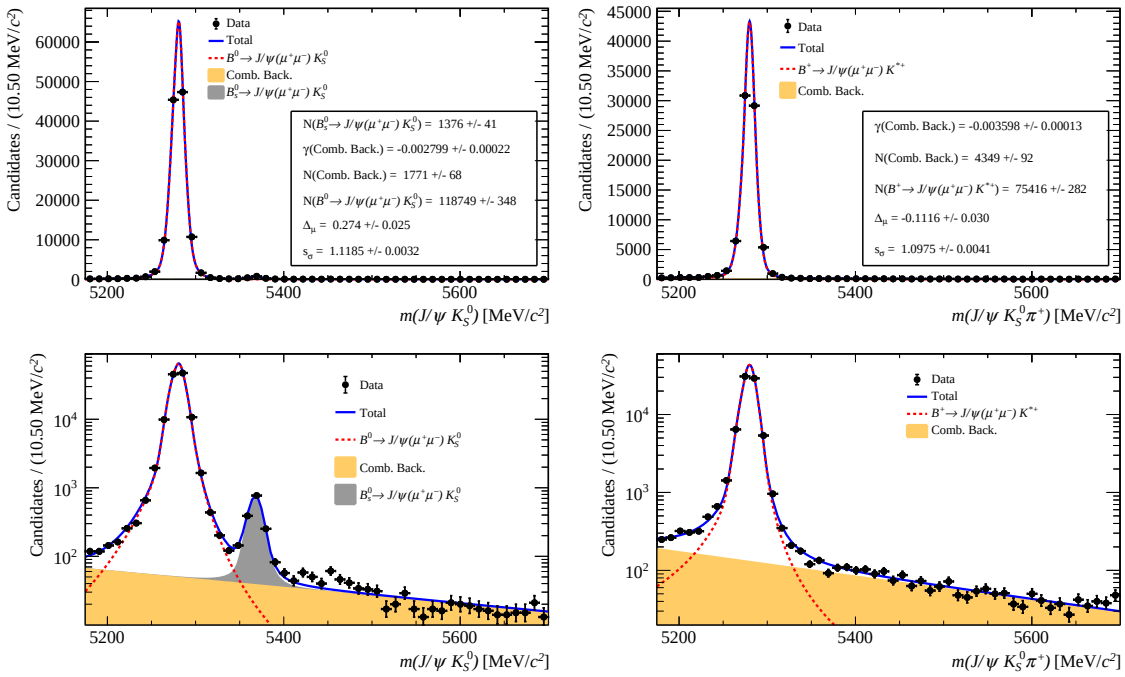


Fig. 4.22: Results of fits to $m(B)$, determining signal yields for $B^0 \to J/\psi(\mu^+\mu^-)K^0_S$ (left) and $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$ (right) candidates in data. These plots are shown with a linear (top) and logarithmic (bottom) y-scale, to highlight low-yield backgrounds. The values of $m(B)$ calculated with dilepton mass constraints are used here.

PDF for $\Lambda_b^0 \to \Lambda J/\psi(\mu^+\mu^-)$ candidates into the mass-fit, which was parameterised using a DCB whose parameters were determined using a RAPIDSIM sample to which the kinematic veto was applied. When this additional component was included in the fit, the signal yield was increased by 0.08%. Therefore, the $\Lambda_b^0 \to \Lambda J/\psi(\mu^+\mu^-)$ contamination can be safely neglected.

4.5.2.2 Muon modes without dilepton constraint

Fits to $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ data are also carried out using $m(B)$ values computed without dilepton mass-constraints. While the mass-resolution is degraded when this constraint is omitted, no additional backgrounds become present in the range $m(B) \in [5175, 5700\,\text{MeV}]$. Therefore, the same mass-fit parameterisation can be used as above, when the dilepton mass constraint *is* used. Results of the fits to $B \to J/\psi(\mu^+\mu^-)K^{(*)}$ MC and data can be seen in Figs. 4.23 and 4.24 respectively.

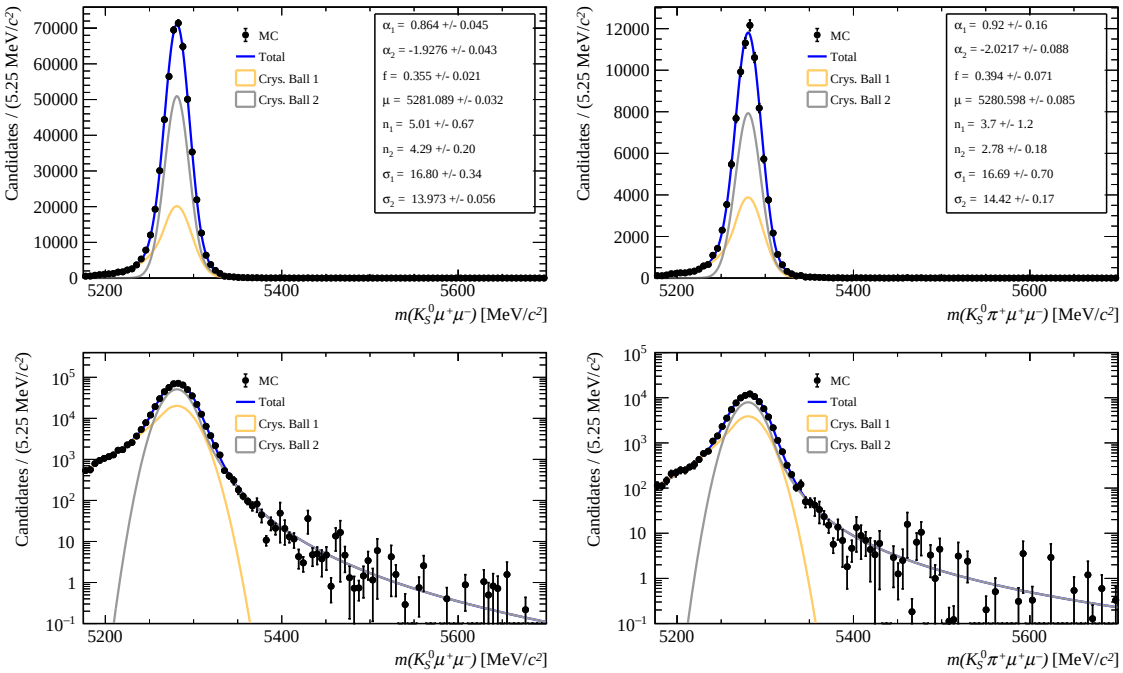


Fig. 4.23: Results of fits to $m(B)$, determining DCB fit parameters from MC for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (right) decays. These plots are shown with a linear (top) and logarithmic (bottom) scale on the y-axis. The values of $m(B)$ calculated without dilepton mass constraints are used here.

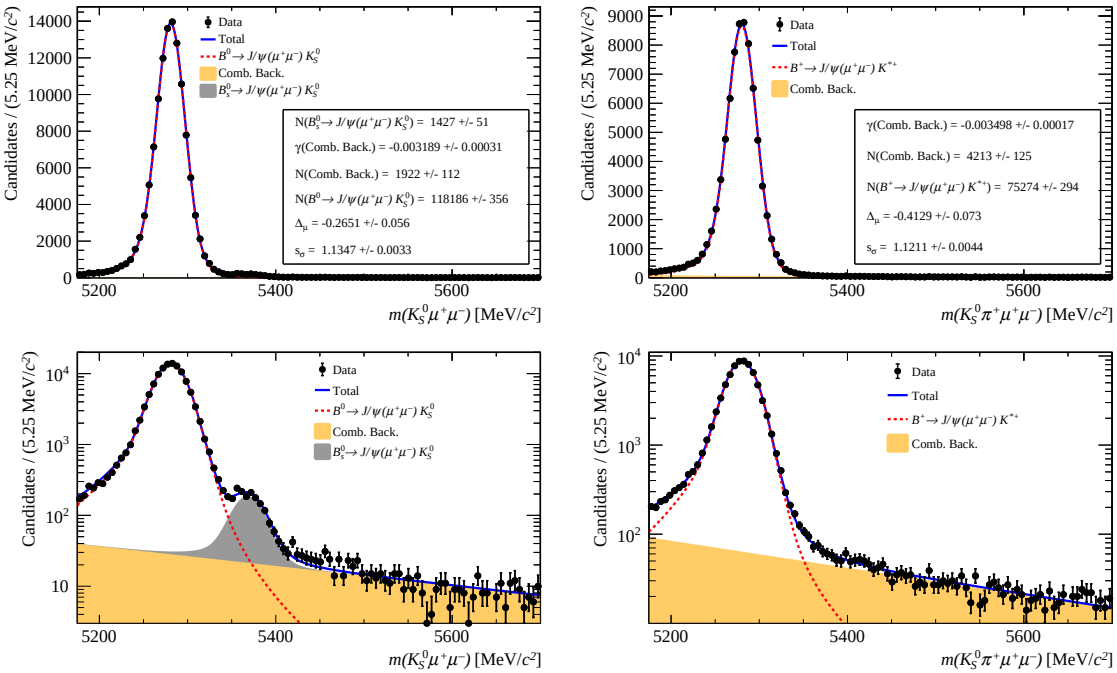


Fig. 4.24: Results of fits to $m(B)$, determining signal yields for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (right) candidates in data. These plots are shown with a linear (top) and logarithmic (bottom) y-scale. The values of $m(B)$ calculated without dilepton mass constraints are used here.

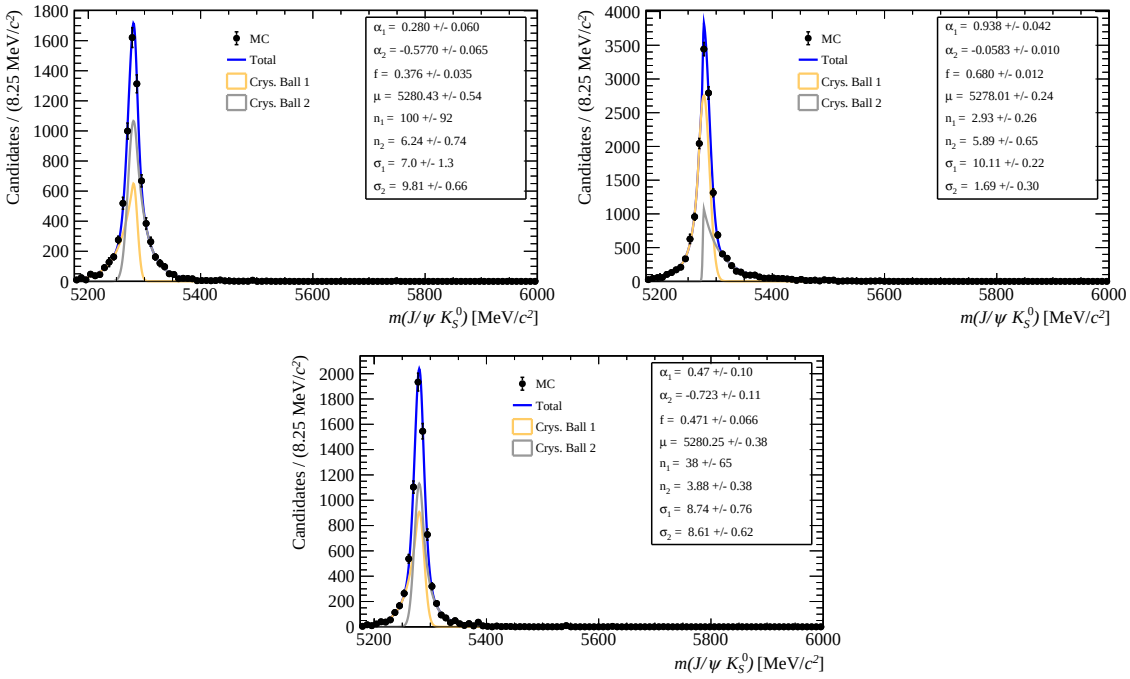


Fig. 4.25: Results of fits to $m(B)$, determining DCB fit parameters from MC for $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ decays with 0 (top left), 1 (top right), and ≥ 2 recovered bremsstrahlung photons (bottom). These plots are shown with a linear scale on the y-axis. Plots with a logarithmic scale can be found in Fig. 4.26. The values of $m(B)$ calculated with dilepton mass constraints are used here.

4.5.2.3 Electron modes with dilepton constraint

While the use of a dilepton mass constraint in `DecayTreeFitter` greatly improves the $m(B)$ resolutions of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays, some differences in mass-shapes can be seen depending on the number of bremsstrahlung photons associated to a candidate. Therefore, the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ signal PDFs are parameterised using a sum of three DCBs, describing the $m(B)$ distributions with 0, 1, or ≥ 2 recovered bremsstrahlung photons, respectively. Examples of such PDFs are shown in Fig. 4.25 and Fig. 4.26. The three DCBs are weighted in proportion to the fractions $f_{i\gamma}$ of fully-selected MC with each bremsstrahlung multiplicity, which have typical values of $f_{0\gamma} \approx 0.25$, $f_{1\gamma} \approx 0.5$, and $f_{\geq 2\gamma} \approx 0.25$. In fits to data, the factors $f_{0\gamma}$ and $f_{1\gamma}$ are Gaussian-constrained, while $f_{\geq 2\gamma}$ is reparameterised as $f_{\geq 2\gamma} = 1 - f_{0\gamma} - f_{1\gamma}$.

Other than the signal PDFs, the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ fits to data have the same compositions as the $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ fits, with combinatorial background and a small peak from $B^0_s \to J/\psi\,(e^+e^-)\,K^0_S$ decays in the B^0-mode fit. No other sources of background fall within the fit region of $m(B) \in [5175, 6000\,\text{MeV}]$. The results of fits to data are shown in Fig. 4.27.

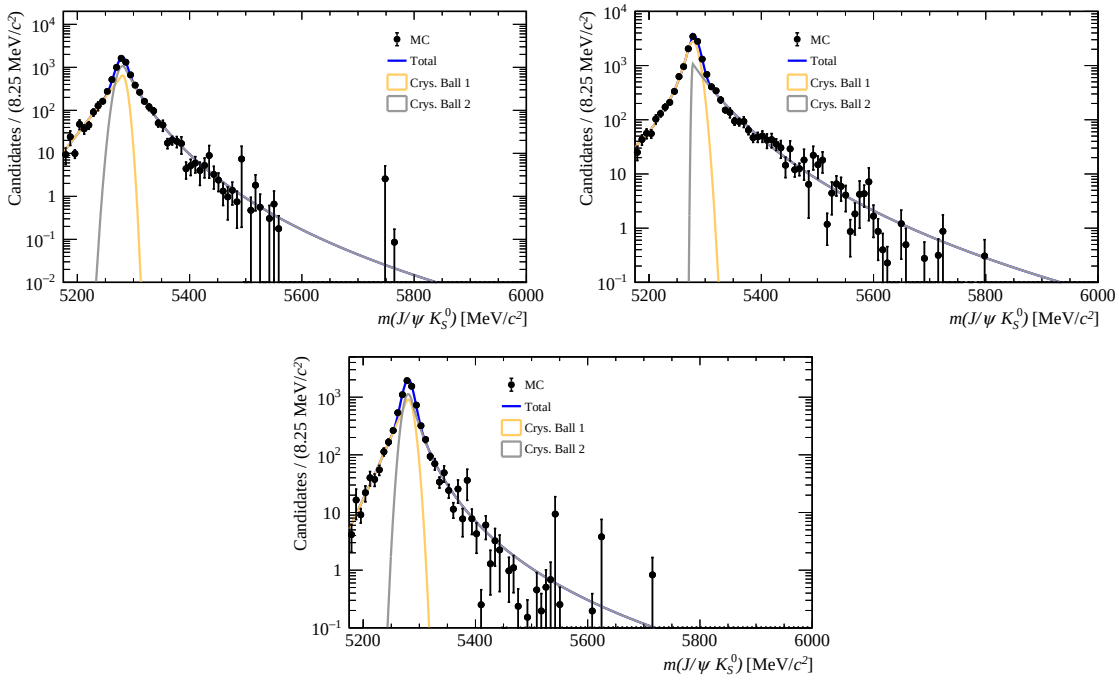


Fig. 4.26: Results of fits to $m(B)$, determining DCB fit parameters from MC for $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ decays with 0 (top left), 1 (top right), and ≥ 2 recovered bremsstrahlung photons (bottom). These plots are shown with a logarithmic scale on the y-axis. Plots with a linear scale can be found in Fig. 4.25. The values of $m(B)$ calculated with dilepton mass constraints are used here.

As in the $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ fit, a small peak from $\Lambda_b^0 \to \Lambda J/\psi\,(e^+e^-)$ decays can be seen in the region $5400 \lesssim m(B^0) \lesssim 5500\,\text{MeV}$ of the $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ channel. This was investigated in the same way, by inserting a DCB component parameterised from a RapidSim sample. The signal yield decreased by 0.15% when this additional component was included in the fit, meaning $\Lambda_b^0 \to \Lambda J/\psi\,(e^+e^-)$ decays can be safely neglected.

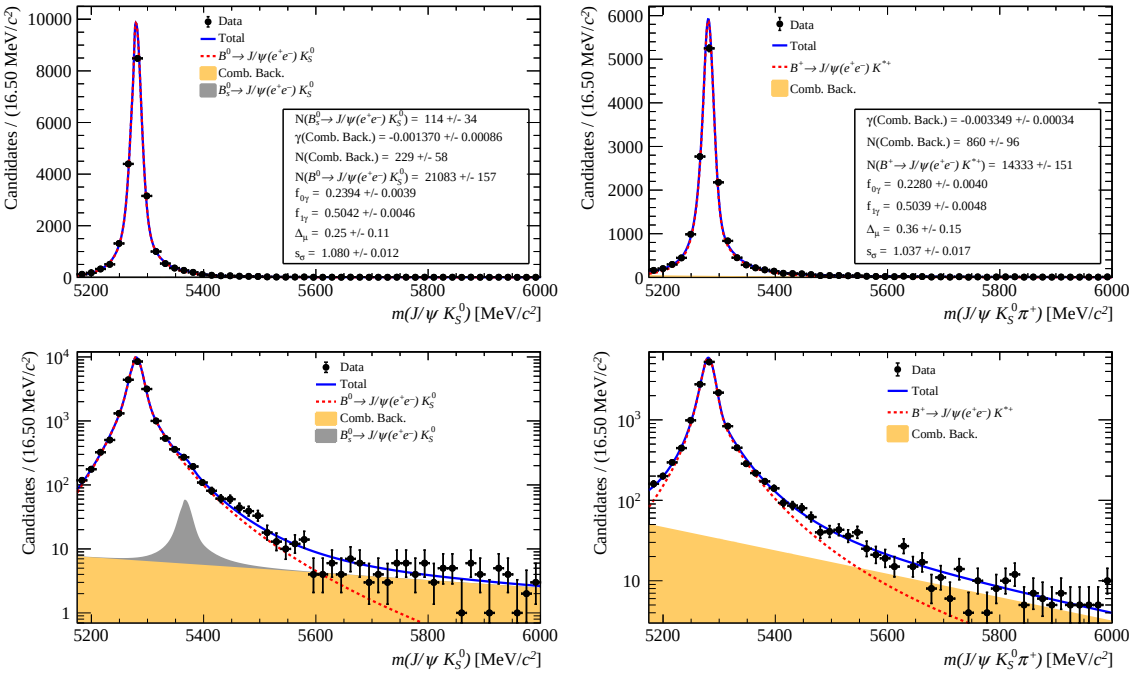


Fig. 4.27: Results of fits to $m(B)$, determining signal yields for $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) candidates in data. These plots are shown with a linear (top) and logarithmic (bottom) y-scale. The values of $m(B)$ calculated with dilepton mass constraints are used here.

4.5.2.4 Electron modes without dilepton constraint

The $m(B)$ resolution is greatly degraded when the dilepton mass-constraint is omitted, meaning a wide range of $m(B) \in [4700, 6000\,\text{MeV}]$ must be used for mass-fits. For signal decays, large variations in $m(B)$ distributions can be seen with different numbers of recovered bremsstrahlung photons. A substantial low-mass tail can be seen with 0 photons, caused by emission of bremsstrahlung photons which are not subsequently recovered. As the number of recovered photons increases, this tail becomes less prominent. However, an upper-mass tail appears, which is caused by spurious recovery of photons which did not actually originate from a candidate electron. As above, the signal is parameterised using a sum of three DCBs, weighted in proportion to the fractions of fully-selected MC with different numbers of recovered photons. Example PDFs are shown in Fig. 4.28.

In addition to combinatorial background and $B^0_s \to J/\psi\,(e^+e^-)\,K^0_S$ decays, various backgrounds pollute the fit region at low $m(B)$ values. Chief among these are partially-reconstructed $B \to J/\psi\,(e^+e^-)\,K^*$ decays with high-mass kaon resonances (see Sections 4.4.1.3 and 4.4.2). In the $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ fit, these are parameterised from $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ MC using a DCB. In the $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ fit, they are described from an MC sample containing

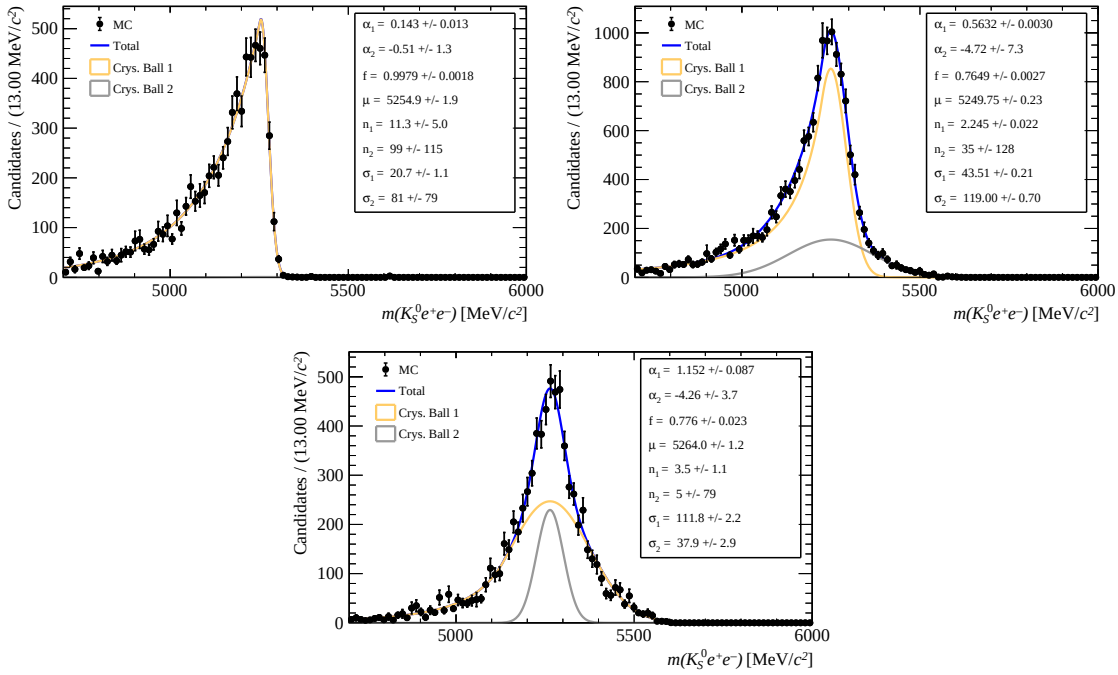


Fig. 4.28: Results of fits to $m(B)$, determining DCB fit parameters from MC for $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ decays with 0 (top left), 1 (top right), and ≥ 2 recovered bremsstrahlung photons (bottom). The values of $m(B)$ calculated without dilepton mass constraints are used here.

a mixture of of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ processes involving $K_1(1270)$, $K_1(1400)$ and $K_1^*(1430)$ resonances, decaying to a $K^0_S\pi^+\pi$ final-state. A DCB was found to be unable to describe this cocktail, so it is instead described using a GKDE with $\rho = 2.0$.

Due to bremsstrahlung emission, some $B \to \psi(2S)\,(e^+e^-)\,K^{(*)}$ events "leak" in to the $q^2_{J/\psi}$ region, at low $m(B)$ values. These are parameterised from $B \to \psi(2S)\,(e^+e^-)\,K^{(*)}$ MC, using CBs. Finally, higher-order $c\bar{c}$ resonances may decay via $c\bar{c} \to J/\psi X$, and the decay product(s) X may be missed during reconstruction. Such partially-reconstructed $c\bar{c}$ resonances are described using MC samples containing cocktails of $B \to c\bar{c}\,(J/\psi X)\,K^{(*)}$ decays involving different $c\bar{c}$ resonances. There are no suitable parameterisations for these decays, so GKDEs with $\rho = 2.0$ are used.

PDFs for these low-mass backgrounds are shown in Fig. 4.29. Yields for all components were left freely-varying in fits to data, which are shown in Fig. 4.30.

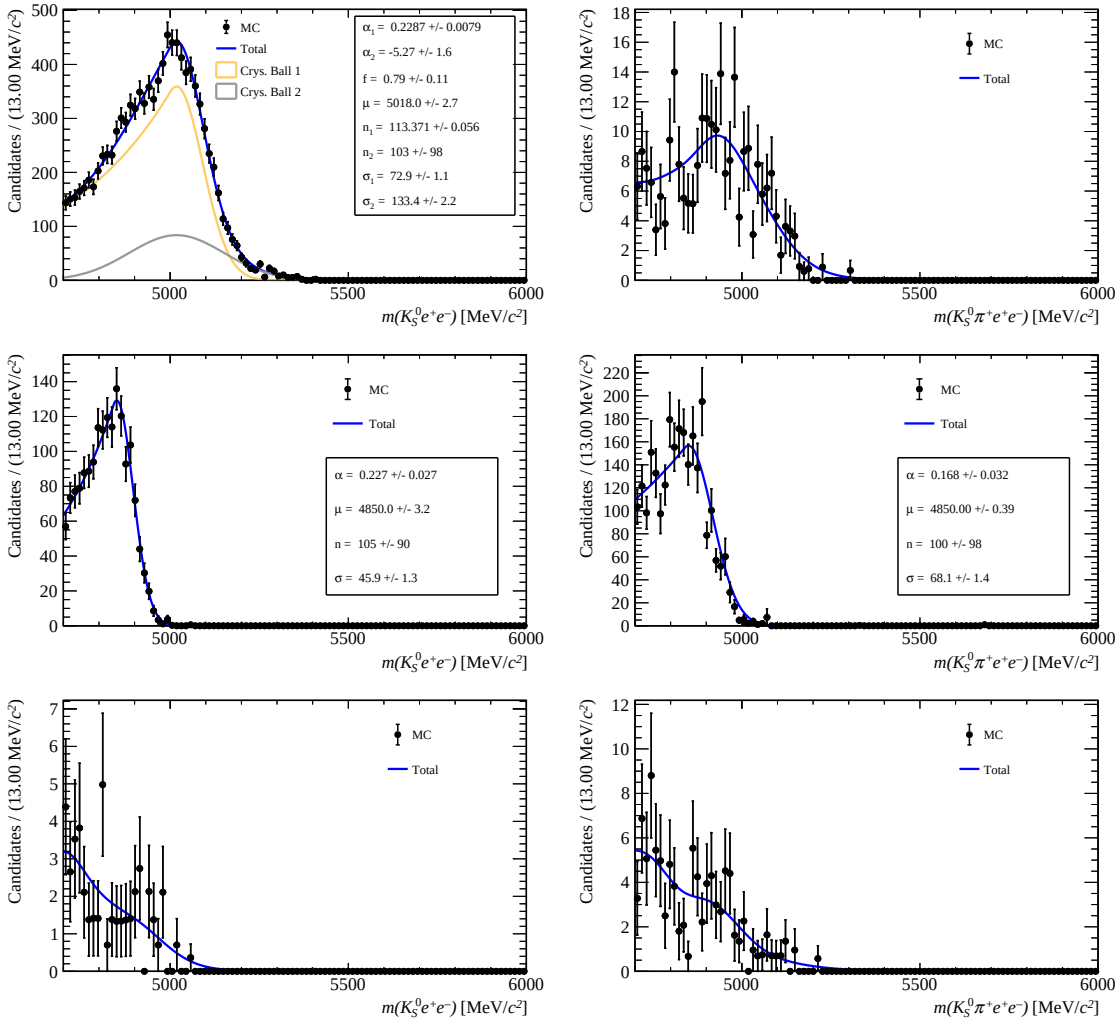


Fig. 4.29: Results of fits to $m(B)$, determining fit PDFs for partially-reconstructed kaon (top), $\psi(2S)$ leakage (middle), and partially-reconstructed $c\bar{c}$ (bottom) backgrounds, in the $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) channels. The values of $m(B)$ calculated without dilepton mass constraints are used here.

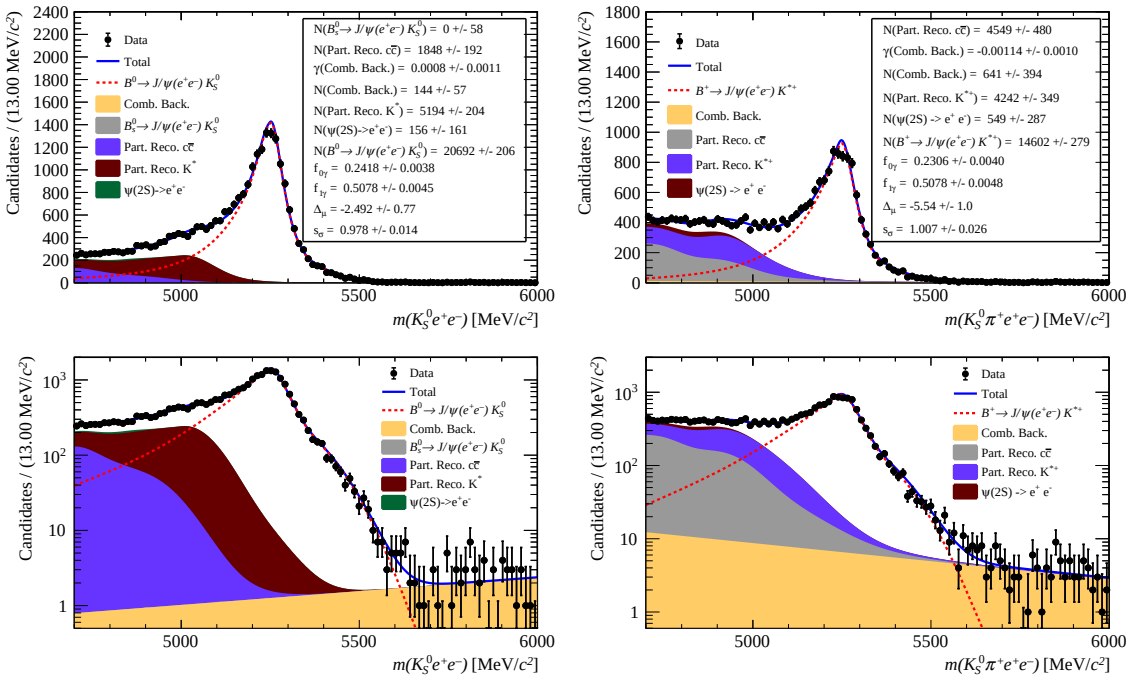


Fig. 4.30: Results of fits to $m(B)$, determining signal yields for $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) candidates in data. These plots are shown with a linear (top) and logarithmic (bottom) y-scale. The values of $m(B)$ calculated without dilepton mass constraints are used here.

Channel	Yields		Yield Diff.	Signif.
	$m(B)_{Cons-K^0_S,J/\psi}$ Fit	$m(B)_{Cons-K^0_S}$ Fit		
B^0 electron	$(2.108 \pm 0.016) \times 10^4$	$(2.069 \pm 0.021) \times 10^4$	$1.6 \pm 1.0\%$	1.5
B^0 muon	$(1.1875 \pm 0.0035) \times 10^5$	$(1.1819 \pm 0.0036) \times 10^5$	$0.15 \pm 0.31\%$	0.49
B^+ electron	$(1.433 \pm 0.015) \times 10^4$	$(1.460 \pm 0.028) \times 10^4$	$-1.0 \pm 1.9\%$	-0.54
B^+ muon	$(7.542 \pm 0.028) \times 10^4$	$(7.527 \pm 0.029) \times 10^4$	$-0.17 \pm 0.25\%$	-0.67

Table 4.29: Tests of the consistency of yields for various $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ decays found from fits to $m(B)_{Cons-K^0_S,J/\psi}$ and $m(B)_{Cons-K^0_S}$, showing the yields found from each fit, the difference in efficiency-corrected yields ($N(m(B)_{Cons-K^0_S,J/\psi}) - N(m(B)_{Cons-K^0_S})$), and the statistical significance of this difference in efficiency-corrected yields.

4.5.2.5 Consistency of fits with and without dilepton constraint

As a self-consistency check of the J/ψ-mode mass-fits, the yields found with and without the dilepton mass-constraint in `DecayTreeFitter` are compared, for each channel. Hereafter, these mass variables are denoted $m(B)_{Cons-K^0_S,J/\psi}$ and $m(B)_{Cons-K^0_S}$, respectively. Since the fits to rare-mode data use $m(B)_{Cons-K^0_S}$, any discrepancies between the $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ yields found with each mass variable may indicate a systematic bias which has not been accounted for, and does not cancel in ratio when calculating rare-mode branching fractions.

The efficiencies of the mass-window requirements on $m(B)_{Cons-K^0_S,J/\psi}$ and $m(B)_{Cons-K^0_S}$ differ, so these are estimated using fully-selected and corrected $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ MC samples. In addition, these MC samples are used to estimate the correlations between the yields from the fits to $m(B)_{Cons-K^0_S,J/\psi}$ and $m(B)_{Cons-K^0_S}$. The $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ yields found with the two mass variables, for each channel, are then normalised by these mass-window efficiencies*. The differences in normalised yields for each channel, and the statistical significances of these differences, are shown in Table 4.29. The largest discrepancy is seen for $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$, where the yield found with $m(B)_{Cons-K^0_S}$ is 1.6% lower than that found with $m(B)_{Cons-K^0_S,J/\psi}$. However, the yields remain in good statistical agreement, at the level of 1.5σ.

*Throughout this study, the UNCERTAINTIES software package in PYTHON is used to keep track of uncertainties for yields and efficiencies, plus any correlations, using standard uncertainty propagation [328].

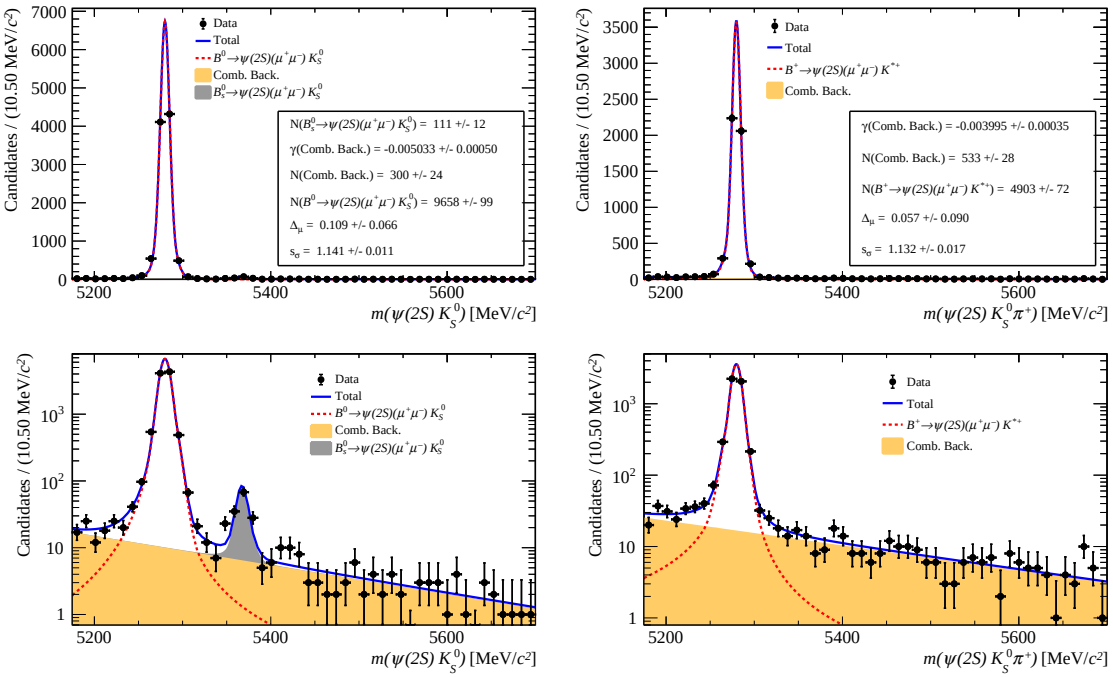


Fig. 4.31: Results of fits to $m(B)$, determining signal yields for $B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$ (right) candidates in data. These plots are shown with a linear (top) and logarithmic (bottom) y-scale.

4.5.3 $\psi(2S)$ control modes

The yields of the $B \to \psi(2S)\,(\ell^+\ell^-)\,K^{(*)}$ control modes are used as inputs to the $R^{-1}_{\psi(2S)}$ cross-check (see Section 6.2). Values of $m(B)$ computed with the dilepton mass constrained to $m(\psi(2S))$ are used in the fits to these channels.

4.5.3.1 Muon modes

The same models are used as in fits to the $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ control modes (see Section 4.5.2.1), namely a DCB to parameterise the signal, an exponential PDF for combinatorial background, and an additional peak for $B^0_s \to \psi(2S)\,(\ell^+\ell^-)\,K^0_S$ decays. The results of the fits to the $B^0 \to \psi(2S)\,(\mu^+\mu^-)\,K^0_S$ and $B^+ \to \psi(2S)\,(\mu^+\mu^-)\,K^{*+}$ channels, in the range $m(B) \in [5175, 5700\,\text{MeV}]$, are shown in Fig. 4.31.

4.5.3.2 Electron modes

Similarly to the fits to the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ channels, the fits to the $B \to \psi(2S)\,(\ell^+\ell^-)\,K^{(*)}$ channels contain a signal peak parameterised with a sum of three DCBs, combinatorial background parameterised with an exponential PDF, and a small $B^0_s \to \psi(2S)\,(e^+e^-)\,K^0_S$ contribution

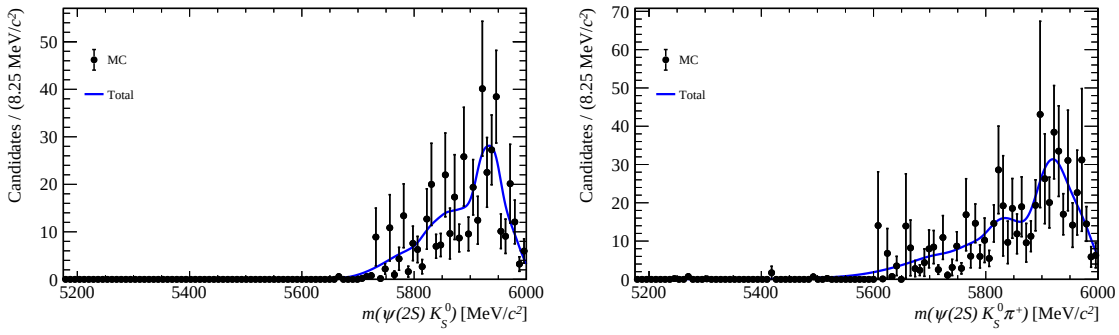


Fig. 4.32: Results of fits to $m(B)$, determining fit PDFs for J/ψ leakage, in the $B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$ (left) and $B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$ (right) channels. The values of $m(B)$ calculated with dilepton mass constraints are used here.

parameterised with a mass-shifted signal peak. In addition, these fits contain contributions from $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays reconstructed in the $q^2_{\psi(2S)}$ bin, due to the spurious recovery of bremsstrahlung photons which did not actually originate from the candidate electrons. These are described using GKDEs with $\rho = 1$, determined using fully-selected $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays reconstructed in $q^2_{\psi(2S)}$, as shown in Fig 4.32. The yields of these "leakage" contributions are Gaussian-constrained using the yields of the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ fits, and the relative proportions of fully-selected $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC reconstructed in the $q^2_{J/\psi}$ and $q^2_{\psi(2S)}$ regions. The widths of these constraints are determined from the uncertainties on the J/ψ-mode yields returned by HESSE, and the uncertainties on the relative proportions of MC in different q^2 regions.

The results of the fits to $B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$ and $B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$ data, in the range $m(B) \in [5175, 6000\,\text{MeV}]$, are shown in Fig. 4.33.

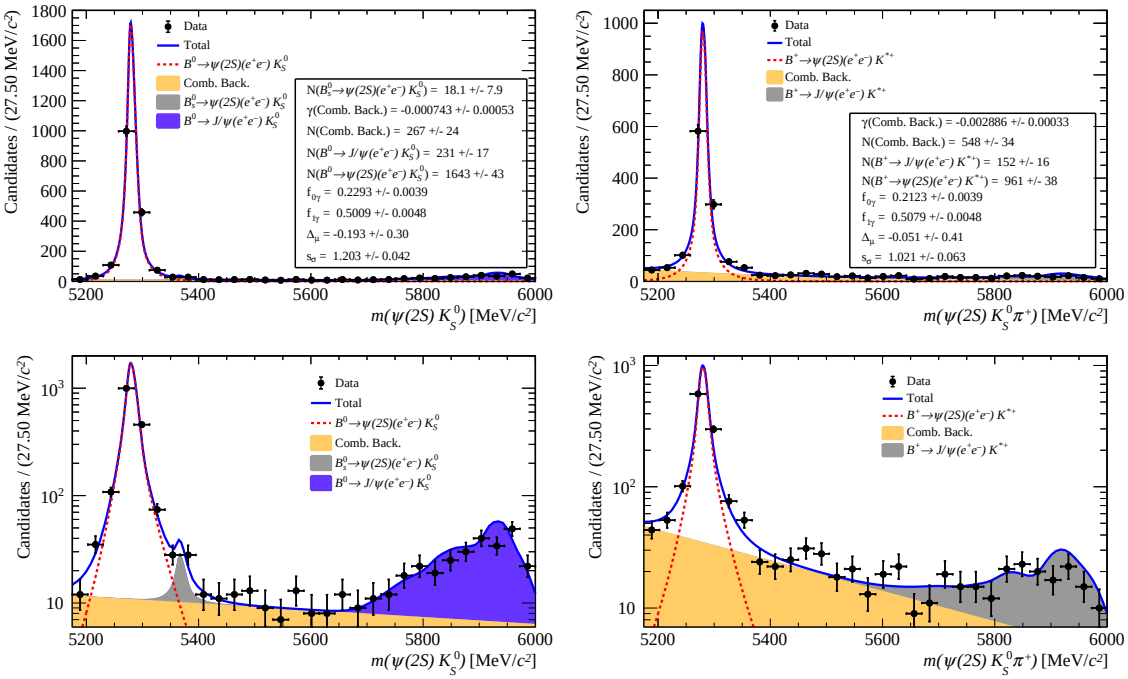


Fig. 4.33: Results of fits to $m(B)$, determining signal yields for $B^0 \to \psi(2S)\,(e^+e^-)\,K^0_S$ (left) and $B^+ \to \psi(2S)\,(e^+e^-)\,K^{*+}$ (right) candidates in data. These plots are shown with a linear (top) and logarithmic (bottom) y-scale, to highlight low-yield backgrounds. The values of $m(B)$ calculated with dilepton mass constraints are used here.

4.5.4 Signal modes

The yields for the rare-mode $B \to K^{(*)}\ell^+\ell^-$ decays are used to measure the ratios R_X^{-1}. In addition, they are used to measure differential branching fractions for these decays (though only as a cross-check in the case of the muon modes). As the leptons don't originate from any resonance, the dilepton invariant mass cannot be constrained when computing $m(B)$ with `DecayTreeFitter`. Therefore, the resolutions of $m(B)$ are degraded for signal decays, similarly to the fits to $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ data without dilepton mass constraints (see Sections 4.5.2.2 and 4.5.2.4).

4.5.4.1 Muon modes

As in the $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ fit without dilepton mass constraints, low-mass background is not present in the range $m(B) \in [5175, 5700\,\text{MeV}]$. Therefore, the $B \to K^{(*)}\mu^+\mu^-$ mass-fits are parameterised using a DCB for the signal peak, and an exponential PDF for the combinatorial background. However, the small numbers of candidates in these channels mean the signal resolution parameters Δ_μ and s_σ must be determined from outside sources, to ensure the fits are

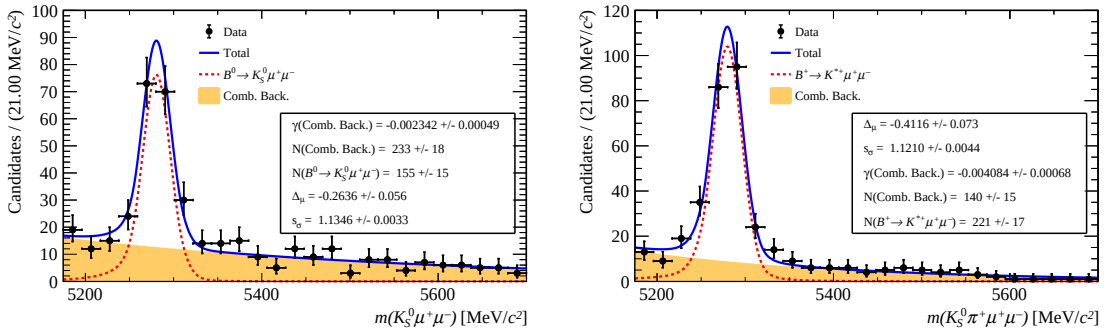


Fig. 4.34: Results of fits to $m(B)$, determining signal yields for $B^0 \to K^0_S \mu^+\mu^-$ (left) and $B^+ \to K^{*+}\mu^+\mu^-$ (right) candidates in data.

stable. They are Gaussian-constrained using the values found in fits to $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ data without dilepton mass constraints applied (see the legends of Fig. 4.24). The results of fits to $B^0 \to K^0_S \mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ data are shown in Fig. 4.34.

4.5.4.2 Electron modes

As in the fit to $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ data without any dilepton mass constraints, a wide range of $m(B) \in [4700, 6000\,\text{MeV}]$ is used in the fits to $B \to K^{(*)}e^+e^-$ data, to ensure signal is not cut away. The signal is parameterised with three DCBs for different numbers of recovered bremsstrahlung photons, and combinatorial background is parameterised using an exponential PDF. To ensure the fits are stable, the signal resolution parameters Δ_μ and s_σ are Gaussian-constrained using the values found in fits to $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ data without dilepton mass constraints applied (see the legends of Fig. 4.30)

In addition, various physical sources of background fall into the signal mass-window, and must be parameterised. Due to bremsstrahlung emission, some $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays are reconstructed in the q^2_{Rare} bins. These leakage components are described using GKDEs with $\rho = 1.5$, and their yields are Gaussian-constrained using the yields of the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ fits, and the relative proportions of J/ψ-mode MC reconstructed in the $q^2_{J/\psi}$ and q^2_{Rare} regions (similarly to the leakage in the $B \to \psi(2S)\,(e^+e^-)\,K^{(*)}$ fits, documented in Section 4.5.3.2).

As found in Section 4.4.1.5 and 4.4.2.3, a non-negligible number of $B \to K^{(*)}\pi^+\pi^-$ decays are expected to survive the selection requirements. Each forms a sharp peak, just below the nominal value of $m(B)$ due to the $\pi \to e$ misidentifications, which is parameterised using a DCB. These PDFs are fitted to $B \to K^{(*)}\pi^+\pi^-$ MC samples, passing all selection requirements except for the trigger, and with PID weights applied (including trigger-unbiased weights for misidentified pions). The yields of these components are Gaussian-constrained according

to the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ yields, and relative selection efficiencies for $B \to K^{(*)}\pi^+\pi^-$ and $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays computed in the sections mentioned above.

Finally, partially-reconstructed $B \to K^{(*)}e^+e^-$ decays with high-mass kaon resonances, where a final-state pion is missed in reconstruction, form a "shoulder" below the nominal signal mass (see Sections 4.4.1.3 and 4.4.2). In the $B^0 \to K^0_S e^+e^-$ fit, these backgrounds are parameterised using a DCB, fitted to MC samples of $B^+ \to K^{*+}e^+e^-$ decays passing full B^0-mode selection requirements. However, no such MC samples exist for the higher-order $B \to K^{(*)}e^+e^-$ decays which pollute the $B^+ \to K^{*+}e^+e^-$ fit. Therefore, they are instead described using MC samples for the analogous $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays, on the assumption that the differing q^2 distributions have a sub-dominant effect on the $m(B)$ shape for such decays[*]. An MC sample including $K_1(1270)$, $K_1(1400)$ and $K_1^*(1430)$ resonances decaying to a $K^0_S\pi^+\pi$ final-state, and passing full reconstruction requirements in the $q^2_{J/\psi}$ region, is used[†]. A PDF for this component is found using a GKDE with $\rho = 2.0$. The yields for partially-reconstructed decays are permitted to float freely in fits to data.

PDFs for all the physical background components are shown in Fig. 4.35. The results of the fits to $B^0 \to K^0_S e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ data are shown in Fig. 4.36. As these decays had not been observed before this analysis, these mass-fits were only unblinded after permission was granted by the review committee (see Section 3.4).

[*]A systematic uncertainty associated to this assumption is computed in Section 5.14.2.

[†]Note that the same MC sample was used to describe partially-reconstructed backgrounds in the $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ fit (see Section 4.5.2.4).

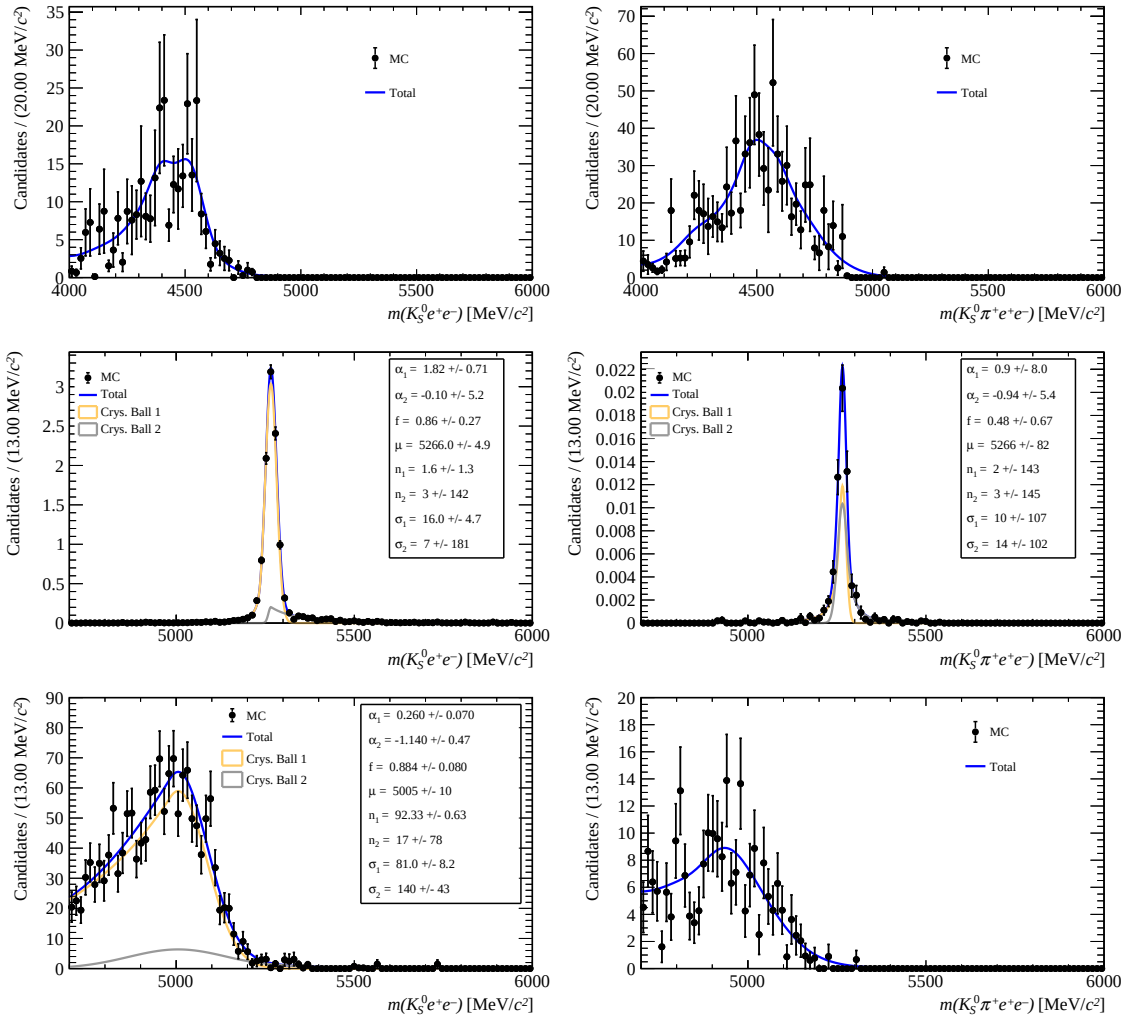


Fig. 4.35: Results of fits to $m(B)$, determining fit PDFs for J/ψ leakage (top), misidentified $B \to K^{(*)}\pi^+\pi^-$ (middle), and partially-reconstructed kaon (bottom) backgrounds, in the $B^0 \to K^0_S e^+ e^-$ (left) and $B^+ \to K^{*+} e^+ e^-$ (right) channels.

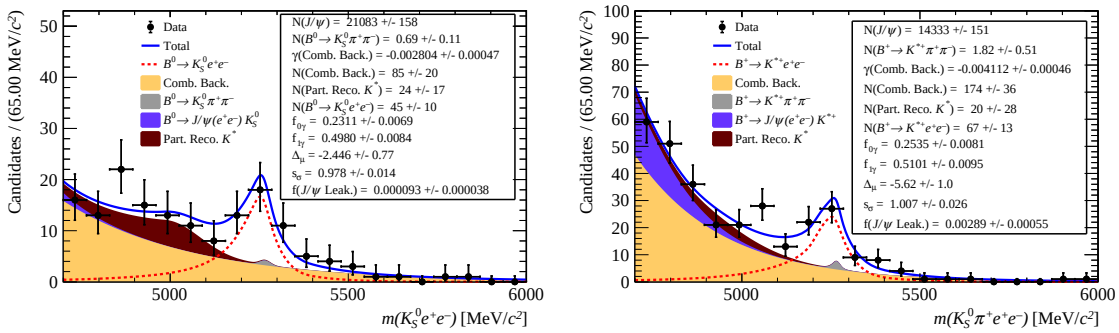


Fig. 4.36: Results of fits to $m(B)$, determining signal yields for $B^0 \to K^0_S e^+e^-$ (left) and $B^+ \to K^{*+}e^+e^-$ (right) candidates in data. Note that the yields for the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ leakage are given by $N(J/\psi) \times f(J/\psi\ \text{Leak.})$

Chapter 5

Systematic uncertainties

A number of systematic uncertainties are associated with the estimates of efficiencies, and the maximum likelihood fits used to extract yields. These are outlined in the following section, beginning with the general methodologies used to calculate systematic uncertainties, then listing the specific sources of systematic uncertainty and how they are assessed, and concluding with summaries of the values of systematic uncertainties affecting R_X^{-1}.

5.1 Common methods

Several methods and techniques are used throughout the systematic uncertainty calculations in this section. They are outlined in the following sub-sections.

5.1.1 Covariance matrices

Most of the systematic uncertainties outlined in this section are found using ensembles of "toy" calculations. First a set of selection efficiencies (and/or f_{q^2} factors) are calculated using the nominal method outlined in Section 4.3, applying the corrections outlined in Section 4.2 to MC samples. Then, this calculation method is varied in some manner, and the efficiencies are recalculated. Using these alternative efficiency values, the corresponding change to an observable (such as a rare-mode branching fraction, or R_X^{-1} ratio) is computed, relative to its nominal value. A variety of alternative calculation methods may be used, and propagated to the measured observables, to find ensembles of such toy values.

The covariances of these ensembles are taken as the values of a given systematic uncertainty for different observables, with off-diagonal elements giving the correlations between the systematics for different observables in different data-taking years [291, p. 67]:

$$cov(X,Y) = \frac{1}{N}\sum_{i}^{N}\sum_{j}^{N}(X_i - \hat{X})(Y_j - \hat{Y}) \tag{5.1}$$

Here, X and Y represent different observables (or the same observable if $X = Y$), with the indices i and j denoting the N different alternative toy values. The nominal observable values are denoted using $\hat{X}$ and $\hat{Y}$. This equation is used to calculate the systematic uncertainties explained throughout this section, unless otherwise specified.

The nature of the variation (or variations) in calculation method depends on the source of systematic uncertainty being examined. For example, to test uncertainties from TISTOS tag biases, a set of L0 correction weights computed with alternative tag requirements may be applied to MC samples. Then, selection efficiencies would be recalculated with these new weights. Note that all correlations between systematic uncertainties in different years are manually set to zero, unless there is some physical reason why such correlations would exist.

5.1.2 Bootstrapping

Many of the systematic uncertainties associated to the finite sizes of MC and data samples used to compute corrections, are calculated using "bootstrapping" [329]. This is where entries in a sample are randomly re-sampled to construct a new "bootstrapped sample", with comparable size to the original. Note that each entry may be re-sampled two or more times, or not be re-sampled at all. Then, the bootstrapped sample is used to re-compute some component of the analysis, such as a set of correction weights or a mass-fit PDF. The selection efficiencies or signal yields are then re-computed with this bootstrapped set of weights or PDF. By repeatedly bootstrapping a sample, an ensemble of observable values can be calculated, and used to assign systematic uncertainty values.

For ease of implementation, samples are bootstrapped in this analysis by assigning each entry a random weight, drawn from a Poisson distribution with unit mean [330].

5.1.3 Assessing fit model systematics

Toy mass-fits are used to estimate systematic uncertainties associated to the fit model. First, an alternative mass-fit model is constructed for a given channel, which is varied somehow from the nominal model. For example, the PDF used to describe the signal may be varied, or a PDF for an additional background component may be introduced. Then, a "toy dataset" is generated from this model, using the methods outlined in Section 7.1.3. This consists of a set of $m(B)$ values, of comparable size to the real data in that channel. After this, fits are carried out to this toy dataset using both the nominal and alternative fit models, and the resulting signal yields

$N_{Sig.}^{Nom.}$ and $N_{Sig.}^{Alt.}$ are recorded. By repeatedly generating and fitting to toy datasets, an ensemble of $\Delta_N \equiv N_{Sig.}^{Alt.} - N_{Sig.}^{Nom.}$ values is built up. Finally, a range across Δ_N covering 68.3% of these toys is assigned as a systematic uncertainty on the fitted yield in this channel.

Note that as data from all years is combined in the mass-fits to data, uncertainties relating to the fit model are not quoted for individual years. Instead, a single value is quoted across all years. The uncertainties for yields in each channel are then propagated to find the total uncertainties on each measured observable.

5.2 PID efficiency corrections

Several systematic uncertainties are associated with the use of PID weights when calculating selection efficiencies, as outlined in Section 4.2.1.

5.2.1 Calibration sample sizes

Each PID weight has a statistical uncertainty associated to it, due to the finite sizes of the calibration datasets. These statistical uncertainties lead to obtain systematic uncertainties on the selection efficiency estimates.

The PIDCALIB package automatically calculates the uncertainties associated to the weights in each kinematic bin, using the normal approximation (see Section 3.2). For the fit-and-count technique used to compute electron PID efficiencies, these uncertainties are calculated from the maximum likelihood fits in each kinematic bin, using HESSE. Note that the weights in each kinematic bin are assigned to multiple candidates across different MC samples in a given year. For example, a map of PID efficiency weights will be applied to MC samples for both $B^0 \to K^0_S e^+ e^-$ and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ decays. Therefore, the estimates of systematic uncertainties resulting from the finite PID calibration sample sizes must account for these correlations between weights in different candidates.

To do this, the PID weights in each bin of each efficiency map are randomly varied according to their Gaussian uncertainties. These varied PID weights are then folded into the relevant MC samples, and selection efficiencies are recalculated. This process is repeated 100 times, and systematic uncertainties of $\sim 0.2 - 0.8\%$ are obtained for R_X^{-1} in different years.

5.2.2 Binning schemes

As outlined in Section 4.2.1.2, an algorithm is used to find binning schemes which capture variations in PID efficiencies with p, p_T, and `nTracks`, when calculating PID weights. However,

any finite binning schemes will still give imperfect descriptions of these variations, resulting in systematic uncertainties for the selection efficiencies.

To estimate these, the number of initial bins and re-binning tolerances given in Table 4.11 are randomly varied, between half and double their nominal values. Binning schemes are calculated with these varied parameters, which are then used to create new PID weight maps. These new PID weights are folded into the relevant MC samples, and selection efficiencies are recalculated. This process is repeated 10 times, and systematic uncertainties of $\sim 0.2 - 1.5\%$ are found for R_X^{-1} in different years.

5.2.3 Electron PID factorisation

The efficiencies of PID requirements for different tracks are assumed to factorise, such that the total PID efficiency for a candidate can be calculated by multiplying together the PID efficiency weights associated to different tracks, as in eq. 4.2. However, this assumption breaks down if the PID performances for different tracks are correlated. This could occur when multiple tracks pass through a PID sub-detector. For example, if two tracks are close together when they pass through the RICH, it may be difficult to distinguish the Cherenkov rings produced by them. Therefore, the PID variables assigned to the two tracks could be correlated. Note that such correlations are not accounted for with the PID weights used in this analysis, but *are* accounted for by the simulated PID responses in MC samples.

Such an effect can be seen in the PID responses for final-state electrons in $b \to s e^+ e^-$ decays. This is primarily due to energy clusters in the ECAL overlapping when the two candidate electron tracks are close together [331]. Such tracks will have correlated ECAL energy measurements, which are used to compute electron PID variables (see Section 2.2.2.4). The size of this effect is assessed by computing the PID efficiencies for the two electrons in MC with and without the assumption that these two efficiencies factorise:

$$\varepsilon_{PID} \equiv \frac{N(e^+,\ e^-\ both\ passed)}{N(Total)}, \qquad \varepsilon^{'}_{PID} \equiv \prod_i^{e^+,e^-} \frac{N(i\ passed)}{N(Total)} \tag{5.2}$$

The ratio $r_{fact} \equiv \varepsilon_{PID}/\varepsilon^{'}_{PID}$ is known as the "factorisation bias". In Fig. 5.1, the factorisation biases in $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ MC are plotted as a function of the distance between the two candidate electrons in the ECAL, and can be seen to peak when this separation is below $\sim 200\,\text{mm}$.

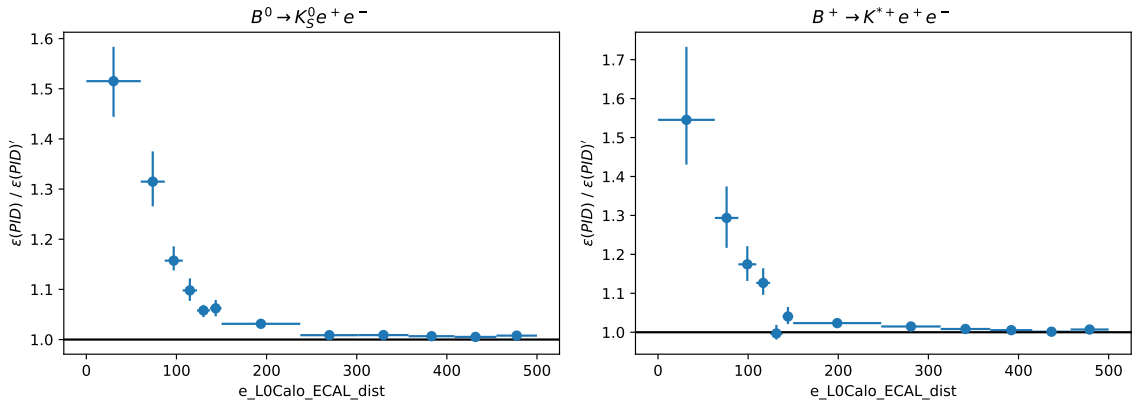


Fig. 5.1: The factorisation biases for electron PID performances seen in MC for $B^0 \to K^0_S e^+ e^-$ (left) and $B^+ \to K^{*+} e^+ e^-$ (right) decays, in bins of the distance between the e^+ and e^- in the ECAL (in mm). These biases are calculated using MC samples passing preselection, across all of q^2, and from all years.

5.2.3.1 Assigning a systematic uncertainty

For each electron-mode channel, in each year, r_{fact} is calculated using the simulated PID responses in a full-selected MC sample, with all corrections applied except for electron PID weights. This is taken as the change in total selection efficiency for that channel, due to the non-factorisation of electron PID responses. The MC samples for each channel are bootstrapped 500 times, and the calculations of r_{fact} carried out for each bootstrapped sample. The resulting ensemble of r_{fact} in different channels, and the corresponding changes in observables, are used to assign systematic uncertainties. Values of $\sim 0.2 - 1.5\%$ are found for R_X^{-1} in different years. As the non-factorisation of PID is common across different channels and years, correlations between the systematics in different years, found from this ensemble, are preserved.

5.3 Electron tracking efficiency corrections

A finite number of data and MC candidates were used when calculating the electron tracking efficiency correction weights [314]. This leads to systematic uncertainties on the selection efficiencies estimated for the electron modes. The centrally-produced maps of electron tracking weights include estimated statistical uncertainties in each bin, meaning the resulting systematic uncertainties can be calculated in a similar way to those for the PID weights (see Section 5.2.1). Within each bin, each tracking weight is randomly varied according to its Gaussian uncertainty. These varied weights are then folded into the corresponding electron-mode MC samples, and

selection efficiencies are recalculated. This process is repeated 100 times, to find systematic uncertainties of $0.06-0.2\%$ for R_X^{-1} in different years.

5.4 Muon tracking efficiencies

While better modelled then for electrons, there are some differences between the tracking efficiencies modelled in MC for muons, compared to those found in data [241]. A systematic uncertainty is assigned to the muon-mode efficiencies, to account for this mismodelling.

Maps containing the ratios of muon tracking efficiencies in data and MC, as a function of the track p and η, are folded into the muon-mode MC samples. These maps were centrally computed by an LHCb tracking performance group, using a similar tag-and-probe method to that used for the electron tracking weights [241]. Maps are available for each data-taking year except for 2011, so the map for 2012 is applied to all run-1 MC samples.

In order to account for statistical uncertainties on the tracking corrections themselves, the correction weights in each bin of each map are randomly varied according to their Gaussian uncertainties, and these randomly-varied weights are then folded into the MC samples. The total selection efficiencies are re-computed with the muon tracking weights imposed, under the assumption that the tracking efficiencies factorise for the two muons*. This process is repeated 100 times to find an ensemble of efficiency values found with randomly-varied tracking weights applied. Systematic uncertainties of $0.05-0.25\%$ for R_X^{-1} in different years were found using this ensemble. These systematic uncertainties are assumed to be uncorrelated between different years, with the exception of run-1, as the same correction map is applied in these two years.

5.5 Generated kinematic and detector occupancy corrections

The $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ control-mode MC and data samples used to compute corrections for generator mismodelling (see Section 4.2.3) are of finite size, meaning these weights will have statistical uncertainties associated to them. These statistical uncertainties must be estimated, and used to quantify the resulting systematic uncertainties on selection efficiency estimates. However, relatively few $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ data and MC candidates fall into some bins of the weight maps, meaning uncertainties on these weights may not be accurately described with the normal approximation. In addition, each candidate is used to calculate weights across several k-folds, meaning weights for different k-folds will be correlated. These correlations also need to be accounted for when estimating systematic uncertainties.

*Under this assumption, the total correction weight for a candidate i is given by $w^i_{Trk.}(Tot.) = w^i_{Trk.}(\mu^+) \times w^i_{Trk.}(\mu^-)$

To account for these effects, the $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ MC and data samples are bootstrapped, and these bootstrapped samples are used to calculate new sets of generator mismodelling weights. Note that the data samples are bootstrapped before calculating *sWeight*s, meaning a new mass-fit is carried out using the bootstrapped sample. The new weights are then applied to the corresponding MC samples, and selection efficiencies are recalculated. This process is repeated 100 times, to obtain systematic uncertainties of $0.7 - 5.3\%$ for R_X^{-1}.

5.6 K^0_S reconstruction category corrections

As outlined in Section 4.2.4, the weights for the proportions of K^0_S-mesons in each reconstruction category are calculated from $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ MC and data samples. These samples have finite sizes, meaning the weights have statistical uncertainties associated to them. To assess their impacts on selection efficiency estimates, the proportions of LongLong K^0_S candidates in $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ MC and *sWeight*ed data are varied according to their uncertainties, found using the normal approximation. Then, the corresponding fraction of DownDown K^0_S candidates is calculated according to $f_{DD} = 1 - f_{LL}$. These fractions are used to calculate randomised K^0_S category correction weights, which are folded into MC samples, and used to recalculate the selection efficiencies. This process is repeated 100 times, to find an ensemble of selection efficiency estimates, from which systematic uncertainties of $0.009 - 0.61\%$ are found for R_X^{-1} in different data-taking years.

5.7 Trigger performance corrections

As outlined in Section 4.2.5, the TISTOS method is used to examine the performance of the L0 trigger in data, and hence correct the performance simulated in MC. Some aspects of this method lead to systematic uncertainties, which are outlined in the following pages.

5.7.1 Control-mode sample sizes

Each trigger efficiency correction weight has a statistical uncertainty associated to it, arising from the limited sizes of the $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ control-mode data and MC samples used for its calculation. Uncertainties for the MC and data efficiencies in each bin of the efficiency profiles are calculated using the normal approximation. These statistical uncertainties lead to systematic uncertainties on the estimated selection efficiencies, which must be quantified.

First, the MC and data efficiencies in each bin of the efficiency profiles are varied within their Gaussian uncertainties. Then, the randomised efficiency profiles for the `L0Muon-TOS`,

L0Electron-TOS, and L0Hadron-TOS requirements are fitted using the functional parameterisations outlined in 4.2.5. New sets of L0 efficiency correction weights are calculated for each candidate in MC, and all selection efficiencies are recalculated. This process is repeated 100 times, to find an ensemble of changes to observables.

A small number of fits to the randomised efficiency maps do not converge properly, meaning any correction weights found using these parameterised maps may not have physical values. This can lead to spuriously large changes to the measured observables. While the fits to the nominal efficiency maps are monitored to ensure they converge properly, this is not feasible for the large numbers of fits used in the toy studies here. Therefore, the outlier efficiency values from these bad fits could lead to unphysically large systematic uncertainty values, if they are estimated using eq. 5.1.

Instead, the systematic uncertainties on each observable are calculated from each set of toy values $\vec{\varepsilon}$ using their inner 68.3% quantiles, *i.e.* as $\sigma = (q_{84.15\%}(\vec{\varepsilon}) - q_{15.85\%}(\vec{\varepsilon}))/2$. Any correlations between different observables are still estimated using the off-diagonal elements from eq. 5.1. Systematic uncertainties of $0.07 - 0.5\%$ are found for R_X^{-1} in different years.

5.7.2 Efficiency parameterisation

In order to mitigate statistical noise, the L0Muon-TOS, L0Electron-TOS, and L0Hadron-TOS efficiency profiles are fitted using various sums of error functions before calculating L0 correction weights. However, these choices of parameterisation may not fully capture the relationships between particle kinematics and trigger efficiencies. Systematic uncertainties are assigned for this effect, which are calculated by re-computing the L0 correction weights using the unparameterised, binned efficiency profiles, and applying these new weights to the relevant MC samples. Then, the selection efficiencies were recalculated, and systematic uncertainties were calculated from the corresponding changes to measured observables using eq. 5.1. This method gives systematic uncertainties of $0.05 - 1.8\%$ for R_X^{-1} in different data-taking years.

5.7.3 L0Global-TIS tag bias

As discussed in Section 4.2.5, it is impossible to retrieve a sample of LHCb data which has not passed some set of trigger requirements. Therefore, any sample used to assess the efficiency of some "probe" trigger requirement must be "tagged" using a different, uncorrelated trigger requirement. If the performances of the tag and probe requirements are correlated, this can bias an estimate of the probe trigger efficiency.

There is possible evidence of such a bias in the L0Global-TIS efficiency profiles, where the efficiencies are seen to differ between events tagged with L0Muon-TOS and L0Electron-TOS

(see Section 4.2.5.4). A systematic uncertainty is assigned to account for this effect. Instead of the nominal efficiency profiles tagged with `L0Muon-TOS`, a set of correction weights are calculated using the profiles tagged with `L0Electron-TOS`. To account for the limited sizes of the samples used to calculate the `L0Electron-TOS`-tagged efficiency profiles, the efficiencies in each bin were randomly varied according to their Gaussian uncertainties before computing L0 correction weights, as in Section 5.7.1 above. Using eq. 5.1 with an ensemble of 100 such toy efficiency calculations, systematics uncertainties of $0.13-0.94\%$ were obtained for R_X^{-1} in different years.

5.8 Residual corrections

Limited quantities of $B\to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ MC and data were used when calculating weights for the residual discrepancies between data and MC, as outlined in 4.2.6. As the `GBReweighter` algorithm was used to compute these weights, rather than a binned method, the statistical uncertainties on the correction weights cannot be calculated using the normal approximation. Therefore, these uncertainties were assessed using a bootstrapping method, as used for the generator mismodelling weights in Section 5.5 above.

The $B\to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ MC and data samples are bootstrapped, and these bootstrapped samples are used to train new sets of `GBReweighters`. The weights from these new `GBReweighters` are then folded into the relevant MC samples, and selection efficiencies are recalculated. This process was repeated 50 times to find uncertainties of $0.8-4.7\%$ for R_X^{-1} in different years.

5.9 q^2 resolution smearing

As described in Section 4.2.7, a pair of "smearing parameters" μ_{Smear} and σ_{Smear} are calculated using $B\to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ data and MC, to account for any differences in resolution for q^2. Then, the q^2 value of each MC candidate is smeared by an offset Δ_{q^2}, sampled from a Gaussian distribution with mean μ_{Smear} and width σ_{Smear}. However, the smearing parameters have statistical uncertainties associated to them, which would lead to uncertainties on the smeared values of q^2, and hence to systematic uncertainties on the selection efficiencies.

To calculate this, each value of μ_{Smear} and σ_{Smear} is randomly varied according to the Gaussian uncertainties found using HESSE in the fits to data. Then, the q^2 values in MC samples are smeared using these randomised smearing parameters, and selection efficiencies are recalculated. With an ensemble of 100 such toys, systematic uncertainties of $0.16-0.73\%$ were found for R_X^{-1} in different years. These systematics were assumed not to be correlated

between different years, except for run-1, where MC and data from each year were combined when computing the smearing parameters.

5.10 Residual BDT mismodelling

Even after all the MC corrections are applied, there may be some mismodelling of the $\text{BDT}_{Comb.}$ output distribution, and hence some biases on the efficiencies of the cuts to these variables. There are two likely causes for this: Firstly, in order to obtain stable weight values, some of the $\text{BDT}_{Comb.}$ input variables were not considered when calculating the weights for residual discrepancies between MC and data, as outlined in Section 4.2.6. Any mismodelling of these omitted variables would lead to mismodelling of the $\text{BDT}_{Comb.}$ output variable. Secondly, these residual discrepancy weights were calculated using *sWeight*ed data. However, the *sPlot* method assumes the control variables are not correlated with the discriminating variable, for both signal and background. Contrary to this assumption, the efficiencies of cuts to $\text{BDT}_{Comb.}$ can be seen to vary smoothly with $m(B)$ for combinatorial background, in the sculpting studies outlined in Section 4.1.5.3. This suggests the $\text{BDT}_{Comb.}$ input variables are also likely correlated with $m(B)$, meaning their *sWeight*ed distributions in data may be biased.

To account for such possible mismodelling, the distributions of $\text{BDT}_{Comb.}$ output are compared between data and MC. Note that the arguments above mean that the *sPlot* method cannot be used to unfold the distribution in data, so a fit-and-count method is used instead. For each $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ control mode, in each year, a binning scheme is constructed across $\text{BDT}_{Comb.}$. This consists of ten bins populated approximately equally by MC. Mass-fits are carried out to fully-selected data falling within each bin, using the models outlined in Sections 4.5.2.1 and 4.5.2.3. The proportions of data falling within each bin of $\text{BDT}_{Comb.}$ are calculated using the signal yields from these fits, and compared to the proportions of fully-corrected MC in each bin. Examples of such comparisons are shown in Fig. 5.2. Sets of correction weights are calculated from the relative proportions of MC and data in each bin of $\text{BDT}_{Comb.}$, and are then applied to the relevant rare-mode and control-mode MC samples. The efficiencies of the cuts to $\text{BDT}_{Comb.}$ are recalculated with these weights applied.

To account for the statistical uncertainties on these $\text{BDT}_{Comb.}$ correction weights, MC and data samples are bootstrapped 100 times, and the procedure above is repeatedly carried out. Systematic uncertainties are calculated from the resulting ensemble of changes to $\text{BDT}_{Comb.}$ cut efficiencies, to give values of $0.09 - 1.9\%$ for the systematic uncertainties on R_X^{-1} in different years.

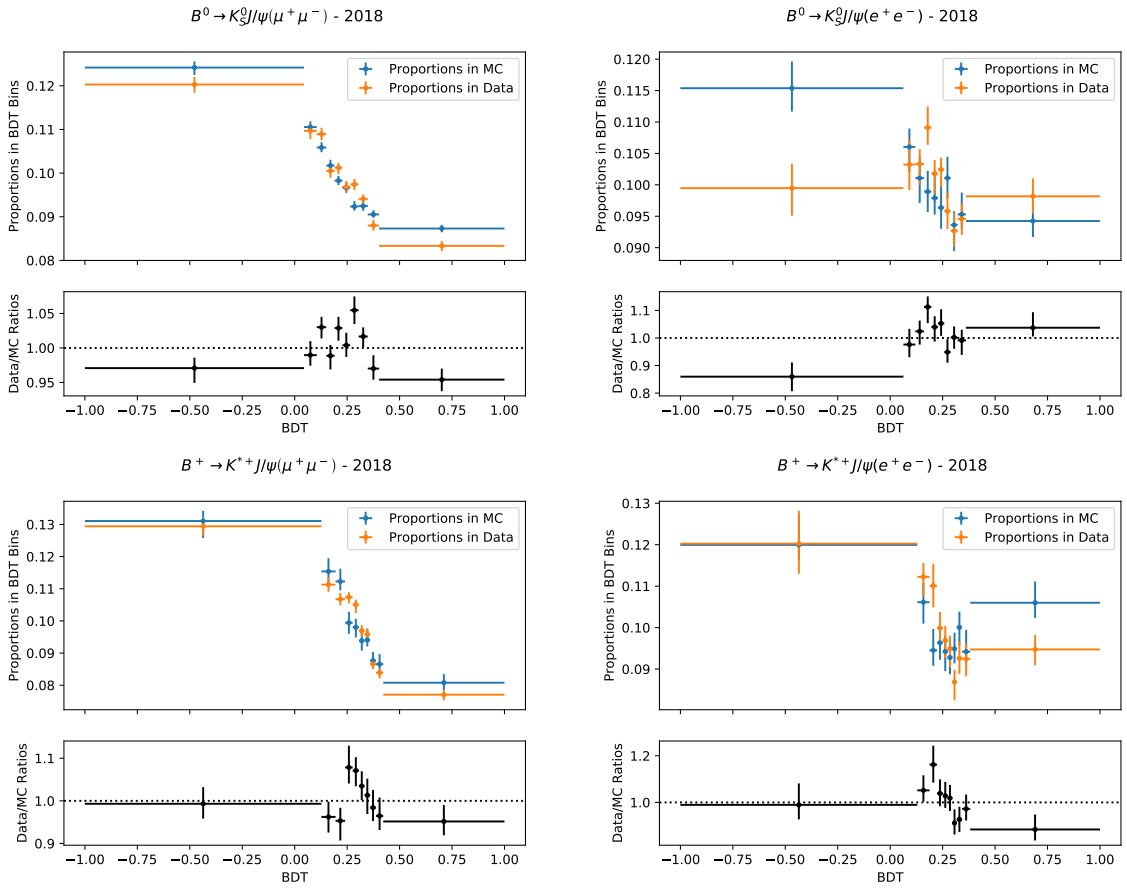
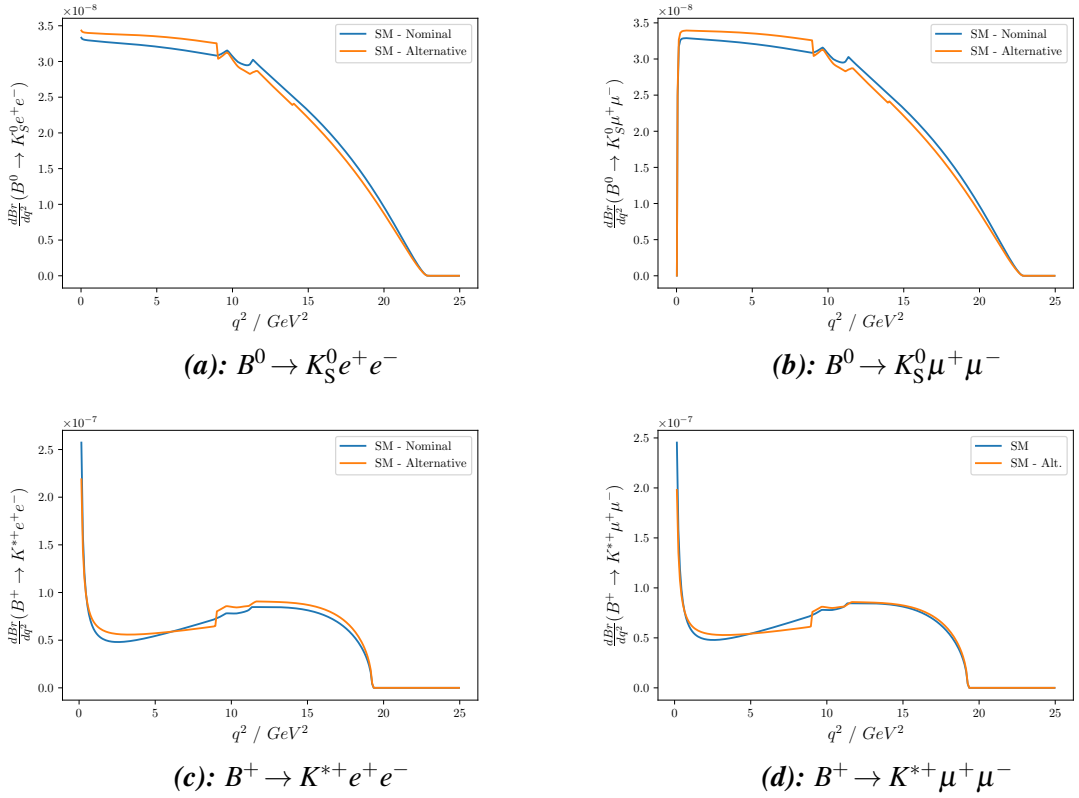


Fig. 5.2: Comparisons of proportions of MC and data found in bins of BDT output for $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (top left), $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (top right), $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (bottom left), $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (bottom right) decays, recorded in 2018. The lower proportions of the plots show ratios of the proportions in data and MC, used to calculate sets of correction weights from which systematic uncertainties are computed.

5.11 Signal decay model

The MC samples for $B \to K^{(*)}\ell^+\ell^-$ rare-mode decays were generated using a set of hadronic form-factors computed with light-cone sum rules [95]. However, as outlined in Section 1.2.2, there are substantial theoretical uncertainties associated to these form-factors. Such uncertainties mean the q^2 distributions used in these MC samples may not be physically accurate, meaning the efficiencies of the cuts to $q^2_{Reco.}$ which select signal decays could be biased. As this analysis considers the differential branching fractions for these decays in bins of q^2, such form-factor dependencies are cancelled to first order by the f_{q^2} factors (see Section 4.3.1). However, the proportions of candidates migrating between bins of q^2 is still dependent



(a): $B^0 \to K^0_S e^+ e^-$ *(b):* $B^0 \to K^0_S \mu^+ \mu^-$

(c): $B^+ \to K^{*+} e^+ e^-$ *(d):* $B^+ \to K^{*+} \mu^+ \mu^-$

Fig. 5.3: Plots of the nominal SM distributions of $\frac{d\Gamma}{dq^2}$ for different rare decay modes, plus the distributions found with random variations to the hadronic form-factors, as computed using FLAVIO [332].

on the underlying q^2_{True} distribution, meaning some form-factor dependencies still remain for R_X^{-1}. Systematic uncertainties are assigned to account for these.

The hadronic form-factors (and other parameters used to compute $B \to K^{(*)} \ell^+ \ell^-$ differential branching fractions) are varied within their theoretical uncertainties under the SM, and used to recalculate the q^2_{True} distributions for each $B \to K^{(*)} \ell^+ \ell^-$ decay. These predictions are computed using FLAVIO; a PYTHON software package designed to compute the theoretical values of flavour physics observables, and to carry out global fits to experimental measurements (see Section 1.2.4) [332]. Examples of such randomised predictions of the q^2_{True} distributions are shown in Fig. 5.3. These randomised and nominal predictions are used to calculate sets of weights across q^2_{True}, which are applied to both reconstruction-level and generator-level $B \to K^{(*)} \ell^+ \ell^-$ MC samples. Then, the selection efficiencies and f_{q^2} values are re-computed with these reweighted q^2_{True} distributions. This process is repeated 250 times, and the corresponding ensemble of changes to measured observables is used to compute systematic uncertainties. Values of $0.1 - 0.43\%$ and $0.33 - 1.5\%$ are found for $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$, respectively,

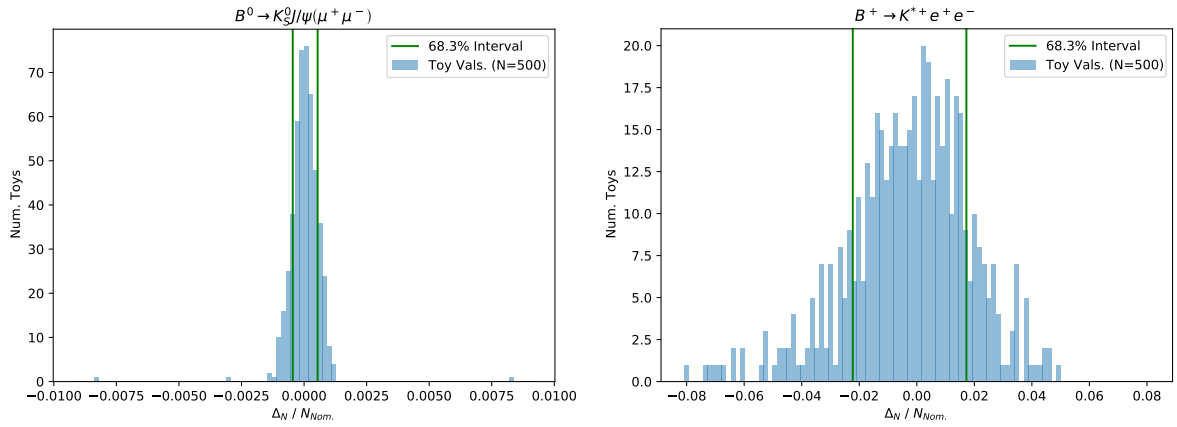


Fig. 5.4: Normalised distribution of Δ_N found from toy fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^+ \to K^{*+}e^+e^-$ (right) datasets, used to assess the systematic uncertainties arising from finite MC sample sizes when parameterising signal PDFs. The 68.3% inner quantiles of these distributions, used to assign these systematic uncertainties, are shown by the vertical green lines.

with the increased uncertainties in the latter resulting from the photon pole at low q^2. Large correlations between the uncertainties in different years were retained, as the decay model is common to all data-taking years.

5.12 Fit model signal parameterisation

Various sums of DCB PDFs are used to parameterise signal contributions in the mass-fits, as documented in Section 4.5.1. For each mass-fit channel, the tail parameters α_i and n_i for this PDF are fixed from fits to fully-selected MC. Therefore, these parameters will have uncertainties associated to them, due to the limited size of the corresponding MC sample. The signal PDF shape has an impact on the fitted signal yield in data, meaning any uncertainties on the PDF tail parameters lead to a systematic uncertainty on this signal yield.

To quantify this systematic uncertainty for a given mass-fit channel, the MC used to fix the signal PDF tail parameters is bootstrapped. Then, a new signal PDF - constructed from a sum of DCBs as usual - is parameterised using this bootstrapped MC sample. This new signal PDF is used to construct an alternative mass-fit model. As outlined in Section 5.1.3, toys are generated from this alternative model, which are then fitted using both the nominal and alternative fit models, to extract a pair of signal yields.

This process is repeated 500 times for each mass-fit channel, to find distributions of $\Delta_N \equiv N^{Alt.}_{Sig.} - N^{Nom.}_{Sig.}$. The half-width of the inner 68.3% quantile for each distribution is

assigned as the systematic uncertainty in that channel. The smallest and largest resulting systematics are 0.05% and 2% relative to the signal yield, for the $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ and $B^+ \to K^{*+}e^+e^-$ channels respectively. Distributions of Δ_N for these channels are shown in Fig. 5.4.

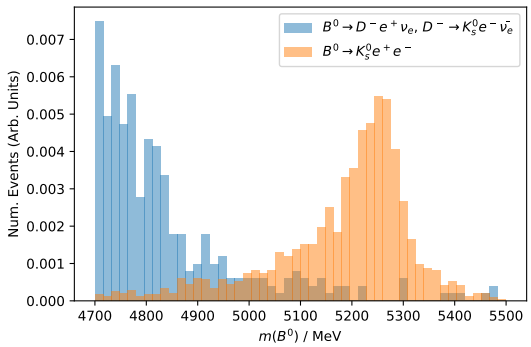
5.13 Residual open charm cascade decays

As outlined in Section 4.4.1.4, cascade processes involving $D^- \to K^0_S e^- \overline{\nu}_e$ decays may contaminate the $B^0 \to K^0_S e^+ e^-$ signal region. Given the nominal mass-fit model for this channel does not parameterise these cascade decays, such contamination would bias the fitted yields of the signal and various other sources of background. In order to assess these biases, the distributions of cascade decays in $m(B^0)$ must be determined, accounting for any sculpting effects due to the kinematic vetoes for such decays. The resulting bias to the $B^0 \to K^0_S e^+ e^-$ signal yield is then taken as a systematic uncertainty.

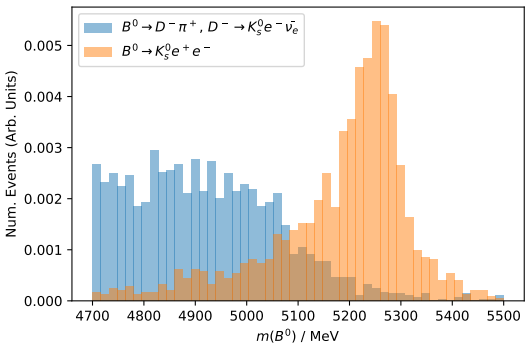
As shown in Fig. 5.5, $D^- \to K^0_S e^- \overline{\nu}_e$ cascade background candidates in REDECAY MC samples, passing preselection, lie at fairly low values of $m(B^0)$. However, very few such MC candidates pass the cascade kinematic vetoes, meaning they cannot be used to infer the $m(B^0)$ distributions for fully-selected decays. Instead, samples of $D^- \to K^0_S e^- \overline{\nu}_e$ decays were produced using RAPIDSIM (described in Section 2.2.4.1). Cascade decays which survive the kinematic veto of $m\left(K^0_S, e^\pm\right) > m\left(D^-\right)$ most likely do so due to the recovery of false bremsstrahlung photons, so it is crucial to include such an effect in the particle momentum modelling. However, RAPIDSIM's treatment of electron momenta appears to only model bremsstrahlung emission - not recovery.

To deal with this, a fully-selected sample of $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ MC candidates, from nominal LHCb simulation, is used to calculate a set of alternative smearing parameters. For each electron candidate in this sample, a smearing factor $\Delta_p \equiv p_{Reco} - p_{True}$ is calculated. Then, each electron in the RAPIDSIM sample is randomly assigned one of these Δ_p values. Both of the samples are first divided into ten bins of p_{True}, to account for correlations between Δ_p and p_{True} seen in the $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ sample. The distribution of $m(B^0)$ for a RAPIDSIM sample of $B^0 \to K^0_S e^+ e^-$ candidates is shown in Fig. 5.6, with both the default electron momentum smearing, and the custom approach outlined here. A prominent upper-mass tail can be seen when this custom smearing is used, which corresponds to the recovery of false bremsstrahlung photons.

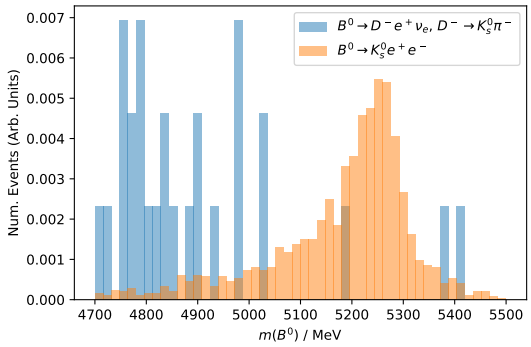
The distributions of $m(B^0)$ in the $D^- \to K^0_S e^- \overline{\nu}_e$ RAPIDSIM samples are shown in Fig. 5.7, both before and after the kinematic cascade vetoes are applied. When these vetoes are applied,



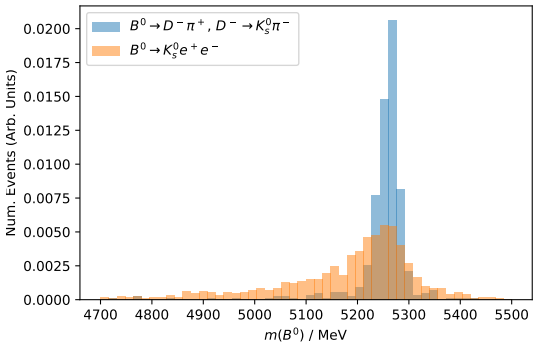
(a): $B^0 \to D^- e^+ \nu_e, D^- \to K^0_S e^- \overline{\nu}_e$



(b): $B^0 \to D^- \pi^+, D^- \to K^0_S e^- \overline{\nu}_e$



(c): $B^0 \to D^- e^+ \nu_e, D^- \to K^0_S \pi^-$



(d): $B^0 \to D^- \pi^+, D^- \to K^0_S \pi^-$

Fig. 5.5: Reconstructed $m(B^0)$ distributions for different cascade background processes, from REDECAY MC candidates passing preselection requirements, and falling within signal mass window. The signal mass distribution, from MC candidates passing full selection, is also shown in each plot.

the average values of $m(B^0)$ are increased. In addition, a broad peaking structure can be seen for $B^0 \to D^- \pi^+$ decays, close to the signal mass of 5280 MeV.

5.13.1 Systematic calculation

Toy mass-fits are used to calculate the systematic uncertainty arising from residual $D^- \to K^0_S e^- \overline{\nu}_e$ contamination. The PDFs for $B^0 \to D^- e^+ \nu_e$ and $B^0 \to D^- \pi^+$ decays are both described using GKDE PDFs, with $\rho = 1.0$, determined using RAPIDSIM samples with custom-smeared electron momenta. Any purely-kinematic requirements in the stripping and preselection are applied to these samples, as are the cascade background vetoes. Then, toy data samples are generated with these cascade background components included, the expected yields of which are taken from the upper-limits to the contamination rates in Table 4.20.

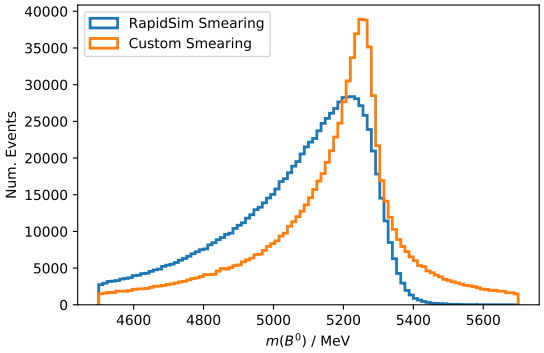
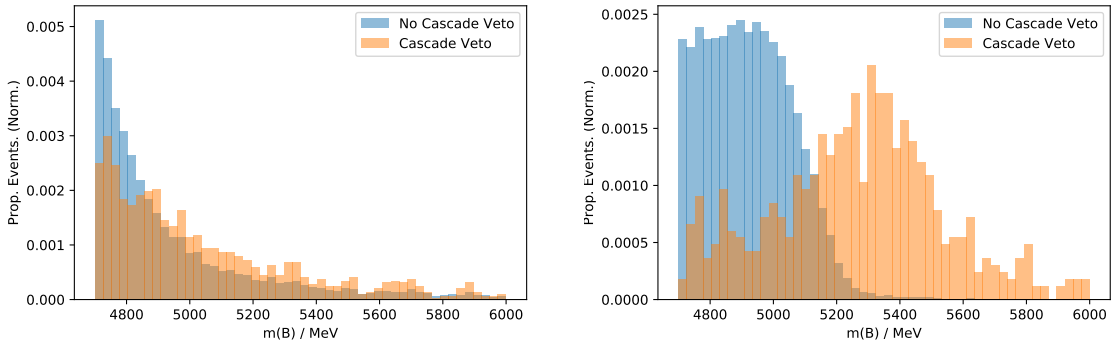


Fig. 5.6: Reconstructed $m(B^0)$ distributions, from a RAPIDSIM MC sample of $B^0 \to K^0_S e^+ e^-$ decays, with the default electron momentum smearing used in RAPIDSIM, and a customised procedure which accounts for bremsstrahlung recovery.



(a): $B^0 \to D^- e^+ \nu_e$, $D^- \to K^0_S e^- \overline{\nu}_e$ ***(b):*** $B^0 \to D^- \pi^+$, $D^- \to K^0_S e^- \overline{\nu}_e$

Fig. 5.7: Reconstructed $m(B^0)$ distributions for $D^- \to K^0_S e^- \overline{\nu}_e$ cascade background processes, from RAPIDSIM MC samples, before and after cascade background veto requirements are imposed. The electron momenta in these samples are smeared using a custom procedure, and are required to pass all kinematic requirements in the stripping and preselection.

These toy data samples are fitted using the nominal fit model, and the alternative fit model with the two $D^- \to K^0_S e^- \overline{\nu}_e$ components included (the yields of which are fixed to their expected values). An ensemble of 1000 toys are generated and fitted using this method, to find a distribution of Δ_N values. As shown in Fig. 5.8, the average value of Δ_N is above zero, indicating a systematic bias to the nominal fitted value of the signal yield. Therefore, the 68.3% upper-quantile of the Δ_N distribution is taken as a systematic uncertainty, with a value of 0.41% relative to the signal yield.

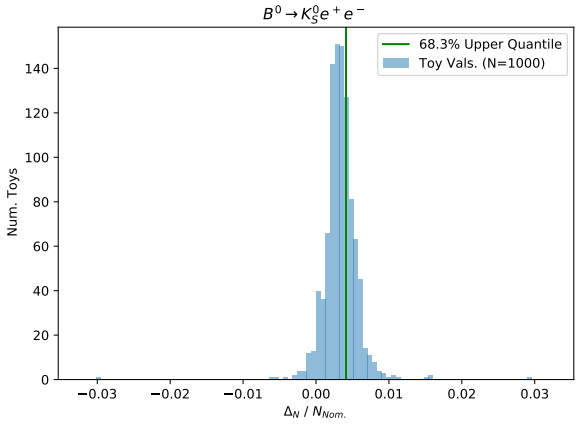


Fig. 5.8: Normalised distribution of Δ_N found from fits to toy datasets with $D^- \to K^0_S e^- \overline{\nu}_e$ background contributions. The 68.3% upper-quantile for this distribution, taken as the systematic uncertainty associated to these residual decays, is shown by the vertical green line.

5.14 Partially-reconstructed background mass shapes

Both the $B \to K^{(*)} e^+ e^-$ mass-fits contain components for partially-reconstructed decays involving higher-order kaon resonances, where the pion from the decay $K^{(*)'} \to K^{(*)}\pi$ is missed during reconstruction. The PDFs for such decays are determined using MC samples. However, the real $m(B)$ distributions in data will depend on the dynamics of the hadronic systems, which are not accurately modelled in these MC samples. Systematic uncertainties are assigned to account for such mismodelling.

5.14.1 Higher-order kaon resonances in the $B^0 \to K^0_S e^+ e^-$ fit

Partially-reconstructed backgrounds to the $B^0 \to K^0_S e^+ e^-$ signal are modelled using MC samples for $B^+ \to K^{*+} e^+ e^-$ decays, which only include the $K^*(892)^+$ resonance in the hadronic system. However, other kaon resonances decaying by $K^* \to K^0_S \pi$ could also contribute, including $K(1270)$, $K(1400)$, and $K^*(1430)$ mesons. Non-resonant $B \to K^0_S \pi e^+ e^-$ decays are also key contributors. The partially-reconstructed $m(B^0)$ values for these processes may differ from those for $B^+ \to K^{*+} e^+ e^-$, leading to a systematic uncertainty on the fitted $B^0 \to K^0_S e^+ e^-$ signal yield.

In order to account for these effects, the generator-level distributions of $m(K^0_S, \pi^+)$ in the $B^+ \to K^{*+} e^+ e^-$ MC samples are reweighted to match the distribution of $m(K^+, \pi^-)$ found in a previous analysis of $B^0 \to K^+ \pi^- \mu^+ \mu^-$ decays by LHCb* [333]. Hereafter, the weights

*By isospin symmetry, the distribution of $m(K^+, \pi^-)$ in $B^0 \to K^+ \pi^- \ell^+ \ell^-$ decays should closely match the distribution of $m(K^0_S, \pi^+)$ in $B^+ \to K^0_S \pi^+ \ell^+ \ell^-$ decays.

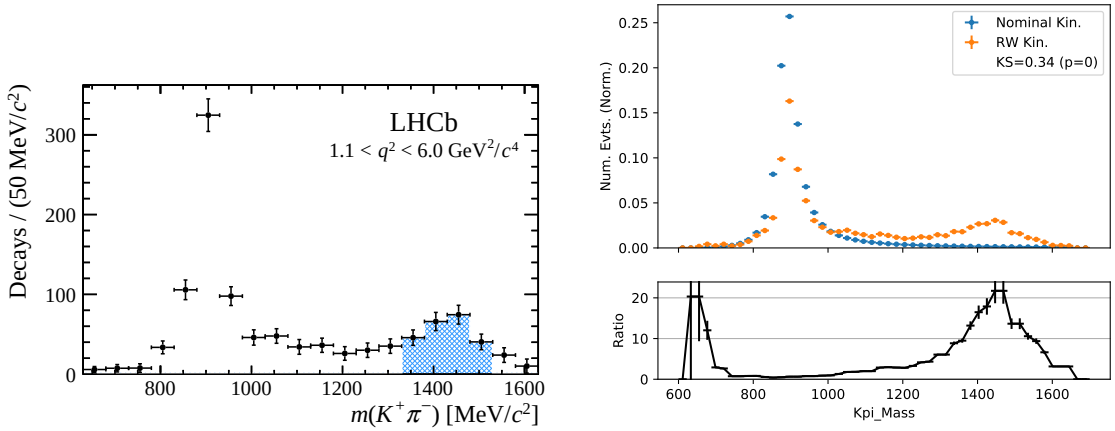


Fig. 5.9: Left - Figure from [333], showing the distribution of $m(K^+,\pi^-)$ in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays as measured by LHCb. Right - The generator-level distribution of $m(K^0_S,\pi^+)$ found in $B^+ \to K^{*+}e^+e^-$ MC from all years before (blue) and after (orange) it is reweighted to match data. The lower portion of this plot shows the ratio of the reweighted and original distributions.

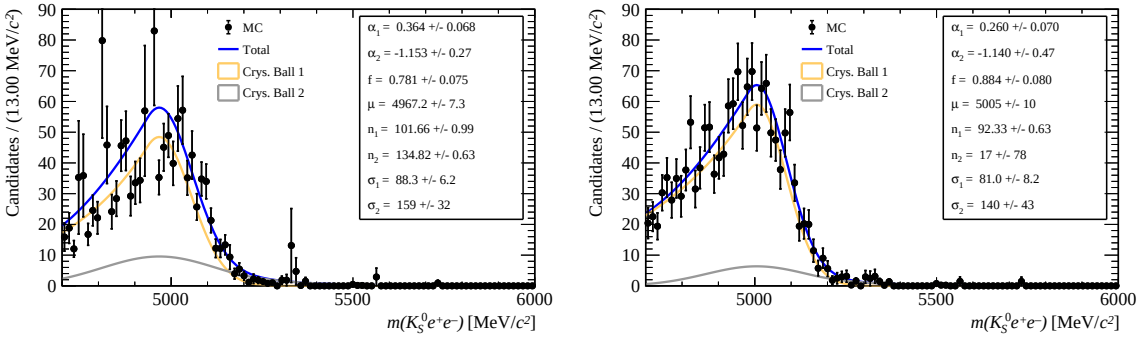


Fig. 5.10: DCB PDFs used to parameterise $m(B^0)$ distributions in partially-reconstructed $B^+ \to K^{*+}e^+e^-$ decays, with (left), and without (right) kaon kinematic weights

which correct the distribution of $m(K^0_S,\pi^+)$ are referred to as "kaon kinematic weights". The distribution of $m(K^+,\pi^-)$ found in that analysis is shown in Fig. 5.9, as is the distribution of $m(K^0_S,\pi^+)$ seen in $B^+ \to K^{*+}e^+e^-$ generator-level MC before and after these weights are applied.

Toy fits are used to assess the systematic uncertainty on the $B^0 \to K^0_S e^+e^-$ signal yield. As shown in Fig. 5.10, DCB PDFs are fitted to the distributions of $m(B^0)$ for partially-reconstructed $B^+ \to K^{*+}e^+e^-$ decays with and without the kaon kinematic weights applied. The former is used to construct an alternative data mass-fit model, while the latter is part of the nominal model outlined in Section 4.5.4.2. An ensemble of 1000 toy data samples is generated from

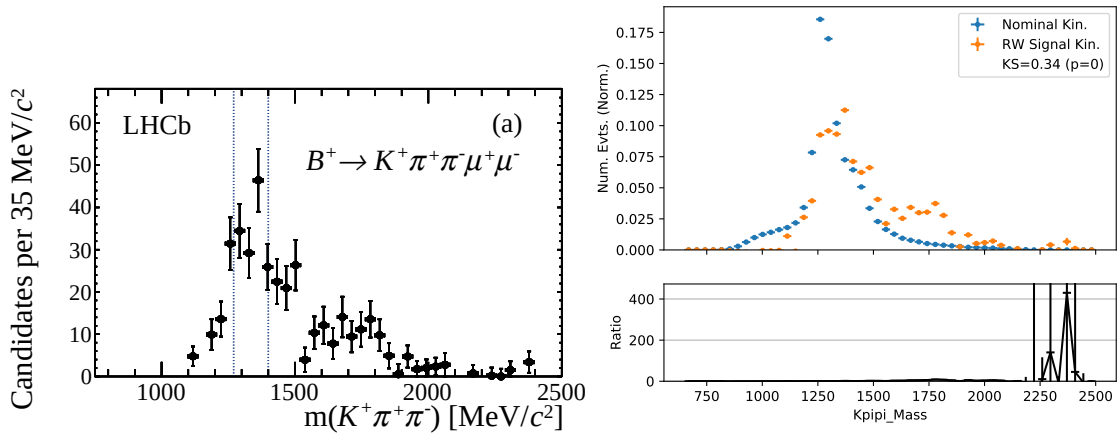


Fig. 5.11: Left - Figure from [334], showing the distribution of $m(K^+,\pi^+,\pi^-)$ in $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decays as measured by LHCb. Right - The generator-level distribution of $m(K^0_S,\pi)$ found in $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC before (blue) and after (orange) it is reweighted to match data. The lower portion of this plot shows the ratio of the reweighted and original distributions.

the alternative fit model, with each being fitted using the nominal and alternative fit models. The distribution of Δ_N has an average value below zero, so the systematic uncertainty on the $B^0 \to K^0_S e^+e^-$ signal yield is taken as the 68.3% lower-quantile of this distribution. This gives a value of 2.2% relative to the signal yield.

5.14.2 Higher-order kaon resonances in the $B^+ \to K^{*+}e^+e^-$ fit

Analogously to above, the $m(B^+)$ distribution for partially-reconstructed $B \to K^0_S\pi^+\pi e^+e^-$ decays is also dependent on the dynamics of the final-state hadrons. As outlined in Section 4.5.4.2, this component is described in the nominal mass-fit model from a sample of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays, containing a mixture of $K_1(1270)$, $K_1(1400)$ and $K^*_1(1430)$ resonances decaying to a $K^0_S\pi^+\pi$ final-state. This mixture may not reflect the true dynamics found in data, where additional contributions from non-resonant $B \to K^0_S\pi^+\pi e^+e^-$ decays will also be found. Note that the differing dilepton invariant masses for the $q^2_{low|central}$ and $q^2_{J/\psi}$ regions also affects the kinematic phase-space available to the hadronic system, causing further discrepancies between the $m(B^+)$ distribution in rare-mode data compared to $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC.

Similarly to the approach taken for PR backgrounds in the $B^0 \to K^0_S e^+e^-$ fit, the generator-level distribution of $m(K^0_S,\pi^+,\pi)$ in the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC sample is reweighted to match the distribution of $m(K^+,\pi^+,\pi^-)$ observed in data during a previous analysis of

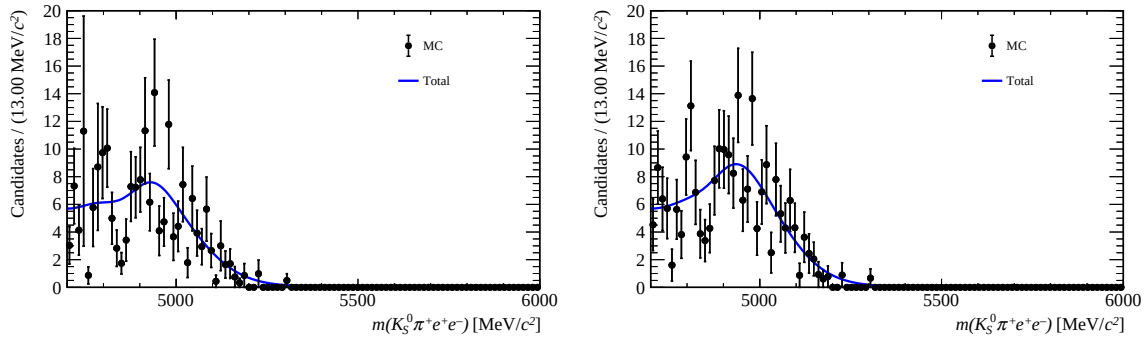


Fig. 5.12: PDFs used to parameterise $m(B^+)$ distributions in partially-reconstructed $B \to K^{(*)}e^+e^-$ decays, with (left), and without (right) kaon kinematic weights.

$B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ decays by LHCb [334]. These distributions are shown in Fig. 5.11. Then, two PDFs are constructed to model the $m(B^+)$ distributions with and without these new kaon kinematic weights applied, as shown in Fig. 5.12. A GKDE PDF description is used, with $\rho = 2.0$ as in the nominal fit. The PDF found with the weights applied is used to construct an alternative data mass-fit model, from which 1000 toy data samples are generated. These are each fitted using the nominal and alternative data mass-fit models, to extract a distribution of Δ_N. Once again, this distribution has an average value below zero, so the 68.3% lower-quantile is assigned as the systematic uncertainty on the $B^+ \to K^{*+}e^+e^-$ signal yield resulting from mismodelling of the partially-reconstructed background kinematics. This gives a value of 1.0% relative to the signal yield.

5.14.3 Use of kernel density estimation

In the $B^+ \to K^{*+}e^+e^-$ mass-fit, the background from partially-reconstructed kaon decays is parameterised using a GKDE rather than any physically-motivated parametric model. A relatively small MC sample is used to describe this background, meaning this GKDE could over-fit or under-fit the $m(B^+)$ distribution, depending on the chosen bandwidth parameter ρ. By default, $\rho = 2.0$ is used, as this appears to give a smoothly-varying PDF without being affected by statistical fluctuations. However, this choice is still somewhat arbitrary. A systematic uncertainty is assigned to the fitted $B^+ \to K^{*+}e^+e^-$ signal yield, to account for the effects of the chosen value of ρ, and of the limited MC sample size for PR background decays.

Toy fits are used for this purpose. First, the sample of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays used to describe the PR background is bootstrapped. Then, a GKDE is constructed using this bootstrapped sample, with a randomly chosen value of $\rho \in [1.0, 2.5]$. A toy dataset is generated with this alternative model for the PR background, and fitted using the alternative and nominal mass-fit models. This process is repeated 500 times, to find a distribution of Δ_N values. The

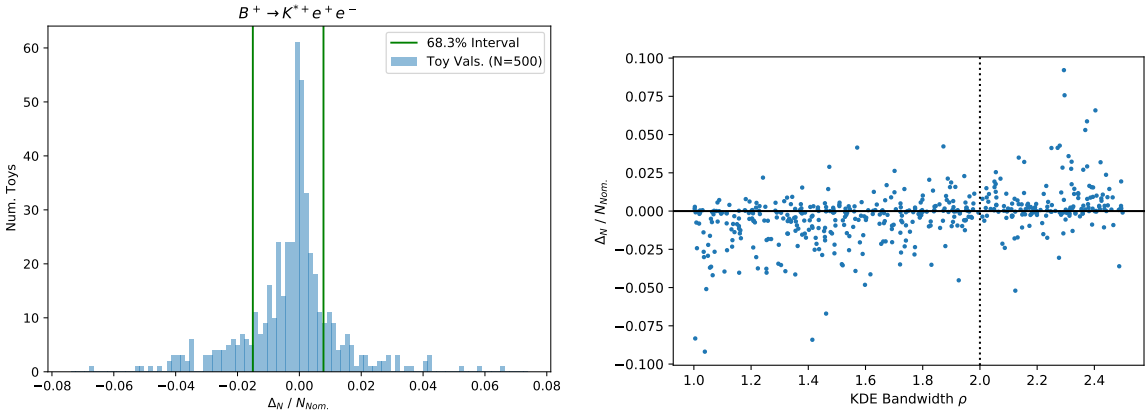


Fig. 5.13: Left - Normalised distribution of Δ_N, found from fits to $B^+ \to K^{*+} e^+ e^-$ toy datasets with variations in the KDE bandwidth ρ, and bootstrapping of the MC samples used to parameterise the partially-reconstructed background. Right - Values of Δ_N found in different toy fits, against the values of ρ used for each toy. The vertical dotted line indicates the nominal bandwidth value of $\rho = 2.0$.

resulting distribution is shown in Fig. 5.13, together with the relationship between ρ and Δ_N. The half-width of the inner 68% quantile for Δ_N is taken as a systematic uncertainty, with a value of 1.1% relative to the $B^+ \to K^{*+} e^+ e^-$ signal yield.

5.15 J/ψ leakage mass shapes

The $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ leakage components in the $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ mass-fits to data are also both described using GKDE PDFs. By default, a bandwidth of $\rho = 1.5$ is chosen, which appears to give smoothly-varying PDFs, but as above there is a risk of under-fitting or over-fitting due to the limited numbers of $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC candidates surviving the rare-mode selection requirements.

A similar method to above is used to find the systematic uncertainties associated to this possible mismodelling of the leakage components. In each channel the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ MC sample is bootstrapped, and a random value of $\rho \in [1.0, 2.5]$ is used when constructing a new GKDE PDF. A toy dataset is generated using this PDF, and fitted using the nominal and alternative mass-fit models to find a value of Δ_N. For each $B \to K^{(*)} e^+ e^-$ channel, 500 such toy fits are carried out, and the half-widths of the inner 68.3% quantiles for Δ_N are taken as systematic uncertainties. These have values of 0.094% and 1.1% in the $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ channels, respectively, relative to the signal yields.

5.16 Summary of systematic uncertainties

Values of the various systematic uncertainties on the R_X^{-1} ratios are shown in Tables 5.1 and 5.2, for each data-taking year. Tables of systematic uncertainties on the individual branching fractions, and the cross-checks, are given in Appendix A. All tables also show the value of each systematic uncertainty for all years combined, found with averages weighted by the recorded luminosities and $b\bar{b}$ cross-sections for each year* [286]. The total systematic uncertainties in each year, and for all years combined, are found by summing the individual systematics from different sources in quadrature. Note that the fit-model systematics are only evaluated across all years, rather than in individual years. Hence, their values are only quoted for all years combined.

There are also some systematic uncertainties resulting directly from the finite MC statistics when calculating selection efficiencies and f_{q^2} values, which are found using the normal approximation (see Section 3.2). In contrast to other systematic uncertainties, these can be calculated trivially when examining sub-sets of the MC, so are generally quoted as part of the "statistical" uncertainties for the cross-checks in Chapter 6. Their values are quoted in the last lines of the tables in this section. Finally, other sources of systematic uncertainty are incorporated as Gaussian constraints in the mass-fit likelihoods, such as the relative proportions of different Bremsstrahlung categories in signal MC, and the yields of misidentified $B \to K^{(*)}\pi^+\pi^-$ decays. The impacts of each of these constraints on the fitted signal yields are difficult to unfold individually, but they are considered when calculating the total systematic uncertainties on R_X^{-1}, the electron-mode branching fractions, and the cross-checks $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$.

*In practice, this average is carried out using the UNCERTAINTIES software package in PYTHON, which also accounts for any correlations between systematic uncertainties in different years [328].

	$\sigma\left(R^{-1}\left(K^0_S\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.43	0.38	0.25	0.56	0.6	0.24
PIDCalib Bin.	0.3	0.21	1.5	0.76	0.23	0.4
Elec. PID Fact.	0.64	0.25	1	0.59	1.4	0.82
Elec. Trk. Stats.	0.21	0.21	0.07	0.077	0.058	0.047
Muon Trk.	0.048	0.11	0.21	0.058	0.014	0.054
Gen. Stats.	3	2.8	1.1	1.2	0.7	0.67
KS Stats.	0.0088	0.013	0.0057	0.029	0.053	0.018
L0 Stats.	0.26	0.34	0.23	0.18	0.066	0.091
L0I Tag	0.57	0.55	0.13	0.2	0.14	0.12
L0 Param.	1.8	0.89	0.19	0.75	0.054	0.27
BDT Mismodelling	0.97	1.9	0.33	1.2	0.55	0.45
Reco. Weight Stats.	2.4	4.7	1.1	2.2	0.8	1
q2 Smear. Stats.	0.49	0.63	0.39	0.33	0.24	0.18
Decay Model	0.26	0.35	0.43	0.1	0.15	0.24
Cascade Back.	-	-	-	-	-	0.41
Part. Reco. $m(K^{(*)})$	-	-	-	-	-	2.2
Leakage KDE	-	-	-	-	-	0.094
Sig. PDF MC Stats.	-	-	-	-	-	1.5
Total of above	4.5	6	2.6	3.1	2	3.2
Sig. MC Stats	8.5	9.1	5.7	5.4	3.9	2.6

Table 5.1: Summary of systematic uncertainties on $R^{-1}_{K^0_S}$.

	$\sigma\left(R^{-1}\left(K^{*+}\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.8	0.36	0.29	0.63	0.23	0.19
PIDCalib Bin.	0.24	0.41	0.76	0.45	0.3	0.23
Elec. PID Fact.	0.26	0.23	1.4	0.36	0.76	0.57
Elec. Trk. Stats.	0.22	0.17	0.084	0.082	0.056	0.045
Muon Trk.	0.16	0.077	0.25	0.08	0.069	0.068
Gen. Stats.	3.8	5.1	5.3	2.3	1.3	1.6
KS Stats.	0.61	0.22	0.08	0.016	0.036	0.061
L0 Stats.	0.32	0.53	0.44	0.45	0.18	0.18
L0I Tag	0.72	0.94	0.17	0.29	0.16	0.18
L0 Param.	0.36	0.19	0.19	0.15	0.33	0.12
BDT Mismodelling	0.091	0.53	0.85	0.85	0.31	0.31
Reco. Weight Stats.	1.3	1.9	2.1	1.7	0.94	0.77
q2 Smear. Stats.	0.42	0.73	0.42	0.42	0.16	0.18
Decay Model	1.1	0.33	1.5	0.69	0.84	0.81
Part. Reco. $m(K^{(*)})$	-	-	-	-	-	1
Part. Reco. KDE	-	-	-	-	-	1.1
Leakage KDE	-	-	-	-	-	1.1
Sig. PDF MC Stats.	-	-	-	-	-	2.1
Total of above	4.4	5.7	6.2	3.3	2.1	3.6
Sig. MC Stats	7.7	9.2	7.2	6	3.9	2.8

Table 5.2: Summary of systematic uncertainties on $R^{-1}_{K^{*+}}$.

Chapter 6

Cross-checks

A series of cross-checks were performed before the measurements of R_X^{-1} were unblinded, to check whether selection efficiencies are well estimated. These consist of measurements of the single and double ratios $r_{J/\psi}^{-1}$ and $R_{\psi(2S)}^{-1}$, followed by blinded measurements of the $B \to K^{(*)}\mu^+\mu^-$ differential branching fractions. After unblinding, the $B \to K^{(*)}e^+e^-$ datasets were examined for any excesses which could be caused by larger contamination from $B \to K^{(*)}\pi^+\pi^-$ decays than anticipated. All these cross-checks are documented in this section.

6.1 $r_{J/\psi}^{-1}$

As discussed in Chapter 3, the $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ decays used as control modes in this analysis are lepton-universal. The ratio of muon-mode and electron-mode branching fractions should be equal to unity, and can be measured using a ratio of efficiency-adjusted yields:

$$r_{J/\psi}^{-1} = \frac{N_{Sel}\left(B \to XJ/\psi\,(e^+e^-)\right)}{N_{Sel}\left(B \to XJ/\psi\,(\mu^+\mu^-)\right)} \times \frac{\varepsilon\left(B \to XJ/\psi\,(\mu^+\mu^-)\right)}{\varepsilon\left(B \to XJ/\psi\,(e^+e^-)\right)} \tag{6.1}$$

If a measured value of $r_{J/\psi}^{-1}$ deviates from unity, this would indicate that the estimates of efficiencies or yields are biased for muon-mode decays, electron-mode decays, or both. Note also that any systematic biases on the estimates of the muon-mode or electron-mode efficiencies are partially cancelled in the double-ratio used to measure R_X^{-1}. However, this is not the case for $r_{J/\psi}^{-1}$, making its measurement a very stringent cross-check.

In addition, the value of $r_{J/\psi}^{-1}$ should be equal to unity in all sub-sets of data, for example datasets recorded in different years or with different L0 trigger lines. If the measured values of $r_{J/\psi}^{-1}$ vary between different sub-sets of data, this could indicate which aspects of the efficiency calculation, if any, are poorly understood.

6.1.1 Differential measurements

A key aspect of this analysis is ensuring the selection efficiencies are well described as a function of numerous continuous variables, describing the kinematic and geometric behaviour of particles in the decay chain. For many such variables, the underlying distributions differ between rare-mode and control-mode decays (for example, the momentum of a final-state lepton depends on the candidate q^2 value). If the selection efficiency is not well described as a function of this variable, it could lead to a systematic bias on R_X^{-1} and the rare-mode differential branching fractions.

To check this, $r_{J/\psi}^{-1}$ is measured as a function of sixteen different variables. These studies are performed separately for the $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$ and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ decay-modes. For each variable, a binning scheme is constructed with five bins which are approximately equally populated by MC. Fully-selected $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ and $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ data, recorded across all years, is then split according to this binning scheme. Within each bin, the control-mode yields are extracted using the mass-fit models outlined in Section 4.5.2.1 and 4.5.2.3, and the corresponding selection efficiencies are calculated using fully-corrected MC, as in Section 4.3. The values of $r_{J/\psi}^{-1}$ are then calculated using eq. 6.1, and plotted. The values of $r_{J/\psi}^{-1}$, and the efficiency-adjusted yields for each J/ψ control-mode, should be consistent across different bins of a given variable. If not, this could indicate that its efficiency profile is mismodelled.

One can quantify the size of potential systematic biases resulting from this mismodelling. First, the rare-mode yields in each bin, for a given lepton, can be estimated using the control-mode yield and relative efficiencies (*i.e.* as $N_\ell^{Rare} \propto \varepsilon_\ell^{Rare} \cdot N_\ell^{J/\psi} / \varepsilon_\ell^{J/\psi}$). By summing these estimates of the rare-mode yields, and the measured control-mode yields, in each bin i, a "flatness parameter" is computed [335, Sec. 8.1.2]:

$$d_f^\ell = \left(\frac{\sum_i \varepsilon_\ell^{Rare,i} \cdot \frac{N_\ell^{J/\psi,i}}{\varepsilon_\ell^{J/\psi,i}}}{\sum_i \varepsilon_\ell^{J/\psi,i} \cdot \frac{N_\ell^{J/\psi,i}}{\varepsilon_\ell^{J/\psi,i}}} \cdot \frac{\sum_i \varepsilon_\ell^{J/\psi,i}}{\sum_i \varepsilon_\ell^{Rare,i}} \right) - 1 \tag{6.2}$$

If one assumes that any trend in efficiency-adjusted yields for a given variable is due to genuine mismodelling of its efficiency profile, d_f^ℓ gives the value of the systematic bias induced on a measurement of the differential branching fraction for $B \to K^{(*)}\ell^+\ell^-$. By using a double-ratio of yield estimates, one can also compute a flatness parameter showing the potential bias on R_X^{-1}:

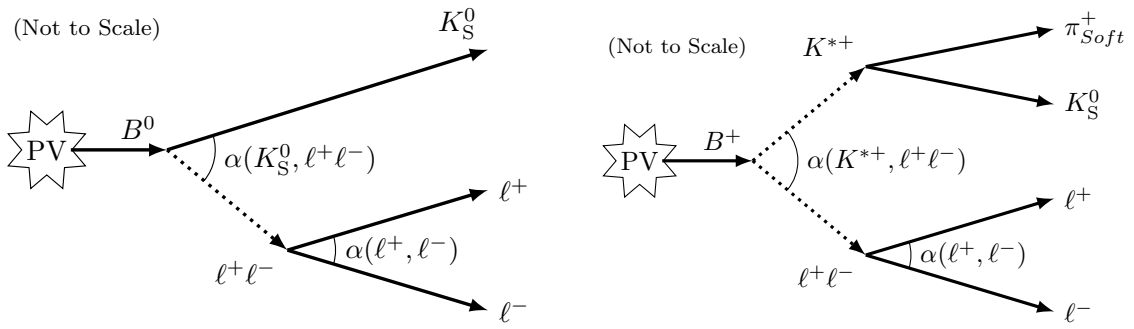


Fig. 6.1: Diagrams showing the definitions of the angles $\alpha(\ell^+, \ell^-)$ and $\alpha(\ell^+\ell^-, K)$ for $B^0 \to K_S^0 \ell^+\ell^-$ decays (left) and $B^+ \to K^{*+}\ell^+\ell^-$ decays (right), according to the lab-frame momenta of final-state particles. Adapted from [335, Fig. 8.4].

$$d_f = \left(\frac{\sum_i \varepsilon_\mu^{Rare,i} \cdot \frac{N_\mu^{J/\psi,i}}{\varepsilon_\mu^{J/\psi,i}}}{\sum_i \varepsilon_\mu^{J/\psi,i} \cdot \frac{N_\mu^{J/\psi,i}}{\varepsilon_\mu^{J/\psi,i}}} \cdot \frac{\sum_i \varepsilon_\mu^{J/\psi,i}}{\sum_i \varepsilon_\mu^{Rare,i}} \Bigg/ \frac{\sum_i \varepsilon_e^{Rare,i} \cdot \frac{N_e^{J/\psi,i}}{\varepsilon_e^{J/\psi,i}}}{\sum_i \varepsilon_e^{J/\psi,i} \cdot \frac{N_e^{J/\psi,i}}{\varepsilon_e^{J/\psi,i}}} \cdot \frac{\sum_i \varepsilon_e^{J/\psi,i}}{\sum_i \varepsilon_e^{Rare,i}} \right) - 1 \tag{6.3}$$

6.1.1.1 Choice of variables

These differential measurements of $r_{J/\psi}^{-1}$ are carried out in bins of sixteen different variables, describing the kinematic and geometric properties of particles in the decay chain, including:

- $\alpha(\ell^+, \ell^-)$ - The opening angle of the dilepton pair in the laboratory frame, as illustrated in Fig. 6.1.
- $\alpha(K, \ell^+\ell^-)$ - The opening angle of the dilepton pair and the kaon (the K_S^0 or K^{*+} in the B^0 and B^+ modes, respectively) in the laboratory frame. This is also illustrated in Fig. 6.1.
- $\max(p_T(\ell^+), p_T(\ell^-))$, $\min(p_T(\ell^+), p_T(\ell^-))$ - The maximum and minimum of the lepton transverse momenta.
- $p(B)$, $p_T(B)$ - The reconstructed momentum and transverse momentum of the B-meson.
- $\eta(B)$, $\eta(J/\psi)$, $\eta(K)$ - The reconstructed pseudorapidities of the B-meson, dilepton pair, and kaon (the K_S^0 or K^{*+} in the B^0 and B^+ modes, respectively).
- $\max(\eta(\ell^+), \eta(\ell^-))$, $\min(\eta(\ell^+), \eta(\ell^-))$ - The maximum and minimum of the lepton pseudorapidities.

- $\phi(B)$, $\phi(J/\psi)$, $\phi(K)$ - The reconstructed azimuthal angles of the B-meson, dilepton pair, and kaon (the K^0_S or K^{*+} in the B^0 and B^+ modes, respectively).

- $z_{VTX}(K^0_S)$ - The displacement of the K^0_S decay vertex, along the z-axis. This is important, as discrepancies have been previously seen between the K^0_S reconstruction efficiencies in data and MC, the sizes of which depend on the K^0_S decay position [315].

Assuming selection efficiencies are uniform as a function of azimuthal angle ϕ, one can show that the dynamics of a $b \to s\ell^+\ell^-$ decay (and equivalent control-modes) can be described using four variables; $\alpha(\ell^+, \ell^-)$, $\alpha(K, \ell^+\ell^-)$, and the momenta of the two leptons [336, Sec. 10.3]. Rare-mode and control-mode candidates are differently distributed in these variables, meaning any trends in $r^{-1}_{J/\psi}$ across them would be a strong indication of a non-cancelling bias on R^{-1}_X.

In the latest R_{K^+} measurement by LHCb, $r^{-1}_{J/\psi}$ was measured in sets of two-dimensional bins across combinations of these variables, to account for correlations between them [147]. However, it would be better in principle to carry out a check across all four variables simultaneously, as this would provide maximal separation between the rare-modes and control-modes, and hence the best sensitivity to any biases on R^{-1}_X. Using four-dimensional bins would be impractical for this purpose, due to the limited sizes of the $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ control-mode samples.

Instead, taking inspiration from a cross-check in an angular analysis of $B^+ \to K^+e^+e^-$ decays at LHCb, BDTs are trained to separate rare-mode and control-mode decays [337, Sec. 4.5.3]. Then, a differential study of $r^{-1}_{J/\psi}$ is carried out in bins of the BDT output scores (denoted BDT$_{Shell}$). Four such BDTs are trained; one for each lepton-mode and each B-meson. These classifiers are trained using fully-selected samples of rare-mode and control-mode MC from all years as training samples, with $\alpha(\ell^+, \ell^-)$, $\alpha(K, \ell^+\ell^-)$, $\min(p_T(\ell^+), p_T(\ell^-))$, and $\max(p_T(\ell^+), p_T(\ell^-))$ used as training features. These variables are denoted "shell variables" hereafter. The CatBoost boosting algorithm is used, as this was found to give good separation between rare-mode and control-mode decays with default hyper-parameter values [338, 339]. In order to make differential plots of $r^{-1}_{J/\psi}$ more legible, the BDT output scores are then "flattened" by measuring the cumulative distribution function of the control-mode MC sample, then using a rank transformation to map the score to a uniform distribution across BDT$_{Shell} \in [0, 1]$. This is performed using the `QuantileTransformer` class from the scikit-learn machine-learning software package in Python [340, 341]. To prevent efficiency biases from potential overtraining, the BDTs are trained in five k-folds. Candidates in data, and candidates in MC which are not used in training, are assigned BDT output scores from randomly-chosen folds.

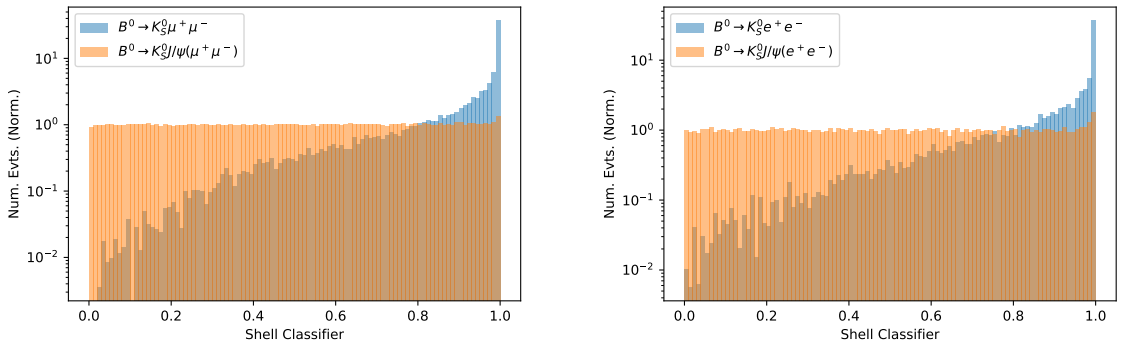


Fig. 6.2: Output distributions of BDT$_{Shell}$, as found in fully-selected $B^0 \to K^0_S \mu^+ \mu^-$ (left) and $B^0 \to K^0_S e^+ e^-$ (right) MC samples, plus the corresponding J/ψ control-mode decays.

The distributions of BDT$_{Shell}$ found in fully-selected $B^0 \to K^0_S \ell^+ \ell^-$ and $B^0 \to J/\psi\,(\ell^+ \ell^-)\, K^0_S$ MC are shown in Fig. 6.2.

6.1.1.2 Results

The results of the $r_{J/\psi}^{-1}$ measurements in bins of the shell variables and BDT$_{Shell}$ are shown in Fig. 6.3 and 6.4, for $B^0 \to J/\psi\,(\ell^+ \ell^-)\, K^0_S$ and $B^+ \to J/\psi\,(\ell^+ \ell^-)\, K^{*+}$ decays, respectively. Measurements in bins of the other variables are given in Appendix B, while measurements of efficiency-adjusted yields for the muon and electron modes are given in Appendix C. Values of $r_{J/\psi}^{-1}$ in each bin are calculated with two efficiency estimates; one using the nominal method outlined in Section 4.3 with all MC corrections applied (denoted *All Corrections* in the plots), and the other without any corrections applied (denoted *No Corrections*). Signal PDFs used in the mass-fits are also parameterised using MC with and without corrections applied. Similar comparisons of $r_{J/\psi}^{-1}$ are presented throughout this section.

The flatness parameters associated to differential fits are mostly below $1-2\%$, and are generally reduced when MC correction weights are applied. There are some variables where the flatness parameters are higher, notably BDT$_{Shell}$ for $B^0 \to J/\psi\,(\ell^+ \ell^-)\, K^0_S$ decays, with $d_f = 4.5 \pm 1.8\%$. One should note, however, that systematic uncertainties are not accounted for in the calculation of flatness parameters. Moreover, all flatness parameters are well below the measured statistical uncertainties on $R_{K^0_S}^{-1}$ and $R_{K^{*+}}^{-1}$ of $\sim 20-25\%$ (see Section 7.2), and the uncertainties of $25-30\%$ anticipated from toy fits under SM assumptions, prior to unblinding (see Section 7.1.3).

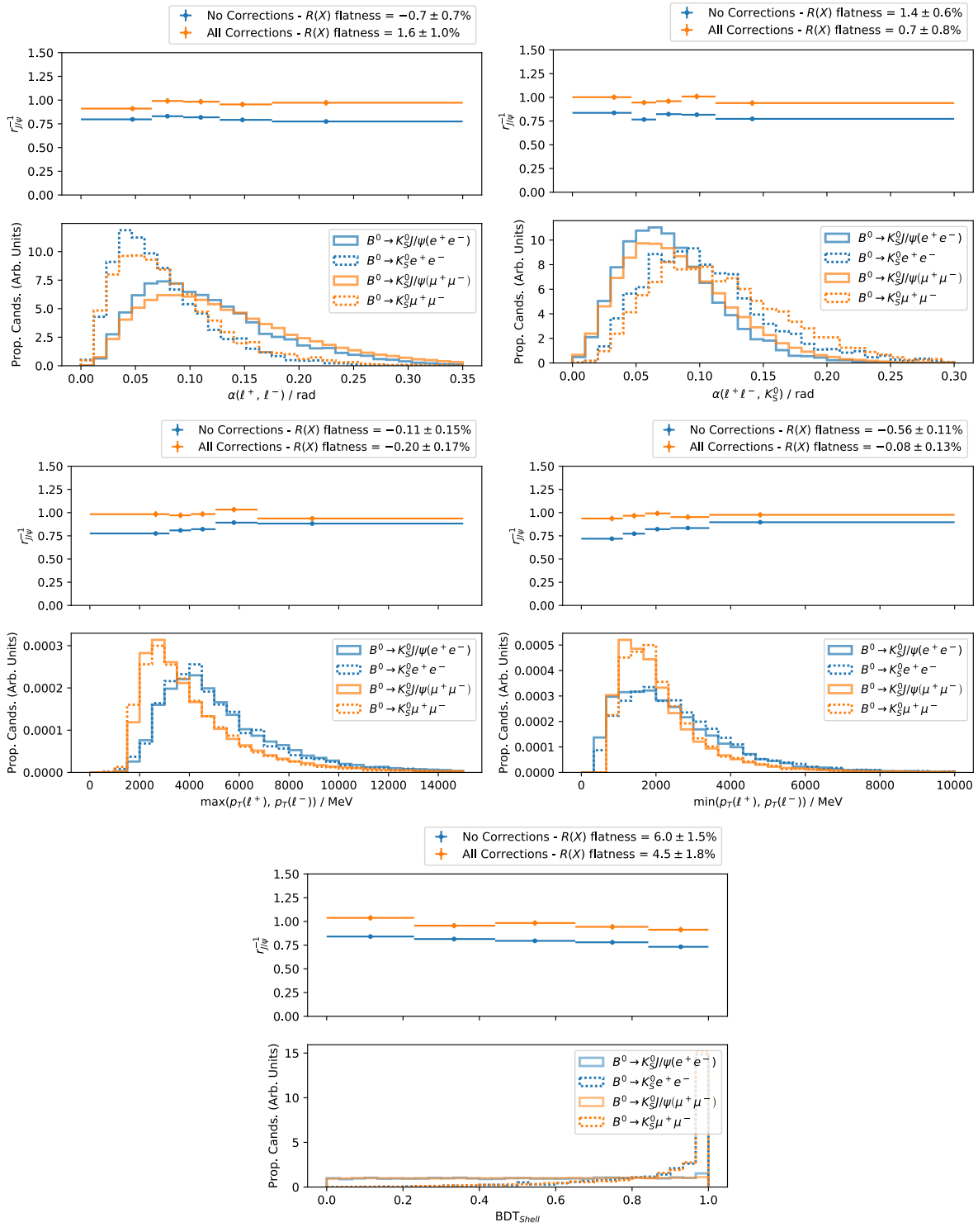


Fig. 6.3: Differential measurements of $r^{-1}_{J/\psi}$ across different variables, found from fits to $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$ data recorded across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots.

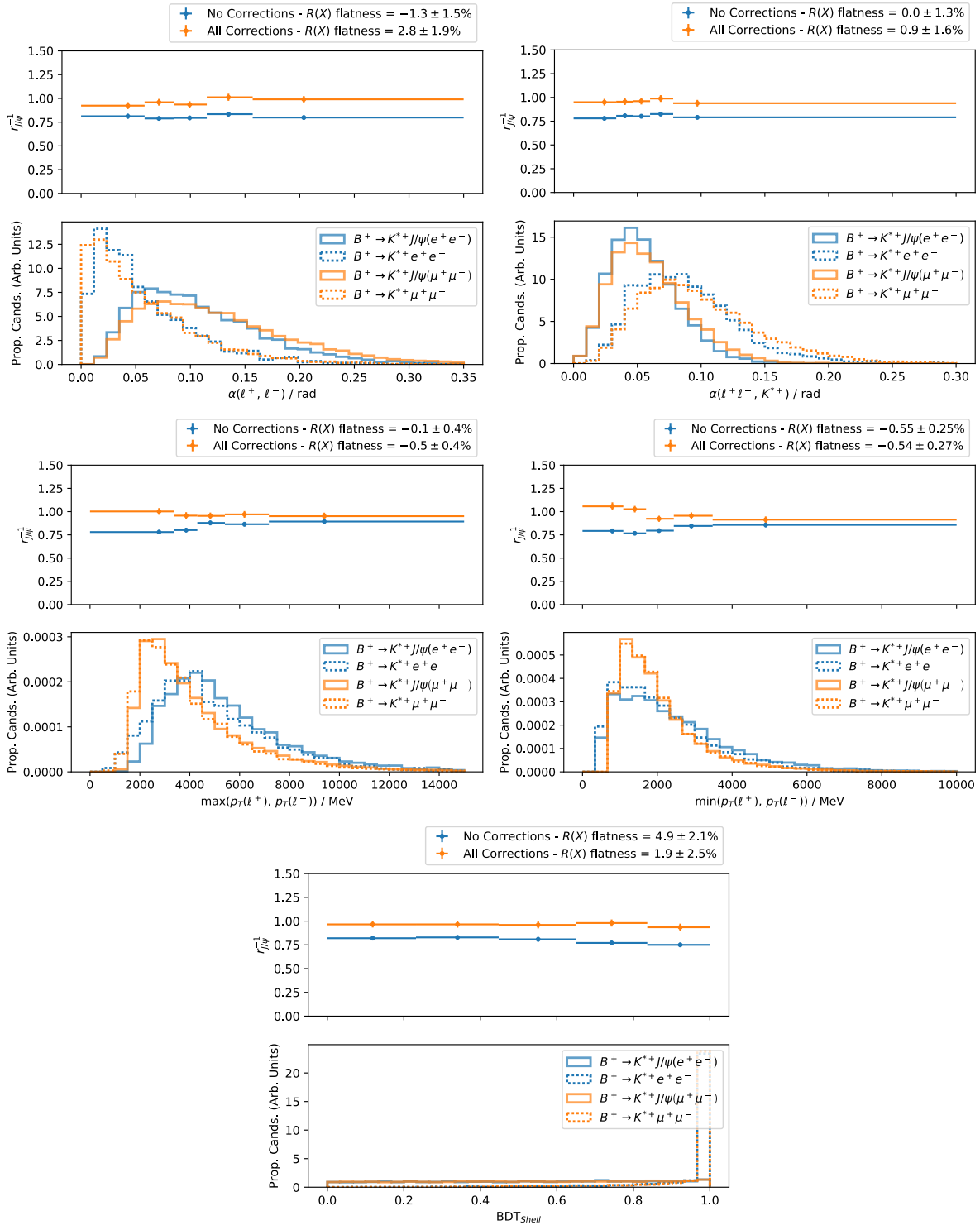


Fig. 6.4: Differential measurements of $r^{-1}_{J/\psi}$ across different variables, found from fits to $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ data recorded across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots.

Decay mode	Year	L0 Cat. L0E	L0E	L0I
$B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$	2011	2%	6.7%	17%
	2012	2.4%	6.8%	9.9%
	2016	0.82%	2.3%	3.3%
	2017	0.96%	3.2%	4.9%
	2018	1%	3.9%	4.1%
$B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$	2011	2.1%	7.3%	10%
	2012	2.3%	7.6%	8.6%
	2016	0.83%	4.4%	3.1%
	2017	0.95%	3.4%	4.1%
	2018	1%	2.8%	2.9%

Table 6.1: Summary of systematic uncertainties on $r^{-1}_{J/\psi}$ related to the L0 correction weights in each trigger category, calculated according to Section 5.7, and used when assessing the compatibility of $r^{-1}_{J/\psi}$ measurements in different L0 categories.

6.1.2 Measurements in trigger categories

The consistency of $r^{-1}_{J/\psi}$ is examined between the three different electron-mode L0 trigger categories (L0E, L0H, and L0I), for each data-taking year. This is accomplished using a χ^2 test. A covariance matrix V is calculated, which describes the total uncertainties - statistical and systematic - on each measurement of $r^{-1}_{J/\psi}$ (*i.e.* in each year and L0 category). Together with a set of measured $r^{-1}_{J/\psi}$ values (denoted $\vec{x}$), and a set of average $r^{-1}_{J/\psi}$ values for each year and B-meson (denoted $\vec{\mu}$), a χ^2 value is calculated:

$$\chi^2 = (\vec{x}-\vec{\mu})^T V^{-1} (\vec{x}-\vec{\mu}) \tag{6.4}$$

The average values $\vec{\mu}$ are then found by numerically minimising the value of χ^2, using a minimisation algorithm supplied in the SciPy software library [342]. By using independent average values μ for each year and B-meson, only the compatibility of the different trigger categories is assessed (the compatibility of different data-taking years is assessed later in this section).

Only the systematic uncertainties related to the L0 correction weights (see Section 5.7), plus the statistical uncertainties on selection efficiencies calculated using the normal approximation, are considered in this study. These uncertainties are calculated independently using fully-selected MC samples in each L0 category. All other systematic uncertainties are assumed to

	Stat Only		With Systs.	
	$\chi^2/d.o.f.$	$p(\chi^2, d.o.f.)$	$\chi^2/d.o.f.$	$p(\chi^2, d.o.f.)$
No Corrections	2.722	0.000	-	-
All Corrections	1.088	0.354	0.848	0.656

Table 6.2: Compatibility tests of measurements of $r_{J/\psi}^{-1}$ in different trigger categories, in different years and B-meson decays, with and without the parameterisation of systematic uncertainties. the compatibilities of measurements carried out with and without the use of correction weights are shown.

	Stat Only		With Systs.	
	$\chi^2/d.o.f.$	$p(\chi^2, d.o.f.)$	$\chi^2/d.o.f.$	$p(\chi^2, d.o.f.)$
No Corrections	3	0.0173	-	-
All Corrections	0.432	0.786	0.412	0.8

Table 6.3: Compatibility tests of measurements of $r_{J/\psi}^{-1}$ in different trigger categories, for different B-meson decays, using mass-fits to data sets recorded across all years, with and without the parameterisation of systematic uncertainties. the compatibilities of measurements carried out with and without the use of correction weights are shown.

correlate perfectly between different L0 categories, for each given year and B-meson. The total systematic uncertainties on $r_{J/\psi}^{-1}$ for each year and L0 category are shown in Table 6.1.

The values of $r_{J/\psi}^{-1}$ in each L0 category are shown in Fig. 6.5, while the minimised values of χ^2 (normalised by numbers of degrees of freedom, and with corresponding p-values) are shown in Table 6.2. The number of degrees of freedom is taken as the total number of $r_{J/\psi}^{-1}$ measurements $\vec{x}$, minus the number of average values $\vec{\mu}$. In addition, this check is carried out without systematic uncertainties included in V (*i.e.* only considering the statistical uncertainties on the yields and efficiency values), and using $r_{J/\psi}^{-1}$ values found without any MC corrections applied when calculating efficiencies.

Finally, all the checks are repeated using values of $r_{J/\psi}^{-1}$ found from fits to data recorded in all years combined. The values of $r_{J/\psi}^{-1}$ are shown in the bottom rows of Fig. 6.5, and the corresponding χ^2 values in Table 6.3.

Very poor compatibility is seen between the different L0 categories when corrections are not applied to MC. However, a very good level of compatibility is seen when the corrections are applied, even when systematic uncertainties are not accounted for. This suggests that the L0

		All Corrections / %	No Corrections / %
2011	B^0	−0.041	0.22
	B^+	−1.1	−1.3
2012	B^0	−0.24	0.78
	B^+	−2.4	1.5
2016	B^0	−0.0039	0.045
	B^+	0.98	0.25
2017	B^0	0.045	−0.35
	B^+	0.3	0.035
2018	B^0	−0.31	−0.7
	B^+	1.3	−0.33
Total	B^0	−0.078	−0.16
	B^+	0.28	−0.2

Table 6.4: Flatness parameters found using measurements of $r^{-1}_{J/\psi}$ in different trigger categories, for different B-mesons, using mass-fits to data recorded across various different years. The flatness parameter values are presented both with and without the use of correction weights.

correction weights outlined in Section 4.2.5 are working as intended, and reducing any biases on the estimates of electron-mode trigger efficiencies to negligible levels.

6.1.2.1 Flatness measurement

The measured $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ yields in each L0 category, and the corresponding selection efficiencies, are used to calculate flatness parameters according to eq. 6.3. This check is carried out for each B-meson and data-taking year, plus all years combined. The obtained flatness parameter values are shown in Table 6.4. When all years are combined, the flatness parameters are $\lesssim 0.3\%$ when MC corrections are applied. The largest flatness parameter in individual years is seen for $B^+ \to K^{*+}e^+e^-$ in 2012, at 2.4%. These flatness parameters are of comparable size to the systematic uncertainties on R^{-1}_X related to the L0 correction weights, and negligible compared to the statistical uncertainties on the $B \to K^{(*)}e^+e^-$ differential branching fractions and R^{-1}_X ratios. The flatness parameters are similarly small even without any MC corrections applied, illustrating the stability offered by using $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ decays as a control-mode.

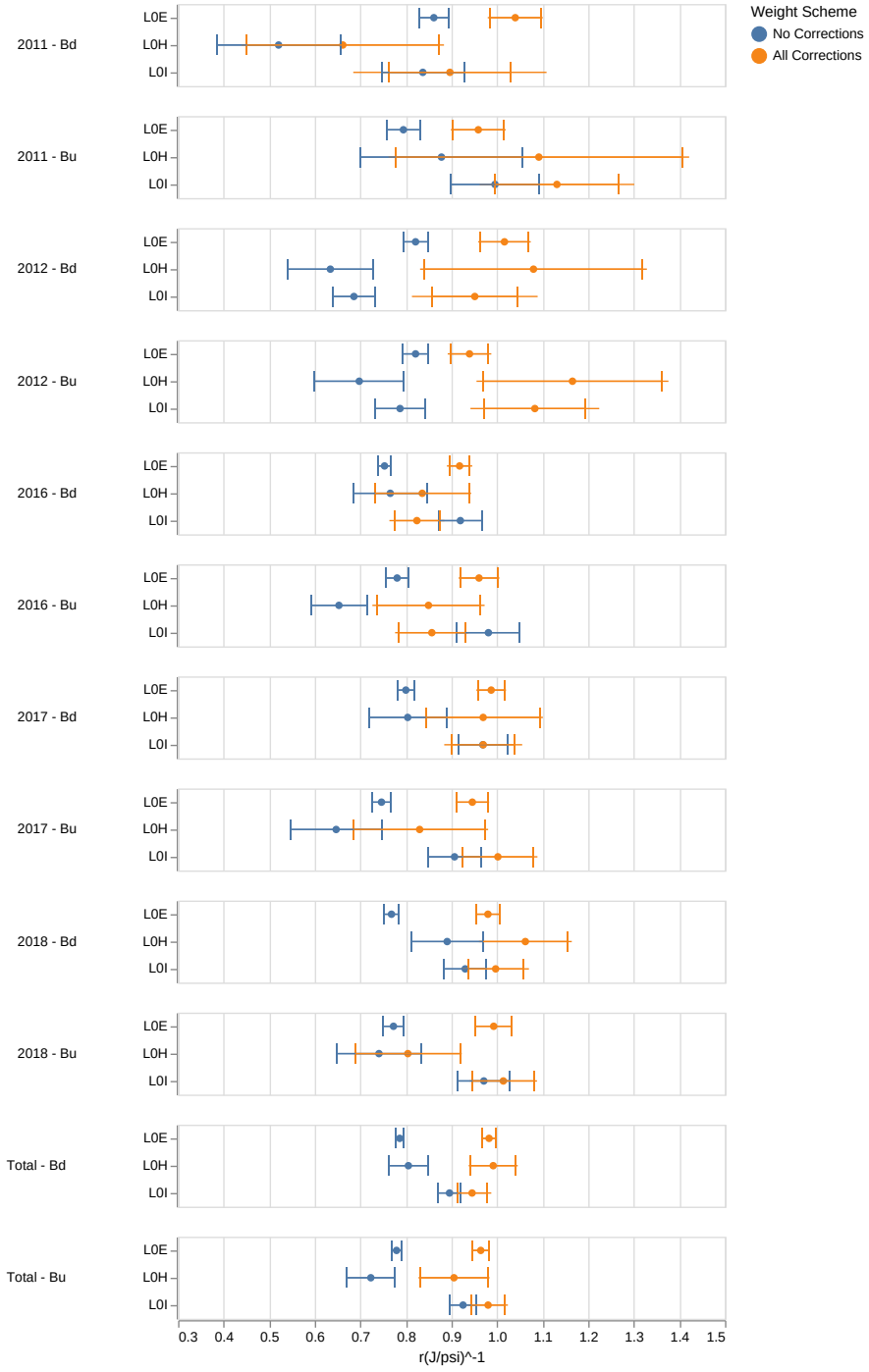


Fig. 6.5: Values of $r_{J/\psi}^{-1}$ measured for each trigger category, for each year of data-taking, found from $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$, and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ decays. The values found both with and without the use of correction weights are shown. Values found from fits to data from all years combined are also shown. The inner, capped error bars show statistical uncertainties, while the outer, uncapped error bars show the total statistical and systematic uncertainties.

Fig. 6.6: Values of $r_{J/\psi}^{-1}$ measured for each K_S^0-meson reconstruction category, for each year of data-taking, found from $B^0 \to J/\psi\,(\ell^+\ell^-)\,K_S^0$, and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ decays. The values found both with and without the use of correction weights are shown. Values found from fits to data from all years combined are also shown.

6.1.3 Measurements in K_S^0 reconstruction categories

In addition, $r_{J/\psi}^{-1}$ is measured in each K_S^0-meson reconstruction category, and compared. These measurements are carried out for each B-meson and data-taking year, plus all years combined, resulting in the values shown in Fig. 6.6. The associated flatness parameters are also calculated, and are given in Table 6.5. While poor compatibility between the DownDown and

		All Corrections / %	No Corrections / %
2011	B^0	0.11	0.47
	B^+	0.93	−0.041
2012	B^0	−0.024	−0.89
	B^+	0.11	0.15
2016	B^0	−0.16	0.37
	B^+	0.093	0.016
2017	B^0	0.4	0.22
	B^+	−0.06	−0.11
2018	B^0	−0.54	0.061
	B^+	0.028	−0.23
Total	B^0	0.05	0.14
	B^+	0.052	−0.068

Table 6.5: Flatness parameters found using measurements of $r_{J/\psi}^{-1}$ in different K_S^0-meson reconstruction categories, for different B-mesons, using mass-fits to data recorded across various different years. The flatness parameter values are presented both with and without the use of correction weights.

LongLong categories can be seen in some individual data-taking years, even when MC correction weights are applied, the values of $r_{J/\psi}^{-1}$ found from fits to all years combined are in good agreement. In addition, all flatness parameters are at the $\mathcal{O}(0.1\%)$ level or below, indicating that any discrepancies in $r_{J/\psi}^{-1}$ will not lead to significant deviations in the corresponding R_X^{-1} measurements.

6.1.4 Measurements in different years

Values of $r_{J/\psi}^{-1}$ are measured for each B-meson, in each data-taking year, resulting in the values shown in Fig. 6.7. Their compatibility is then assessed using a χ^2 test, similar to the previous test of the compatibility of different trigger categories. In this case, a single average value of $r_{J/\psi}^{-1}$ is used across all years and B-hadrons, when minimising the χ^2 function given by eq. 6.4. In addition to statistical uncertainties on the yields and efficiency estimates, all the efficiency-related systematic uncertainties outlined in Chapter 5 are included when calculating χ^2 values. The values of these systematic uncertainties are shown in Tables A.5 and A.6, for the B^0 and B^+ modes, respectively. Note that systematic uncertainties related to the mass-fit model are not considered in this test, as they have not been evaluated for individual years.

The values of χ^2 found from this test, and corresponding p-values, are given in Table 6.6. In addition to the parameterisation of systematic uncertainties outlined above, this compatibility

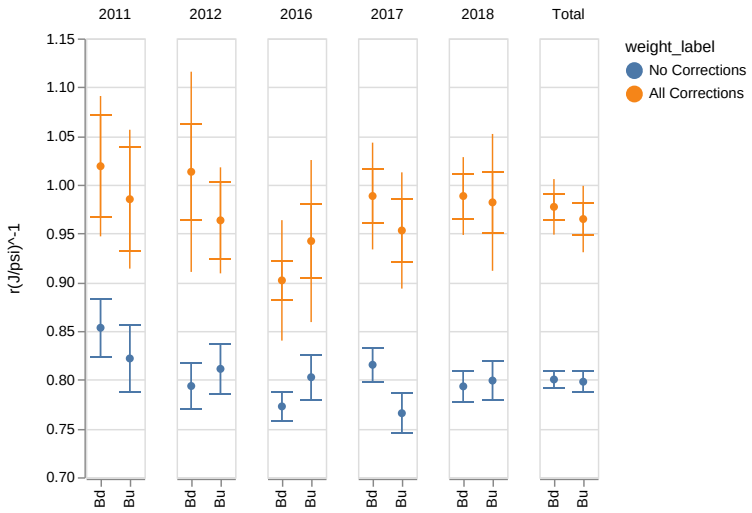


Fig. 6.7: Values of $r^{-1}_{J/\psi}$ measured for each year of data-taking, found from $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$, and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ decays. The values found both with and without the use of correction weights are shown. Values found from fits to data from all years combined are also shown. The inner, capped error bars show statistical uncertainties, while the outer, uncapped error bars show the total statistical and systematic uncertainties.

	Stat Only		With Systs.	
	$\chi^2/d.o.f.$	$p(\chi^2, d.o.f.)$	$\chi^2/d.o.f.$	$p(\chi^2, d.o.f.)$
No Corrections	1.192	0.295	-	-
All Corrections	1.577	0.116	0.267	0.983

Table 6.6: Compatibility tests of measurements of $r^{-1}_{J/\psi}$ in different years and B-meson decays, with and without the parameterisation of systematic uncertainties. the compatibilities of measurements carried out with and without the use of correction weights are shown.

test is carried out using statistical uncertainties alone, and also without the use of MC correction weights. A reasonable level of compatibility is seen both with and without the use of MC correction weights, with very good agreement when systematic uncertainties are accounted for.

6.1.5 Integrated value across years

Finally, $r^{-1}_{J/\psi}$ is measured using fits to data from all years combined, for each decay-mode, resulting in the values given in Table 6.7. The statistical uncertainties on the yields and efficiency estimates are shown here, as are total systematic uncertainties found according to Chapter 5. Very poor compatibility with unity is seen when MC correction weights are not

		$\mu(r^{-1}_{J/\psi})$	$\sigma(Stat.)$	$\sigma(Syst.)$	$\sigma(Tot.)$
No Corrections	B^0	0.800	0.008	0.000	0.008
	B^+	0.798	0.011	0.000	0.011
All Corrections	B^0	0.977	0.013	0.025	0.028
	B^+	0.965	0.017	0.030	0.034

Table 6.7: Values of $r^{-1}_{J/\psi}$ found from fits to combined datasets across all years. Note that systematic uncertainties have not been calculated when MC corrections are not applied, so are not quoted.

	$\mu(r^{-1}_{J/\psi})$	$\sigma(Stat.)$	$\sigma(Syst.)$	$\sigma(Tot.)$
B^0	0.977	0.008	0.027	0.028
B^+	0.965	0.011	0.032	0.034

Table 6.8: Values of $r^{-1}_{J/\psi}$ found from a fits to combined datasets across all years, with all corrections applied to MC when calculating efficiencies. In this table, the "true" statistical uncertainty is quoted, via mass-fits with all Gaussian constraints fixed to their central values, in line with the strategy used to separate statistical and systematic uncertainties on R^{-1}_X. The statistical uncertainties on the efficiencies are here included as part of $\sigma(Syst.)$

applied, with values of $r^{-1}_{J/\psi} \approx 0.8$. However, good compatibility with unity is seen when MC weights are applied and systematic uncertainties are accounted for. This suggests that the MC weights lead to accurate estimates of the relative efficiencies in the muon and electron modes, and that any residual systematic effects have been well captured by the studies in Chapter 5.

In addition, the $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ mass-fits are carried out with the relative fractions of signal in different Bremsstrahlung categories fixed to their central values, where they would normally be Gaussian-constrained. This allows the "true" statistical uncertainties on $r^{-1}_{J/\psi}$ to be unfolded from all systematic effects. The resulting values and uncertainties on $r^{-1}_{J/\psi}$ are shown in Table 6.8. Note that here, the statistical uncertainties on the efficiency estimates are included as part of $\sigma(Syst.)$. This form of uncertainties is used when quoting the $r^{-1}_{J/\psi}$ results in the analysis paper [5].

6.2 $R^{-1}_{\psi(2S)}$

In addition to the J/ψ resonance, a second control-mode decay is considered in this analysis - $B \to \psi(2S)\,(\ell^+\ell^-)\,K^{(*)}$. The relative branching fractions of the electron

	Stat Only		With Systs.	
	$\chi^2/d.o.f.$	$p(\chi^2,d.o.f.)$	$\chi^2/d.o.f.$	$p(\chi^2,d.o.f.)$
No Corrections	0.461	0.901	-	-
All Corrections	0.796	0.62	0.743	0.67

Table 6.9: Compatibility tests of measurements of $R^{-1}_{\psi(2S)}$ in different years and B-meson decays, with and without the parameterisation of systematic uncertainties. the compatibilities of measurements carried out with and without the use of correction weights are shown.

and muon modes for these decays have not been measured to great precision, with $\mathcal{B}(\psi(2S)\to e^+e^-)/\mathcal{B}(\psi(2S)\to \mu^+\mu^-) = 0.991 \pm 0.077$ [13]. However, $\psi(2S)\to \ell^+\ell^-$ decays are expected to be lepton-universal under the SM. In addition, as they proceed via tree-level processes at leading order, they are expected to remain lepton-universal even in NP scenarios. Therefore, the ratio of branching fractions is expected to be unity, and can be used as a cross-check. Analogously to R_X^{-1}, this is measured as a double-ratio normalised by the J/ψ-mode branching fractions:

$$R^{-1}_{\psi(2S)} = \frac{N_{Sel}\left(B\to X\psi(2S)\left(e^+e^-\right)\right)}{N_{Sel}\left(B\to XJ/\psi\left(e^+e^-\right)\right)}\frac{N_{Sel}\left(B\to XJ/\psi\left(\mu^+\mu^-\right)\right)}{N_{Sel}\left(B\to X\psi(2S)\left(\mu^+\mu^-\right)\right)}\times \\ \frac{\varepsilon\left(B\to XJ/\psi\left(e^+e^-\right)\right)}{\varepsilon\left(B\to X\psi(2S)\left(e^+e^-\right)\right)}\frac{\varepsilon\left(B\to X\psi(2S)\left(\mu^+\mu^-\right)\right)}{\varepsilon\left(B\to XJ/\psi\left(\mu^+\mu^-\right)\right)} \tag{6.5}$$

Note that many systematic biases are expected to partially cancel in this ratio, providing a more realistic view of any residual systematics on R_X^{-1} compared to the $r^{-1}_{J/\psi}$ cross-check. However, any deviations of $R^{-1}_{\psi(2S)}$ from unity may be difficult to resolve, due to the low branching fractions of the $\psi(2S)$ modes compared to the J/ψ modes.

The yields of the $B\to \psi(2S)\left(\ell^+\ell^-\right)K^{(*)}$ decays are estimated using the mass-fits documented in Section 4.5.3, with efficiencies calculated according to Section 4.3. Separate values of $R^{-1}_{\psi(2S)}$ are calculated for the B^0 and B^+ decay-modes.

6.2.1 Comparison of years

Measurements of $R^{-1}_{\psi(2S)}$ are carried out using data from each year, and also from all years combined. The values found with and without the use of MC correction weights when calculating efficiencies and parameterising signal PDFs are shown in Fig. 6.8.

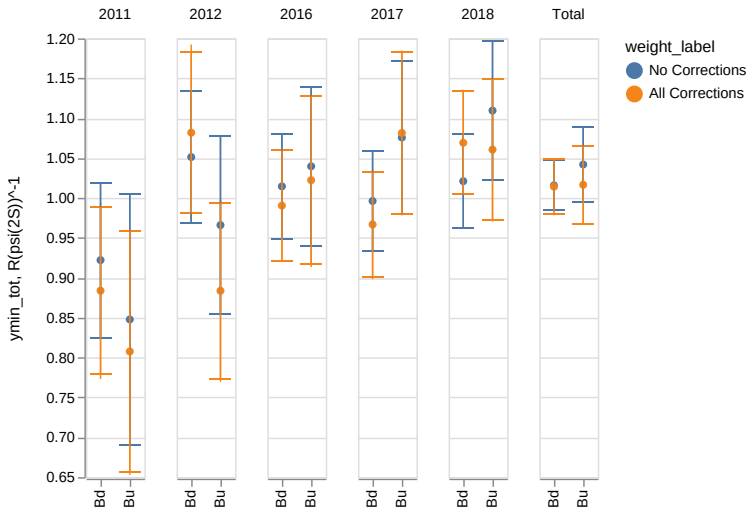


Fig. 6.8: Values of $R^{-1}_{\psi(2S)}$ measured for each year of data-taking, found from $B^0 \to \psi(2S)\,(\ell^+\ell^-)\,K^0_S$, and $B^+ \to \psi(2S)\,(\ell^+\ell^-)\,K^{*+}$ decays. The values found both with and without the use of correction weights are shown. Values found from fits to data from all years combined are also shown. The inner, capped error bars show statistical uncertainties, while the outer, uncapped error bars show the total statistical and systematic uncertainties.

As for $r^{-1}_{J/\psi}$ in Section 6.1.4, a χ^2 test is carried out to assess the compatibility of $R^{-1}_{\psi(2S)}$ measurements in each year, and for each B-meson. The systematic uncertainties given in Tables A.7 and A.8 are combined with the statistical uncertainties on the estimated yields and efficiencies to construct a covariance matrix. This is used to calculate a χ^2 value according to eq. 6.4, which is then minimised. The values of χ^2 found with and without the use of MC correction weights and systematic uncertainties are shown in Table 6.9, together with the corresponding p-values. Very good compatibility is seen between the values of $R^{-1}_{\psi(2S)}$ in different years, both with and without the use of MC correction weights and inclusion of systematic uncertainties.

6.2.2 Integrated value across years

The values of $R^{-1}_{\psi(2S)}$ found from mass-fits to all years combined are shown in Table 6.10, along with statistical and systematic uncertainties. These measurements show good compatibility with unity, both with and without the use of MC correction weights, once again illustrating the strength of the double-ratio approach.

As for $r^{-1}_{J/\psi}$, the $B \to \psi(2S)\,(\ell^+\ell^-)\,K^{(*)}$ mass-fits are also carried out with all parameters which would normally be Gaussian-constrained instead fixed to their central values, to un-

		$\mu(R^{-1}_{\psi(2S)})$	$\sigma(Stat.)$	$\sigma(Syst.)$	$\sigma(Tot.)$
No Corrections	B^0	1.016	0.031	0.000	0.031
	B^+	1.042	0.047	0.000	0.047
All Corrections	B^0	1.014	0.034	0.011	0.036
	B^+	1.017	0.049	0.012	0.050

Table 6.10: Values of $R^{-1}_{\psi(2S)}$ found from fits to combined datasets across all years. Note that systematic uncertainties have not been calculated when MC corrections are not applied, so are not quoted.

	$\mu(R^{-1}_{\psi(2S)})$	$\sigma(Stat.)$	$\sigma(Syst.)$	$\sigma(Tot.)$
B^0	1.014	0.030	0.020	0.036
B^+	1.017	0.045	0.023	0.050

Table 6.11: Values of $R^{-1}_{\psi(2S)}$ found from a fits to combined datasets across all years, with all corrections applied to MC. In this table, the "true" statistical uncertainty is quoted, via mass-fits with all Gaussian constraints fixed to their central values. The statistical uncertainties on the efficiencies are here included as part of $\sigma(Syst.)$

fold the "true" statistical uncertainties from all systematic effects. The resulting values and uncertainties for $R^{-1}_{\psi(2S)}$ are shown in Table 6.11, and are quoted in the analysis paper [5].

6.3 Muon-mode branching fractions

While the $B \to K^{(*)}\mu^+\mu^-$ differential branching fractions are not reported in this analysis, they are measured as cross-checks, the failure of which could indicate some non-cancelling systematic bias on R_X^{-1} which has not been accounted for. First, the consistency of differential branching fraction measurements across different data-taking years is tested. Then, the differential branching fractions are compared to the values previously measured by LHCb, using run-1 data [100]. In both of these tests, the absolute values of the differential branching fraction values are kept blind, to avoid biasing potential future measurements of them - or related observables such as isospin asymmetries - by LHCb.

The differential branching fractions are measured according to eq. 3.2, with the yields and efficiencies estimated according to Sections 4.5.4.1 and 4.3, respectively. In order to directly extract measurements of the branching fractions, the signal yield in each mass-fit is reparameterised in terms of the $B \to K^{(*)}\mu^+\mu^-$ differential branching fraction,

$B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ yield and branching fraction, and the relative efficiencies. The efficiencies and $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ yield form a single "normalisation factor" for each fit channel, which is Gaussian-constrained.

6.3.1 Compatibility between years

The compatibility between different years is assessed using a hypothesis test, where two different fit models are used. In each fit model, the rare-mode and J/ψ-mode data is split between run-1 and each year of run-2. Then, separate fits are carried out to the J/ψ-mode data in each dataset, and corresponding J/ψ-mode and rare-mode efficiencies are calculated. These, along with the systematic uncertainties outlined in Tables A.3 and A.4, are used to calculate normalisation factors for each dataset.

Following this, two sets of simultaneous fits are carried out to the rare-mode data falling within each dataset. Correlations between normalisation factors in each dataset, arising from correlated systematic uncertainties and the $B \to J/\psi\,(\mu^+\mu^-)\,K^{(*)}$ branching fraction, are accounted for by using multivariate Gaussian constraints in the likelihood [291, p. 33]. In the first fit - taken as the *alternative hypothesis* - separate values of the $B \to K^{(*)}\mu^+\mu^-$ branching fraction are allowed to vary for each dataset (*i.e.* separate branching fractions are obtained for run-1, 2016, 2017, and 2018). In the second - taken as the *null hypothesis* - a single value of $B \to K^{(*)}\mu^+\mu^-$ branching fraction is shared between all datasets.

The difference in log-likelihood (Δ_{LL}) between these two fits is taken as a test-statistic, as according to the *Neyman-Pearson Lemma* this provides the most powerful hypothesis test which is possible for any given significance level [291, p. 50] [343]. If a value of Δ_{LL} observed in data gives a p-value above 1.24% (corresponding to $< 2.5\sigma$ significance), the test is considered to be passed. This threshold was agreed with the analysis' review committee, who also oversaw the unblinding process. In order to calculate a p-value, the distribution of Δ_{LL} expected under the null hypothesis must be determined. This is found using an ensemble of 2000 toy mass-fits, with toys generated under the null hypothesis using the procedures outlined in Section 7.1.3. Note that the expected signal yields used for toy generation were taken from the branching fraction fitted in data, under the null hypothesis.

Following this procedure, the expected and observed values of Δ_{LL} are shown in Fig. 6.9, together with the corresponding p-values. In both the $B^0 \to K^0_S\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ modes, $p > 1.24\%$, meaning the test is considered to be passed.

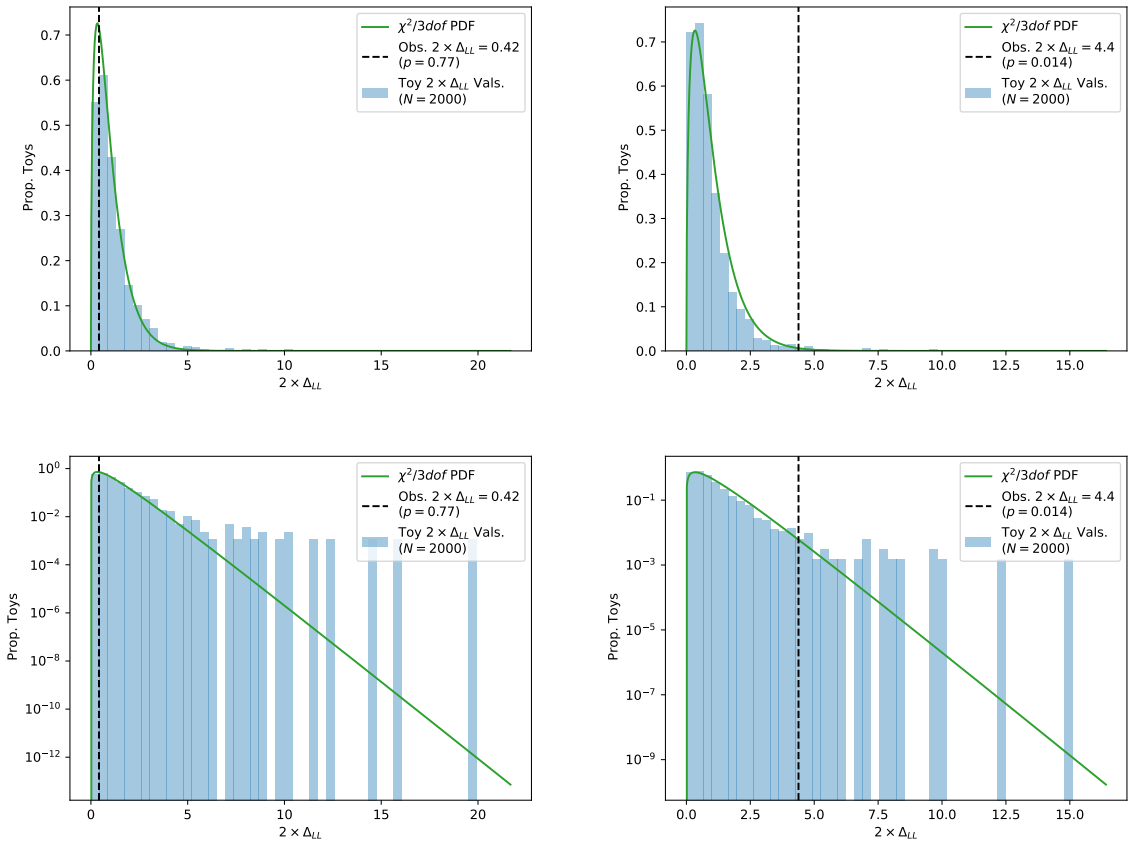


Fig. 6.9: Comparisons of observed $2 \times \Delta_{LL}$ values, and those expected in toys generated from the null hypothesis, used to assess the compatibility of $B^0 \to K^0_S \mu^+ \mu^-$ (left) and $B^+ \to K^{*+} \mu^+ \mu^-$ (right) differential branching fraction measurements in different years. These distributions are shown with a linear (top) and logarithmic (bottom) scale on the y-axis. The observed values of Δ_{LL}, and corresponding p-values, are shown in the legends. A χ^2 distribution with 3 degrees-of-freedom is also shown, though this is not used when computing the p-values.

6.3.2 Consistency with run-1 measurement

The differential branching fractions for $B^0 \to K^0 \mu^+ \mu^-$ and $B^+ \to K^{*+} \mu^+ \mu^-$ decays were previously measured by LHCb, using data recorded in run-1, in a variety of q^2 regions above and below the $m^2(c\bar{c})$ region. Results relevant to this analysis are summarised in Table 6.12. These results are compared to the branching fractions found in this analysis, to further validate the efficiency and yield estimates. In order to minimise the statistical uncertainty in this check, data recorded across run-1 and run-2 are combined and fitted, as in the nominal fits for R_X^{-1}. The test is considered to be passed if these $B \to K^{(*)} \mu^+ \mu^-$ branching fraction measurements are within 2.5σ of the previously-measured values - a threshold agreed with the review committee.

Decay Mode	q^2 range / GeV^2	Diff. Branching Frac. / 10^{-8}			
		Mean	Stat. Err. Low	Stat. Err. High	Syst Err.
$B^0 \to K^0\mu^+\mu^-$	$[1.1, 6.0]$	1.87	0.32	0.35	0.09
$B^+ \to K^{*+}\mu^+\mu^-$	$[0.1, 2.0]$	5.92	1.30	1.44	0.40
	$[2.0, 4.0]$	5.59	1.44	1.59	0.38
	$[4.0, 6.0]$	2.49	0.96	1.10	0.17
	$[0.1, 6.0]$	4.65	0.72	0.80	0.19

Table 6.12: Values for differential branching fractions of $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ decays in relevant bins of q^2, found in the run-1 LHCb isospin analysis [100]. Note that this analysis did not report a value for $B^+ \to K^{*+}\mu^+\mu^-$ in $0.1 < q^2 < 6.0\,\text{GeV}^2$. The value given in this table was found by summing the differential branching fractions in the three separate q^2 bins for this region, then normalising to account for the change in q^2 range.

Note that as both analyses use run-1 data, their measured branching fraction values will be correlated. However, this analysis and the previous analysis use different selection requirements. For example, different stripping lines are used, and different multivariate classifiers are trained to suppress combinatorial background. Due to the differing selection requirements, and the fact that the newly-analysed run-2 data contains ~ 4 times the number of B-meson decays as in run-1, the correlations in branching fraction measurements should be minimal. Furthermore, the ROOT files containing the previous analysis' processed data and MC samples have not been retained, meaning it is not practically possible to accurately estimate these correlations. For these reasons, the measurements are assumed to be uncorrelated.

The previous analysis used a lower q^2 bound of $0.1\,\text{GeV}^2$ when examining $B^+ \to K^{*+}\mu^+\mu^-$ decays, compared to the lower bound of $0.045\,\text{GeV}^2$ used in this analysis. This difference must be accounted for when comparing the two branching fraction measurements. Using FLAVIO, the $B^+ \to K^{*+}\mu^+\mu^-$ differential branching fraction is computed in ranges of $q^2 \in [0.1, 6.0\,\text{GeV}^2]$ and $q^2 \in [0.045, 6.0\,\text{GeV}^2]$, under the SM. Due to the photon pole at low-q^2, the differential branching fraction is found to be 3.9% higher when a lower bound of $0.045\,\text{GeV}^2$ is used. Therefore, the differential branching fraction measured in this analysis is scaled down by 3.9% before comparison with the previous measurement. This scaling factor is conservatively assigned a 3.9% uncertainty, to account for the model-dependence of its predicted value.

The differential branching fractions for $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ decays both fall within 2.5σ of their previously-measured values, meaning this cross-check is considered to

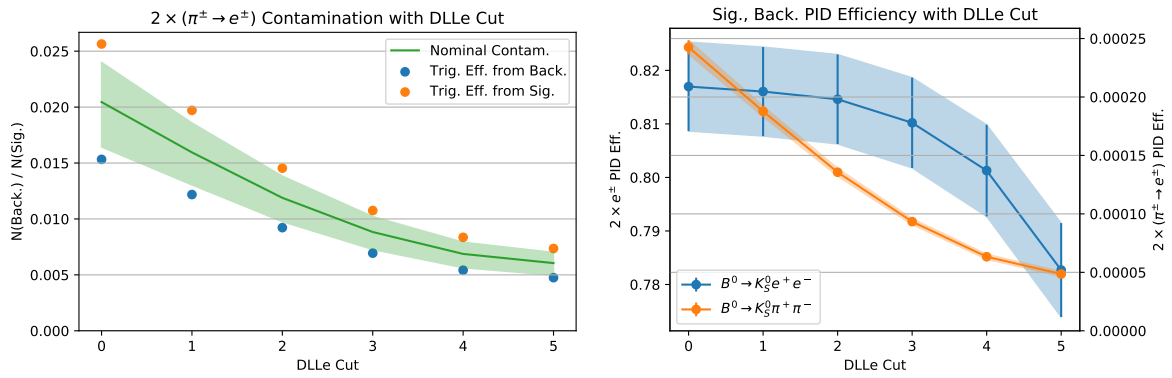


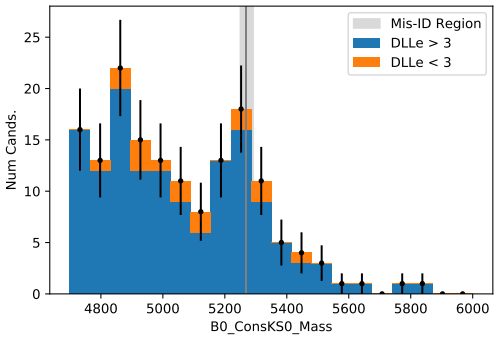
Fig. 6.10: Left - Total expected contamination-rates (relative to expected signal yields) from $B^0 \to K^0_S \pi^+ \pi^-$ background modes when additional cuts to $\text{DLL}_{e\pi}(e^\pm)$ are imposed in run-1, on top of nominal PID requirements. These plots show the expected contamination calculated using the "trigger-unbiased" method outlined in 4.4.1.5 with the trigger efficiency taken from background-mode MC (blue points), signal MC (yellow points), and the average of these, taken as the nominal contamination value (green band). Right - Efficiencies of PID requirements on the two reconstructed electrons, averaged across run-1 and run-2, when additional cuts are imposed on $\text{DLL}_{e\pi}(e^\pm)$ in run-1. The efficiencies are shown for $B^0 \to K^0_S e^+ e^-$ and $B^0 \to K^0_S \pi^+ \pi^-$ decays.

be passed. Note that the *values* of the differences in branching fractions, and the corresponding significances, are kept blind. All that is checked is whether these significances fall below 2.5σ.

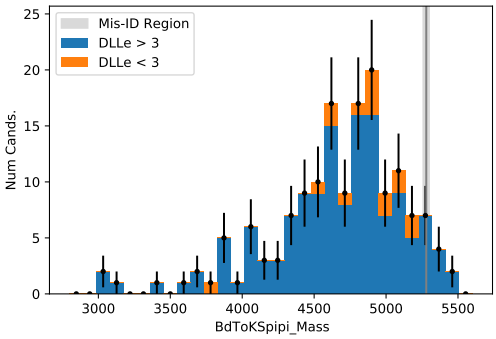
6.4 Checks of misidentified pion background in electron-mode fits

The $B \to K^{(*)} e^+ e^-$ datasets are examined for any excesses which might originate from misidentified $B \to K^{(*)} \pi^+ \pi^-$ decays. If the contamination from such decays is under-estimated, they could artificially increase the measured $B \to K^{(*)} e^+ e^-$ yields, biasing the measured differential branching fractions and R_X^{-1} ratios.

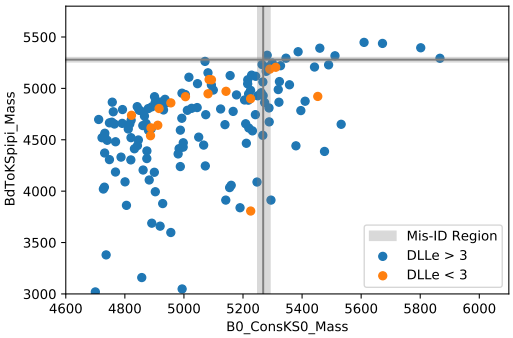
First, an additional PID cut is imposed on the reconstructed electrons, and the mass distributions of the discarded candidates are examined. Note that with the default selection requirements, the $B \to K^{(*)} \pi^+ \pi^-$ contamination rates are ~ 10 times greater in run-1 than in run-2. Therefore, the additional cut is only imposed on run-1 data. To choose a suitable PID cut for this study, the expected $B^0 \to K^0_S \pi^+ \pi^-$ and $B^+ \to K^{*+} \pi^+ \pi^-$ contamination rates were computed with various additional cuts of $\text{DLL}_{e\pi}(e^\pm) > [1,2,3,4,5]$, on top of the nominal PID requirements. At each of these working-points, new sets of PID weights were produced and added to MC samples, and the contamination estimates in Section 4.4.1.5 and 4.4.2.3 were



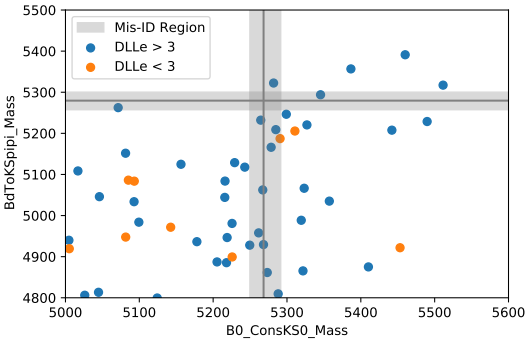
(a): Signal $m(B^0)$



(b): Mis-ID $m(B^0)$



(c): Mis-ID $m(B^0)$ vs. signal $m(B^0)$

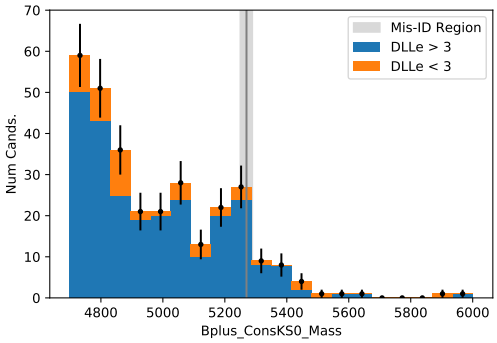


(d): Mis-ID $m(B^0)$ vs. signal $m(B^0)$ (zoomed ranges)

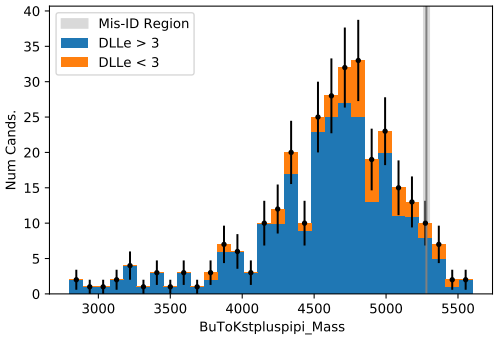
Fig. 6.11: Distributions of signal $m(B^0)$ and mis-ID $m(B^0)$ in $B^0 \to K^0_S e^+ e^-$ data. The events in blue pass additional $\text{DLL}_{e\pi}(e^\pm) > 3$ requirements in run-1, while those in orange fail it. The grey bands indicate the $m(B^0)$ distributions seen in $B^0 \to K^0_S \pi^+ \pi^-$ MC (showing the inner 68.3% quantiles).

repeated. The contamination-rates and selection-efficiencies at each working-point are shown for the $B^0 \to K^0_S e^+ e^-$ mode in Fig. 6.10. A cut of $\text{DLL}_{e\pi}(e^\pm) > 3$ - imposed in run-1 - was chosen for the following study, as this reduces the total $B \to K^{(*)} \pi^+ \pi^-$ contamination across all years by a factor of ~ 2, while reducing the total signal efficiency across all years by only $\sim 1\%$.

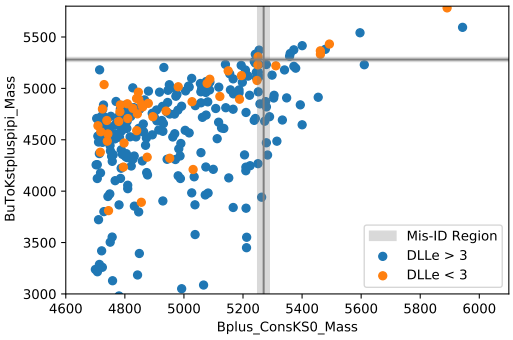
Following this, the nominal "signal" $m(B)$ is plotted for candidates in data which pass or fail the additional electron PID cut. In addition, $m(B)$ is re-computed without bremsstrahlung recovery, and with pion mass hypotheses for the reconstructed "electrons". The distributions of this "mis-ID $m(B)$" are also examined in data. Plots showing these distributions are given in Fig. 6.11 and 6.12, for the $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ channels, respectively. These plots also indicate the regions of $m(B)$ occupied by $B \to K^{(*)} \pi^+ \pi^-$ MC. One can see that none of the candidates discarded by $\text{DLL}_{e\pi}(e^\pm) > 3$ have values of signal and mis-ID $m(B)$



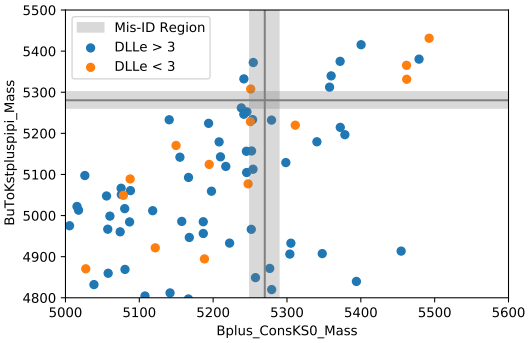
(a): Signal $m(B^+)$



(b): Mis-ID $m(B^+)$



(c): Mis-ID $m(B^+)$ vs. signal $m(B^+)$



(d): Mis-ID $m(B^+)$ vs. signal $m(B^+)$ (zoomed ranges)

Fig. 6.12: Distributions of signal $m(B^+)$ and mis-ID $m(B^+)$ in $B^+ \to K^{*+}e^+e^-$ data. The events in blue pass additional $\mathrm{DLL}_{e\pi}(e^\pm) > 3$ requirements in run-1, while those in orange fail it. The grey bands indicate the $m(B^+)$ distributions seen in $B^+ \to K^{*+}\pi^+\pi^-$ MC (showing the inner 68.3% quantiles).

consistent with those of $B \to K^{(*)}\pi^+\pi^-$ decays. In addition, the discarded candidates lie across broad ranges of $m(B)$, suggesting they likely originate from combinatorial background rather than any physical process.

In addition, the distributions of mis-ID $m(B)$ seen for the signal, partially-reconstructed background, $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ leakage, and $B \to K^{(*)}\pi^+\pi^-$ background, normalised to their fitted yields in Section 4.5.4.2, are plotted. They are compared to the mis-ID $m(B)$ distributions seen in data passing the nominal selection requirements (*i.e.* without the additional cut of $\mathrm{DLL}_{e\pi}(e^\pm) > 3$). Note that the $m(B)$ distribution for combinatorial background was not considered, due to a lack of any MC sample from which it could be inferred.

These $m(B)$ distributions are shown in Fig. 6.13. One can see that there are no signs of any excesses in data at the nominal value of $m(B) \approx 5280\,\mathrm{MeV}$ which could indicate an under-

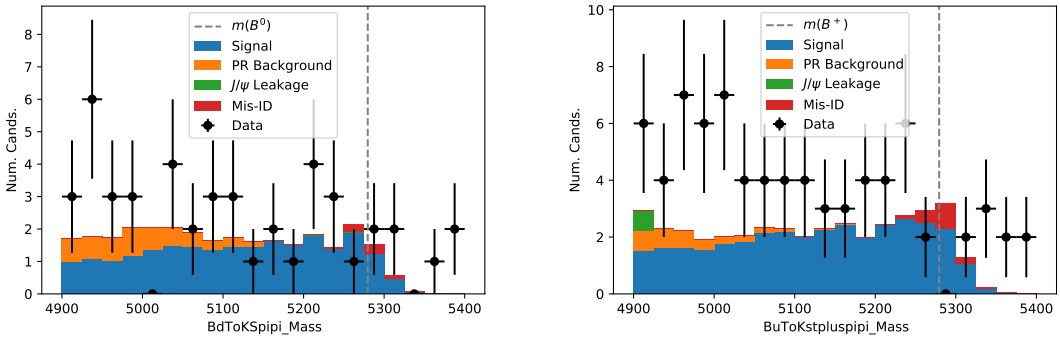


Fig. 6.13: Distributions of mis-ID $m(B)$ in $B^0 \to K^0_S e^+ e^-$ (left) and $B^+ \to K^{*+} e^+ e^-$ (right) data, as compared to the expected distributions seen for various fit-components, normalised to their fitted yields.

estimation of the $B \to K^{(*)} \pi^+ \pi^-$ decay yields*. Combined with the lack of candidates failing $\mathrm{DLL}_{e\pi}(e^\pm) > 3$ found above, this suggests the $B \to K^{(*)} \pi^+ \pi^-$ yields have been accurately estimated in this analysis.

*To check the consistency between the expected distributions of mis-ID $m(B)$ and data in each channel, an additional exponential PDF was added to represent the combinatorial background. Its shape was determined using a binned maximum likelihood fit. Goodness-of-fit tests were then carried out, using a likelihood-based test statistic. These indicated good agreement between the expected distributions and data when combinatorial background was accounted for, in both channels.

Chapter 7

Results

This chapter presents the results of the new measurements of $R^{-1}_{K^0_S}$, $R^{-1}_{K^{*+}}$, and other observables. To begin with, the methods used to extract the observables from efficiency and yield estimates are outlined. Then, the final numerical results are given. The implications of these results are discussed in the following chapter.

7.1 Extracting observables

This section outlines the methods by which the estimates of signal-mode and control-mode yields and selection efficiencies are translated into measured observables. To begin with, the electron-mode differential branching fractions, and associated signal significances, are discussed. Then, the R^{-1}_X ratios are outlined, together with tests of their consistencies with SM expectations. To conclude, the methodology used for toy studies of the mass-fits is discussed. These toy studies are used to validate and debug the mass-fits, and assess any residual biases in the fit models. They were also used to estimate the expected sensitivities of this analysis prior to unblinding.

7.1.1 Electron-mode branching fractions

The differential branching fractions for the rare electron modes are measured according to eq. 3.2, using the efficiency estimates outlined in Section 4.3, and the mass-fit models outlined in Section 4.5. In each $B \to K^{(*)}e^+e^-$ mass-fit model, the signal yield is reparameterised in terms of the signal differential branching fraction, the control-mode yield and branching fraction, and the signal and control-mode efficiencies. The differential branching fraction can hence be directly extracted from the maximum likelihood fit, with its uncertainty found using MINOS (see Section 3.3). All other parameters used to normalise the branching fraction are Gaussian-

constrained, with the widths of these constraints incorporating all sources of systematic uncertainty. This includes the control-mode yield, found from a prior fit to $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ data. The statistical uncertainty returned by HESSE in the fit to $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ data is incorporated in the width of this Gaussian constraint for the control-mode yield.

7.1.1.1 Signal significances

The same maximum likelihood fit model is used when examining the significances of the $B \to K^{(*)}e^+e^-$ signals in data. In each channel, these significances are found with a hypothesis test, calculating the consistency of the data with a null hypothesis of $\mathcal{B}(B \to K^{(*)}e^+e^-) = 0$ (*i.e.* the mass-fit model without any signal component). A test-statistic q_0 is used, which is constructed for the discovery of a positive signal [293]. For a positive best-fit value of the branching fraction ($\hat{\mathcal{B}}$), this is defined as $q_0 = 2 \times \Delta_{PLL}$, where Δ_{PLL} is the profile log-likelihood (discussed in Section 3.3), evaluated at the null-hypothesis $\mathcal{B} = 0$. For negative $\hat{\mathcal{B}}$, it is defined as $q_0 = 0$.

In order to determine a p-value (and hence a significance) for a value of q_0 observed in data, one needs to know the distribution of q_0 expected under the null hypothesis. This can be found by generating an ensemble of toy datasets under the null hypothesis, then using this to calculate an ensemble of q_0 values. The p-value can then easily be found, by calculating the proportion of toys with q_0 greater than the value observed in data. However, this becomes impractical when significances must be calculated at the $\gtrsim 5\sigma$ level, as $\mathcal{O}(10^8)$ toys would be required for an accurate estimate of the p-value. Instead, Wilks' theorem can be used [344]. This states that in the asymptotic limit of $N_{Sig.} \to \infty$, the profile log-likelihood is distributed as a χ^2 distribution with one degree-of-freedom. Under this assumption, one finds that q_0 is distributed as the equal sum of such a χ^2 distribution and a Dirac delta-function at zero [293]. Hence, the p-value and significance (S) for an observed value of q_0 can be found using the cumulative distribution function (CDF) for a χ^2 distribution, and the quantile function (QF) for a Gaussian distribution:

$$p = \frac{1}{2}\left(1 - CDF_{\chi^2}\left(q_0,\ 1\,d.o.f.\right)\right) \qquad S = QF_{Gaus.}\left(1-p\right) \tag{7.1}$$

7.1.1.2 Choice of K^0_S, π^+ invariant mass window

The differential branching fraction for $B^+ \to K^{*+}e^+e^-$ decays reported in this analysis does not correspond to the p-wave component alone, due to significant contributions from non-resonant

$B^+ \to K^0_S \pi^+ e^+ e^-$ decays in the mass-window of $|m(K^0_S, \pi^+_{Soft}) - m(K^{*+})| < 300\,\text{MeV}$*. In truth, the differential branching fraction for $B^+ \to K^0_S \pi^+ e^+ e^-$ decays which fall within this window of $m(K^0_S, \pi^+_{Soft})$ is reported. To do this, the signal and control-mode selection efficiencies should have the $m(K^0_S, \pi^+_{Soft})$ cut in both the numerator and the denominator. This way, the efficiencies estimate the proportion of decays occurring in this mass-window which are subsequently *reconstructed* in the same mass-window. Analogously to the f_{q^2} factors documented in Section 4.3.1, such efficiencies can be recovered by normalising the nominal values in Section 4.3 by factors $f_{m(K\pi)}$, defined as the *true* proportions of $B^+ \to K^{*+} e^+ e^-$ and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ decays which fall within the chosen $m(K^0_S, \pi^+_{Soft})$ window (*i.e.* in generator-level samples, before any reconstruction or selection effects).

However, there is another caveat associated with this measurement. For any B-meson decay-mode, the number of selected candidates $N_{Sel.}$ is given by the product of the total number of those mesons produced at the detector N_B, the decay-mode's branching fraction $\mathcal{B}$, and the selection efficiency ε. If the ratio of $N_{Sel.}$ is taken for two different decay-processes, the factor N_B cancels. One recovers an expression similar to eq. 3.2, enabling some branching fraction of interest to be measured relative to another normalisation mode (in this case $B^+ \to K^{*+} e^+ e^-$ relative to $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$). The value of $\mathcal{B}(B^+ \to J/\psi K^{*+})$ used in this analysis was taken from a world-average dominated by a measurement from the BaBar collaboration [13, 345]. However, this measurement was for $B^+ \to K^0_S \pi^+ J/\psi$ decays falling within a "narrow" window of $m(K^0_S, \pi^+_{Soft}) \in [792, 992\,\text{MeV}]$, which does not match the "wide" window of $|m(K^0_S, \pi^+_{Soft}) - m(K^{*+})| < 300\,\text{MeV}$ used in this analysis. Hence, if the default selection requirements are used for the control mode, a biased estimate of N_B is obtained, which does not cancel in ratio with the signal mode. Therefore, the control-mode selection must be altered, to align with the selection requirements used in the BaBar measurement of $\mathcal{B}(B^+ \to J/\psi K^{*+})$, and allow the factor of N_B to cancel and an accurate measurement of $\mathcal{B}(B^+ \to K^{*+} e^+ e^-)$ to be made.

Accounting for both these effects, one finds that the differential branching fraction measured with the nominal mass-fits and efficiency-estimates must be adjusted by a factor:

$$\frac{BF_{New}}{BF_{Old}} = \frac{N(Cont.\,|\,Wide)}{N(Cont.\,|\,Nar.)} \times \frac{\varepsilon_{K\pi}(Nar.\,|\,Cont.\,Full-Sel.)}{f_{K\pi}(Nar.\,|\,Cont.\,Gen.-Level)} \times f_{K\pi}(Wide\,|\,Sig.\,Gen.-Level) \tag{7.2}$$

*This 300 MeV mass-window is applied to K^* candidates in LHCb stripping lines, ensuring the the tails of the p-wave resonance are well-covered. This enables them to be more easily disentangled from other resonant and non-resonant processes, if required. Many other LHCb analyses then apply tighter mass-windows, to enhance the fraction of p-wave decays in their sample. However, the 300 MeV window was retained in this analysis to increase the signal sample sizes.

	Value
$N(Cont. \mid Wide)$	$(1.433 \pm 0.015) \times 10^4$
$N(Cont. \mid Nar.)$	$(1.181 \pm 0.013) \times 10^4$
$\varepsilon_{K\pi}(Nar. \mid Cont.\, Full-Sel.)$	0.893
$f_{K\pi}(Nar. \mid Cont.\, Gen-Level)$	0.870
$f_{K\pi}(Wide \mid Sig.\, Gen-Level)$	0.960
$\frac{BF_{New}}{BF_{Old}}$	1.195

Table 7.1: Different terms used to when measuring the differential branching fraction for $B^+ \to K^{*+}e^+e^-$ decays, to account for the $m(K^0_S, \pi^+_{Soft})$ used in this analysis, *vs.* that used in the branching fraction of $B^+ \to J/\psi K^{*+}$ decays used to normalise this measurement.

The different terms are defined as follows, with values given in Table 7.1:

- $N(Cont. \mid Wide)$ - The fitted control-mode yield in the wide $m(K^0_S, \pi^+_{Soft})$ window (*i.e.*, the nominal $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ yield).
- $N(Cont. \mid Nar.)$ - The fitted control-mode yield in the narrow $m(K^0_S, \pi^+_{Soft})$ window. This is found by re-fitting the $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ channel with the usual parameterisation, but only fitting data within the narrow window.
- $\varepsilon(Nar. \mid Cont.\, Full-Sel.)$ - The fraction of control-mode events passing the full, nominal selection requirements (*i.e.* with the wide window applied), which then fall within the narrow window. This is evaluated from fully-selected and corrected MC.
- $f_{K\pi}(Wide \mid Sig.\, Gen.-Level)$ - The fraction of generator-level rare-mode events which fall within the wide window. This is evaluated from generator-level MC.
- $f_{K\pi}(Nar. \mid Cont.\, Gen.-Level)$ - As above, but for $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ decays in the narrow window.

The resulting value of BF_{New}/BF_{Old} - also shown in this table - is accounted for in the $B^+ \to K^{*+}e^+e^-$ branching fraction values reported in Section 7.2.

7.1.2 Lepton-universality ratios

The R_X^{-1} double-ratios are measured according to eq. 3.1. In this case, for each B-meson, a simultaneous fit is carried out to both the rare muon-mode and rare electron-mode channels, by summing the NLL functions in each channel before minimisation. Similarly to the

Observable	SM Prediction	Flavio Err.	QED Err.	Tot. Err.
$R^{-1}(K^0_S)$	0.999	0.000 289	0.01	0.01
$R^{-1}(K^{*+})$	1.02	0.003 09	0.01	0.0105

Table 7.2: Standard Model predictions for $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ computed using FLAVIO, with theory uncertainties from FLAVIO, and imperfect simulation of final-state radiation by PHOTOS [143].

$B \to K^{(*)} e^+ e^-$ differential branching fractions, the mass-fit model is reparameterised such that the rare electron-mode yield is expressed in terms of R_X^{-1}, the rare muon-mode yield, and the relevant efficiencies and control-mode yields. The latter terms are Gaussian-constrained according to the nominal efficiency estimates and $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ yields, incorporating all statistical and systematic uncertainties. Then, the value of R_X^{-1} is extracted directly from the maximum likelihood fit, with uncertainties quantified using MINOS.

7.1.2.1 Consistency with Standard Model predictions

As well as measuring the absolute values of $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$, their compatibilities with SM expectations are assessed using hypothesis tests. The SM predictions for these observables are calculated using FLAVIO, which also provides estimates of the theory uncertainties arising from hadronic form-factors. An additional 1% uncertainty is also considered for each observable, to account for imperfections in the simulation of FSR by PHOTOS [143]. These predictions are shown in Table 7.2.

To carry out these hypothesis tests, the fits for $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ are modified, such that each ratio is reparameterised in terms of its predicted value under the SM, and the deviation from this value (*i.e.* $R_X^{-1} = R^{-1}_{X-SM} + \Delta_{R_X^{-1}}$). The profile log-likelihood (Δ_{PLL}) is then calculated for a null hypothesis of $\Delta_{R_X^{-1}} = 0$, and $t_\mu = 2 \times \Delta_{PLL}$ is taken as a test-statistic. As for the $B \to K^{(*)} e^+ e^-$ significances, the p-value corresponding to the value of t_μ observed in data is found using Wilks' theorem, which states that t_μ is expected to be distributed according to a χ^2 distribution with one degree-of-freedom. Hence, the p-value and corresponding significance are found as follows (using the same terminology as in eq. 7.1):

$$p = \left(1 - CDF_{\chi^2}\left(t_\mu,\ 1\,d.o.f.\right)\right) \qquad S = QF_{Gaus.}\left(1 - \frac{p}{2}\right) \tag{7.3}$$

7.1.3 Toy studies

maximum likelihood fits to ensembles of toy datasets are used for various purposes throughout this analysis. Such toy datasets are constructed by randomly drawing $m(B)$ values from the PDFs which model for each fit component. The number of values drawn for each component, for each toy, is randomly decided according to a Poisson distribution with mean equal to the component's expected yield (outlined in the next sub-section). Each toy dataset is then fitted using the nominal fit model for an observable of interest. Prior to fitting each toy dataset, Gaussian-constraints on parameters are randomly varied according to their uncertainties. This effectively amounts to repeating the "external" measurements of these parameters, just as the "internal" measurements of the signal yields are repeated with each toy mass-fit [346].

Prior to unblinding, toy fits were used to validate the fit models outlined above, ensuring they had a high probability of converging properly, would give reasonably-unbiased results, and that the uncertainties would have good coverage. The results of these fits were also used to compute the expected sensitivities and significances of the R_X^{-1} and $B \to K^{(*)} e^+ e^-$ branching fraction measurements.

Toy fits are also used to assess any residual biases or under-coverage in the fit models for R_X^{-1} and the differential branching fractions. The reported results are then adjusted accordingly. Additionally, some of the systematic uncertainties and cross-checks reported in Chapters 5 and 6 are evaluated using toy datasets generated according the methodologies outlined here.

7.1.3.1 Calculation of yields

The expected yields and shape parameters for the background components in the $B \to K^{(*)} e^+ e^-$ and $B \to K^{(*)} \mu^+ \mu^-$ channels were determined using the nominal mass-fit models for these channels, but with the signal yield kept blind (see Section 3.4). In addition, rare electron-mode data in the region $m(B) \in [4900, 5400\,\text{MeV}]$ - encompassing $\mathcal{O}(90\%)$ of the signal - was not plotted, though it was included in the mass-fit. This was also the case for the signal PDF.

Prior to unblinding, the expected signal yields were determined with eq. 3.2, using the previously-measured $B \to K^{(*)} \mu^+ \mu^-$ branching fractions and an assumption of lepton-universality (*i.e.* that $\mathcal{B}(B \to K^{(*)} e^+ e^-) = \mathcal{B}(B \to K^{(*)} \mu^+ \mu^-)$) [13]:

$$\mathcal{B}(B^0 \to K^0 \mu^+ \mu^-) = (3.39 \pm 0.34) \times 10^{-7} \qquad \mathcal{B}(B^+ \to K^{*+} \mu^+ \mu^-) = (9.6 \pm 1.0) \times 10^{-7} \tag{7.4}$$

The nominal efficiency estimates and $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ control-mode yields were used in these calculations*.

After the analysis was unblinded, the signal yields found from the nominal mass-fits could be examined safely, without the potential of causing methodological bias. Therefore, these measured signal yields were used when carrying out the final toy fit studies, assessing the sizes of any biases in the nominal mass-fit models (see Section 7.2.1).

7.1.3.2 Toy study results

Prior to unblinding, ensembles of 5000 toy datasets were generated and fitted for each rare-mode channel, following the methodology outlined above. After this, a set of "pulls" were calculated for each observable, according to the observable's fitted value in a toy i, the uncertainty on this fitted value, and its true value (*i.e.* the expected value used in toy generation) [346, 347].

$$p_i = \frac{x^i_{Fit} - x^i_{True}}{\sigma(x^i_{Fit})} \tag{7.5}$$

Note that as MINOS gives asymmetric uncertainties, the lower value of $\sigma(x_{Fit})$ is used when $x^i_{Fit} > x^i_{True}$, while the upper value is used when $x^i_{Fit} < x^i_{True}$. For an unbiased fit with good uncertainty coverage, p_i should be distributed as a "unit-Gaussian" with a mean of zero, and a width of unity. If the p_i distribution's mean deviates from zero, this could indicate a bias in the fit for that observable. Likewise, if the distribution's standard deviation differs from unity, this could show under/overcoverage of the calculated uncertainties.

The pre-unblinding toy fits show good stability, with $\sim 99.3\%$ and $\sim 95\%$ of fits for $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ converging, respectively, with MINOS successfully calculating 1σ uncertainty bounds. In addition, the pull distributions for all observables (the plots of which are omitted for brevity) were reasonably close to a unit-Gaussian, suggesting the mass-fit strategy was a sound one†. The means and standard deviations of the pulls for the $B \to K^{(*)}\mu^+\mu^-$ yields were all consistent with unit-Gaussian expectations, suggesting the fits for these yields were unbiased and with good uncertainty coverage. In contrast, the pull distribution mean for $R^{-1}_{K^{*+}}$ was inconsistent with zero, with a value $\sim 5\%$, indicating a probable bias in the fitted central value for this

*These nominal $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ yields are also used when carrying out toy fits for the control-mode channels, as part of the systematic uncertainty studies in Section 5.12

†These toy studies also helped to determine the fitting strategy. For example, very biased, asymmetric pull distributions were found when fitting for the conventional ratio R_X, as the small $B \to K^{(*)}e^+e^-$ yield is in the denominator of this observable. In addition, very small amounts of combinatorial background fell into the upper-mass sideband of the $B^+ \to K^{*+}e^+e^-$ channel when using a cut to $\text{BDT}_{Comb.}$ suggested by the studies in Section 4.1.5.2. The shape of the combinatorial background was then poorly-constrained by the fit, leading to biased pull distributions for the signal. When the cut is loosened to the nominal working-point of $\text{BDT}_{Comb.} > 0.250$, there is increased data in this sideband, and the biases disappear.

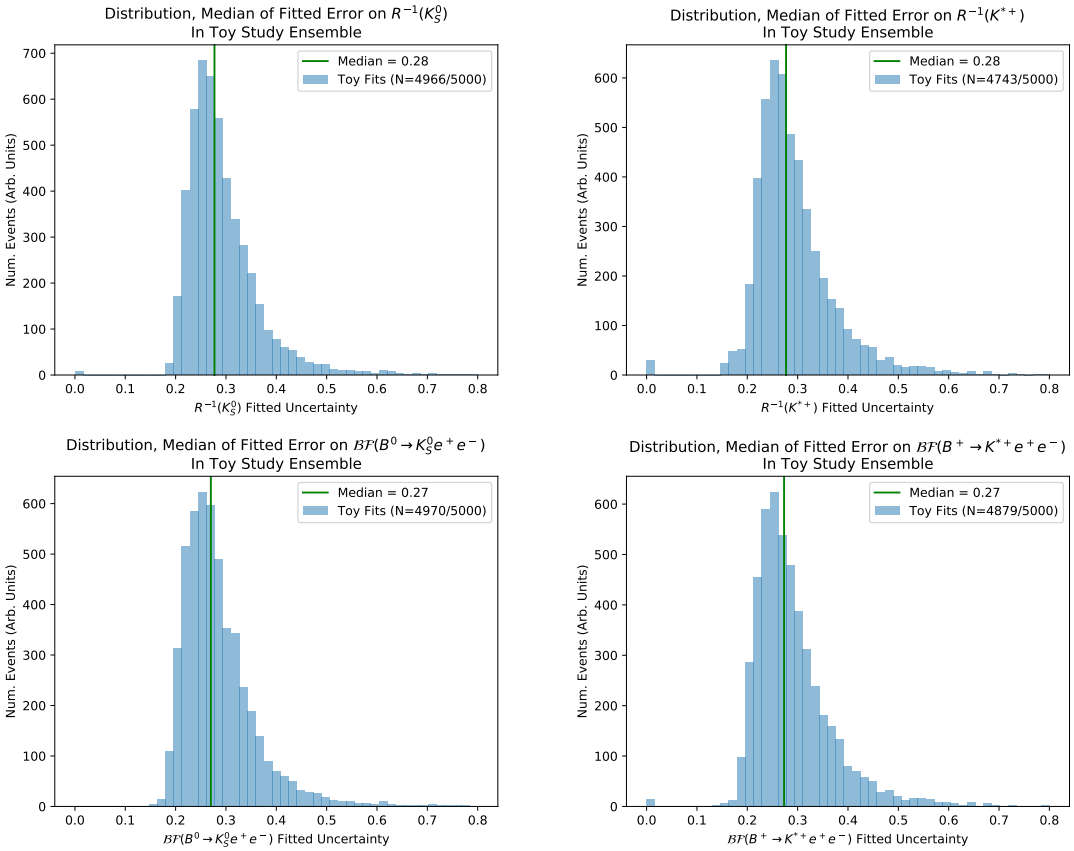


Fig. 7.1: Distributions of fitted uncertainties for $R^{-1}_{K^0_S}$ (top left), $R^{-1}_{K^{*+}}$ (top right) and the differential branching fractions of $B^0 \to K^0 e^+ e^-$ (bottom left) and $B^+ \to K^{*+} e^+ e^-$ (bottom right) decays, as found from fits to ensembles of toy datasets generated under SM expectations. Recall that the world-average branching fractions for $B \to K^{(*)} \mu^+ \mu^-$ were used when calculating the signal yields in these toy datasets [13].

observable. In addition, the pull distributions for all R^{-1}_X and $B \to K^{(*)} e^+ e^-$ branching fraction fits had standard deviations inconsistent with unity, suggesting slight undercoverage by the uncertainty estimates. Therefore, strategies are needed to account for biases and undercoverage in the reported values of all observables, which are documented in Section 7.2.1. Note also that when all expected yields in these toy fits were increased by a factor of 10, the pull distributions showed good agreement with unit-Gaussians, suggesting the deviations seen in the nominal studies result purely from the limited amounts of data in the mass-fits.

The distributions of uncertainties on R^{-1}_X and the $B \to K^{(*)} e^+ e^-$ branching fractions, relative to the fitted values in each toy, are shown in Fig. 7.1. One can see that median uncertainties of $27 - 28\%$ were expected under SM conditions, though with considerable variation about these medians. In addition, the significances of the $B \to K^{(*)} e^+ e^-$ signals were measured for

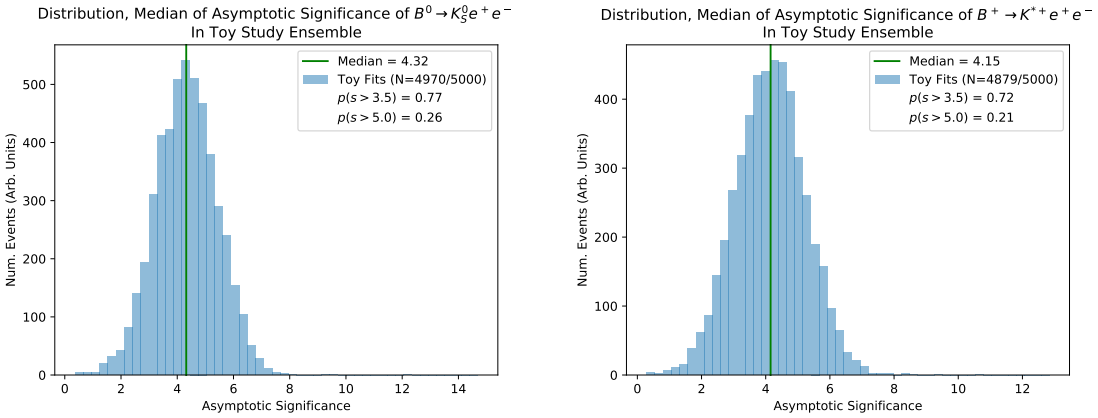


Fig. 7.2: Distributions of significances for $B^0 \to K^0_S e^+ e^-$ (left) and $B^+ \to K^{*+} e^+ e^-$ (right) decays as found from fits to ensembles of toy datasets generated under SM expectations. Recall that the world-average branching fractions for $B \to K^{(*)} \mu^+ \mu^-$ were used when calculating the signal yields in these toy datasets [13].

each toy, with the distributions shown in Fig. 7.2. Here, median significances of $\sim 4\sigma$ were anticipated, with a $\sim 20-25\%$ chance of discovering each decay-mode.

Observable	Separate Stat., Syst. Uncs.
$R^{-1}(K^0_S)$	$1.51\ ^{+0.40}_{-0.35}\ (Stat.)\ ^{+0.09}_{-0.04}\ (Syst.)$
$R^{-1}(K^{*+})$	$1.44\ ^{+0.32}_{-0.29}\ (Stat.)\ ^{+0.09}_{-0.06}\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^0 \to K^0_S e^+e^-)\ /\ 10^{-8}\ GeV^{-2}$	$2.59\ ^{+0.61}_{-0.56}\ (Stat.)\ ^{+0.14}_{-0.10}\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^+ \to K^{*+} e^+e^-)\ /\ 10^{-8}\ GeV^{-2}$	$9.2\ ^{+1.9}_{-1.7}\ (Stat.)\ ^{+0.8}_{-0.6}\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^0 \to K^0_S e^+e^-)/\mathcal{B}(Cont.)\ /\ 10^{-4}\ GeV^{-2}$	$4.9\ \pm 1.1\ (Stat.)\ \pm 0.2\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^+ \to K^{*+} e^+e^-)\ /\mathcal{B}(Cont.)\ /\ 10^{-3}\ GeV^{-2}$	$1.08\ ^{+0.22}_{-0.20}\ (Stat.)\ ^{+0.06}_{-0.04}\ (Syst.)$

Table 7.3: Table showing the fitted values of $R^{-1}_{K^0_S}$, $R^{-1}_{K^{*+}}$, and the differential branching fractions for $B^0 \to K^0_S e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ decays. Results are shown with the statistical and systematic uncertainties separated out. Here, $\mathcal{B}(Cont.)$ denotes the corresponding $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ control-mode branching fraction for each decay. Note that these results have not been adjusted to account for intrinsic fit biases.

7.2 Analysis Results

The results of the fits for $R^{-1}_{K^0_S}$, $R^{-1}_{K^{*+}}$, and the differential branching fractions for $B^0 \to K^0 e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ decays, are reported in this section. First, initial results are given, and residual intrinsic fit biases are assessed. Then, the final results, adjusted for such biases, are reported. In addition, the significances of the electron-mode signals, and consistencies of $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ with SM expectations, are given.

7.2.1 Initial results and intrinsic fit biases

In addition to the nominal fit strategies outlined in the previous sections, each fit is repeated with all parameters which would normally be Gaussian-constrained instead fixed to their central values. This way, the uncertainties for each observable, found using MINOS, only account for statistical effects. Therefore, the total uncertainties associated to systematic effects can be found by subtracting these "stat. only" uncertainties in quadrature from the total uncertainties found in the nominal fits. All such results are summarised in Table 7.3. This table also shows ratios of the $B \to K^{(*)}e^+e^-$ branching fractions relative to the corresponding $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ branching fractions, allowing them to be combined with an alternative value for the control-mode branching fraction if desired.

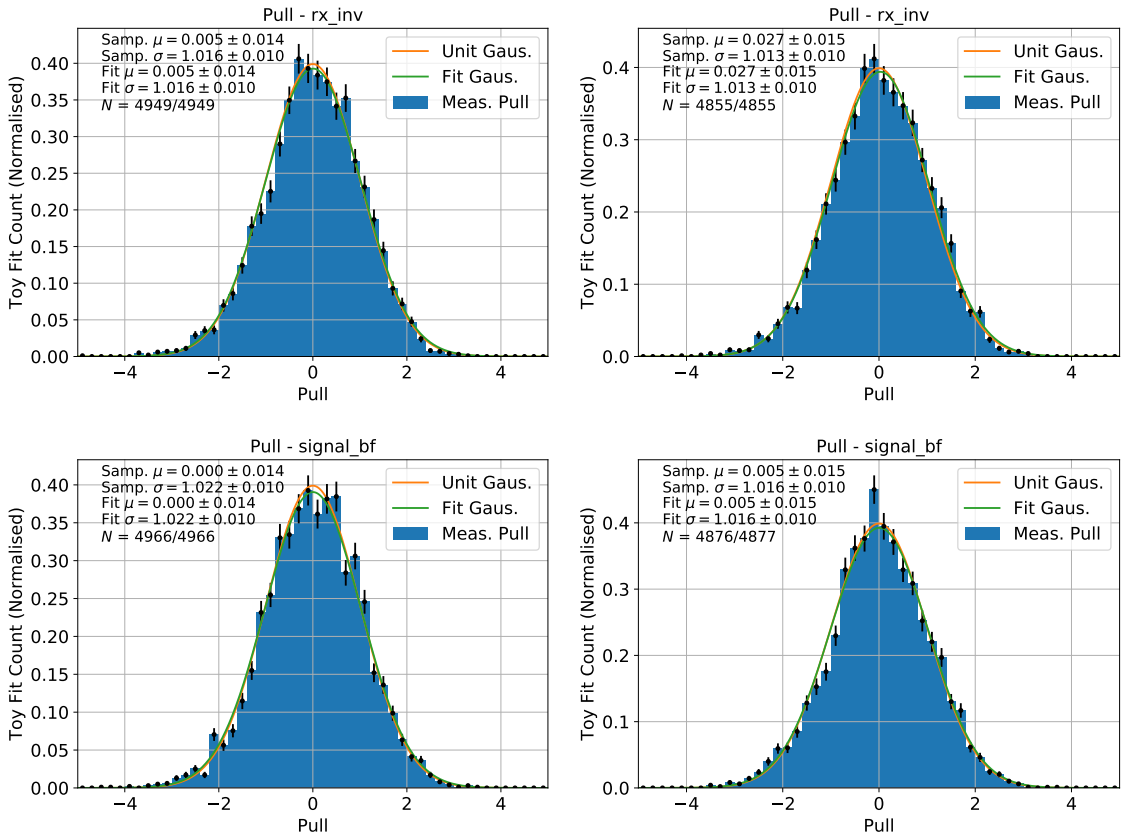


Fig. 7.3: Distributions of pull values for $R^{-1}_{K^0_S}$ (top left), $R^{-1}_{K^{*+}}$ (top right), and the differential branching fractions for $B^0 \to K^0 e^+ e^-$ (bottom left) and $B^+ \to K^{*+} e^+ e^-$ (bottom right), as found from post-unblinding toy fits for these observables. Unit Gaussian PDFs are superimposed on these distributions, as are Gaussian PDFs found from a maximum likelihood fit to the pull distribution. Note that ensembles of 5000 toys were used to produce each of these pull distributions. The numbers in the legends (formatted $N = x/y$) indicate how many of these 5000 toy fits converged properly, and how many of these fall within a range of $-5 < Pull < 5$.

After the measurements were unblinded, toy studies were used to assess residual biases in the maximum likelihood fits. The methodologies documented in Section 7.1.3 were used, but with signal yields generated according to their measured values, rather than the estimates calculated prior to unblinding. The distributions of pulls found from these studies are shown in Fig. 7.3, for the different observables.

If the mean of a pull distribution is not consistent with zero*, this could indicate a bias in the fitted value (μ) of the corresponding observable (θ). In such cases, an additional systematic uncertainty is assigned to θ, given by $\sigma_{Bias\,Syst.} = \mu \times \sigma_{Fit.}$. As asymmetric uncertainties

*Here, the mean is considered incompatible with zero if the standard error on the mean does not cover zero. Likewise when considering the compatibility of the standard deviation with unity.

Observable	Bias	Scale Fact.	Unadj. Coverage	Adj. Coverage
$R^{-1}_{K^0_S}$	0.000	1.016	0.677 ± 0.007	0.683 ± 0.007
$R^{-1}_{K^{*+}}$	0.027	1.013	0.675 ± 0.007	0.681 ± 0.007
$\frac{d\mathcal{B}}{dq^2}(B^0 \to K^0 e^+ e^-)$	0.000	1.022	0.680 ± 0.007	0.693 ± 0.007
$\frac{d\mathcal{B}}{dq^2}(B^+ \to K^{*+} e^+ e^-)$	0.000	1.016	0.681 ± 0.007	0.690 ± 0.007

Table 7.4: Table showing the systematic bias and scaling factors applied to the uncertainties quoted for various observables, as calculated from toy pull distributions. The coverage fractions of the uncertainties, found in toys before and after these systematic uncertainties and scaling factors are applied, are also shown.

are output by MINOS, separate values of $\sigma_{Bias\,Syst.}$ are computed using the upper and lower uncertainties. These are then summed in quadrature with each value of $\sigma_{Fit.}$. In addition, the hypothesis tests used to calculate the significances of the $B \to K^{(*)} e^+ e^-$ signals, and the agreement of R_X^{-1} with SM expectations, are repeated with this systematic uncertainty included. Here, for ease of computation, a single value of $\sigma_{Bias\,Syst.}$ is calculated using the value of $\sigma_{Fit.}$ returned by HESSE, and imposed as an additional Gaussian constraint in the fit. The fit for $R^{-1}_{K^{*+}}$ is the only one with a significant bias in the pull distribution, meaning this procedure needs to be used.

If the standard deviation of the pull distribution is greater than unity (and not consistent with unity), this indicates undercoverage by the uncertainties found with MINOS. In such cases - namely all the observable fits - the upper and lower uncertainties are scaled up by the standard deviation of the pull distribution. In addition, the "widths" of the profile log-likelihoods used in the hypothesis tests are increased by this scaling parameter. In practise, this is done by changing the point at which the profile log-likelihood is evaluated:

$$\Delta_{PLL}(x_{Test}) \to \Delta_{PLL}\left(\mu + \frac{\theta_{Test} - \mu}{s}\right) \tag{7.6}$$

Here, θ_{Test} represents the parameter value representing the null hypothesis (*e.g.* $\mathcal{B}\left(B^0 \to K^0_S e^+ e^-\right) = 0$), μ represents the fitted value of that parameter, and s represents the scaling factor calculated from the pull distribution. The systematic uncertainties and scaling

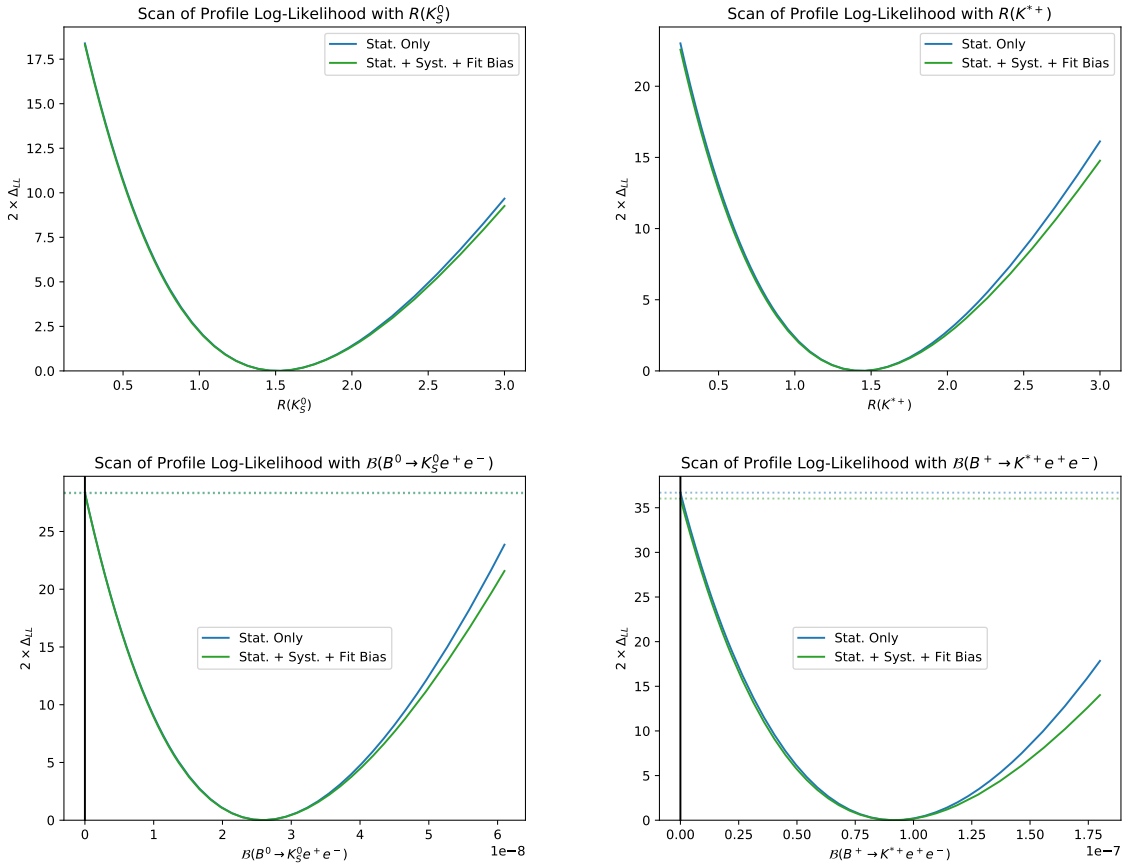


Fig. 7.4: Scans over profile log-likelihoods for $R^{-1}_{K^0_S}$ (top left), $R^{-1}_{K^{*+}}$ (top right), and the differential branching fractions for $B^0 \to K^0 e^+ e^-$ (top left) and $B^+ \to K^{*+} e^+ e^-$ (top right) decays, with and without parameters allowed to float within Gaussian constraints to account for systematic uncertainties.

factors found via this procedure are given in Table 7.4. One can see that good coverage is obtained when these are applied to the results from the toy fits*.

7.2.2 Profile log-likelihoods

In addition to the numerical values reported in this section, scans of the profile log-likelihoods are carried out across R^{-1}_X and the differential branching fractions for $B \to K^{(*)} e^+ e^-$. These scans are also repeated with all Gaussian-constrained parameters fixed to their central values, resulting in scans of Δ_{PLL} which only account for statistical effects. All scans have additional

*Note that the coverage fractions without these adjustments applied already show good consistency with 0.683. Therefore, these adjustments represent a conservative choice, giving further reassurance that the reported uncertainties have good coverage.

Observable	Separate Stat., Syst. Uncs.
$\frac{d\mathcal{B}}{dq^2}(B^0 \to K^0_S e^+e^-)$ / 10^{-8} GeV^{-2}	$2.59\ ^{+0.62}_{-0.57}\ (Stat.)\ ^{+0.14}_{-0.10}\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^+ \to K^{*+} e^+e^-)$ / 10^{-8} GeV^{-2}	$9.2\ ^{+1.9}_{-1.8}\ (Stat.)\ ^{+0.8}_{-0.6}\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^0 \to K^0_S e^+e^-)/\mathcal{B}(Cont.)$ / 10^{-4} GeV^{-2}	$4.9\ ^{+1.2}_{-1.1}\ (Stat.)\ \pm 0.2\ (Syst.)$
$\frac{d\mathcal{B}}{dq^2}(B^+ \to K^{*+} e^+e^-)\ /\mathcal{B}(Cont.)$ / 10^{-3} GeV^{-2}	$1.08\ ^{+0.22}_{-0.21}\ (Stat.)\ ^{+0.06}_{-0.04}\ (Syst.)$

Table 7.5: Table showing the measured differential branching fractions for $B^0 \to K^0 e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ decays. This table also shows the results with statistical and systematic uncertainties separated out. Here, $\mathcal{B}(Cont.)$ denotes the corresponding $B \to J/\psi\,(e^+e^-)\,K^{(*)}$ control-mode branching fraction for each decay. Systematic uncertainties and scaling factors have been applied to these results, to account for intrinsic fit biases.

systematic uncertainties and scaling factors applied, to account for intrinsic fit biases, and are shown in Fig. 7.4.

7.2.3 Final results

After applying systematics and scaling factors to account for residual fit biases, the following results for the lepton universality ratios are found:

$$R^{-1}_{K^0_S} = 1.51\,^{+0.40}_{-0.35}(\text{stat.})\,^{+0.09}_{-0.04}(\text{syst.}), \quad R^{-1}_{K^{*+}} = 1.44\,^{+0.32}_{-0.29}(\text{stat.})\,^{+0.09}_{-0.06}(\text{syst.})$$

By inverting these central values and confidence intervals, the "conventional" ratios are found:

$$R_{K^0_S} = 0.66\,^{+0.20}_{-0.14}\,(\text{stat.})\,^{+0.02}_{-0.04}\,(\text{syst.}), \quad R_{K^{*+}} = 0.70\,^{+0.18}_{-0.13}\,(\text{stat.})\,^{+0.03}_{-0.04}\,(\text{syst.})$$

These are reported in the analysis paper [5], as are the differential branching fractions of the rare electron modes, shown in Table 7.5.

In addition, the significances of the rare electron-mode decays are calculated using the methodology outlined in Section 7.1.1.1, with scaling factors applied to account for intrinsic fit

Decay-mode	Unc. Type	$2 \times \Delta_{LL}$	p-val.	Signif.
$B^0 \to K^0_S e^+ e^-$	Stat. only	28.4	5.03×10^{-8}	5.33
	Stat. + Syst.	28.3	5.17×10^{-8}	5.32
$B^+ \to K^{*+} e^+ e^-$	Stat. only	36.7	6.94×10^{-10}	6.06
	Stat. + Syst.	36.0	9.73×10^{-10}	6.00

Table 7.6: Table showing the significances for signals from $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$ decays, with the corresponding p-values and $2 \times \Delta_{PLL}$ values. Results are shown both with and without the inclusion of systematic uncertainties.

Obs.	$2 \times \Delta_{LL}$	p-val.	Signif.
$R^{-1}_{K^0_S}$	2.19	0.139	1.48
$R^{-1}_{K^{*+}}$	2.03	0.155	1.42

Table 7.7: Table showing levels of agreement between the measurements of $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ and their SM expectations. The p-values from these hypothesis tests are also presented as two-sided Gaussian significances. The values of $2 \times \Delta_{PLL}$ used in the hypothesis tests are also shown.

biases. The values of Δ_{PLL} at $\mathcal{B}(B \to K^{(*)} e^+ e^-) = 0$, together with the corresponding p-values and significances, are shown in Table 7.6. These calculations are also repeated with Gaussian-constrained parameters fixed to their central-values, so that only statistical effects are taken into account when calculating Δ_{PLL}. One can see that, even when systematic uncertainties are accounted for, significances of $> 5\sigma$ are found for both modes, representing the first ever observations of these decays.

7.2.3.1 Consistency with Standard Model predictions

Finally, the consistencies of the observed values of $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ with SM expectations are assessed, using the hypothesis test documented in Section 7.1.2.1. Systematic uncertainties and scaling factors are applied to account for intrinsic fit biases. The values of Δ_{PLL} for each observable under the SM null hypothesis are reported in Table 7.7, together with the corresponding p-values and Gaussian significances. Both the ratios $R^{-1}_{K^0_S}$ and $R^{-1}_{K^{*+}}$ are consistent with SM expectations, at the level of 1.5σ and 1.4σ, respectively. The possible implications of these measurements, and prospects for future measurements, are discussed in the concluding chapter of this dissertation.

Part III

Conclusions

Chapter 8

Conclusions and Prospects

This dissertation documents the first ever observations of the decays $B^0 \to K^0_S e^+ e^-$ and $B^+ \to K^{*+} e^+ e^-$, with measurements of their differential branching fractions in bins of q^2. Measurements of the ratios $R_{K^0_S}$ and $R_{K^{*+}}$ are also documented. As shown in Fig. 8.1, these are the most precise such measurements to date. These measurements remain consistent with SM expectations, and are heavily statistically limited. Due to their well-controlled systematic uncertainties, there are strong prospects for more precise measurements of $R_{K^0_S}$ and $R_{K^{*+}}$ in the future, as further data is recorded by the LHCb experiment.

In this concluding chapter, the impact of these new measurements on the global picture of potential NP is quantified. Then, prospects for future tests of lepton universality in $b \to s\ell^+\ell^-$ decays are reviewed, starting with measurements using data recorded by LHCb in run-1/2. Following this, the impacts of future developments to the LHCb detector are discussed. Finally, future measurements by other experimental collaborations are examined, followed by a brief summary.

8.1 Global fit

Theoretical predictions for $R_{K^0_S}$ and $R_{K^{*+}}$ are compared to the experimental measurements presented in this dissertation, using the EFT framework outlined in Section 1.2.1. Their values are predicted using FLAVIO, in the presence of a new left-handed coupling to muons parameterised by $\mathcal{C}^{NP}_{L\mu} = \mathcal{C}^{NP}_{9\mu} = -\mathcal{C}^{NP}_{10\mu}$. All other Wilson coefficients are held at their SM values. Then, a likelihood is calculated, showing the compatibility between the experimental measurements and theoretical expectations at a given value of $\mathcal{C}^{NP}_{L\mu}$. This likelihood is maximised to find the value of $\mathcal{C}^{NP}_{L\mu}$ which best describes the data, while MINOS is used to calculate the corresponding uncertainty. In addition, the compatibility with SM expectations is quantified using the likelihood at this best-fit value, relative to the SM value of $\mathcal{C}^{NP}_{L\mu} = 0$.

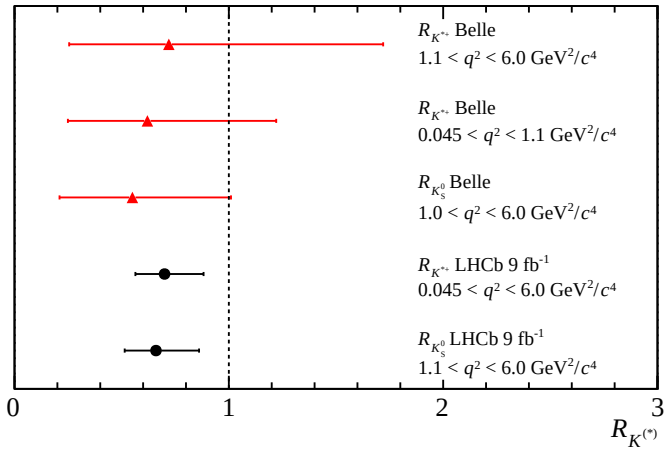


Fig. 8.1: Comparison of measurements of $R_{K^0_S}$ and $R_{K^{*+}}$ by LHCb (this work), and by Belle [154, 105].

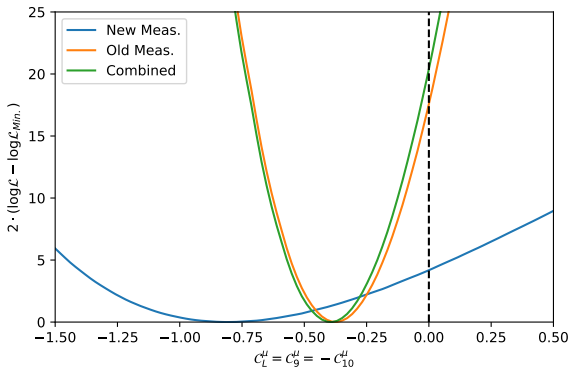


Fig. 8.2: Scans of likelihoods for compatibility between EFT predictions and experimental values from previous LHCb measurements of R_{K^+}, $R_{K^{*0}}$, and $\mathcal{B}(B^0_s \to \mu^+\mu^-)$, the new measurements of $R_{K^0_S}$ and $R_{K^{*+}}$, and all of these combined, as a function of $\mathcal{C}^{NP}_{L\mu}$.

This study is repeated using previous LHCb measurements of R_{K^+}, $R_{K^{*0}}$, and $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ [135, 136, 147, 151], representing theoretically "clean" observables. Scans of the likelihoods found with these "old" measurements, the "new" $R_{K^0_S}$ and $R_{K^{*+}}$ measurements, and both sets of measurements combined are shown in Fig. 8.2, while the minimised values of $\mathcal{C}^{NP}_{L\mu}$ and compatibilities with SM expectations are shown in Table 8.1. While the new measurements are consistent with SM expectations, they are also in good agreement with previous LHCb measurements. Hence, they slightly increase the global tensions with SM

Observables	$\mathcal{C}_L^{Fit.}$	$2\cdot(\log\mathcal{L}_{Min.}-\log\mathcal{L}_{SM})$	Signif.
Old Meas.	$-0.375\,^{+0.083}_{-0.082}$	17.6	4.20
New Meas.	$-0.81\,^{+0.36}_{-0.30}$	4.18	2.04
Combined	$-0.391\,^{+0.079}_{-0.082}$	20.4	4.52

Table 8.1: Best-fit values for $\mathcal{C}_{L\mu}^{NP}$, and associated uncertainties, found by comparing EFT predictions with previous LHCb measurements of R_{K^+}, $R_{K^{*0}}$, and $\mathcal{B}(B_s^0\to\mu^+\mu^-)$, the new measurements of $R_{K_S^0}$ and $R_{K^{*+}}$, and all of these combined. The differences in log-likelihood found at these best-fit values, relative to the SM expectations of $\mathcal{C}_{L\mu}^{NP}=0$, are also shown. These differences in log-likelihood are translated into significances using Wilks' theorem [344]. Note that these significances are only valid under the specific assumption that new physics can be parameterised by $\mathcal{C}_{L\mu}^{NP}$. This basis was chosen because it gives a good fit to data, meaning these significances are inflated by the "look elsewhere effect", and do not give an accurate description of the *global* significance of data compared to SM expectations (*i.e.* how much new physics is favoured over the SM *in general*). See Section 1.2.4 for details.

expectations under the assumption of NP in $\mathcal{C}_{L\mu}^{NP}$. They should prove useful data for theorists constructing models to explain the anomalies in $b\to s\ell^+\ell^-$ decays.

As for other R_X ratio measurements, these measurements of $R_{K_S^0}$ and $R_{K^{*+}}$ favour a deficit of muons relative to electrons. This slightly strengthens the case against the global tensions in $b\to s\ell^+\ell^-$ measurements being simply due to statistical fluctuations. Instead, two possibilities remain. First, the tensions could be due to systematic biases in experimental measurements and/or theoretical predictions. For example, some ratios of efficiencies may be misestimated. Alternatively, there could be an unknown source of background in the electron modes, artificially increasing the measured yields for these processes. Such backgrounds could explain the tensions in R_X ratio measurements, while biases in theoretical estimates of soft QCD effects could explain the tensions in $b\to s\mu^+\mu^-$ observables (see Section 1.2.2.1). The second, more tantalising possibility is that the tensions in $b\to s\ell^+\ell^-$ measurements are caused by genuine contributions from new particles, such as leptoquarks or Z' bosons (see Section 1.2.4.1).

8.2 Future prospects for lepton universality tests

As discussed in Section 1.2.3.6, R_X ratios for different decays are sensitive to different NP effects. Therefore, further such measurements are crucial, if the physics governing $b\to s\ell^+\ell^-$ decays is to be understood. The existing measurements of $R_{K^{*0}}$ and R_{pK} by LHCb are being

Particle	Requirement	
K^{*+}	$p_T > 400\,\text{MeV}$,	$\lvert m - m_{K^{*+}}\rvert < 300\,\text{MeV}$
K^+	$p_T > 400\,\text{MeV}$,	$\chi^2_{\text{IP}} > 9$
Merged π^0	$p_T > 2000\,\text{MeV}$,	$\lvert m - m_{\pi^0}\rvert < 60\,\text{MeV}$
Resolved π^0	$p_T > 600\,\text{MeV}$,	$\lvert m - m_{\pi^0}\rvert < 30\,\text{MeV}$
γ from resolved π^0	$p_T > 200\,\text{MeV}$	

Table 8.2: Stripping selection requirements imposed on $K^{*+} \to K^+\pi^0$ candidates.

updated to include all data from run-1/2, as are the measurements of differential branching fractions and angular observables for $b \to s\mu^+\mu^-$ decays. In addition, tests of lepton universality are planned using decays which have not been previously examined, including $B^0_s \to \phi\ell^+\ell^-$, $B^+ \to \pi^+\ell^+\ell^-$, and $\Lambda^0_b \to \Lambda\ell^+\ell^-$. Angular analyses of $b \to se^+e^-$ decays are also in progress, which provide clean tests of the SM when combined with the equivalent $b \to s\mu^+\mu^-$ measurements [155].

In the short-to-medium term, the methods used to select $b \to s\ell^+\ell^-$ decays at LHCb could be improved, to increase the sensitivities of R_X tests. For example, low-momentum electron tracks constructed from hits in the VELO and TT could be used, the inclusion of which could increase the yields of $b \to se^+e^-$ decays by $\sim 20-30\%$ [348, Sec. 8.4.1] [349]. Other potential developments at LHCb are discussed in the following sections.

8.2.1 Alternative K^{*+}-meson reconstruction

The measurement of $R_{K^{*+}}$ presented in this dissertation used candidates reconstructed via the decay $K^{*+} \to K^0_S\pi^+$. However, $\sim 33\%$ of K^{*+}-mesons decay via an alternative mode - $K^{*+} \to K^+\pi^0$- which could provide an additional data sample for a new measurement of $R_{K^{*+}}$. At LHCb, π^0-mesons are reconstructed from the decay $\pi^0 \to \gamma\gamma$, using energy deposits in the ECAL. The signature left by such decays depends on the momentum of the π^0-meson [350]. At low momentum, the two photons form separate clusters. Once the energies of all photon candidates in an event have been determined, they are combined with each other to build "resolved" π^0 candidates. At high momentum, the π^0-mesons are boosted, meaning the two photons may form a single ECAL cluster. A dedicated algorithm is used to split high-energy ECAL clusters into two sub-clusters, from which "merged" π^0 candidates are built.

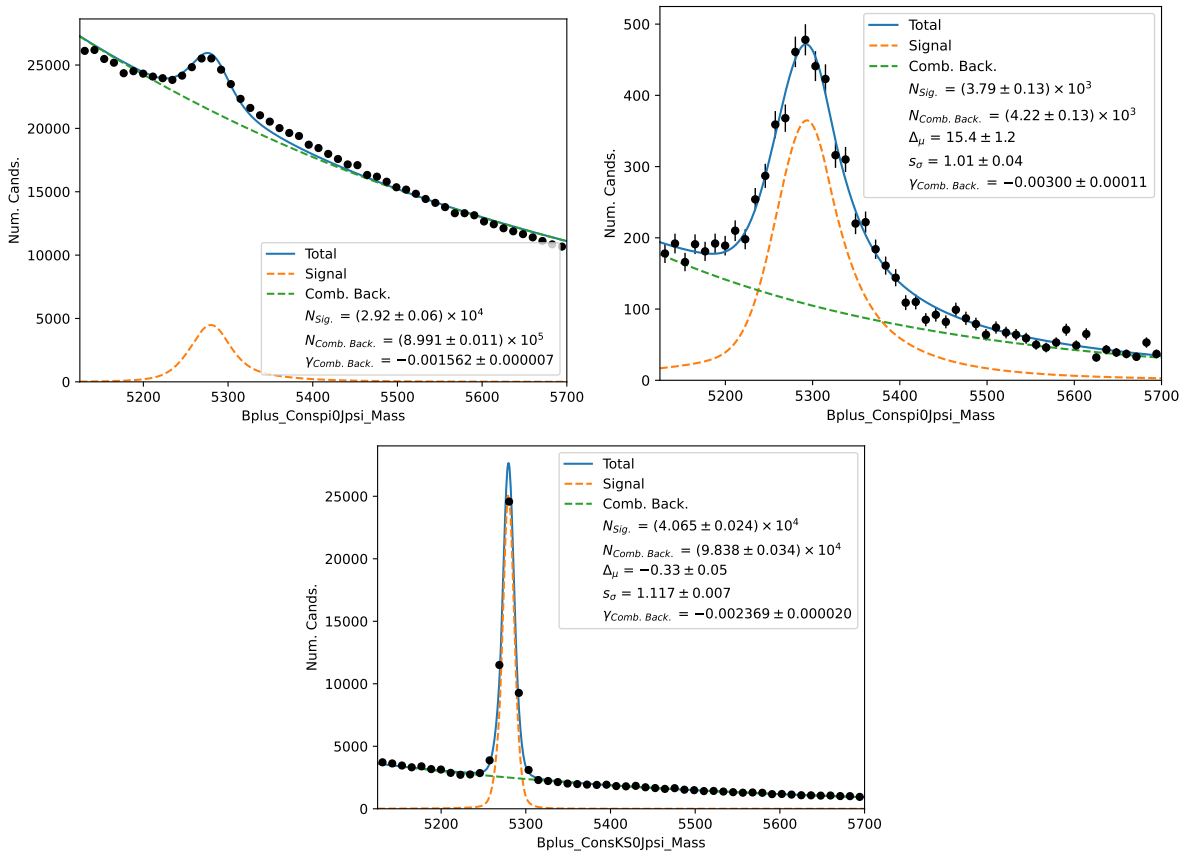


Fig. 8.3: Fits to $m(B^+)$ using $B^+ \to J/\psi K^{*+}$ candidates recorded in 2018 with resolved (top left) and merged (top right) $K^{*+} \to K^+\pi^0$ candidates, and with $K^{*+} \to K^0_S\pi^+$ candidates (bottom). Note that the values of $m(B^+)$ used here were computed using `DecayTreeFitter`, with the dilepton mass constrained to $m(J/\psi)$, and the K^0_S and π^0 masses constrained to their physical values [13].

A new stripping line* was introduced by the author to select $B^+ \to K^{*+}\ell^+\ell^-$ candidates in run-1/2 data using the $K^{*+} \to K^+\pi^0$ decay mode. Both merged and resolved π^0 candidates are included, and the requirements applied to particles in the $K^{*+} \to K^+\pi^0$ decay chain are shown in Table 8.2. The requirements applied to lepton and B-meson candidates are as in the original $K^{*+} \to K^0_S\pi^+$ stripping line (see Section 4.1.2).

To estimate the potential sensitivity of an $R_{K^{*+}}$ measurement using this new stripping line, the number of $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$ candidates recorded in 2018 are examined, using both the old and new K^{*+}-meson reconstructions. The muon-mode trigger requirements outlined in Section 4.1.1 are applied to this data, as are the preselection requirements for muons outlined in Section 4.1.3. None of the preselection requirements for other particles are applied, to ensure

*A stripping line is a set of loose requirements designed to select a certain decay mode, producing data samples which are small enough for use by LHCb physics analysts (see Section 2.2.4).

	$B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$	$B^+ \to K^{*+}e^+e^-$
$K^{*+} \to K^0_S\pi^+$	$(4.065 \pm 0.024) \times 10^4$	67
$K^{*+} \to K^+\pi^0$, Res.	$(2.92 \pm 0.06) \times 10^4$	48
$K^{*+} \to K^+\pi^0$, Mrg.	$(3.79 \pm 0.13) \times 10^3$	6

Table 8.3: Number of $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$ candidates found with different K^{*+}-meson reconstruction methods, plus the previously-measured $B^+ \to K^{*+}e^+e^-$ yield, and extrapolations of the number of $B^+ \to K^{*+}e^+e^-$ decays expected with the new reconstruction methods.

selection requirements are aligned between the two K^{*+}-meson reconstructions. In addition, vetoes for over-reconstructed $B^0 \to J/\psi(\mu^+\mu^-)K^0_S$ decays are applied (see Section 4.4.2.1), as is an analogous veto for over-reconstructed $B^+ \to K^+J/\psi(\mu^+\mu^-)$ decays. Unbinned extended maximum likelihood fits are used to extract $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$ yields for the different reconstructions, following the parameterisations outlined in Section 4.5.2.1 (though implemented using the ZFIT software package) [351]. Three separate fits are carried out; one for $K^{*+} \to K^0_S\pi^+$ candidates, one for the $K^{*+} \to K^+\pi^0$ candidates with resolved π^0-mesons, and the other with merged π^0-mesons. The results of these fits are shown in Fig. 8.3*. Note that a "shoulder" can be seen in the fit with resolved π^0-mesons, at $m(B^+) \sim 5350\,\text{MeV}$, suggesting there is an additional source of background which is not accounted for in the fit model†. However, the fit should still give an estimate of the $B^+ \to J/\psi K^{*+}$ yield which is sufficiently accurate for the sensitivity study outlined in this section.

These $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}$ yields are used to estimate the number of $B^+ \to K^{*+}e^+e^-$ decays which could be found with this new reconstruction, under a naive assumption that the same $N_{Sig.}/N_{Back.}$ ratio could be achieved as with the $K^{*+} \to K^0_S\pi^+$ reconstruction. As shown in Table 8.3, a total of ~ 54 candidates are expected, offering similar sensitivity to the existing measurement of $R_{K^{*+}}$. However, this is unlikely to ultimately be the case. Significant contamination from combinatorial background can be seen with resolved π^0-meson candidates, primarily resulting from the incorrect association of different photons. In addition, some photon candidates actually arise from electrons or hadrons striking the ECAL. These backgrounds would need to be suppressed, the latter using sets of PID algorithms specially developed for

*Note that due to the broad peak and high background contamination, the signal resolution parameters in the fit to resolved $K^{*+} \to K^+\pi^0$ candidates are fixed to $s_\sigma = 1.1$ and $\Delta_\mu = 0$, to improve fit stability. These are similar to the resolution parameters seen for $K^{*+} \to K^0_S\pi^+$ candidates.

†One such possible background is $B^+ \to \pi^+ J/\psi$ decays with a $\pi^+ \to K^+$ misidentification, and spurious addition of a π^0-meson candidate. This is similar to the over-reconstructed backgrounds outlined in Section 4.4.2.1.

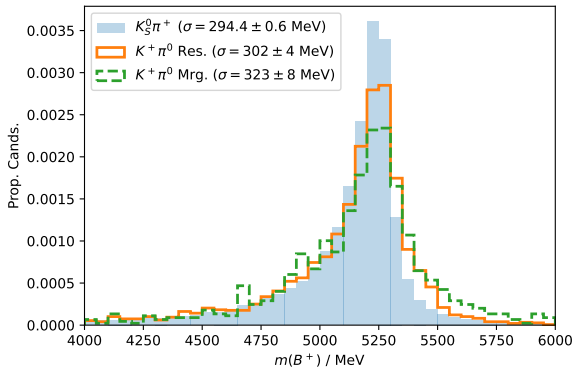


Fig. 8.4: Distributions of $m(B^+)$ found in MC for $B^+ \to K^{*+}e^+e^-$ decays recorded in 2018, with resolved and merged $K^{*+} \to K^+\pi^0$ candidates, and with $K^{*+} \to K^0_S\pi^+$ candidates. Note that the values of $m(B^+)$ used here were computed using `DecayTreeFitter`, with the K^0_S and π^0 masses constrained to their physical values.

π^0-mesons and photons (or "neutral particles") [352, Sec. 3.4.2] [353, 354]. In addition, the mass resolutions for π^0-mesons are poor compared to K^0_S-mesons, leading to broad $m(B^+)$ distributions for signal which could prove problematic in a maximum likelihood fit (see Fig. 8.4). Further studies are required into whether an efficient selection can be achieved, assessing the potential sensitivity of an $R_{K^{*+}}$ measurement.

8.2.2 Ruling out potential systematic biases

While statistical fluctuations are all but ruled out, there remains the possibility that the R_X anomalies are caused by systematic effects. Upcoming measurements at LHCb will aim to rule this out.

For example, all existing R_X analyses by LHCb have been carried out in low regions of q^2, while the J/ψ and $\psi(2S)$ resonances were used for normalisation and cross-checks. Therefore, if there were an unknown source of systematic bias which does not cancel between these signal and control regions, all measurements could be affected, mimicking an NP effect. However, such a bias would be unlikely to affect regions of q^2 above $m^2(c\bar{c})$. A measurement of R_{K^+} in this region is under investigation [355]. The separation of $B^+ \to K^+e^+e^-$ decays from background is challenging here, due to leakage from $B^+ \to K^+\psi(2S)\,(e^+e^-)$ decays and kinematic sculpting of the combinatorial background. If this result is compatible with previous measurements, it will provide evidence against an unknown systematic effect.

In addition, lepton-universality ratios are being measured using $D^+_{(s)} \to \pi^+\phi\,(\ell^+\ell^-)$ decays. The tree-level decays $\phi \to \ell^+\ell^-$ are measured to be lepton-universal, and as $m(\phi) = 1020\,\text{MeV}$,

LHC run	Calendar Years	Detector config.	Max. $\mathcal{L}$ / cm^2 s^{-1}	Cumul. $\mathcal{L}_{Int.}$ / fb^{-1}
1-2	2010-'12, '15-'18	Original	4×10^{32}	9 fb^{-1}
3	2022-'25	Upgrade-1	2×10^{33}	23 fb^{-1}
4	2029-'32	Upgrade-1b	2×10^{33}	50 fb^{-1}
5 $\rightarrow$	2035 $\rightarrow$	Upgrade-2	2×10^{34}	300 fb^{-1}

Table 8.4: Anticipated long-term schedule for LHCb data-gathering operations. The calendar years for each data-taking run are shown [357], along with the detector configuration which will be used, and the maximum instantaneous luminosity [358]. The size of the *total* LHCb data sample expected to be accumulated by the end of each run is also shown. Note that each run is succeeded by a long shutdown assigned the same number (*e.g.* run-3 will be succeeded by LS3). This schedule has been repeatedly pushed back over the previous 3 years, primarily due to delays to upgrade activities caused by the COVID-19 pandemic [359–361].

this measurement will be a stringent cross-check of efficiency estimates in the low-q^2 region [356].

8.2.3 The future of LHCb

The LHCb physics programme is still in its adolescence, and much more data is set to be gathered over the coming years, as outlined in Table 8.4. This table shows the schedules for upcoming data-taking runs and long shutdowns, plus the instantaneous luminosities the detector is expected to operate at, and the volumes of data which will be recorded.

8.2.3.1 Upgrade-1

Many measurements at LHCb are statistically-limited, and would benefit from significantly increased data sample sizes. To gather such samples, the LHCb detector will operate with a five-fold increase in instantaneous luminosity when it re-starts for run-3 [362, 363]. However, preserving good detector performance with such luminosity presents some difficulties, due to the increased detector occupancy. In particular, a large volume of background events would pass the L0 hardware triggers - particularly the `L0Hadron` line - meaning HLT bandwidth would be overwhelmed. To deal with this, LHCb will move to a completely software-based trigger, using tracking information to impose more exclusive selection requirements for decays of interest. To implement this, the detector electronics have been replaced during LS2 to enable fast read-out, and many sub-detectors have been upgraded to increase their granularity. The ECAL hardware from run-1/2 is being re-used, though the SPD and PS have been removed. This may impact the

PID performance for electrons and neutral particles. However, some material upstream of the ECAL has been removed, which may reduce bremsstrahlung emission, and the software-based trigger may mean trigger efficiencies can be increased for electron-mode decays. Studies of improved PID algorithms for electrons and neutral particles in run-3 are ongoing [364, 354]. Low-energy pp collisions took place in October 2021, to test the upgraded LHCb system, and full data-taking is set to commence in mid-2022 [365].

8.2.3.2 Upgrade-2

A second upgrade of the LHCb detector has been proposed, which would allow it to operate with a further ten-fold increase in instantaneous luminosity, gathering $\sim 50\,\text{fb}^{-1}$ of data per year [366, 358]. To implement this, detector systems with further-increased granularity would be installed during LS4 (after some preliminary activities during LS3). In addition, precision timing capabilities would be added to all sub-detectors, enabling them to time signals within $\mathcal{O}(10-100\,\text{ps})$ windows, and hence differentiate between particles from the ~ 55 different PVs expected with each LHC bunch-crossing. A high-granularity ECAL would be installed, using a material which gives reduced radii for electromagnetic showers, such as tungsten. This is expected to restore the performance for electrons and neutral particles to the levels in run-1/2, though may bring further improvements beyond this. In addition, tracking stations would be added to the walls of the dipole magnet, improving the reconstruction of upstream track momenta.

8.2.3.3 Sensitivity estimates

Table 8.5 shows the statistical uncertainties anticipated for various R_X measurements at LHCb, with different volumes of data. Note that the measurements of $R_{K^0_S}$ and $R_{K^{*+}}$ only become statistically limited with $300\,\text{fb}^{-1}$ of data, so could offer considerable constraints to future models of NP. In addition, many of the systematic uncertainties listed in Chapter 5 would also be reduced with larger data samples. Others would not, however, and would require closer study (such as the non-factorisation of electron PID performance, and residual decay-model dependencies).

8.2.4 Other experiments

Independent verification is essential if the cause of the $b \to s\ell^+\ell^-$ anomalies is to be confirmed as a new physical effect. As discussed in the following sections, the most likely experiments to offer this are Belle II, ATLAS, and CMS.

	Lumi. / fb^{-1}				
Observable	3	9	23	50	300
R_{K^+}	$0.745 \pm 0.090 \pm 0.036$	0.043	0.025	0.017	0.007
$R_{K^{*0}}$	$0.69 \pm 0.11 \pm 0.05$	0.052	0.031	0.020	0.008
R_{ϕ}	-	0.130	0.076	0.050	0.020
R_{pK}	-	0.105	0.061	0.041	0.016
R_{π^+}	-	0.302	0.176	0.117	0.047
$R_{K^0_S}$	-	$0.66^{+0.20}_{-0.14}{}^{+0.02}_{-0.04}$	0.099	0.065	0.026
$R_{K^{*+}}$	-	$0.70^{+0.18}_{-0.13}{}^{+0.03}_{-0.04}$	0.091	0.060	0.024

Table 8.5: The expected statistical uncertainties for various R_X ratios with different volumes of LHCb data. The uncertainties are extrapolated from run-1/2 results (where central values and systematic uncertainties are also shown), or from muon-mode yields if no R_X result is yet available. Uniform detector response is assumed, while $\sigma(pp \to b\bar{b}X)$ is assumed to scale linearly with centre-of-mass energy. This table is taken from *Physics Case for an LHCb Upgrade II* [358], with the addition of extrapolations for $R_{K^0_S}$ and $R_{K^{*+}}$.

8.2.4.1 Belle II

After Belle ceased data-taking in 2008, a new detector - Belle II - was constructed [367, 10]. This detector is fed by an upgraded electron-positron accelerator capable of supplying ~ 40 times the instantaneous luminosity used at Belle. Physics data-recording began in 2019. Over the next decade, Belle II aim to record 50 ab^{-1} of data, roughly 50 times that recorded at Belle [368, 369]. This will enable many B-physics measurements with competitive sensitivity to LHCb. The new detector uses a similar design to Belle, with various upgrades to deal with the increased particle occupancy brought by increased luminosity. It includes a high-resolution electromagnetic calorimeter, providing similar reconstruction efficiencies for electrons as for muons. Therefore, measurements of R_X at Belle II are impacted by different systematic effects to LHCb, so will provide crucial cross-checks of LHCb results.

With the complete 50 ab^{-1} sample, measurements of R_{K^+} and $R_{K^{*+}}$ in the $q^2 \in [1, 6\,\text{GeV}^2]$ bin are anticipated to have total uncertainties of 4% and 3%, respectively [10, Tab. 5]. In addition, because the initial state is known in electron-positron collisions, Belle II will be able to measure an "inclusive" ratio R_{X_s}, using decays $B \to X_s \ell^+ \ell^-$ involving any hadronic final-state. To do this, dilepton candidates will be reconstructed, with q^2 calculated from the recoil between the B-meson and leptons. Similar statistical uncertainties to the "exclusive"

ratios are anticipated, and R_{X_s} would provide complementary information on potential NP effects.

8.2.4.2 ATLAS and CMS

The ATLAS and CMS collaborations both have active B-physics programmes [370, 371]. Extensive upgrades to both detectors are planned during LS3, to increase the instantaneous luminosities they operate at, with a view to eventually collecting $3000\,\text{fb}^{-1}$ of data each [372, 373]. Their measurements of $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$ and angular observables in $B^0 \to K^{*0}\mu^+\mu^-$ decays are projected to benefit greatly from this increased luminosity, as well as improved mass resolutions offered by new high-precision tracking systems [374–377]. However, they still fall well short of the sensitivities expected at LHCb.

In light of the anomalies seen at LHCb, both collaborations are making efforts towards R_X measurements of their own. The ATLAS collaboration have developed new trigger schemes for $B \to K^{(*)}e^+e^-$ decays [378]. These were deployed in 2018, and will also be used during run-3. Meanwhile, the CMS collaboration have recorded events containing pairs of B-hadrons, using a novel triggering technique to retain them for offline processing [379, 380]. In total, $\sim 1 \times 10^{10}$ B-hadron decays were recorded in 2018 using this approach, including $\mathcal{O}(1000)$ $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays. Further such samples are to be recorded in run-3.

Direct searches for particles such as leptoquarks and Z' bosons at CMS and ATLAS also provide useful constraints on models of NP which could explain the $b \to s\ell^+\ell^-$ anomalies [381, 382]. If these anomalies are confirmed as NP, such searches will be extremely important when characterising the nature of new particles, including at future experiments such as the proposed Future Circular Collider at CERN [383, 384].

8.3 Summary

In this dissertation, new measurements of the differential branching fractions for the decays $B^0 \to K^0e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ were presented, in bins of q^2. The first ever observations of these processes were reported. In addition, new measurements of the ratios $R_{K^0_S}$ and $R_{K^{*+}}$ were presented. These are the first such measurements of these decays at LHCb, and the most precise globally to date.

These new measurements of $R_{K^0_S}$ and $R_{K^{*+}}$ are both consistent with SM expectations, and are heavily statistically limited. Under certain assumptions, they lead to very mild increases in global tensions between measurements of $b \to s\ell^+\ell^-$ observables and SM expectations. They have been warmly received by theoretical flavour physicists, and have been incorporated into global fits and models of NP (see *e.g.* [385–388]).

With well-controlled systematic uncertainties, there are strong prospects for more precise measurements of $R_{K^0_S}$ and $R_{K^{*+}}$ in the future, using larger LHCb data samples. Time will tell whether these future measurements will remain consistent with SM predictions, or begin to exhibit the tensions currently seen in other $b \to s\ell^+\ell^-$ observables. We live in an exciting era for high-energy physics, where numerous tensions with SM predictions are becoming apparent. If these anomalies persist, it is possible that they are giving humanity its first glimpses of physics beyond the SM; a peek at the machinery of the universe at scales never seen before.

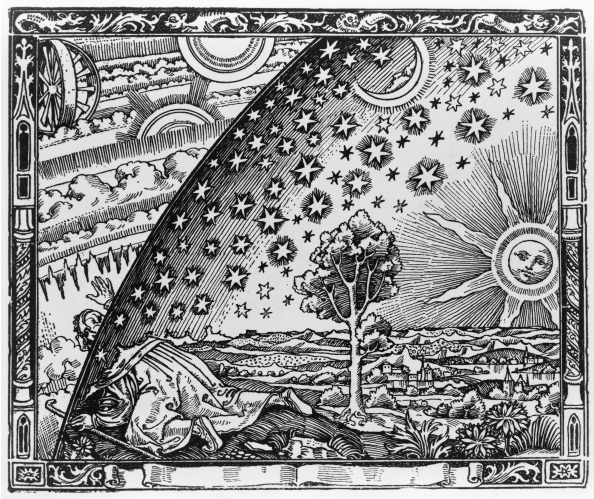


Fig. 8.5: "The Flammarion engraving" [389]

Part IV

Back Matter

Appendix A

Additional tables of systematic uncertainties

This appendix contains additional tables showing the systematic uncertainties for different measured observables. See Section 5.16 for details on the contents of these tables.

	$\sigma\left(\mathcal{B}\left(B^0 \to K^0_S e^+ e^-\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.33	0.36	0.24	0.55	0.26	0.17
PIDCalib Bin.	0.29	0.18	1.5	0.76	0.21	0.39
Elec. PID Fact.	0.64	0.25	1	0.59	1.4	0.82
Elec. Trk. Stats.	0.21	0.21	0.07	0.077	0.058	0.047
Gen. Stats.	2.7	2.3	0.73	0.94	0.62	0.53
KS Stats.	0.052	0.028	0.045	0.0044	0.078	0.026
L0 Stats.	0.23	0.29	0.094	0.11	0.051	0.061
L0I Tag	0.57	0.55	0.13	0.2	0.14	0.12
L0 Param.	1.5	0.8	0.16	0.44	0.02	0.2
BDT Mismodelling	0.87	1.8	0.21	0.93	0.61	0.41
Reco. Weight Stats.	1.8	5.3	1.1	2.2	0.83	1.1
q2 Smear. Stats.	0.39	0.43	0.3	0.26	0.24	0.14
Decay Model	0.21	0.26	0.28	0.082	0.16	0.19
Cascade Back.	-	-	-	-	-	0.41
Part. Reco. $m(K^{(*)})$	-	-	-	-	-	2.2
Leakage KDE	-	-	-	-	-	0.094
Sig. PDF MC Stats.	-	-	-	-	-	1.4
Total of above	3.8	6.2	2.3	2.8	1.9	3.1
Sig. MC Stats	6.6	7.9	4.1	4.6	3.2	2.1

Table A.1: Summary of systematic uncertainties on $\mathcal{B}(B^0 \to K^0 e^+ e^-)$.

	$\sigma\left(\mathcal{B}\left(B^+ \to K^{*+} e^+ e^-\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.76	0.33	0.29	0.63	0.23	0.19
PIDCalib Bin.	0.29	0.25	0.76	0.48	0.3	0.23
Elec. PID Fact.	0.26	0.23	1.4	0.36	0.76	0.57
Elec. Trk. Stats.	0.22	0.17	0.084	0.082	0.056	0.045
Gen. Stats.	2.9	4.2	4.7	1.8	1.2	1.4
KS Stats.	0.34	0.19	0.041	0.0046	0.016	0.04
L0 Stats.	0.33	0.45	0.16	0.17	0.13	0.1
L0I Tag	0.72	0.94	0.17	0.29	0.16	0.18
L0 Param.	0.31	0.047	0.061	0.14	0.25	0.087
BDT Mismodelling	0.12	0.5	0.58	0.74	0.2	0.24
Reco. Weight Stats.	1.2	1.7	2	1.6	0.85	0.7
q2 Smear. Stats.	0.33	0.68	0.35	0.23	0.13	0.15
Decay Model	0.49	1	0.57	0.31	0.32	0.27
Part. Reco. $m(K^{(*)})$	-	-	-	-	-	1
Part. Reco. KDE	-	-	-	-	-	1.1
Leakage KDE	-	-	-	-	-	1.1
Sig. PDF MC Stats.	-	-	-	-	-	2.1
Total of above	3.5	4.9	5.5	2.7	1.8	3.3
Sig. MC Stats	5.9	7.3	6.4	5.3	3.3	2.4

Table A.2: Summary of systematic uncertainties on $\mathcal{B}(B^+ \to K^{*+} e^+ e^-)$.

	$\sigma\left(\mathcal{B}\left(B^0 \to K^0_S \mu^+ \mu^-\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.24	0.12	0.033	0.02	0.54	0.17
PIDCalib Bin.	0.16	0.082	0.06	0.033	0.033	0.026
Muon Trk.	0.048	0.11	0.21	0.058	0.014	0.054
Gen. Stats.	1.5	1.7	0.8	0.97	0.38	0.43
KS Stats.	0.044	0.015	0.051	0.034	0.025	0.016
L0 Stats.	0.095	0.17	0.18	0.13	0.045	0.06
L0 Param.	0.15	0.00098	0.018	0.23	0.029	0.057
BDT Mismodelling	0.12	0.13	0.16	0.17	0.11	0.067
Reco. Weight Stats.	1.6	1.7	0.47	0.5	0.25	0.35
q2 Smear. Stats.	0.32	0.24	0.2	0.15	0.12	0.083
Decay Model	0.048	0.085	0.15	0.025	0.011	0.055
Sig. PDF MC Stats.	-	-	-	-	-	0.52
Total of above	2.3	2.5	1	1.1	0.73	0.79
Sig. MC Stats	5.4	4.6	4	2.9	2.2	1.5

Table A.3: Summary of systematic uncertainties on $\mathcal{B}(B^0 \to K^0 \mu^+ \mu^-)$.

	$\sigma\left(\mathcal{B}\left(B^+ \to K^{*+} \mu^+ \mu^-\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.17	0.15	0.036	0.022	0.014	0.029
PIDCalib Bin.	0.13	0.38	0.08	0.028	0.018	0.065
Muon Trk.	0.16	0.077	0.25	0.08	0.069	0.068
Gen. Stats.	2.8	2.2	2.1	1.2	0.72	0.72
KS Stats.	0.28	0.034	0.039	0.012	0.019	0.023
L0 Stats.	0.19	0.26	0.35	0.38	0.087	0.13
L0 Param.	0.01	0.13	0.11	0.28	0.047	0.075
BDT Mismodelling	0.09	0.076	0.31	0.28	0.16	0.11
Reco. Weight Stats.	1.1	1	0.67	0.34	0.26	0.26
q2 Smear. Stats.	0.32	0.17	0.11	0.21	0.058	0.07
Decay Model	1.2	0.72	0.98	0.93	1.1	0.96
Sig. PDF MC Stats.	-	-	-	-	-	0.53
Total of above	3.3	2.6	2.5	1.7	1.3	1.4
Sig. MC Stats	5	5.6	3.4	2.6	2	1.4

Table A.4: Summary of systematic uncertainties on $\mathcal{B}(B^+ \to K^{*+} \mu^+ \mu^-)$.

	$\sigma\left(r_{J/\psi}^{-1}(B^0)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.41	0.44	0.3	0.38	0.65	0.24
PIDCalib Bin.	0.28	0.21	1.3	0.85	0.5	0.39
Elec. PID Fact.	0.1	0.31	0.33	0.29	0.16	0.2
Elec. Trk. Stats.	0.53	0.34	0.24	0.21	0.22	0.12
Muon Trk.	0.84	1	2.2	0.91	1.1	0.68
Gen. Stats.	1.7	4.1	0.8	0.85	1.8	0.89
KS Stats.	0.00086	0.093	0.08	0.05	0.052	0.031
L0 Stats.	1.4	1.7	1.3	1.3	0.68	0.56
L0I Tag	2.2	2	0.55	0.79	0.74	0.48
L0 Param.	0.6	0.025	0.69	0.61	0.076	0.22
BDT Mismodelling	2.5	5.2	5.2	3.3	1.6	1.7
Reco. Weight Stats.	2.7	5.5	1.4	2.7	1.8	1.3
q2 Smear. Stats.	0.35	0.15	0.036	0.033	0.068	0.049
Sig. PDF MC Stats.	-	-	-	-	-	0.34
Total of above	5	9.1	6.2	4.8	3.4	2.6
Sig. MC Stats	4.1	4.3	1.5	2.3	1.9	1.1

Table A.5: Summary of systematic uncertainties on $r_{J/\psi}^{-1}(B^0)$.

	$\sigma\left(r_{J/\psi}^{-1}(B+)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.48	0.43	0.26	0.42	0.26	0.16
PIDCalib Bin.	0.28	0.14	0.78	0.9	0.42	0.31
Elec. PID Fact.	0.25	0.072	0.21	0.38	0.2	0.17
Elec. Trk. Stats.	0.55	0.35	0.23	0.21	0.19	0.12
Muon Trk.	0.84	1	2.2	0.87	1.1	0.67
Gen. Stats.	2.5	2	2.3	1.1	0.93	0.75
KS Stats.	0.039	0.063	0.24	0.17	0.12	0.076
L0 Stats.	1.7	1.8	1.3	1.2	0.74	0.57
L0I Tag	1.9	1.8	0.57	0.7	0.81	0.45
L0 Param.	0.38	0.62	0.48	0.49	0.087	0.19
BDT Mismodelling	2.3	1.1	6.3	3.9	5.7	2.4
Reco. Weight Stats.	1.9	1.2	2.5	2.4	2.1	1
q2 Smear. Stats.	0.21	0.18	0.13	0.1	0.072	0.057
Sig. PDF MC Stats.	-	-	-	-	-	0.66
Total of above	4.8	3.9	7.7	5.1	6.4	3
Sig. MC Stats	3.7	2.9	3.4	2.6	2.3	1.3

Table A.6: Summary of systematic uncertainties on $r_{J/\psi}^{-1}(B^+)$.

	$\sigma\left(R^{-1}_{\psi(2S)}\left(B^0\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.31	0.5	0.21	0.26	0.61	0.22
PIDCalib Bin.	0.71	0.16	0.57	0.29	0.17	0.17
Elec. PID Fact.	0.26	0.29	0.41	0.29	0.23	0.17
Elec. Trk. Stats.	0.16	0.1	0.057	0.049	0.046	0.03
Muon Trk.	0.15	0.11	0.21	0.048	0.074	0.062
Gen. Stats.	2.5	2.5	0.65	0.71	0.56	0.52
KS Stats.	0.2	0.031	0.018	0.01	0.0094	0.016
L0 Stats.	0.31	0.56	0.41	0.28	0.13	0.15
L0I Tag	0.64	0.57	0.19	0.19	0.22	0.14
L0 Param.	0.59	0.34	0.39	0.44	0.076	0.16
BDT Mismodelling	0.46	0.65	0.23	0.28	0.25	0.16
Reco. Weight Stats.	1.8	3.5	0.61	1.6	1.2	0.79
q2 Smear. Stats.	0.72	0.26	0.18	0.097	0.11	0.095
Sig. PDF MC Stats.	-	-	-	-	-	0.34
Total of above	3.5	4.5	1.3	1.9	1.5	1.1
Sig. MC Stats	5.9	5.6	3.3	3.3	2.7	1.6

Table A.7: Summary of systematic uncertainties on $R^{-1}_{\psi(2S)}(B^0)$.

	$\sigma\left(R^{-1}_{\psi(2S)}\left(B+\right)\right)/\%$					
	2011	2012	2016	2017	2018	Ave.
PIDCalib Stats.	0.31	0.17	0.17	0.29	0.22	0.11
PIDCalib Bin.	0.26	0.3	0.63	0.26	0.26	0.18
Elec. PID Fact.	0.28	0.091	0.25	0.45	0.29	0.2
Elec. Trk. Stats.	0.14	0.095	0.059	0.046	0.047	0.029
Muon Trk.	0.2	0.42	0.3	0.14	0.18	0.12
Gen. Stats.	2.7	2.5	2.3	1.1	1.4	0.84
KS Stats.	0.054	0.0062	0.0074	0.0033	0.072	0.022
L0 Stats.	0.31	0.33	0.48	0.45	0.16	0.17
L0I Tag	0.38	0.35	0.17	0.18	0.21	0.11
L0 Param.	0.38	0.63	0.28	0.19	0.096	0.13
BDT Mismodelling	0.16	0.26	0.22	0.37	0.23	0.13
Reco. Weight Stats.	1	0.91	1.2	0.79	0.93	0.46
q2 Smear. Stats.	0.37	0.31	0.27	0.19	0.14	0.099
Sig. PDF MC Stats.	-	-	-	-	-	0.66
Total of above	3.1	2.9	2.7	1.6	1.8	1.2
Sig. MC Stats	5.3	4.4	4.6	3.7	3.3	1.9

Table A.8: Summary of systematic uncertainties on $R^{-1}_{\psi(2S)}(B^+)$.

Appendix B

Extra differential plots of $r_{J/\psi}^{-1}$

As outlined in Section 6.1.1, $r_{J/\psi}^{-1}$ is measured in bins of sixteen different variables, to test for mismodelling of selection efficiencies. Plots which were not given in that section are shown in this appendix, with Fig. B.1 and B.2 showing the plots from $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$ and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ data respectively.

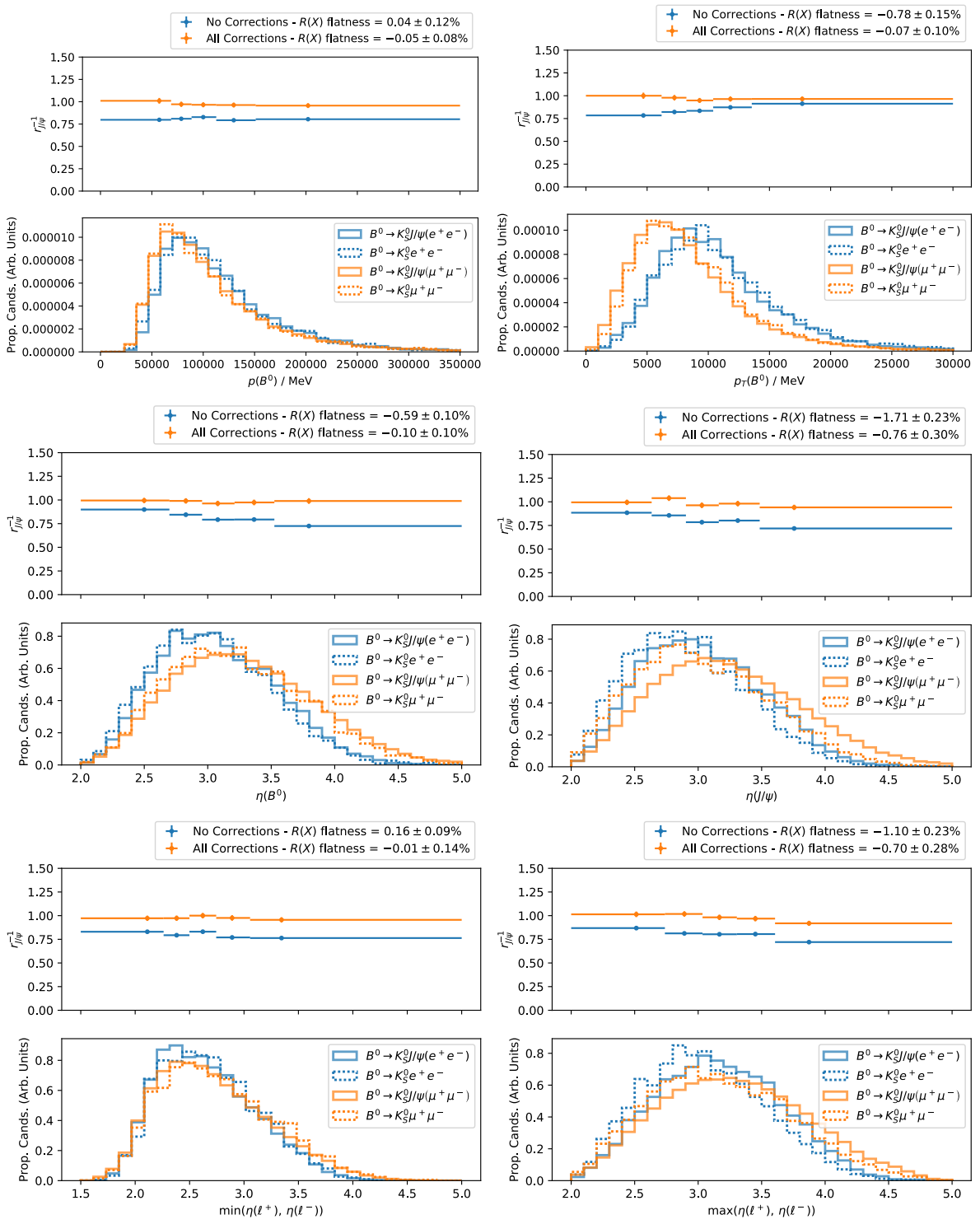


Fig. B.1: Differential measurements of $r^{-1}_{J/\psi}$ across different variables, found from fits to $B^0 \to J/\psi(\ell^+\ell^-)K^0_S$ data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots.

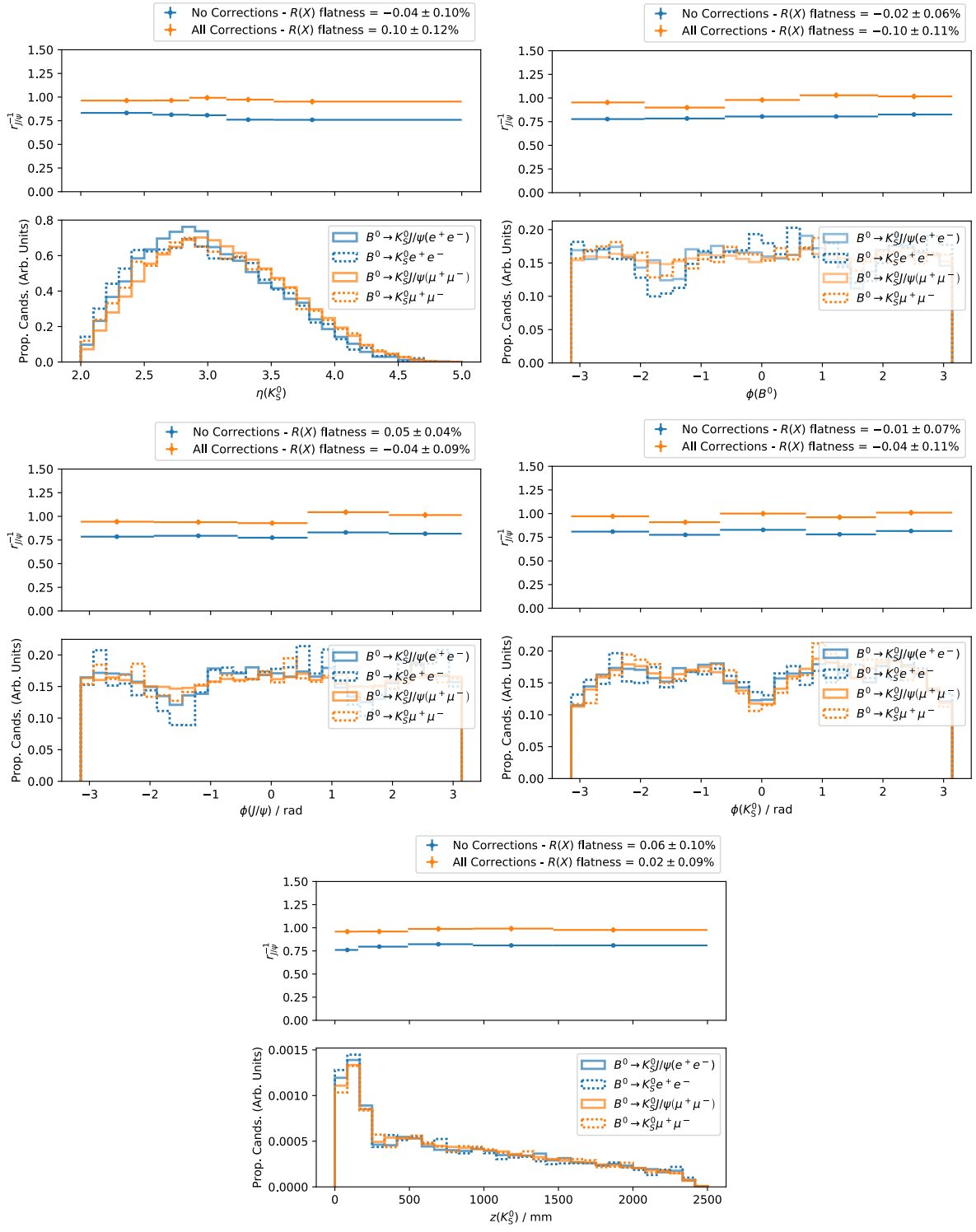


Fig. B.1: Differential measurements of $r_{J/\psi}^{-1}$ across different variables, found from fits to $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$ data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

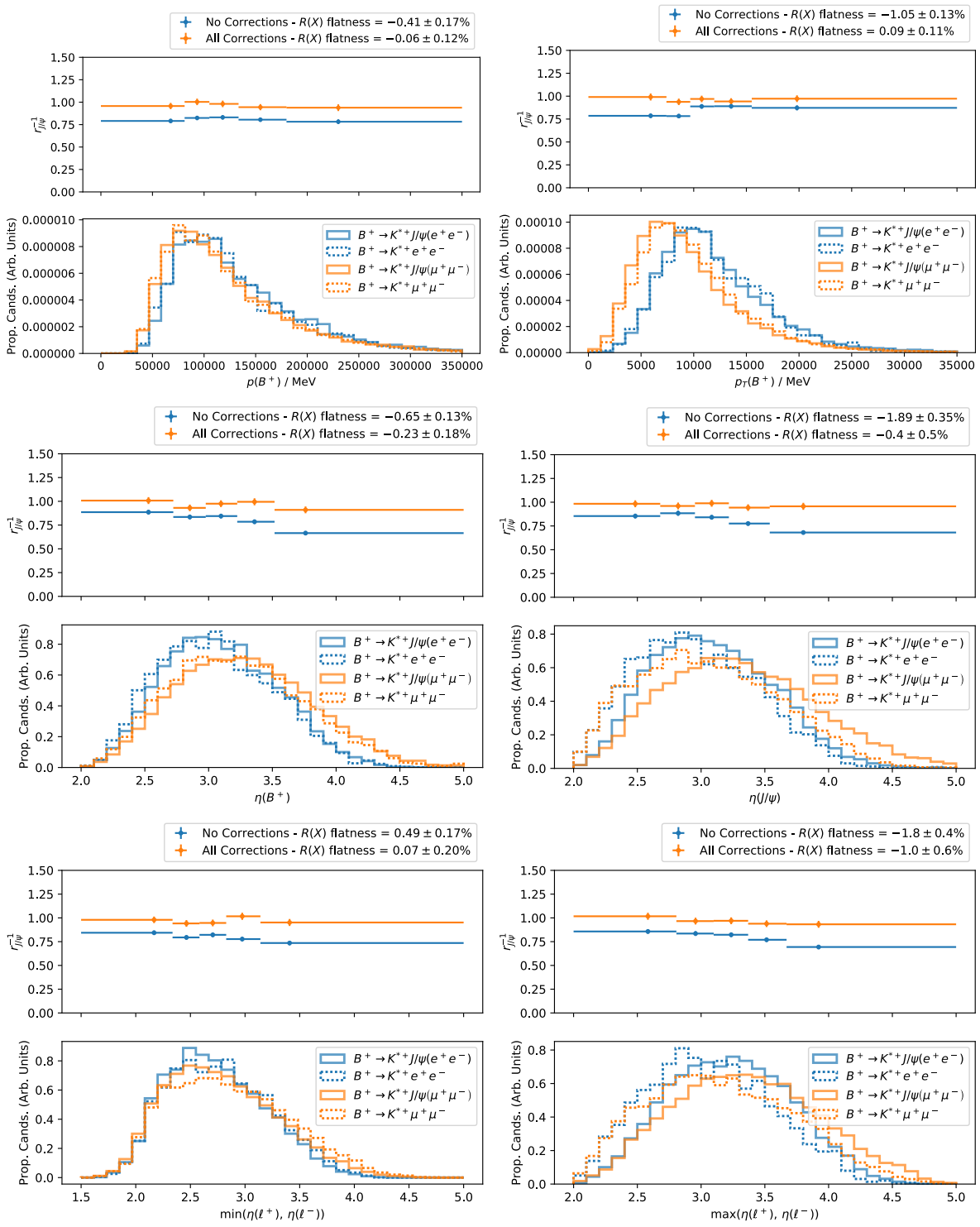


Fig. B.2: Differential measurements of $r^{-1}_{J/\psi}$ across different variables, found from fits to $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots.

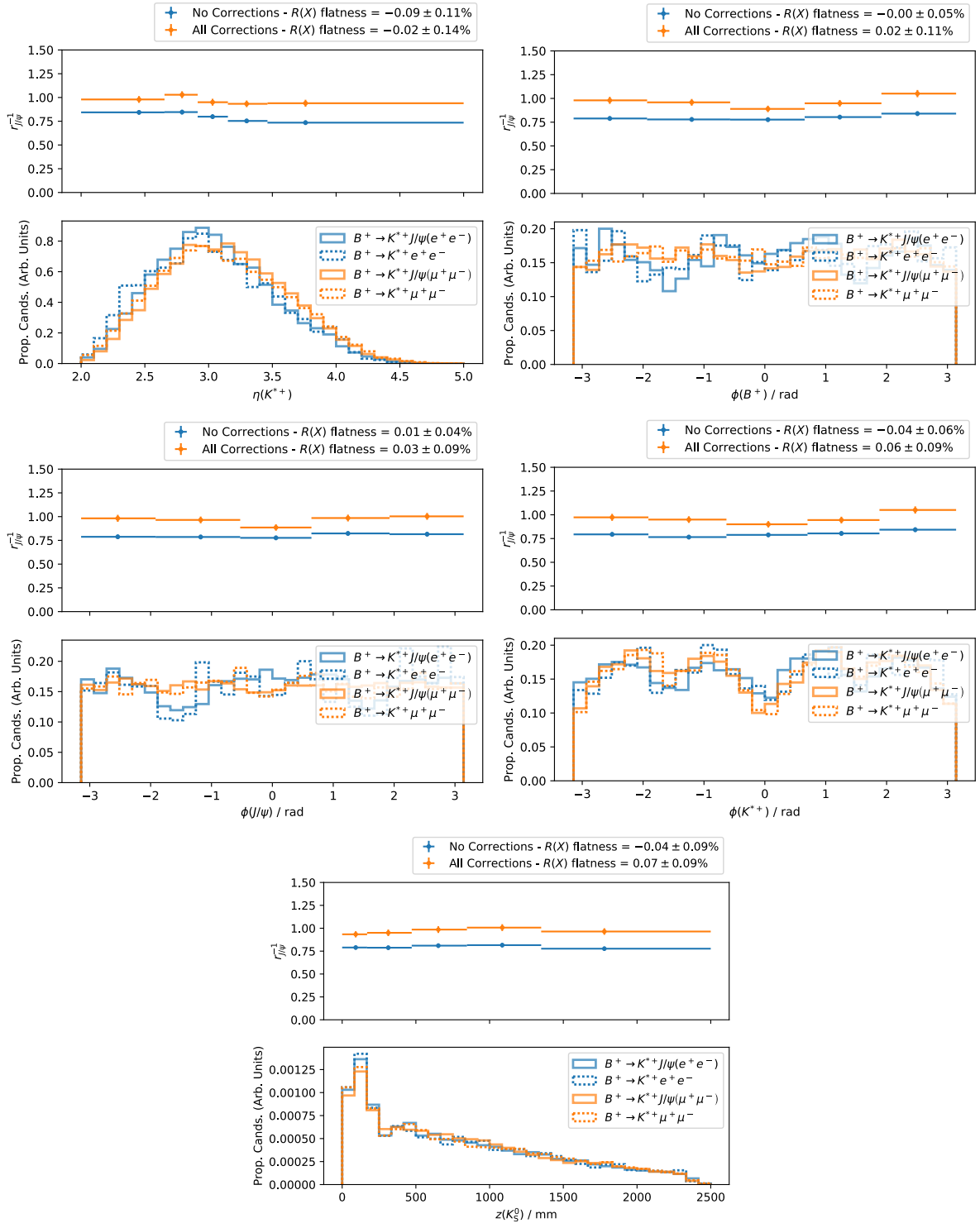


Fig. B.2: Differential measurements of $r^{-1}_{J/\psi}$ across different variables, found from fits to $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

Appendix C

Differential plots of efficiency-adjusted control-mode yields

As outlined in Section 6.1.1, the yields of each $B \to J/\psi\,(\ell^+\ell^-)\,K^{(*)}$ control mode, normalised by the selection efficiencies, are studied in bins of sixteen variables which describe the kinematics and geometries of particles in the decay chain. Any trends in these efficiency-adjusted yields could indicate mismodelling of the selection efficiencies, and the resulting biases on differential branching fraction measurements are quantified with a flatness parameter calculated using eq. 6.2. These plots of efficiency-adjusted yields are given in this appendix, in Fig. C.1 and C.2 for $B^0 \to J/\psi\,(\ell^+\ell^-)\,K^0_S$ and $B^+ \to J/\psi\,(\ell^+\ell^-)\,K^{*+}$ decays respectively.

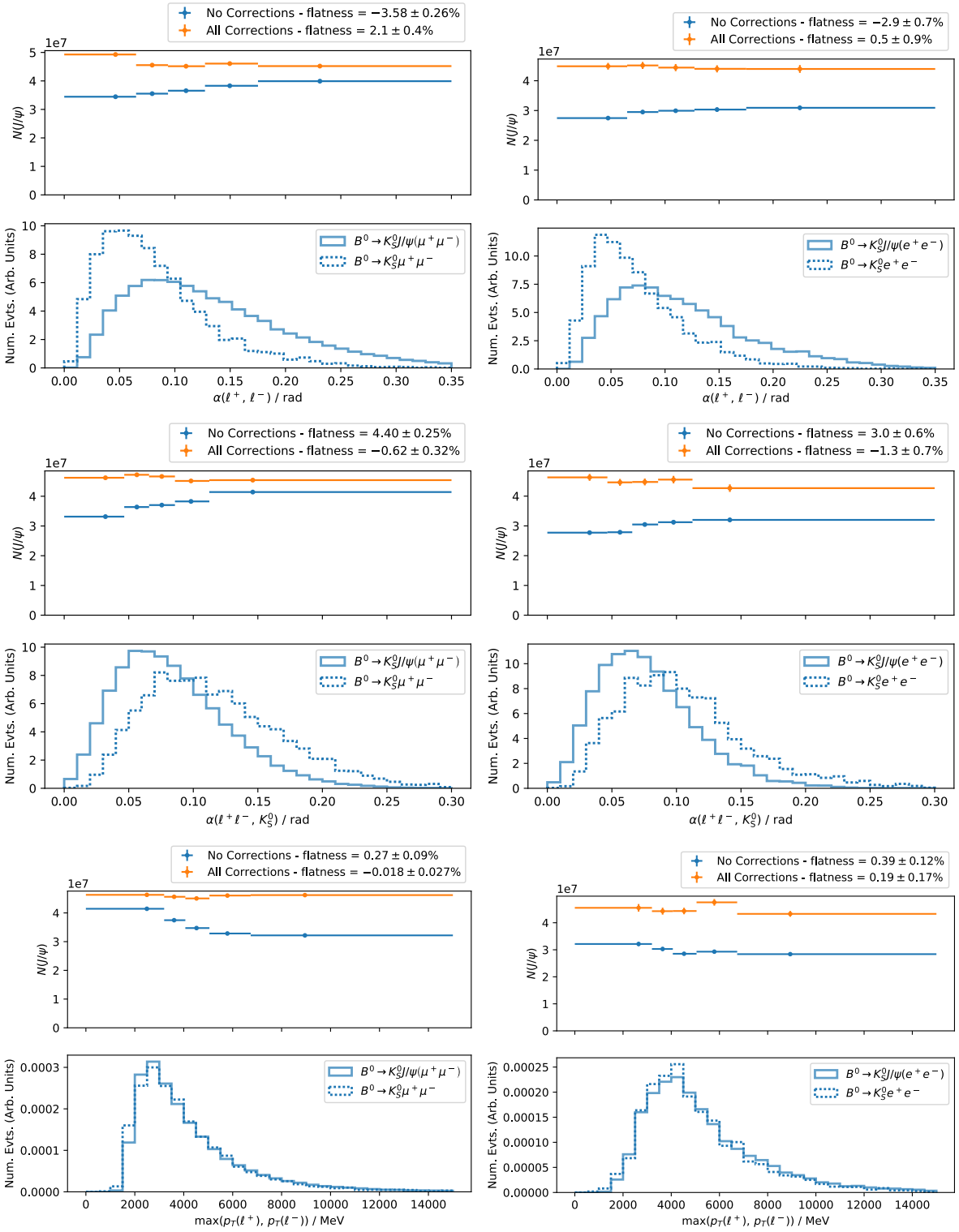


Fig. C.1: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots.

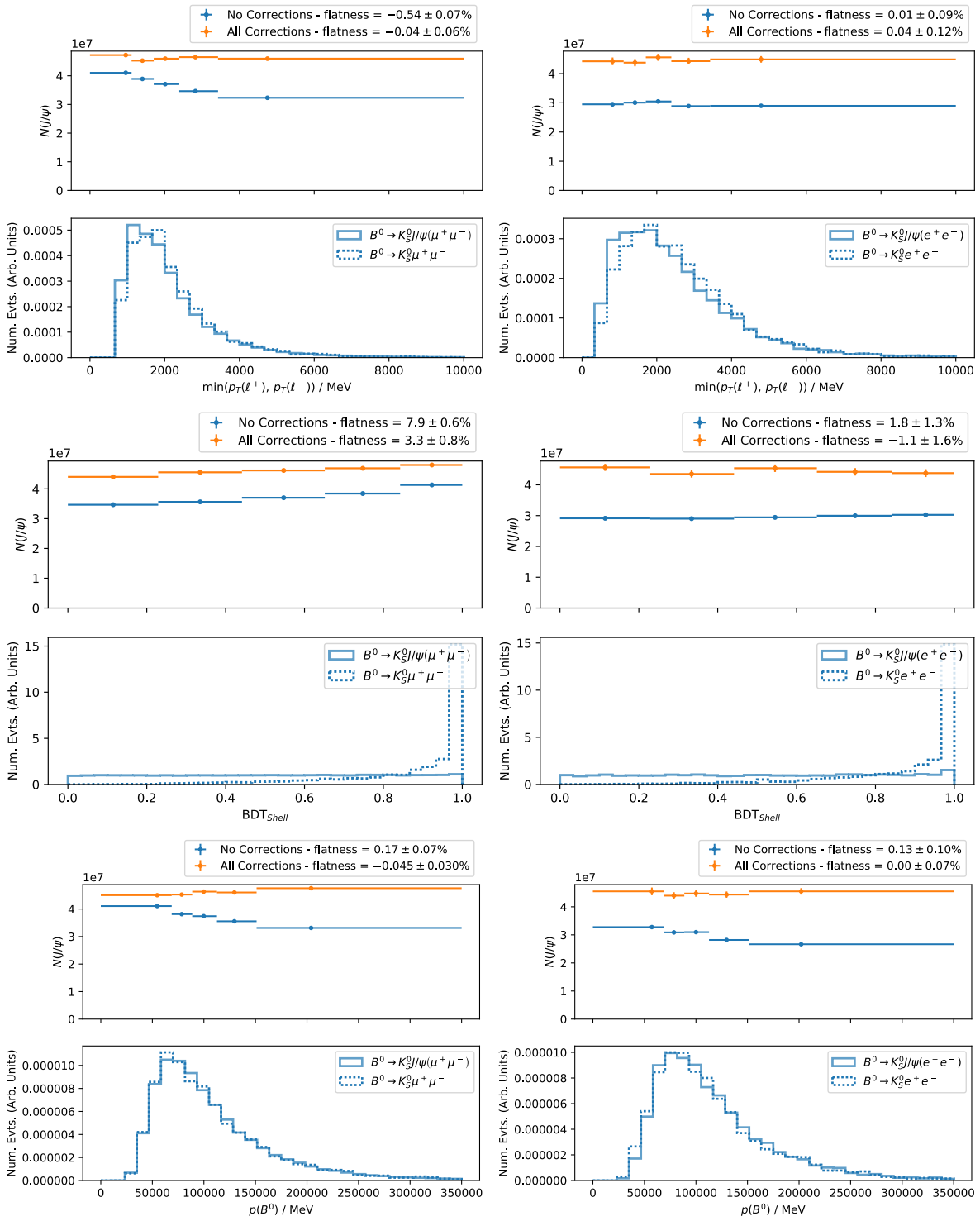


Fig. C.1: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

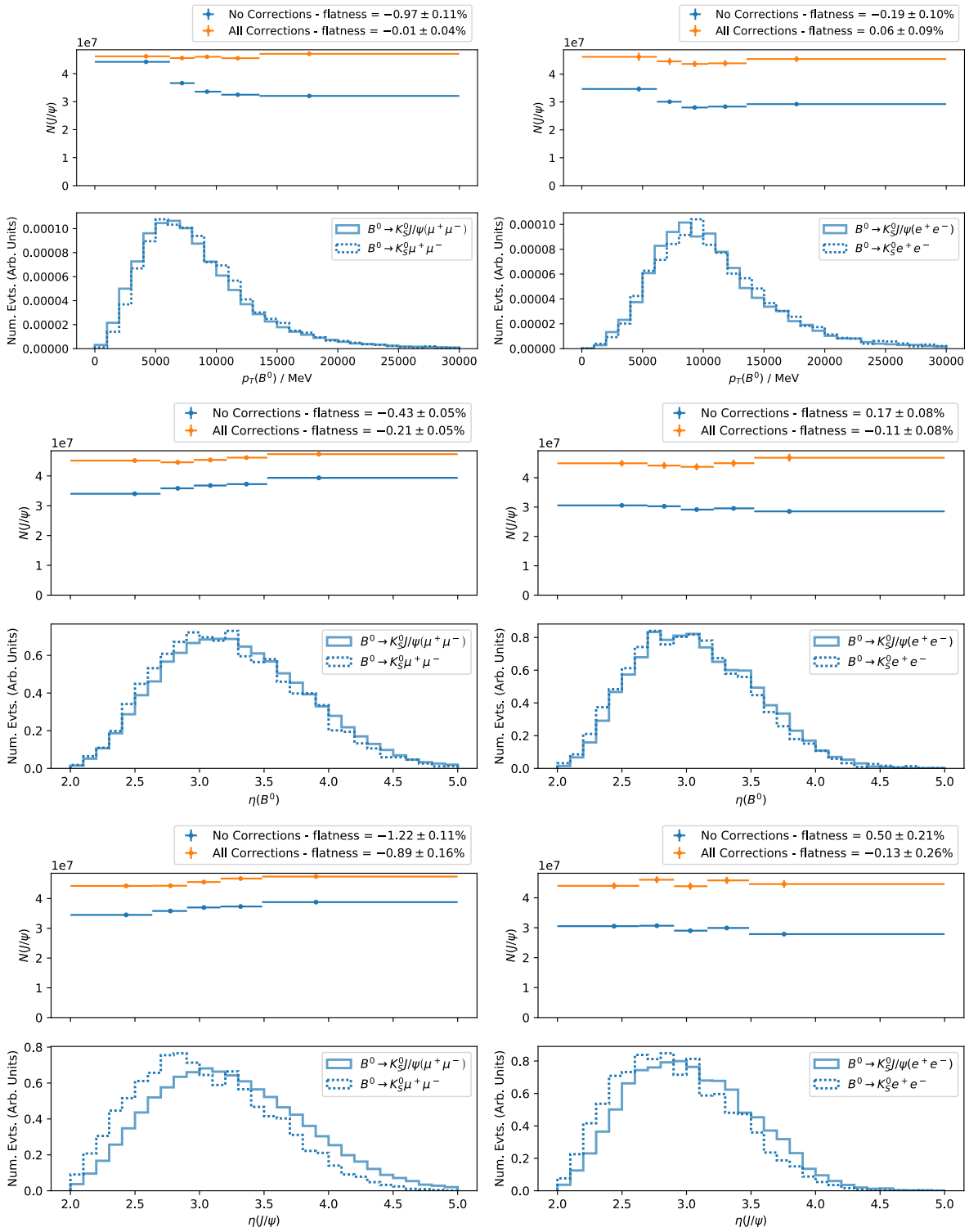


Fig. C.1: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

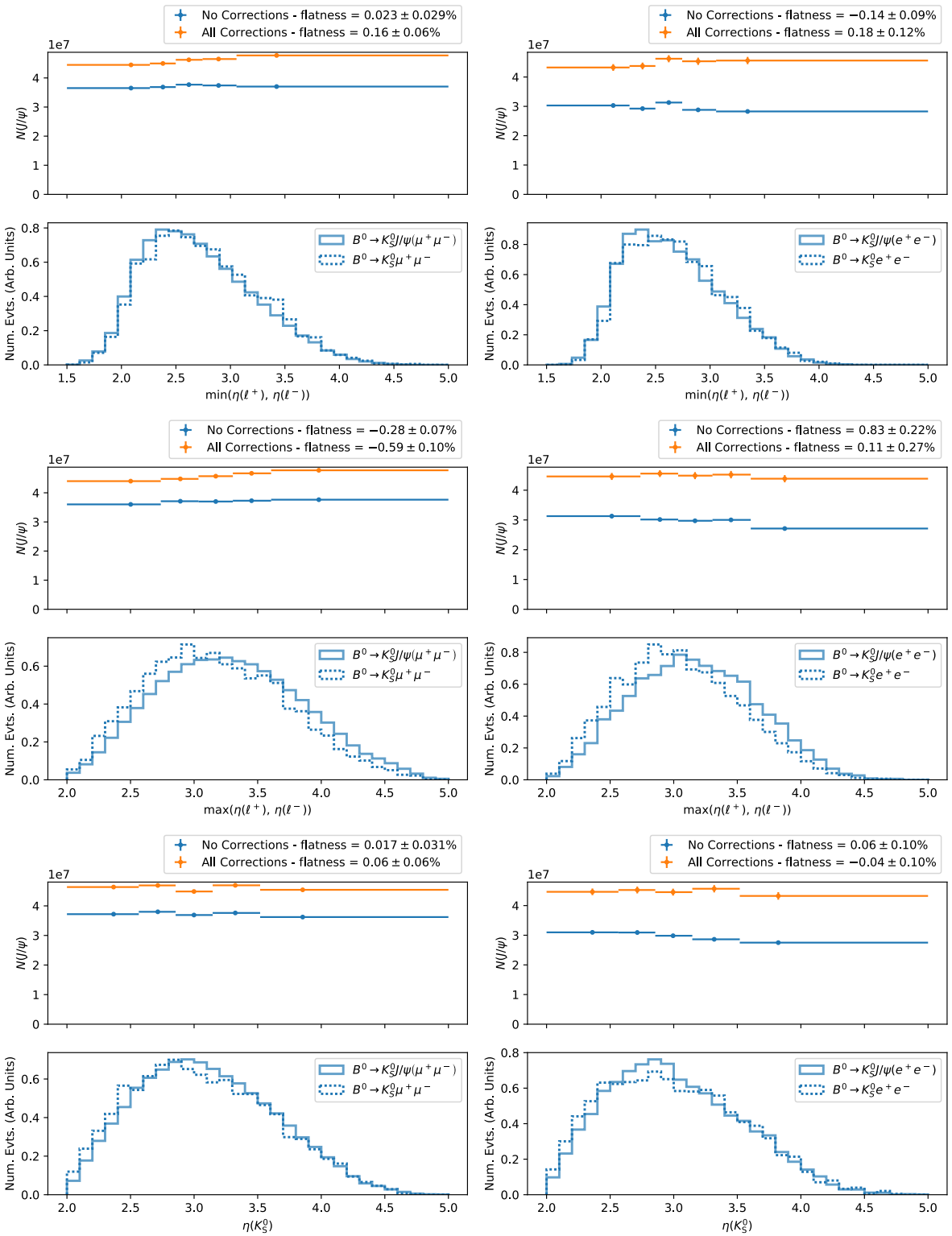


Fig. C.1: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

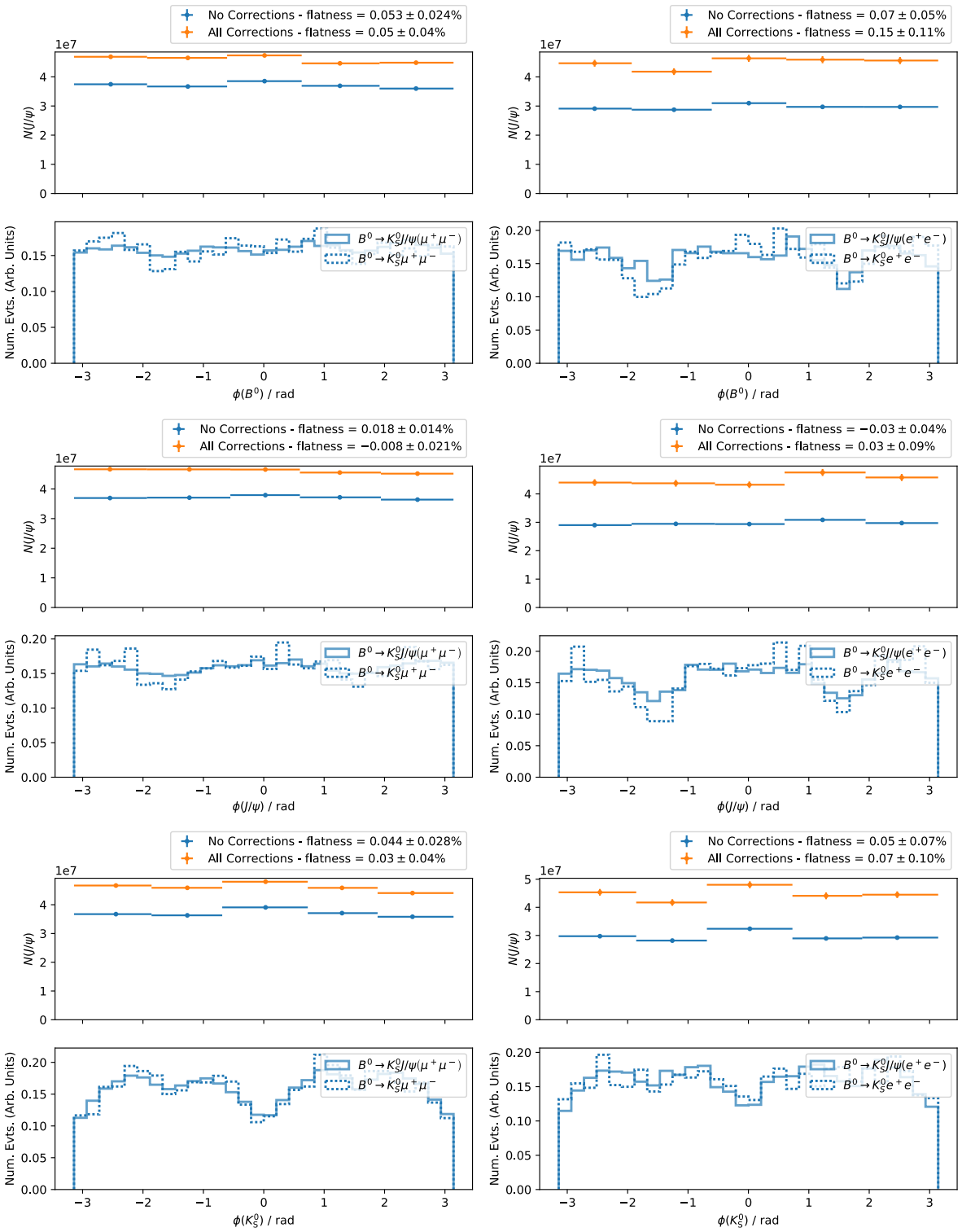


Fig. C.1: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

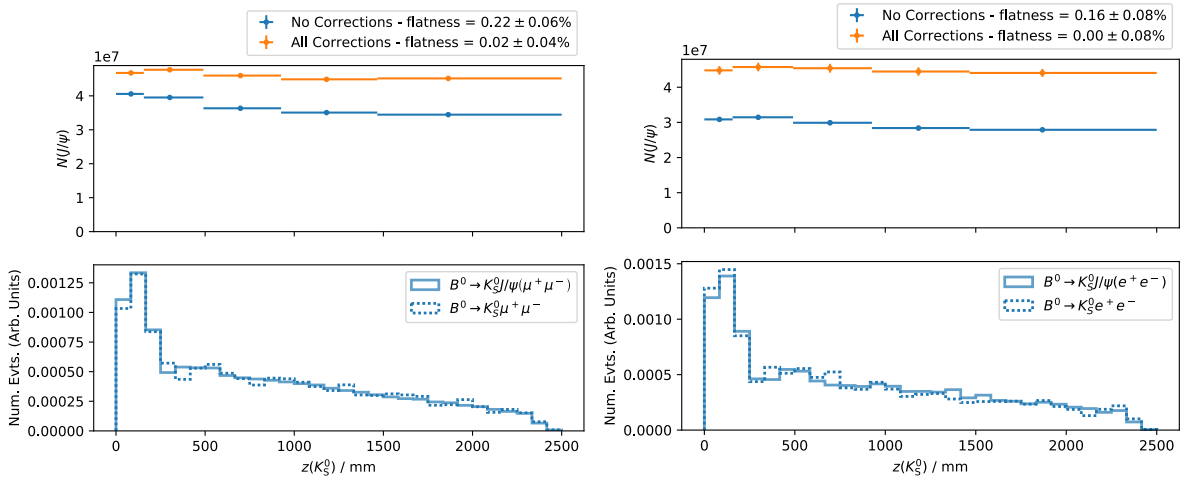


Fig. C.1: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^0 \to J/\psi\,(\mu^+\mu^-)\,K^0_S$ (left) and $B^0 \to J/\psi\,(e^+e^-)\,K^0_S$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

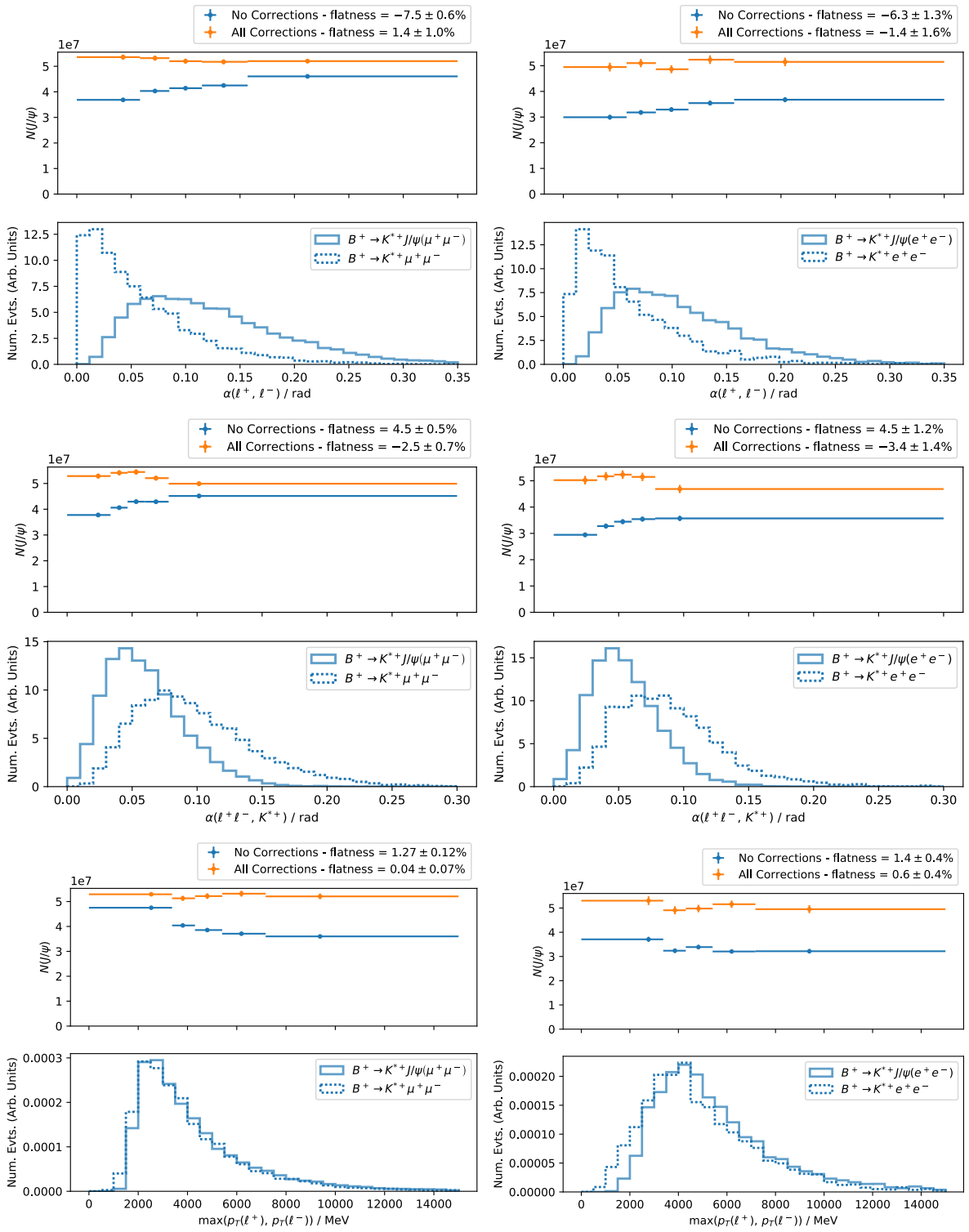


Fig. C.2: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots.

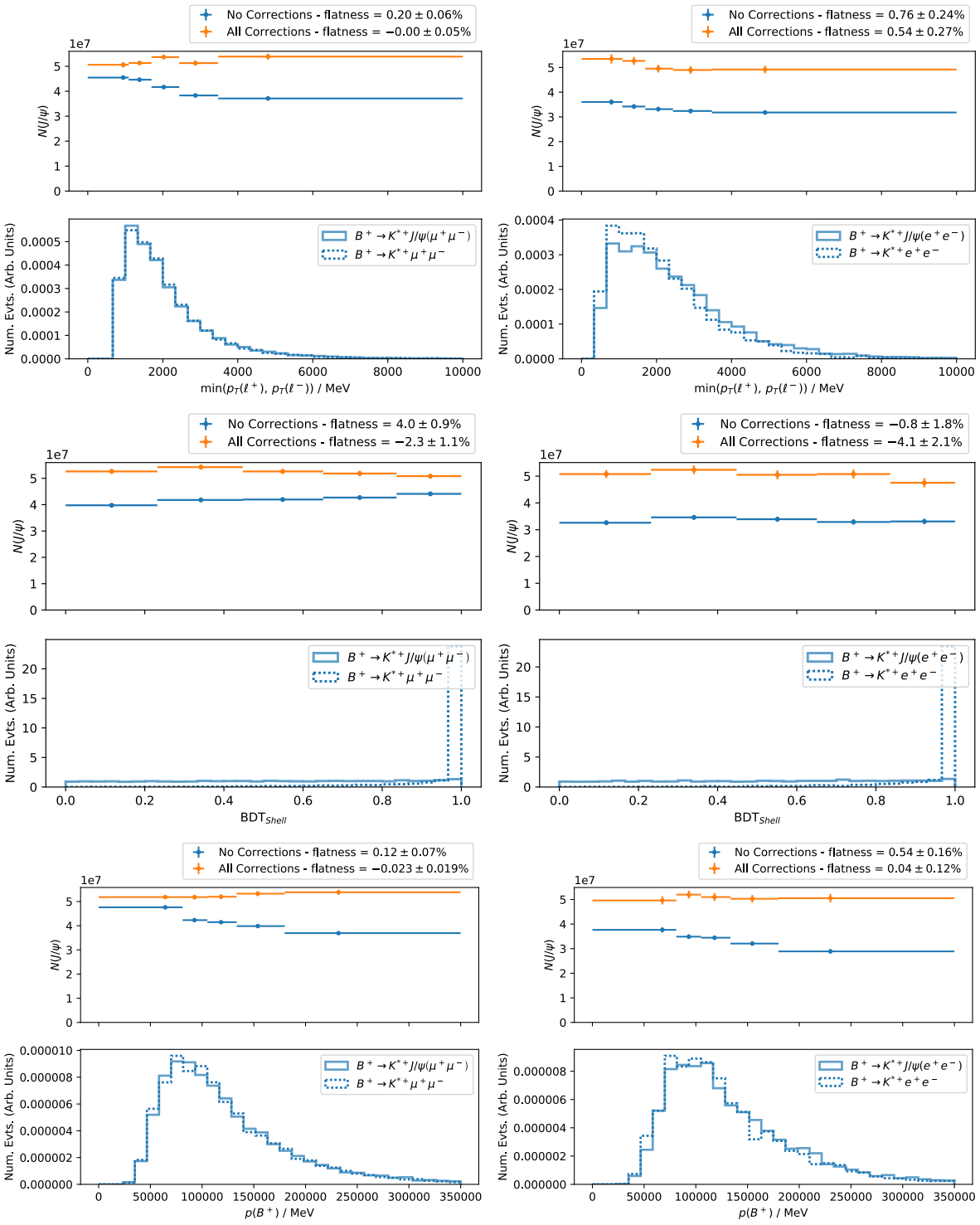


Fig. C.2: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

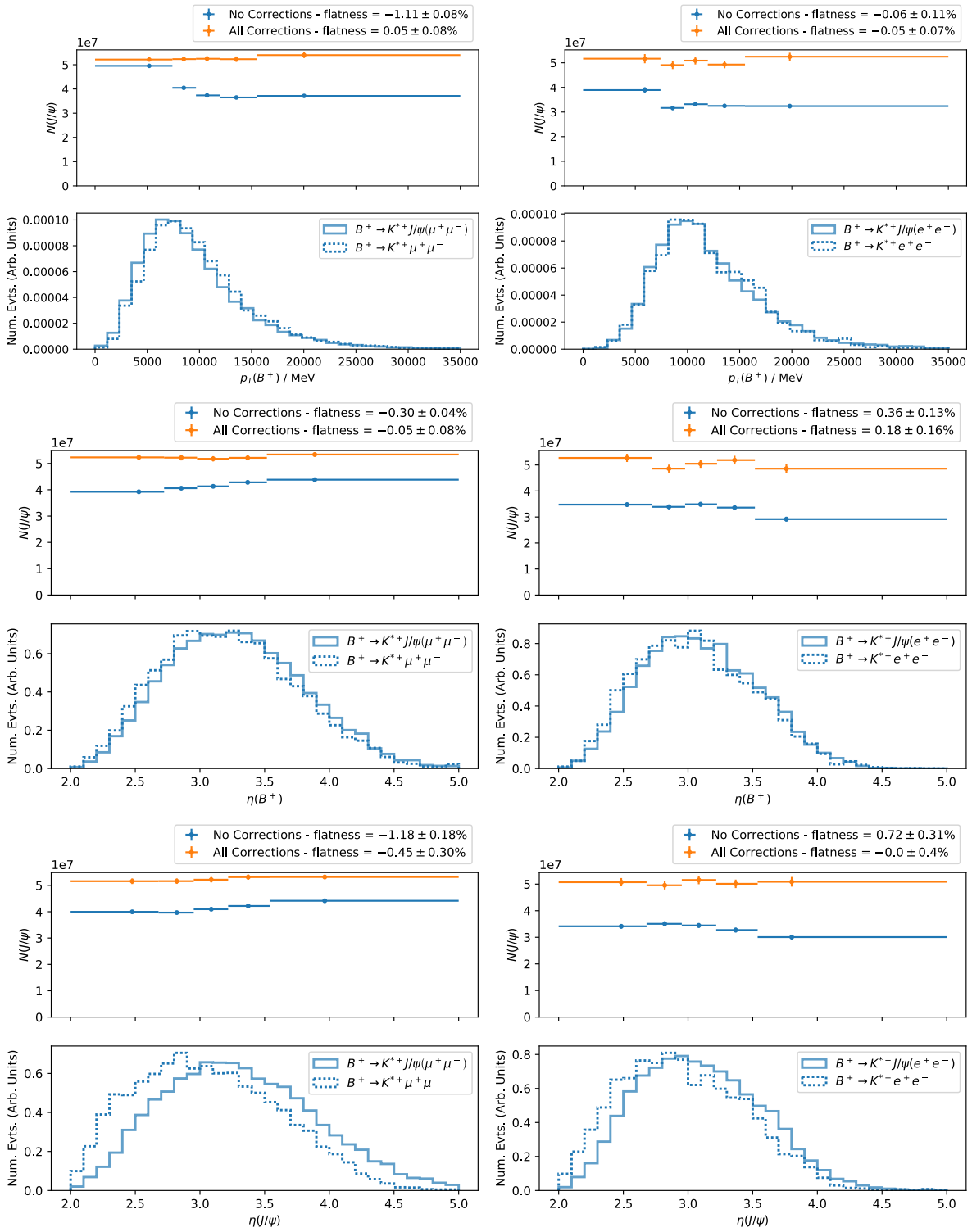


Fig. C.2: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

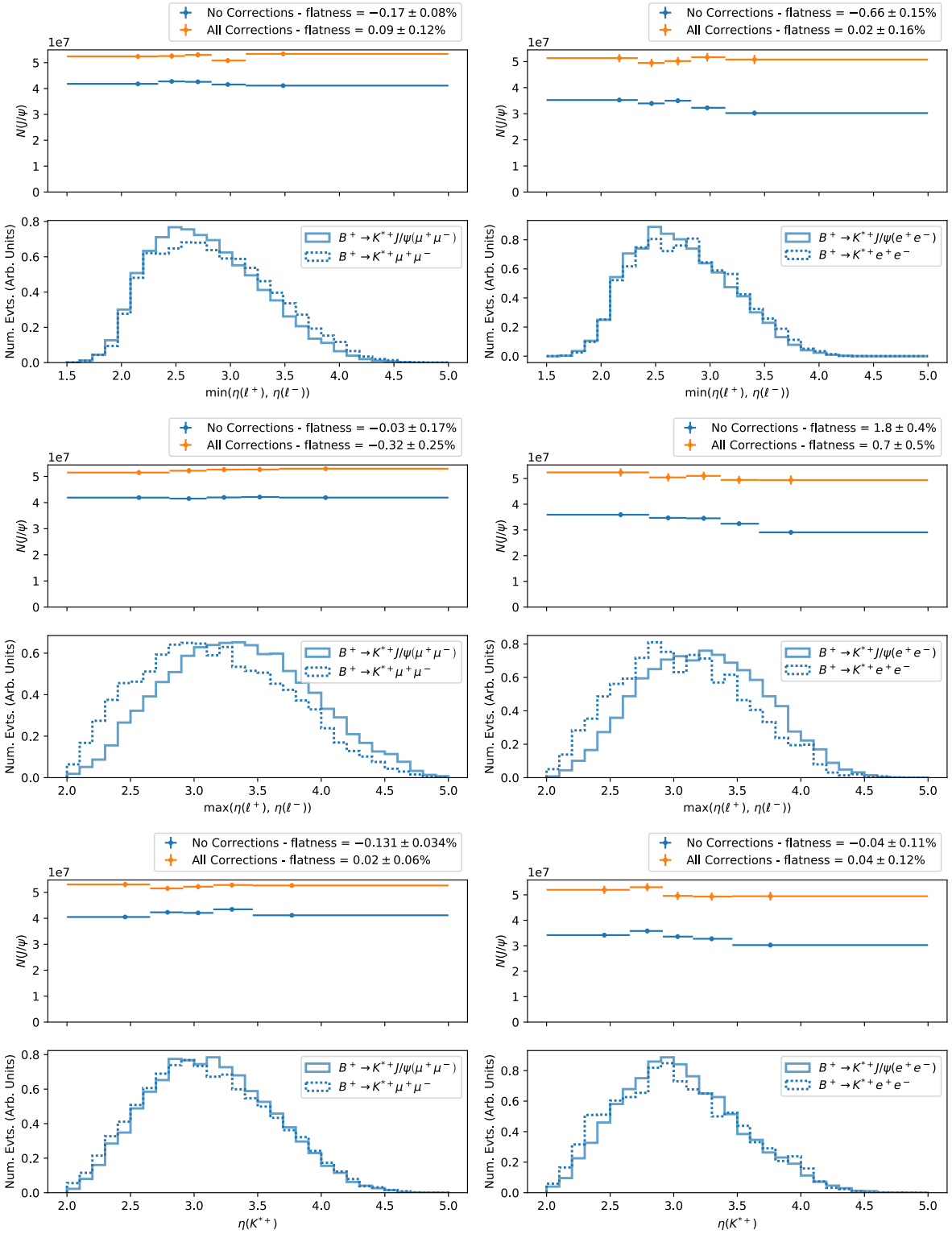


Fig. C.2: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

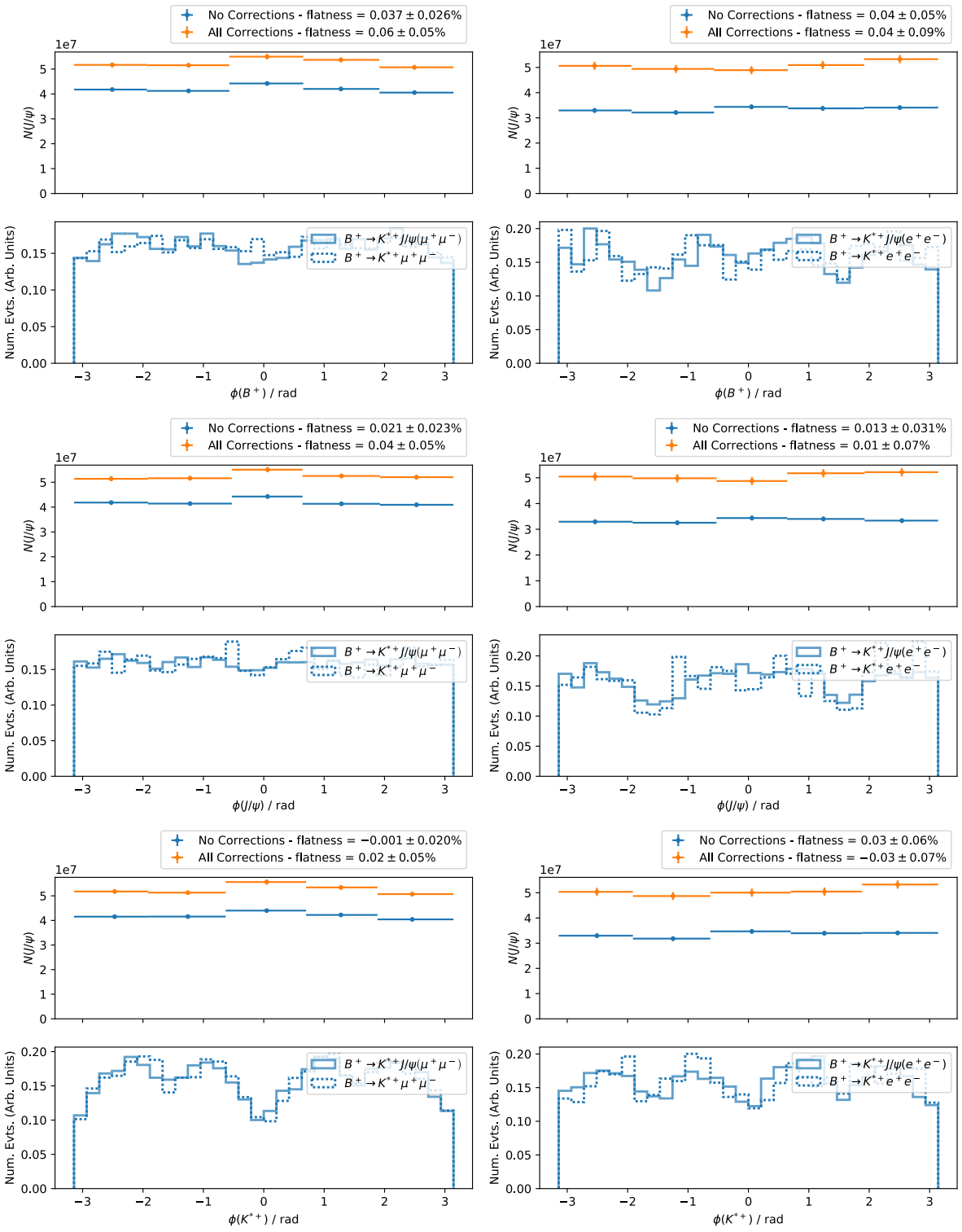


Fig. C.2: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

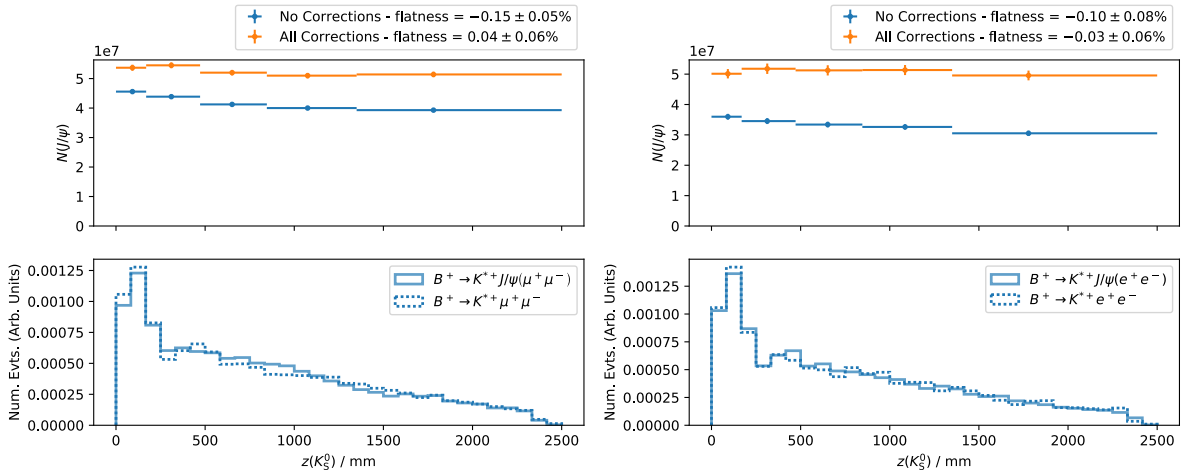


Fig. C.2: Differential measurements of efficiency-adjusted yields across different variables, found from fits to $B^+ \to J/\psi\,(\mu^+\mu^-)\,K^{*+}$ (left) and $B^+ \to J/\psi\,(e^+e^-)\,K^{*+}$ (right) data, across all years, with different weighting schemes applied when calculating efficiencies (top halves of plots), and the distributions of binned variables in the fully-selected and corrected rare-mode and control-mode MC (bottom halves of plots). Flatness parameter values are shown in the legends above the plots. *(Cont.)*

References

[1] J. Smeaton and L. Christie, *AI and healthcare*, POST-PN-0637, Parliamentary Office of Science and Technology (POST), London, UK, 2021.

[2] M. Thomson, *Modern Particle Physics*, Cambridge University Press, Cambridge, 2013.

[3] J. Smeaton, *New lepton universality results from LHCb and other rare decays*, in *32nd Rencontres de Blois on Particle Physics and Cosmology*, (Blois, France), 2021. LHCb-TALK-2021-304, on behalf of the LHCb Collaboration.

[4] H. V. Cliff, *New tests of lepton universality using rare B decays with K^0_S mesons in the final state at LHCb*, in *CERN LHC Seminar*, (Geneva), 2021. LHCb-TALK-2021-366, on behalf of the LHCb Collaboration.

[5] LHCb collaboration, R. Aaij *et al.*, *Tests of lepton universality using $B^0 \to K^0_S \ell^+ \ell^-$ and $B^+ \to K^{*+} \ell^+ \ell^-$ decays*, arXiv:2110.09501, submitted to PRL.

[6] J. Woithe, G. J. Wiener, and F. F. V. der Veken, *Let's have a coffee with the Standard Model of particle physics!*, Phys. Educ. **52** (2017) 034001.

[7] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse, *Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment*, Phys. Rev. **A83** (2011) 052122, arXiv:1009.4831.

[8] A. J. Buras, *Gauge Theory of Weak Decays: The Standard Model and the Expedition to New Physics Summits*, Cambridge University Press, Cambridge, 2020.

[9] T. Mannel, *Effective Field Theories in Flavour Physics*, vol. 203 of *Springer Tracts in Modern Physics*, Springer International Publishing, Heidelberg, 2004.

[10] Belle II collaboration, E. Kou *et al.*, *The Belle II Physics Book*, PTEP **2019** (2019) , arXiv:1808.10567.

[11] A. Lenz, *Heavy flavour physics and effective field theories*, notes for Cambridge HEP graduate student lectures, https://www.precision.hep.phy.cam.ac.uk/wp-content/people/mitov/lectures/GraduateLectures/Flavour-Lenz.pdf (accessed 11/11/2021), 2020.

[12] D. Galbraith and C. Burgard, *Example: Standard model of physics*, 2016. https://texample.net/tikz/examples/model-physics/ (accessed 29/10/2021).

[13] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, Prog. Theor. Exp. Phys. **2020** (2020) .

[14] Y.-R. Liu *et al.*, *Pentaquark and Tetraquark States*, Prog. Part. Nucl. Phys. **107** (2019) 237, arXiv:1903.11976.

[15] M. E. Peskin and D. V. Schroeder, *An Introduction To Quantum Field Theory*, Westview Press, Boulder, Colorado, US, 1995.

[16] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579.

[17] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[18] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[19] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

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

[21] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[22] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.

[23] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[24] T. W. B. Kibble, *Symmetry Breaking in Non-Abelian Gauge Theories*, Phys. Rev. **155** (1967) 1554.

[25] J. Erler and M. Schott, *Electroweak precision tests of the Standard Model after the discovery of the Higgs boson*, Prog. Part. Nucl. Phys. **106** (2019) 68, arXiv:1902.05142.

[26] 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. **B716** (2012) 1, arXiv:1207.7214.

[27] 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. **B716** (2012) 30, arXiv:1207.7235.

[28] F. Wilczek, *Decays of Heavy Vector Mesons into Higgs Particles*, Phys. Rev. Lett. **39** (1977) 1304.

[29] F. Wilczek, *Origins of mass*, Cent. Eur. J. Phys. **10** (2012) 1021, arXiv:1206.7114.

[30] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531.

[31] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[32] J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. **C41** (2005) 1–131, arXiv:hep-ph/0406184, updated results and plots available at http://ckmfitter.in2p3.fr (accessed 10/11/2021).

[33] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[34] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285.

[35] A. M. Baldini *et al.*, *Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment*, Eur. Phys. J. **C76** (2016) 434, arXiv:1605.05081.

[36] SINDRUM collaboration, U. Bellgardt *et al.*, *Search for the decay $\mu^+ \to e^+e^+e^-$*, Nucl. Phys. **B299** (1988) 1.

[37] ATLAS Collaboration, G. Aad *et al.*, *Search for the lepton flavor violating decay $Z \to e\mu$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Rev. **D90** (2014) 072010, arXiv:1408.5774.

[38] CMS collaboration, A. M. Sirunyan *et al.*, *Search for lepton flavour violating decays of the Higgs boson to $\mu\tau$ and $e\tau$ in proton-proton collisions at $\sqrt{s} = 13$TeV*, JHEP **06** (2018) 001, arXiv:1712.07173.

[39] BNL E871 collaboration, D. Ambrose *et al.*, *New limit on muon and electron lepton number violation from $K_L^0 \to \mu^\pm e^\mp$ decay*, Phys. Rev. Lett. **81** (1998) 5734, arXiv:hep-ex/9811038.

[40] LHCb collaboration, R. Aaij *et al.*, *Search for the lepton flavour violating decay $B^+ \to K^+\mu^-\tau^+$ using B_{s2}^{*0} decays*, JHEP **06** (2020) 129, arXiv:2003.04352.

[41] OPAL collaboration, G. Abbiendi *et al.*, *Precise determination of the Z resonance parameters at LEP: "Zedometry"*, Eur. Phys. J. **C19** (2001) 587, arXiv:hep-ex/0012018.

[42] ATLAS collaboration, G. Aad *et al.*, *Test of the universality of τ and μ lepton couplings in w-boson decays with the ATLAS detector*, Nature Physics **17** (2021) 813, arXiv:2007.14040.

[43] ATLAS collaboration, M. Aaboud *et al.*, *Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 367, arXiv:1612.03016.

[44] PIENU collaboration, A. Aguilar-Arevalo *et al.*, *Improved measurement of the $\pi \to e\nu$ branching ratio*, Phys. Rev. Lett. **115** (2015) 071801, arXiv:1506.05845.

[45] D. Tong, *Lectures on Cosmology*, notes for University of Cambridge Part II Mathematical Tripos, http://www.damtp.cam.ac.uk/user/tong/cosmo.html (accessed 24/05/2022), 2019.

[46] A. Arbey and F. Mahmoudi, *Dark matter and the early Universe: A review*, Prog. Part. Nucl. Phys. **119** (2021) 103865, arXiv:2104.11488.

[47] P. Villanueva-Domingo, O. Mena, and S. Palomares-Ruiz, *A Brief Review on Primordial Black Holes as Dark Matter*, Front. Astron. Space Sci. **8** (2021) , arXiv:2103.12087.

[48] F. Kahlhoefer, *Review of LHC dark matter searches*, Int. J. Mod. Phys. **A32** (2017) 1730006, arXiv:1702.02430.

[49] XENON collaboration, E. Aprile *et al.*, *First Dark Matter Search Results from the XENON1T Experiment*, Phys. Rev. Lett. **119** (2017) 181301, arXiv:1705.06655.

[50] AMS collaboration, M. Aguilar *et al.*, *Towards Understanding the Origin of Cosmic-Ray Positrons*, Phys. Rev. Lett. **122** (2019) 041102.

[51] A. D. Sakharov, *Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32–35.

[52] N. S. Manton, *Topology in the Weinberg-Salam theory*, Phys. Rev. **D28** (1983) 2019.

[53] F. R. Klinkhamer and N. S. Manton, *A saddle-point solution in the Weinberg-Salam theory*, Phys. Rev. **D30** (1984) 2212.

[54] D. E. Morrissey and M. J. Ramsey-Musolf, *Electroweak baryogenesis*, New J. Phys. **14** (2012) 125003, arXiv:1206.2942.

[55] G. R. Farrar and M. E. Shaposhnikov, *Baryon asymmetry of the Universe in the standard model*, Phys. Rev. **D50** (1994) 774, arXiv:hep-ph/9305275.

[56] A. Hook, *TASI Lectures on the Strong CP Problem and Axions*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Theory in an Era of Data (TASI 2018)*, (Boulder, Colorado, US), 2018, arXiv:1812.02669.

[57] M. Dine, *Naturalness Under Stress*, Annu. Rev. Nucl. Part. Sci. **65** (2015) 43, arXiv:1501.01035.

[58] J. M. Pendlebury *et al.*, *Revised experimental upper limit on the electric dipole moment of the neutron*, Phys. Rev. **D92** (2015) 092003, arXiv:1509.04411.

[59] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Pseudoparticles*, Phys. Rev. Lett. **38** (1977) 1440.

[60] R. D. Peccei and H. R. Quinn, *Constraints imposed by CP conservation in the presence of pseudoparticles*, Phys. Rev. **D16** (1977) 1791.

[61] P. W. Graham *et al.*, *Experimental searches for the axion and axion-like particles*, Annu. Rev. Nucl. Part. Sci. **65** (2015) 485, arXiv:1602.00039.

[62] Y. Farzan and M. Tórtola, *Neutrino Oscillations and Non-standard Interactions*, Front. Phys. **6** (2018) 10, arXiv:1710.09360.

[63] Y. Cai, T. Han, T. Li, and R. Ruiz, *Lepton Number Violation: Seesaw Models and Their Collider Tests*, Front. Phys. **6** (2018) 40, arXiv:1711.02180.

[64] G. Esposito, *An introduction to quantum gravity*, arXiv:1108.3269, invited contribution to Section 6.7.17 of the Encyclopedia of Life Support Systems (EOLSS).

[65] S. Weinberg, *Implications of dynamical symmetry breaking*, Phys. Rev. **D13** (1976) 974.

[66] S. Weinberg, *Implications of dynamical symmetry breaking: An addendum*, Phys. Rev. **D19** (1979) 1277.

[67] L. Susskind, *Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory*, Phys. Rev. **D20** (1979) 2619.

[68] E. Gildener, *Gauge-symmetry hierarchies*, Phys. Rev. **D14** (1976) 1667.

[69] S. P. Martin, *A supersymmetry primer*, in *Perspectives on Supersymmetry* (G. L. Kane, ed.), vol. 18 of *Advanced Series on Directions in High Energy Physics*, pp. 1–98. World Scientific, 1998. arXiv:hep-ph/9709356.

[70] A. Canepa, *Searches for supersymmetry at the Large Hadron Collider*, Rev. Phys. **4** (2019) 100033.

[71] J. Bain, *Why be Natural?*, Found. Phys. **49** (2019) 898.

[72] S. Hossenfelder, *Screams for explanation: Finetuning and naturalness in the foundations of physics*, Synthese **198** (2021) 3727, arXiv:1801.02176.

[73] P. Williams, *Naturalness, the autonomy of scales, and the 125GeV Higgs*, Stud. Hist. Philos. Mod. Phys. **51** (2015) 82.

[74] *Cloud-Chamber Test Finds Neutrino 'Real': Drs. Crane and Halpern Decide It Is No Mere Hypothesis*, New York Times (1938), p44., 22nd. May 1938 edition.

[75] E. Fermi, *Tentativo di una teoria dei raggi β*, Il Nuovo Cimento (1924-1942) **11** (1934) 1.

[76] E. Fermi, *Versuch einer theorie der β-strahlen. I*, Z. Physik **88** (1934) 161.

[77] F. L. Wilson, *Fermi's Theory of Beta Decay*, Am. J. Phys. **36** (1968) 1150.

[78] S. Charley, *Through a muon's eyes*, Symmetry Magazine (2012), https://www.symmetrymagazine.org/article/june-2012/through-a-muons-eyes (accessed 07/02/2022).

[79] K. G. Wilson, *Non-Lagrangian Models of Current Algebra*, Phys. Rev. **179** (1969) 1499.

[80] K. G. Wilson and W. Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Commun. Math. Phys. **24** (1972) 87.

[81] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, Rev. Mod. Phys. **68** (1996) 1125, arXiv:hep-ph/9512380.

[82] K. Chetyrkin, M. Misiak, and M. Münz, *Weak radiative B-meson decay beyond leading logarithms*, Phys. Lett. **B400** (1997) 206, arXiv:hep-ph/9612313.

[83] T. Blake, G. Lanfranchi, and D. M. Straub, *Rare B decays as tests of the Standard Model*, Prog. Part. Nucl. Phys. **92** (2017) 50, arXiv:1606.00916.

[84] C. Bobeth, P. Gambino, M. Gorbahn, and U. Haisch, *Complete NNLO QCD analysis of $B \to X_s \ell^+ \ell^-$ and higher order electroweak effects*, JHEP **04** (2004) 071, arXiv:hep-ph/0312090.

[85] T. Huber, E. Lunghi, M. Misiak, and D. Wyler, *Electromagnetic logarithms in $B \to X_s \ell^+ \ell^-$*, Nucl. Phys. **B740** (2006) 105, arXiv:hep-ph/0512066.

[86] LHCb collaboration, *Rare Decay Conference Material*, https://twiki.cern.ch/twiki/bin/view/LHCbPhysics/RareDecayConferenceMaterial (internal, accessed 24/05/2022).

[87] HPQCD collaboration, C. Bouchard *et al.*, *Rare decay $B \to K \ell^+ \ell^-$ form factors from lattice QCD*, Phys. Rev. **D88** (2013) 054509, arXiv:1306.2384.

[88] HPQCD collaboration, C. Bouchard *et al.*, *Standard model predictions for $B \to K \ell^+ \ell^-$ with form factors from lattice QCD*, Phys. Rev. Lett. **111** (2013) 162002, arXiv:1306.0434.

[89] Fermilab Lattice and MILC collaborations, J. A. Bailey *et al.*, *$B \to K l^+ l^-$ decay form factors from three-flavor lattice QCD*, Phys. Rev. **D93** (2016) 025026, arXiv:1509.06235.

[90] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Lattice QCD calculation of form factors describing the rare decays $B \to K^* \ell^+ \ell^-$ and $B_s \to \phi \ell^+ \ell^-$*, Phys. Rev. **D89** (2014) 094501, arXiv:1310.3722.

[91] C. DeTar and S. Gottlieb, *Lattice Quantum Chromodynamics Comes of Age*, Phys. Today **57** (2004) 45.

[92] A. Khodjamirian and R. Rückl, *QCD Sum Rules for Exclusive Decays of Heavy Mesons*, in *Heavy Flavours II* (A. J. Buras, ed.), vol. 15 of *Advanced Series on Directions in High Energy Physics*, pp. 345–401. World Scientific, 1998. arXiv:hep-ph/9801443.

[93] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form factors from light-cone sum rules reexamined*, Phys. Rev. **D71** (2005) 014029, arXiv:hep-ph/0412079.

[94] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V \ell^+ \ell^-$ in the standard model from light-cone sum rules*, JHEP **08** (2016) 098, arXiv:1503.05534.

[95] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *Comparative study of the decays $B \to (K, K^*) l^+ l^-$ in the standard model and supersymmetric theories*, Phys. Rev. **D61** (2000) 074024, arXiv:hep-ph/9910221.

[96] W. Altmannshofer and D. M. Straub, *New physics in $b \to s$ transitions after LHC run 1*, Eur. Phys. J. **75** (2015) 382, arXiv:1411.3161.

[97] N. Gubernari, A. Kokulu, and D. van Dyk, *$B \to P$ and $B \to V$ form factors from B-meson light-cone sum rules beyond leading twist*, JHEP **01** (2019) 150, arXiv:1811.00983.

[98] A. Khodjamirian, T. Mannel, A. A. Pivovarov, and Y.-M. Wang, *Charm-loop effect in $B \to K^{(*)} l^+ l^-$ and $B \to K^{(*)} \gamma$*, JHEP **09** (2010) 089, arXiv:1006.4945.

[99] J. Lyon and R. Zwicky, *Resonances gone topsy turvy - the charm of QCD or new physics in $b \to s \ell^+ \ell^-$?*, EDINBURGH-14-10, CP3-ORIGINS-2014-021, DIAS-2014-21, University of Edinburgh, Edinburgh, UK, 2014. arXiv:1406.0566.

[100] LHCb collaboration, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP **06** (2014) 133, arXiv:1403.8044.

[101] LHCb collaboration, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016) 047, Erratum ibid. **04** (2017) 142, arXiv:1606.04731.

[102] 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** (2021) 151801, arXiv:2105.14007.

[103] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, JHEP **06** (2015) 115, Erratum ibid. **09** (2018) 145, arXiv:1503.07138.

[104] Belle collaboration, J.-T. Wei *et al.*, *Measurement of the differential branching fraction and forward-backward asymmetry for $B \to K^{(*)}l^+l^-$*, Phys. Rev. Lett. **103** (2009) 171801, arXiv:0904.0770.

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

[106] BaBar collaboration, J. P. Lees *et al.*, *Measurement of branching fractions and rate asymmetries in the rare decays $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. **D86** (2012) 032012, arXiv:1204.3933.

[107] CDF collaboration, T. Aaltonen *et al.*, *Observation of the baryonic flavor-changing neutral current decay $\Lambda_b^0 \to \Lambda\mu^+\mu^-$*, Phys. Rev. Lett. **107** (2011) 201802, arXiv:1107.3753.

[108] CMS collaboration, S. Chatrchyan *et al.*, *Angular analysis and branching fraction measurement of the decay $B^0 \to K^{*0}\mu + \mu-$*, Phys. Lett. **B727** (2013) 77–100, arXiv:1308.3409.

[109] CMS collaboration, V. Khachatryan *et al.*, *Angular analysis of the decay $B^0 \to K^{*0}\mu + \mu-$ from pp collisions at $\sqrt{s}$=8 TeV*, Phys. Lett. **B753** (2016) 424–448, arXiv:1507.08126.

[110] C. Bobeth, G. Hiller, and G. Piranishvili, *Angular distributions of $\bar{B} \to K\bar{l}l$ decays*, JHEP **12** (2007) 040, arXiv:0709.4174.

[111] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of charged and neutral $B \to K\mu^+\mu^-$ decays*, JHEP **05** (2014) 082, arXiv:1403.8045.

[112] CMS collaboration, A. M. Sirunyan *et al.*, *Angular analysis of the decay $B^+ \to K^+\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Rev. **D98** (2018) 112011, arXiv:1806.00636.

[113] CDF collaboration, T. Aaltonen *et al.*, *Measurements of the angular distributions in the decays $B \to K^{(*)}\mu^+\mu^-$ at CDF*, Phys. Rev. Lett. **108** (2012) 081807, arXiv:1108.0695.

[114] W. Altmannshofer *et al.*, *Symmetries and asymmetries of $B \to K^* \mu^+ \mu^-$ decays in the Standard Model and beyond*, JHEP **01** (2009) 019, arXiv:0811.1214.

[115] S. Descotes-Genon, J. Matias, M. Ramon, and J. Virto, *Implications from clean observables for the binned analysis of $B \to K^* \mu^+ \mu^-$ at large recoil*, JHEP **01** (2013) 048, arXiv:1207.2753.

[116] 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** (2020) 011802, arXiv:2003.04831.

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

[118] Belle collaboration, S. Wehle *et al.*, *Lepton-flavor-dependent angular analysis of $B \to K^* \ell^+ \ell^-$*, Phys. Rev. Lett. **118** (2017) 111801, arXiv:1612.05014.

[119] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay $B^0 \to K^{*0} \mu^+ \mu^-$ in proton–proton collisions at $\sqrt{s}$=8 TeV*, Phys. Lett. **B781** (2018) 517–541, arXiv:1710.02846.

[120] ATLAS collaboration, M. Aaboud *et al.*, *Angular analysis of $B_d^0 \to K^* \mu^+ \mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **010** (2018) 047, arXiv:1805.04000.

[121] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the rare decay $B_s^0 \to \phi \mu^+ \mu^-$*, JHEP **11** (2021) 043, arXiv:2107.13428.

[122] LHCb collaboration, R. Aaij *et al.*, *Strong constraints on the $b \to s\gamma$ photon polarisation from $B^0 \to K^{*0} e^+ e^-$ decays*, JHEP **12** (2020) 081, arXiv:2010.06011.

[123] T. Feldmann and J. Matias, *Forward-backward and isospin asymmetry for $B \to K^* l^+ l^-$ decay in the standard model and in supersymmetry*, JHEP **01** (2003) 074, arXiv:hep-ph/0212158.

[124] J. Lyon and R. Zwicky, *Isospin asymmetries in $B \to (K^*, \rho)\gamma / l^+ l^-$ and $B \to K l^+ l^-$ in and beyond the standard model*, Phys. Rev. **D88** (2013) 094004, arXiv:1305.4797.

[125] A. Khodjamirian, T. Mannel, and Y.-M. Wang, *$B \to K l^+ l^-$ decay at large hadronic recoil*, JHEP **02** (2013) 010, arXiv:1211.0234.

[126] M. Ahmady, S. Lord, and R. Sandapen, *Isospin asymmetry in $B \to K^* \mu^+ \mu^-$ using AdS/QCD*, Phys. Rev. **D90** (2014) 074010, arXiv:1407.6700.

[127] H. Ishida and Y. Koide, *Effective valence quark model and a possible dip in $dBr(B \to K\ell\bar{\ell})/dq^2$*, Int. J. Mod. Phys. **A29** (2014) 1450040, arXiv:1305.7342.

[128] H.-B. Fu *et al.*, *Asymmetries of the $B \to K^* \mu^+ \mu^-$ decay and the search of new physics beyond the standard model*, Phys. Rev. **D97** (2018) 055037, arXiv:1801.04808.

[129] A. K. Alok *et al.*, *New physics in $b \to s \mu^+ \mu^-$: CP-violating observables*, JHEP **11** (2011) 122, arXiv:1103.5344.

[130] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetries in the decays* $B^0 \to K^{*0}\mu^+\mu^-$ *and* $B^+ \to K^+\mu^+\mu^-$, JHEP **09** (2014) 177, arXiv:1408.0978.

[131] LHCb collaboration, R. Aaij *et al.*, *Observation of the decay* $\Lambda_b^0 \to pK^-\mu^+\mu^-$ *and search for CP violation*, JHEP **06** (2017) 108, arXiv:1703.00256.

[132] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the* $B^0 \to K^{*0}\mu^+\mu^-$ *decay using* $3\,fb^{-1}$ *of integrated luminosity*, JHEP **02** (2016) 104, arXiv:1512.04442.

[133] A. Carvunis *et al.*, *On the effective lifetime of* $B_s \to \mu\mu\gamma$, arXiv:2102.13390.

[134] CMS and LHCb collaborations, V. Khachatryan *et al.*, *Observation of the rare* $B_s^0 \to \mu^+\mu^-$ *decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68, arXiv:1411.4413.

[135] LHCb collaboration, R. Aaij *et al.*, *Analysis of neutral B-meson decays into two muons*, Phys. Rev. Lett. **128** (2022) 041801, arXiv:2108.09284.

[136] LHCb collaboration, R. Aaij *et al.*, *Measurement of the* $B_s^0 \to \mu^+\mu^-$ *decay properties and search for the* $B^0 \to \mu^+\mu^-$ *and* $B_s^0 \to \mu^+\mu^-\gamma$ *decays*, Phys. Rev. **D105** (2022) 012010, arXiv:2108.09283.

[137] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of properties of* $B_s^0 \to \mu^+\mu^-$ *decays and search for* $B^0 \to \mu^+\mu^-$ *with the CMS experiment*, JHEP **04** (2020) 188, arXiv:1910.12127.

[138] ATLAS collaboration, M. Aaboud *et al.*, *Study of the rare decays of* B_s^0 *and* B^0 *mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04** (2019) 098, arXiv:1812.03017.

[139] W. Altmannshofer and P. Stangl, *New physics in rare B decays after Moriond 2021*, Eur. Phys. J. **C81** (2021) 952, arXiv:2103.13370.

[140] S. Aoki *et al.*, *FLAG Review 2019*, Eur. Phys. J. **C80** (2020) 113, arXiv:1902.08191.

[141] K. De Bruyn *et al.*, *Probing new physics via the* $B_s^0 \to \mu^+\mu^-$ *effective lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv:1204.1737.

[142] G. Hiller and F. Krüger, *More model-independent analysis of* $b \to s$ *processes*, Phys. Rev. **D69** (2004) 074020, arXiv:hep-ph/0310219.

[143] M. Bordone, G. Isidori, and A. Pattori, *On the standard model predictions for* R_K *and* R_{K^*}, Eur. Phys. J. **C76** (2016) 440, arXiv:1605.07633.

[144] G. Isidori, S. Nabeebaccus, and R. Zwicky, *QED corrections in* $\overline{B} \to \overline{K}\ell^+\ell^-$ *at the double-differential level*, JHEP **012** (2020) 104, arXiv:2009.00929.

[145] G. Hiller and M. Schmaltz, *Diagnosing lepton-nonuniversality in* $b \to s\ell\ell$, JHEP **02** (2015) 055, arXiv:1411.4773.

[146] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using* $B^+ \to K^+\ell^+\ell^-$ *decays*, Phys. Rev. Lett. **113** (2014) 151601, arXiv:1406.6482.

[147] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769, to appear in Nature Physics.

[148] S. Knapton, *Key to how universe works may have been discovered*, The Telegraph (2021), https://www.telegraph.co.uk/science/2021/03/23/key-universe-works-may-have-discovered/ (accessed 06/11/2021).

[149] P. Ghosh, *Machine finds tantalising hints of new physics*, BBC News (2021), https://www.bbc.com/news/science-environment-56491033 (accessed 06/11/2021).

[150] I. Sample, *Cern experiment hints at new force of nature*, The Guardian (2021), https://www.theguardian.com/science/2021/mar/23/large-hadron-collider-scientists-particle-physics (accessed 06/11/2021).

[151] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, JHEP **08** (2017) 055, arXiv:1705.05802.

[152] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality using* $\Lambda_b^0 \to pK^-\ell^+\ell^-$ *decays*, JHEP **05** (2020) 040, arXiv:1912.08139.

[153] G. Isidori *et al.*, *A general effective field theory description of* $b \to sl^+l^-$ *lepton universality ratios*, arXiv:2110.09882.

[154] Belle collaboration, S. Wehle *et al.*, *Test of lepton-flavor universality in* $B \to K^*\ell^+\ell^-$ *decays at Belle*, Phys. Rev. Lett. **126** (2021) 161801, arXiv:1904.02440.

[155] B. Capdevila, *Assessing lepton-flavour non-universality from* $B \to K^*\ell\ell$ *angular analyses*, J. Phys. : Conf. Ser. **873** (2017) 012039, arXiv:1605.03156.

[156] D. London and J. Matias, *B flavour anomalies: 2021 theoretical status report*, arXiv:2110.13270, to be published in the Annual Review of Nuclear and Particle Science.

[157] M. Algueró *et al.*, $b \to s\ell\ell$ *global fits after Moriond 2021 results*, in *55th Rencontres de Moriond on QCD and High Energy Interactions*, 2021, arXiv:2104.08921. (v1).

[158] M. Ciuchini *et al.*, *Lessons from the* $B^{0,+} \to K^{*0,+}\mu^+\mu^-$ *angular analyses*, Phys. Rev. **D103** (2021) 015030, arXiv:2011.01212.

[159] T. Hurth, F. Mahmoudi, D. Martínez Santos, and S. Neshatpour, *More indications for lepton nonuniversality in* $b \to s\ell^+\ell^-$, Phys. Lett. **B824** (2022) 136838, arXiv:2104.10058.

[160] M.-H. Schune and D. van Dyk, *Beyond bumps*, CERN Courier (2021), https://cerncourier.com/a/beyond-bumps/ (accessed 18/11/2021).

[161] B. Capdevila, M. Fedele, P. Stangl, and S. Neshatpour, *Effective field theory for* $b \to s\ell\ell$ *transitions*, in *Flavour Anomaly Workshop 2021*, (Geneva), 2021. https://indico.cern.ch/event/1055780/contributions/4454282/ (accessed 08/11/2021).

[162] G. Isidori, D. Lancierini, P. Owen, and N. Serra, *On the significance of new physics in* $b \to s\ell^+\ell^-$ *decays*, Phys. Lett. **B822** (2021) 136644, arXiv:2104.05631.

[163] A. Cerri *et al.*, *Opportunities in flavour physics at the HL-LHC and HE-LHC*, in *Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC* (A. Dainese *et al.*, eds.), no. 7 in CERN Yellow Reports: Monographs, pp. 867–1158. CERN, 2019. arXiv:1812.07638.

[164] B. C. Allanach, J. M. Butterworth, and T. Corbett, *Collider constraints on Z' models for neutral current B-anomalies*, JHEP **08** (2019) 106, arXiv:1904.10954.

[165] A. Celis, J. Fuentes-Martín, M. Jung, and H. Serôdio, *Family nonuniversal Z' models with protected flavor-changing interactions*, Phys. Rev. **D92** (2015) 015007, arXiv:1505.03079.

[166] B. Capdevila, *Explaining $b \to s\ell^+\ell^-$ and the Cabibbo angle anomaly with a vector triplet*, Phys. Rev. **D103** (2021) , arXiv:2005.13542.

[167] Z.-L. Han, R. Ding, S.-J. Lin, and B. Zhu, *Gauged $U(1)_{L_\mu - L_\tau}$ scotogenic model in light of $R_{K^{(*)}}$ anomaly and AMS-02 positron excess*, Eur. Phys. J. **C79** (2019) 1007, arXiv:1908.07192.

[168] A. Crivellin, G. D'Ambrosio, and J. Heeck, *Explaining $h \to \mu^\pm\tau^\mp$, $b \to K^*\mu^+\mu^-$, and $B \to K\mu^+\mu^-/B \to Ke^+e^-$ in a two-Higgs-doublet model with gauged $L_\mu - L_\tau$*, Phys. Rev. Lett. **114** (2015) 151801, arXiv:1501.00993.

[169] W. Altmannshofer, S. Gori, S. Profumo, and F. S. Queiroz, *Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model*, JHEP **12** (2016) 106, arXiv:1609.04026.

[170] M. Fernández Navarro and S. F. King, *Fermiophobic Z' model for simultaneously explaining the muon anomalies $r_{K^{(*)}}$ and $(g-2)_\mu$*, Phys. Rev. **D105** (2022) 035015, arXiv:2109.08729.

[171] L. Di Luzio, M. Kirk, and A. Lenz, *Updated B_s-mixing constraints on new physics models for $b \to s\ell^+\ell^-$ anomalies*, Phys. Rev. **D97** (2018) 095035, arXiv:1712.06572.

[172] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, *Neutrino Trident Production: A Powerful Probe of New Physics with Neutrino Beams*, Phys. Rev. Lett. **113** (2014) 091801, arXiv:1406.2332.

[173] ATLAS collaboration, G. Aad *et al.*, *Search for high-mass dilepton resonances using $139 fb^{-1}$ of pp collision data collected at $\sqrt{s} = 13TeV$ with the ATLAS detector*, Phys. Lett. **B796** (2019) 68–87, arXiv:1903.06248.

[174] CMS collaboration, A. M. Sirunyan *et al.*, *Search for resonant and nonresonant new phenomena in high-mass dilepton final states at $\sqrt{s} = 13TeV$*, JHEP **07** (2021) 208, arXiv:2103.02708.

[175] W. Buchmüller, R. Rückl, and D. Wyler, *Leptoquarks in lepton-quark collisions*, Phys. Lett. **B191** (1987) 442.

[176] D. Marzocca, *Minimal explanation of flavor anomalies: B-meson decays, muon magnetic moment, and the cabibbo angle*, Phys. Rev. Lett. **127** (2021) , arXiv:2104.05730.

[177] R. Barbieri, G. Isidori, A. Pattori, and F. Senia, *Anomalies in B-decays and $U(2)$ flavor symmetry*, Eur. Phys. J. **C76** (2016) 67, arXiv:1512.01560.

[178] L. Di Luzio, A. Greljo, and M. Nardecchia, *Gauge leptoquark as the origin of B-physics anomalies*, Phys. Rev. **D96** (2017) 115011, arXiv:1708.08450.

[179] J. Fuentes-Martín and P. Stangl, *Third-family quark-lepton unification with a fundamental composite Higgs*, Phys. Lett. **B811** (2020) 135953, arXiv:2004.11376.

[180] A. Greljo, P. Stangl, and A. Eller Thomsen, *A model of muon anomalies*, Phys. Lett. **B820** (2021) 136554, arXiv:2103.13991.

[181] J. M. Cline, *B decay anomalies and dark matter from vectorlike confinement*, Phys. Rev. **D97** (2018) 015013, arXiv:1710.02140.

[182] R. Barbieri and A. Tesi, *B-decay anomalies in Pati–Salam $SU(4)$*, Eur. Phys. J. **C78** (2018) 193, arXiv:1712.06844.

[183] A. Greljo and B. A. Stefanek, *Third family quark–lepton unification at the TeV scale*, Phys. Lett. **B782** (2018) 131, arXiv:1802.04274.

[184] C. Cornella, J. Fuentes-Martín, and G. Isidori, *Revisiting the vector leptoquark explanation of the B-physics anomalies*, JHEP **07** (2019) 168, arXiv:1903.11517.

[185] HFLAV Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. **C81** (2021) 226, arXiv:1909.12524, updated average of $R(D)$ and $R(D^*)$ for 2021 available at https://hflav-eos.web.cern.ch/hflav-eos/semi/spring21/html/RDsDsstar/RDRDs.html (accessed 10/11/2021).

[186] Muon $g-2$ collaboration, B. Abi *et al.*, *Measurement of the positive muon anomalous magnetic moment to 0.46 ppm*, Phys. Rev. Lett. **126** (2021) 141801, arXiv:2104.03281.

[187] Muon $g-2$ collaboration, G. W. Bennett *et al.*, *Final report of the E821 muon anomalous magnetic moment measurement at BNL*, Phys. Rev. **D73** (2006) 072003, arXiv:hep-ex/0602035.

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

[189] T. Aoyama *et al.*, *The anomalous magnetic moment of the muon in the Standard Model*, Physics Reports **887** (2020) 1, arXiv:2006.04822.

[190] A. Vicente, *Flavor and Dark Matter Connection*, in *16th Conference on Flavor Physics and CP Violation* (A. Giri and R. Mohanta, eds.), Springer Proceedings in Physics, (Cham), 393–400, Springer International Publishing, 2019, arXiv:1812.03028.

[191] O. S. Brüning *et al.*, *LHC Design Report*, CERN-2004-003-V-1, CERN, Geneva, 2004. doi: 10.5170/CERN-2004-003-V-1.

[192] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[193] J. Jowett, *Colliding Heavy Ions in the LHC*, in *9th Int. Particle Accelerator Conf. (IPAC'18)*, (Vancouver, BC, Canada), 584–589, JACOW Publishing, Geneva, Switzerland, 2018.

[194] E. Mobs, *The CERN accelerator complex - August 2018*, 2018. https://cds.cern.ch/record/2636343 (accessed 17/11/2021).

[195] R. Scrivens and M. Vretenar, *The tale of a billion-trillion protons*, CERN Courier (2018), https://cerncourier.com/a/the-tale-of-a-billion-trillion-protons/ (accessed 17/11/2021).

[196] R. Steerenberg, *LHC report: Full house for the LHC*, CERN (2017), https://home.cern/news/news/accelerators/lhc-report-full-house-lhc (accessed 24/05/2022).

[197] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[198] CMS collaboration, S. Chatrchyan *et al.*, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[199] ATLAS collaboration, *The ATLAS Experiment*, https://atlas.cern/about (accessed 18/11/2021).

[200] CMS collaboration, *People Statistics*, https://cms.cern/collaboration/people-statistics (accessed 18/11/2021).

[201] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[202] ALICE collaboration, *ALICE Institutes*, https://alice-collaboration.web.cern.ch/collaboration/alice_institute (accessed 18/11/2021).

[203] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[204] LHCb collaboration, *The LHCb Collaboration*, https://lhcb-public.web.cern.ch/en/Collaboration/Collaboration-en.html (accessed 18/11/2021).

[205] L. Evans, *LHC machine*, in *Proceedings of European Physical Society Europhysics Conference on High Energy Physics — PoS(EPS-HEP 2009)*, **84** 004, SISSA Medialab, 2010.

[206] M. Bajko *et al.*, *Report of the Task Force on the Incident of 19th September 2008 at the LHC*, LHC-PROJECT-Report-1168, CERN LHC, Geneva, 2009.

[207] *First particles collide in the Large Hadron Collider*, Symmetry Magazine (2009), https://www.symmetrymagazine.org/breaking/2009/11/23/first-particles-collide-in-the-large-hadron-collider (accessed 19/11/2021).

[208] B. Mason, *Phew, it works! science begins at the LHC*, Wired (2010), https://www.wired.com/2010/03/science-begins-at-lhc/ (accessed 19/11/2021).

[209] C. Duc, *Long Shutdown 1: Exciting times ahead*, CERN (2013), https://home.cern/news/news/accelerators/long-shutdown-1-exciting-times-ahead (accessed 19/11/2021).

[210] E. Elsen and F. Bordry, *A successful conclusion to Run 2*, CERN (2018), https://home.cern/news/opinion/accelerators/successful-conclusion-run-2 (accessed 19/11/2021).

[211] C. Agrigoroae, *The four lhc experiments are getting ready for pilot beams*, Phys. org (2021), https://phys.org/news/2021-10-lhc-ready.html (accessed 19/11/2021).

[212] LHCb collaboration, *LHCb operations plots*, https://lbgroups.cern.ch/online/OperationsPlots/index.htm (accessed 19/11/2021).

[213] ATLAS collaboration, *Luminosity public results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults (accessed 19/11/2021).

[214] ATLAS collaboration, *Luminosity public results run-2*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2 (accessed 19/11/2021).

[215] CMS collaboration, *Public CMS Luminosity Information*, https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (accessed 19/11/2021).

[216] F. Follin and D. Jacquet, *Implementation and experience with luminosity levelling with offset beam*, in *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, (Geneva), CERN, 2014, arXiv:1410.3667.

[217] B. Muratori and T. Pieloni, *Luminosity levelling techniques for the LHC*, in *ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders*, (Geneva), CERN, 2014, arXiv:1410.5646.

[218] R. Steerenberg, *LHC Report: Another run is over and LS2 has just begun...*, CERN (2018), https://home.cern/news/news/accelerators/lhc-report-another-run-over-and-ls2-has-just-begun (accessed 24/05/2022).

[219] M. Hostettler *et al.*, *β^* levelling using the LHC lumi server (MD 2427)*, CERN-ACC-NOTE-2018-0001, CERN, Geneva, 2020.

[220] Belle collaboration, A. Abashian *et al.*, *The Belle detector*, Nucl. Instrum. Methods Phys. Res. **A479** (2002) 117.

[221] BaBar collaboration, B. Aubert *et al.*, *The BABAR detector*, Nucl. Instrum. Methods Phys. Res. **A479** (2002) 1, arXiv:hep-ex/0105044.

[222] LHCb collaboration, R. Aaij *et al.*, *Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, Phys. Lett. **B694** (2010) 209, arXiv:1009.2731.

[223] T. Abe *et al.*, *Achievements of KEKB*, PTEP **2013** (2013) 03A001.

[224] LHCb collaboration, C. Elsässer, *$b\bar{b}$ production angle plots*, https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html (accessed 16/11/2021).

[225] R. Lindner, *LHCb layout 2*, 2008. http://cds.cern.ch/record/1087860/ (accessed 22/11/2021).

[226] LHCb collaboration, H. Dijkstra *et al.*, *LHCb Letter of Intent*, LHCb-95-001, CERN, 1995.

[227] S. Amato *et al.*, *LHCb: Technical Proposal*, CERN-LHCC-98-004, CERN, Geneva, 1998.

[228] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030, 2003.

[229] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[230] I. Belyaev *et al.*, *The history of LHCb*, Eur. Phys. J. **H46** (2021) 3, arXiv:2101.05331.

[231] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, 2001.

[232] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[233] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007, 2000.

[234] M. Vesterinen, *Considerations on the LHCb dipole magnet polarity reversal*, LHCb-PUB-2014-006, CERN, Geneva, 2014.

[235] C. Abellan Beteta *et al.*, *Monitoring radiation damage in the LHCb Tracker Turicensis*, arXiv:1809.05063.

[236] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, 2002.

[237] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024, 2001.

[238] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[239] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, arXiv:1708.00819.

[240] LHCb collaboration, *Tracking and Alignment Plots for Conferences*, 2019. https://twiki.cern.ch/twiki/bin/view/LHCb/ConferencePlots (internal, accessed 07/03/2022).

[241] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, arXiv:1408.1251.

[242] L. M. Garcia, L. Henry, B. Kishor, and A. Oyanguren, *Tracking performance for long-lived particles at LHCb*, J. Phys. : Conf. Ser. **1525** (2020) 012095, arXiv:1910.06171.

[243] R. Aaij *et al.*, *Centrality determination in heavy-ion collisions with the LHCb detector*, LHCb-DP-2021-001, in preparation.

[244] B. Jones and B. Tompkins, *A physicist's guide to Kalman filters* , 1998. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.1034 (accessed 22/11/2021).

[245] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, J. Basic Eng. **82** (1960) 35.

[246] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Methods Phys. Res. **A262** (1987) 444.

[247] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037, 2000.

[248] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[249] A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, Nucl. Instrum. Methods Phys. Res. **A876** (2017) 221, arXiv:1703.08152.

[250] R. Forty, *Ring-imaging Cherenkov detectors for LHC-B*, Nucl. Instrum. Methods Phys. Res. **A384** (1996) 167.

[251] R. W. Forty and O. Schneider, *RICH pattern recognition*, LHCb-98-040, CERN, Geneva, 1998.

[252] C. Abellan Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, arXiv:2008.11556, submitted to JINST.

[253] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036, 2000.

[254] S. Barsuk *et al.*, *Design and construction of electromagnetic calorimeter for LHCb experiment*, LHCb-2000-043, CERN, Geneva, 2000.

[255] R. Djeliadine, V. F. Obraztsov, and O. Iouchtchenko, *LHCb hadron trigger and HCAL cell size and length optimization*, LHCb-99-035, CERN, Geneva, 1999.

[256] V. Breton, N. Brun, and P. Perret, *A clustering algorithm for the LHCb electromagnetic calorimeter using a cellular automaton*, LHCb-2001-123, CERN, Geneva, 2001.

[257] T. Mombacher, *Beautiful Leptons - Setting Limits to New Physics with the LHCb Experiment*, PhD thesis, Technische Universitaet Dortmund, Dortmund, Germany, 2021.

[258] D. A. Berninghoff, J. Albrecht, and V. Gligorov, *Bremsstrahlung Recovery of Electrons using Multivariate Methods*, LHCb-INT-2016-018, CERN, Geneva, 2016.

[259] A. A. Alves Jr. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, arXiv:1211.1346.

[260] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, 2001.

[261] F. Archilli *et al.*, *Performance of the muon identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[262] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2019) 1, arXiv:1803.00824.

[263] LHCb collaboration, *LHCb Performance Numbers*, https://lhcb.web.cern.ch/speakersbureau/html/PerformanceNumbers.html (accessed 26/11/2021).

[264] LHCb collaboration, *Trigger Schemes*, http://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html (accessed 26/11/2021).

[265] LHCb collaboration, *LHCb trigger system: Technical Design Report*, CERN-LHCC-2003-031, 2003.

[266] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[267] R. Aaij *et al.*, *Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[268] A. Puig, *The LHCb trigger in 2011 and 2012*, LHCb-PUB-2014-046, CERN, Geneva, 2014.

[269] O. Callot, M. Ferro-Luzzi, and P. Perret, *Using the SPD multiplicity in the Level-0 trigger*, LHCb-2003-022, CERN, Geneva, 2003.

[270] V. V. Gligorov, *A single track HLT1 trigger*, LHCb-PUB-2011-003, CERN, Geneva, 2011.

[271] V. V. Gligorov, M. Williams, and C. Thomas, *The HLT inclusive B triggers*, LHCb-PUB-2011-016, CERN, Geneva, 2011.

[272] LHCb collaboration, *LHCb computing: Technical Design Report*, CERN-LHCC-2005-019, 2005.

[273] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[274] R. Brun and F. Rademakers, *ROOT - An object oriented data analysis framework*, Nucl. Instrum. Methods Phys. Res. **A389** (1997) 81.

[275] I. Belyaev *et al.*, *Simulation application for the LHCb experiment*, in *13th Conference on Computing in High Energy and Nuclear Physics (CHEP)*, **C0303241**, TUMT003, 2003, arXiv:physics/0306035.

[276] M. Clemencic *et al.*, *The LHCb Simulation Application, Gauss: Design, Evolution and Experience*, J. Phys. : Conf. Ser. **331** (2011) 032023.

[277] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[278] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[279] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. : Conf. Ser. **331** (2011) 032047.

[280] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Methods Phys. Res. **A462** (2001) 152.

[281] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv:hep-ph/0506026.

[282] M. P. Whitehead and on behalf of the LHCb collaboration, *A palette of fast simulations in LHCb*, in *Proceedings of The 39th International Conference on High Energy Physics — PoS(ICHEP2018)*, **340** 271, SISSA Medialab, 2019. on behalf of the LHCb collaboration.

[283] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, *ReDecay: A novel approach to speed up the simulation at LHCb*, Eur. Phys. J. **C78** (2018) 1009, arXiv:1810.10362.

[284] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: An application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, arXiv:1612.07489.

[285] F. E. James, *Monte Carlo phase space*, CERN-68-15, CERN, Geneva, 1968.

[286] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017) 052002, Erratum ibid. **119** (2017) 169901, arXiv:1612.05140.

[287] F. James, *Statistical Methods in Experimental Physics*, World Scientific, Singapore, second ed., 2006.

[288] M. Paterno, *Calculating efficiencies and their uncertainties*, FERMILAB-TM-2286-CD, Fermi National Accelerator Lab. (FNAL), Batavia, IL (US), 2004.

[289] *ROOT: TEfficiency Class Reference*, https://root.cern.ch/doc/master/classTEfficiency.html (accessed 31/12/2021).

[290] A. Lee, *mcerp package documentation*, 2014. https://pythonhosted.org/mcerp/ (accessed 15/12/2021).

[291] G. Cowan, *Statistical Data Analysis*, Oxford University Press, Oxford, UK, 1998.

[292] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun. **10** (1975) 343.

[293] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, arXiv:1007.1727.

[294] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Methods Phys. Res. **A552** (2005) 566, arXiv:physics/0503191.

[295] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039, CERN, Geneva, 2014.

[296] M. Mulder, *The essence of rare beauty: Studying $B^0_{(s)} \to \mu^+\mu^-$ decays with the LHCb experiment*, PhD thesis, University of Groningen, Groningen, Netherlands, 2021, CERN-THESIS-2020-401, doi: 10.33612/diss.149618058.

[297] M. De Cian, S. Stahl, P. Seyfert, and S. Farry, *Fast neural-net based fake track rejection in the LHCb reconstruction*, LHCb-PUB-2017-011, CERN, Geneva, 2017.

[298] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY (US), 2006.

[299] L. Gavardi, *Search for lepton flavour violation in τ decays at the LHCb experiment*, PhD thesis, Milan Bicocca University, Milan, Italy, 2014, CERN-THESIS-2013-259.

[300] G. Mancinelli and J. Serrano, *Study of muon isolation in the $B^0_s \to \mu^+\mu^-$ channel*, LHCb-INT-2010-011, Geneva, CERN, 2010.

[301] F. Archilli *et al.*, *Background studies for $B^0 \to \mu^+\mu^-$ analysis optimization*, LHCb-INT-2014-047, Geneva, CERN, 2014.

[302] B. P. Roe *et al.*, *Boosted decision trees as an alternative to artificial neural networks for particle identification*, Nucl. Instrum. Methods Phys. Res. **A543** (2005) 577, arXiv:physics/0408124.

[303] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*, Routledge, Boca Raton, FL (US), 1984.

[304] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[305] H. Voss, A. Hocker, J. Stelzer, and F. Tegenfeldt, *TMVA, the Toolkit for Multivariate Data Analysis with ROOT*, PoS **ACAT** (2007) 040.

[306] A. Hoecker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, CERN-OPEN-2007-007, CERN, Geneva, 2009. arXiv:physics/0703039.

[307] S. Geisser, *The Predictive Sample Reuse Method with Applications*, J. Am. Stat. Assoc. **70** (1975) 320.

[308] J. L. Hodges, *The significance probability of the Smirnov two-sample test*, Arkiv för Matematik **3** (1958) 469.

[309] S. I. Bityukov and N. V. Krasnikov, *On the observability of a signal above background*, Nucl. Instrum. Methods Phys. Res. **A452** (2000) 518.

[310] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, CERN, Geneva, 2016.

[311] L. Anderlini *et al.*, *Computing strategy for PID calibration samples for LHCb Run 2*, LHCb-PUB-2016-020, CERN, Geneva, 2016.

[312] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Methods Phys. Res. **A555** (2005) 356, arXiv:physics/0402083.

[313] V. Lisovskyi, *Study of rare b-baryon decays and test of lepton universality at LHCb*, PhD thesis, Université Paris-Saclay, Orsay, France, 2019, CERN-THESIS-2019-202.

[314] LHCb collaboration, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, JINST **14** (2019) P11023, arXiv:1909.02957.

[315] E. Smith, M. Patel, and P. Owen, *Data driven reconstruction efficiency of K^0_s mesons as a function of decay position*, LHCb-INT-2014-034, CERN, Geneva, 2014.

[316] E. W. Weisstein, *Erf*, 2021. https://mathworld.wolfram.com/Erf.html (accessed 07/01/2022).

[317] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. : Conf. Ser. **762** (2016) 012036, arXiv:1608.05806.

[318] A. Rogozhnikov *et al.*, *Arogozhnikov/hep_ml*, Zenodo, 2018. doi: 10.5281/zenodo.1412677.

[319] LHCb collaboration, R. Aaij *et al.*, *Precise measurement of the f_s/f_d ratio of fragmentation fractions and of B^0_s decay branching fractions*, Phys. Rev. **D104** (2021) 032005, arXiv:2103.06810.

[320] BaBar collaboration, B. Aubert *et al.*, *Amplitude analysis of the decay $B^\pm \to \pi^\pm\pi^\pm\pi^\mp$*, Phys. Rev. **D72** (2005) 052002, arXiv:hep-ex/0507025.

[321] LHCb collaboration, R. Aaij *et al.*, *Amplitude analysis of the decay $\bar{B}^0 \to K^0_S\pi^+\pi^-$ and first observation of CP asymmetry in $\bar{B}^0 \to K^*(892)^-\pi^+$*, Phys. Rev. Lett. **120** (2018) 261801, arXiv:1712.09320.

[322] J. Back *et al.*, *Laura++: A Dalitz plot fitter*, Comput. Phys. Commun. **231** (2018) 198, arXiv:1711.09854.

[323] R. J. Barlow, *Practical statistics for particle physics*, in *2018 Asia–Europe–Pacific School of High-Energy Physics*, **5** (Quy Nhon, Vietnam), 149–197, CERN, 2020, arXiv:1905.12362.

[324] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, in *Computing in High Energy and Nuclear Physics (CHEP03)*, **C0303241**, (La Jolla, California (US)), MOLT007, 2003, arXiv:physics/0306116.

[325] T. Skwarnicki, *A Study of the Radiative Cascade Transitions between the Upsilon-Prime and Upsilon Resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, Poland, 1986, DESY-F31-86-02.

[326] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Methods Phys. Res. **A764** (2014) 150, arXiv:1312.5000.

[327] K. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198, arXiv:hep-ex/0011057.

[328] E. O. Lebigot, *Uncertainties: A Python package for calculations with uncertainties*, 2017. https://pythonhosted.org/uncertainties/ (accessed 12/01/2022).

[329] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Stat. **7** (1979) 1.

[330] J. A. Hanley and B. MacGibbon, *Creating non-parametric bootstrap samples using Poisson frequencies*, Comput. Methods Programs Biomed. **83** (2006) 57.

[331] E. A. Smith, S. Schmitt, and C. M. Langenbruch, *On the PID-response dependency for dielectron pairs*, LHCb-INT-2021-008, CERN, Geneva, 2021. (in preparation).

[332] D. M. Straub, *Flavio: A Python package for flavour and precision phenomenology in the Standard Model and beyond*, arXiv:1810.08132.

[333] LHCb collaboration, R. Aaij *et al.*, *Differential branching fraction and angular analysis of the decay $B^0 \to K^+\pi^-\mu^+\mu^-$ in the $K^*_{0,2}(1430)^0$ region*, JHEP **12** (2016) 065, arXiv:1609.04736.

[334] LHCb collaboration, R. Aaij *et al.*, *First observations of the rare decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$*, JHEP **10** (2014) 064, arXiv:1408.1137.

[335] R.-D. Moise, *First Evidence for the Violation of Lepton Flavour Universality in Rare Beauty-Quark Decays*, PhD thesis, Imperial College London, London, UK, 2021, CERN-THESIS-2021-229.

[336] T. Humair, *Testing Lepton Universality in Penguin Decays of Beauty Mesons Using the LHCb Detector*, PhD thesis, Imperial College London, London, UK, 2019, CERN-THESIS-2019-044.

[337] S. W. Maddrell-Mander, *An Angular Analysis of $B^+ \to K^+e^+e^-$ Decays at the LHCb Experiment*, PhD thesis, University of Bristol, Bristol, UK, 2021, CERN-THESIS-2020-317.

[338] L. Prokhorenkova *et al.*, *CatBoost: Unbiased boosting with categorical features*, in *Advances in Neural Information Processing Systems (NeurIPS 2018)*, **31**, (Montréal, Canada), Curran Associates, Inc., 2018, arXiv:1706.09516.

[339] A. V. Dorogush, V. Ershov, and A. Gulin, *CatBoost: Gradient boosting with categorical features support*, in *Workshop on ML Systems at NIPS 2017*, (Long Beach, CA (US)), 2017, arXiv:1810.11363.

[340] F. Pedregosa *et al.*, *Scikit-learn: Machine Learning in Python*, J. Mach. Learn. Res. **12** (2011) 2825, arXiv:1201.0490.

[341] scikit-learn developers, *QuantileTransformer*, 2021. https://scikit-learn/stable/modules/generated/sklearn.preprocessing.QuantileTransformer.html (accessed 21/01/2022).

[342] P. Virtanen *et al.*, *SciPy 1.0: Fundamental algorithms for scientific computing in Python*, Nature Methods **17** (2020) 261, arXiv:1907.10121.

[343] J. Neyman and E. S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, Philos. Trans. R. Soc. **A231** (1933) 289.

[344] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Ann. Math. Stat. **9** (1938) 60.

[345] BaBar collaboration, B. Aubert *et al.*, *Measurement of branching fractions and charge asymmetries for exclusive B decays to charmonium*, Phys. Rev. Lett. **94** (2005) 141801, arXiv:hep-ex/0412062.

[346] T. M. Karbach and M. Schlupp, *Constraints on Yield Parameters in Extended Maximum Likelihood Fits*, arXiv:1210.7141.

[347] L. Demortier and L. Lyons, *Everything you always wanted to know about pulls*, CDF/ANAL/PUBLIC/5776, Fermi National Accelerator Lab. (FNAL), Batavia, IL (US), 2008.

[348] M. van Veghel, *Pursuing forbidden beauty: Search for the lepton-flavour violating decays $B^0 \to e^{\pm}\mu^{\mp}$ and $B_s^0 \to e^{\pm}\mu^{\mp}$ and study of electron-reconstruction performance at LHCb*, PhD thesis, University of Groningen, Groningen, Netherlands, 2020, CERN-THESIS-2020-060, doi: 10.33612/diss.128123609.

[349] M. van Veghel, *Use of upstream electrons in lepton universality tests*, in *Rare Decays Working Group (LHCb Internal)*, (Geneva), 2019. https://indico.cern.ch/event/835385/contributions/3501192/ (accessed 02/02/2022).

[350] O. Deschamps *et al.*, *Photon and neutral pion reconstruction*, LHCb-2003-091, CERN, Geneva, 2003.

[351] J. Eschle, A. Puig Navarro, R. Silva Coutinho, and N. Serra, *Zfit: Scalable pythonic fitting*, SoftwareX **11** (2020) 100508, arXiv:1910.13429.

[352] M. Hoballah, *Measurement of the photon polarization using $B_s^0 \to \phi\gamma$ at LHCb.*, PhD thesis, Université Blaise Pascal, Clermont-Ferrand, France, 2015, CERN-THESIS-2015-073.

[353] M. Calvo Gomez *et al.*, *A tool for γ/π^0 separation at high energies*, LHCb-PUB-2015-016, CERN, Geneva, 2015.

[354] Y. S. Amhis, C. Marin Benito, and M. Calvo Gomez, *State of the art of the Run 1 and 2 calorimeter reconstruction and towards the Upgrade*, LHCb-INT-2019-027, CERN, Geneva, 2019.

[355] A. Marshall, *R_K in the high q^2 region*, in *LHCb-UK Annual Meeting 2022 (open session)*, (Edinburgh, UK), 2022. https://indico.cern.ch/event/1079623/contributions/4542082/ (accessed 31/01/2022).

[356] W. Altmannshofer *et al.*, *Light resonances and the low-q^2 bin of R_{K^*}*, JHEP **03** (2018) 188, arXiv:1711.07494.

[357] *LHC long term schedule*, 2022. https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm (accessed 04/02/2022).

[358] LHCb collaboration, *Physics case for an LHCb Upgrade II — Opportunities in flavour physics, and beyond, in the HL-LHC era*, arXiv:1808.08865.

[359] *A new schedule for the LHC and its successor*, CERN (2019), https://home.cern/news/news/accelerators/new-schedule-lhc-and-its-successor (accessed 03/02/2022).

[360] A. Schaeffer, *New schedule for CERN's accelerators and experiments*, CERN (2020), http://home.cern/news/news/accelerators/new-schedule-cerns-accelerators-and-experiments (accessed 25/01/2022).

[361] M. Lamont and J. Mnich, *LS3 schedule change*, 2022. (email to cern-personnel@cern.ch mailing list).

[362] LHCb collaboration, R. Aaij *et al.*, *Letter of Intent for the LHCb Upgrade*, CERN-LHCC-2011-001, CERN, Geneva, 2011.

[363] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[364] M. van Veghel, *Electron PID in Run 3*, in *81st Analysis and Software Week (LHCb Internal)*, (Geneva), 2021. https://indico.cern.ch/event/1084632/contributions/4564570/ (accessed 28/01/2022).

[365] M. Chalmers, *Protons back with a splash*, CERN Courier (2021), https://cerncourier.com/a/protons-back-with-a-splash/ (accessed 03/02/2022).

[366] LHCb collaboration, *Expression of Interest for a Phase-II LHCb Upgrade: Opportunities in flavour physics, and beyond, in the HL-LHC era*, CERN-LHCC-2017-003, 2017.

[367] Belle II collaboration, T. Abe *et al.*, *Belle II Technical Design Report*, arXiv:1011.0352.

[368] *Kick-off of the Belle II Phase 3 Physics Run*, KEK (2019), https://www.kek.jp/en/newsroom/2019/03/25/2030/ (accessed 03/02/2022).

[369] M. Pieri, *Anomalies intrigue at Moriond*, CERN Courier (2021), https://cerncourier.com/a/anomalies-intrigue-at-moriond/ (accessed 03/02/2022).

[370] P. Onyisi, *Flavour physics at ATLAS*, in *Flavour Anomaly Workshop 2021*, (Geneva), 2021. https://indico.cern.ch/event/1055780/contributions/4454291/ (accessed 31/01/2022) on behalf of the ATLAS collaboration.

[371] Y. Takahashi, *Flavour physics at CMS*, in *Flavour Anomaly Workshop 2021*, (Geneva), 2021. https://indico.cern.ch/event/1055780/contributions/4454291/ (accessed 31/01/2022) on behalf of the CMS collaboration.

[372] ATLAS collaboration, P. Allport *et al.*, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, CERN-LHCC-2012-022, CERN, Geneva, 2012.

[373] CMS collaboration, D. C. Contardo *et al.*, *Technical Proposal for the Phase-II Upgrade of the CMS Detector*, CERN-LHCC-2015-010, CERN, Geneva, 2015.

[374] ATLAS collaboration, *Prospects for the $\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-)$ measurements with the ATLAS detector in the Run 2 and HL-LHC data campaigns*, ATL-PHYS-PUB-2018-005, CERN, Geneva, 2018.

[375] ATLAS collaboration, *$B^0_d \to K^{*0}\mu\mu$ angular analysis prospects with the upgraded ATLAS detector at the HL-LHC*, ATL-PHYS-PUB-2019-003, CERN, Geneva, 2019.

[376] CMS collaboration, *Measurement of rare $B \to \mu^+\mu^-$ decays with the phase-2 upgraded CMS detector at the HL-LHC*, CMS-PAS-FTR-18-013, CERN, Geneva, 2018.

[377] CMS collaboration, *Study of the expected sensitivity to the P_5' parameter in the $B^0 \to K^{*0}\mu^+\mu^-$ decay at the HL-LHC*, CMS-PAS-FTR-18-033, CERN, Geneva, 2018.

[378] ATLAS collaboration, *Trigger menu in 2018*, ATL-DAQ-PUB-2019-001, CERN, Geneva, 2019.

[379] CMS collaboration, *Recording and reconstructing 10 billion unbiased b hadron decays in CMS*, CMS-DP-2019-043, CERN, Geneva, 2019.

[380] S. Mukherjee, *Data Scouting and Data Parking with the CMS High level Trigger*, in *Proceedings of European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2019)*, **364** 139, SISSA Medialab, 2020. (on behalf of the CMS collaboration).

[381] A. Mann, *High p_T searches at ATLAS and CMS*, in *Flavour Anomaly Workshop 2021*, (Geneva), 2021. https://indico.cern.ch/event/1055780/contributions/4454304/ (accessed 15/02/2022).

[382] B. C. Allanach, J. M. Butterworth, and T. Corbett, *Large hadron collider constraints on some simple Z' models for $b \to s\mu^+\mu^-$ anomalies*, Eur. Phys. J. **C81** (2021) 1126, arXiv:2110.13518.

[383] A. Greljo *et al.*, *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, JHEP **07** (2021) 122, arXiv:2104.02723.

[384] B. C. Allanach, B. Gripaios, and T. You, *The case for future hadron colliders from $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP **03** (2018) 021, arXiv:1710.06363.

[385] M. Algueró *et al.*, *$b \to s\ell\ell$ global fits after R_{K_S} and $R_{K^{*+}}$*, arXiv:2104.08921, (v3).

[386] A. K. Alok, N. R. S. Chundawat, and D. Kumar, *Impact of $b \to s\ell\ell$ anomalies on rare charm decays in non-universal Z' models*, Eur. Phys. J **C82** (2022) 30, arXiv:2110.12451.

[387] X.-Q. Li *et al.*, *Explaining the $b \to s\ell^+\ell^-$ anomalies in Z' scenarios with top-FCNC couplings*, arXiv:2112.14215.

[388] S. Iguro, J. Kawamura, S. Okawa, and Y. Omura, *Importance of vector leptoquark-scalar box diagrams in Pati-Salam unification with vector-like families*, arXiv:2201.04638.

[389] C. Flammarion, *"The Flammarion engraving"*, in *L'Atmosphère: Météorologie Populaire*, p. 163. Hachette, Paris, France, 1888. file from Wikipedia (https://en.wikipedia.org/wiki/File:Flammarion.jpg, accessed 15/02/2022).