Angular Analysis of $B^0 \to K^{*0} e^+ e^-$ Decays at LHCb

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Zhenzi Wang

aus der

Volksrepublik China

Promotionskommission

Prof. Dr. Nicola Serra

Dr. Rafael Silva Coutinho

Prof. Dr. Florencia Canelli

Zürich, 2023

Abstract

The subject of this thesis is the binned angular analysis of $B^0 \to K^{*0}e^+e^-$ decays in the dielectron invariant mass squared (q^2) regions of $1.1 < q^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ and $1.1 < q^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ ('central q^2'), using data collected by the LHCb detector at centre-of-mass energies of 7, 8 and 13 TeV, corresponding to a total of $9\,\mathrm{fb}^{-1}$ of integrated luminosity.

The expectation that the Standard Model (SM) is an effective description of nature, and that a more complete model will become necessary at high energies, motivates the search for new particles through direct production, as well as the indirect search for hitherto unknown contributions via precision measurements. Decays featuring the $b \to s\ell^+\ell^-$ transition, such as $B^0 \to K^{*0}e^+e^-$, can only occur through higher order processes in the SM to which (virtual) new particles can contribute. They are therefore expected to be particularly sensitive to New Physics (NP) effects.

A number of measurements of the angular observables (and differential branching fractions) of $b \to s\mu^+\mu^-$ modes in the past decade reported varying levels of tensions with SM predictions. While no single result is robust enough for a claim of discovery, taken together, they seem to hint at common underlying features, which may related to NP contributions or unaccounted for SM effects. The purpose of this analysis is to contribute to the clarification of this picture by providing inputs from the electron mode, which will offer indications on the nature of these tensions.

This thesis presents all the major components of the first angular analysis of the $B^0 \to K^{*0}e^+e^-$ mode in the central q^2 region at the LHCb, which is expected to be the most precise measurement of its kind to date.

Contents

Introduction

The Standard Model of particle physics, developed in the 20th century, is an extraordinarily successful theory that describes the fundamental constituents of physical world and their interactions. Supported by numerous experiments, its predictions have been tested and confirmed with phenomenal precision. Nevertheless, the number of puzzling questions that it leaves unanswered suggest that it cannot be complete. Questions such as why matter dominates over anti-matter, why does the known particle content appear to only describe a tiny fraction of the matter in the universe, and how does the force of gravity, described by the similarly successful theory of General Relativity at large scales, manifest at the quantum level. These questions prompt many ongoing searches for effects that are beyond the SM. In the field of high energy physics, this often involves the exploration of unprecedented high energies and small length scales to find its breaking point.

Two types of approaches are often taken, the first of which is to attempt to find direct evidence of 'New Physics' (NP), and the second is to detect signs that are indicative of its presence. The former corresponds to colliding particles at high energies to search for new particles and interactions that the SM cannot explain. While straightforward and unambiguous, this approach is strongly limited by the hardware capabilities of the experimental apparatus. Alternatively, one can rely on quantum effects, where massive new particles, even ones that cannot be produced directly, can participate virtually in measurable processes, leading to alterations of its characteristics from expectation. Precision measurements of the details of these processes can reveal their presence. This type of 'indirect' search for NP lies at the heart of this work.

A group of processes that are particularly sensitive to NP effects are b hadron decays that feature the $b \to s\ell^+\ell^-$ transition. In recent years, a number of measurements of the properties of $b \to s\mu^+\mu^-$ decays have revealed tensions with SM predictions. While no single result is sufficiently robust statistically or beyond all reasonable doubts from a theoretical perspective, such that a discovery can be claimed, together they hint at a consistent pattern of deviations. One of the earliest and most persistent entries of this set is the angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay. Performing the analogous measurement for the experimentally more challenging electron mode is the subject of this thesis. This analysis is performed using data collected by the LHCb detector, which is located at a collision point of the Large Hadron Collider (LHC) at CERN. The large $b\bar{b}$ production cross-section at the high centre-of-mass energies reached, together with the optimisation of the detector for the study of b and c hadron decays, with its narrow, but well-instrumented coverage of the kinematic region of interest, enables the study of this rare mode in a complex hadronic environment.

This thesis is organised in three parts. Part I provides an overview of the theory background and the present experimental status. The main features of the SM are discussed in Chapter 1, and that of the $B^0 \to K^{*0}e^+e^-$ decay in Chapter 2, where relevant definitions and observables are also introduced, and an overview of the present state of the field is presented. Part II is focused on the experimental apparatus. The LHC is introduced in Chapter 3, and details of the LHCb detector, in particular its subsystems, capabilities, and the origins of quantities relevant to the measurement, are provided in Chapter 4. All major components of the analysis are discussed in Part III, starting with a strategy overview in Chapter 5, followed by the

descriptions of the simulation and data samples used in Chapter 6, the isolation of signal candidates in Chapter 7, and a detailed breakdown of the components necessary for the angular analysis in Chapter 8. Data fits and systematic uncertainties are discussed in Chapter 9, where the preliminary result is presented without displaying the values of the observables of interest. Finally, Chapter 10 concludes the work performed in this thesis.

Part I

Theoretical and experimental overview

1 The Standard Model

The Standard Model of particle physics (SM), which came together in its current form in the 1970s, is the culmination of theories and ideas that arose in the early part of the century, summarising all that is presently known about the fundamental constituents of matter and the strong and electroweak forces. It has held up extraordinarily well to experimental scrutiny over the last ~ 40 years, and is therefore known to provide a good description of nature up to at least the TeV energy scale.

The SM is a quantum field theory, a mathematical framework that combines the principles of quantum mechanics and classical field theory with special relativity. It is characterised by the local symmetry group of $SU(3)_C \times SU(2)_L \times U(1)_Y$, its matter content, and by the Higgs mechanism [1, 2], which generates mass terms for fundamental particles without violating gauge invariance. The dynamics and kinematics of this theory can be summarised by the Lagrangian,

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{kin+int}} \ , \tag{1.1}$$

where the first term, $\mathcal{L}_{\text{gauge}}$, contains the kinetic terms of the gauge bosons responsible for mediating the fundamental interactions, $\mathcal{L}_{\text{Higgs}}$ and $\mathcal{L}_{\text{Yukawa}}$ are related to the Higgs mechanism, and $\mathcal{L}_{\text{kin+int}}$ contains the kinematic terms of the matter particles and their interactions with the gauge bosons. Important features of the SM are elaborated upon in the following sections.

Note that the content of this chapter is based on [3, 4, 5, 6]. Additional references are also provided.

1.1 Particles and symmetries

The SM describes particles and the interactions between them, the forces, as the exchange of (virtual) particles. In the QFT framework, particles are described as excitations of quantum fields. The SM contains three types of fields: spinor, vector and scalar. Spinor fields correspond to spin-1/2 particles, or fermions, which are the fundamental constituents of matter. This group includes quarks, which always hadronise into bound states, and electrons (e), which form atoms with the bound states of up (u) and down (d) quarks, i.e. protons and neutrons. Another type of lepton, the electron neutrino (ν_e) is produced, for example, from fusion reactions that power the sun. These more familiar fermions all belong to the first 'generation'. The SM contains two additional generations of particles, which differ only in mass from the first. Besides the electron, there is the muon (μ), which has a mass that is ~ 200 times that of the electron, and the much heavier tau (τ), the mass of which ~ 3500 times larger. Along with the muon and the tau, there are also the corresponding muon neutrino (ν_μ) and tau neutrino (ν_τ). For the quarks, in addition to the up and down quarks, there are the strange (s) and charm (c), top (t) and bottom (b) quarks. Many second and third generation fermions are encountered mostly in high energy settings, e.g. astrophysical processes or high energy colliders due to their large masses and short lifetimes. The different species of fermions are also referred to as flavours.

Vector fields correspond to spin-1 particles, or gauge bosons, which are the mediators of

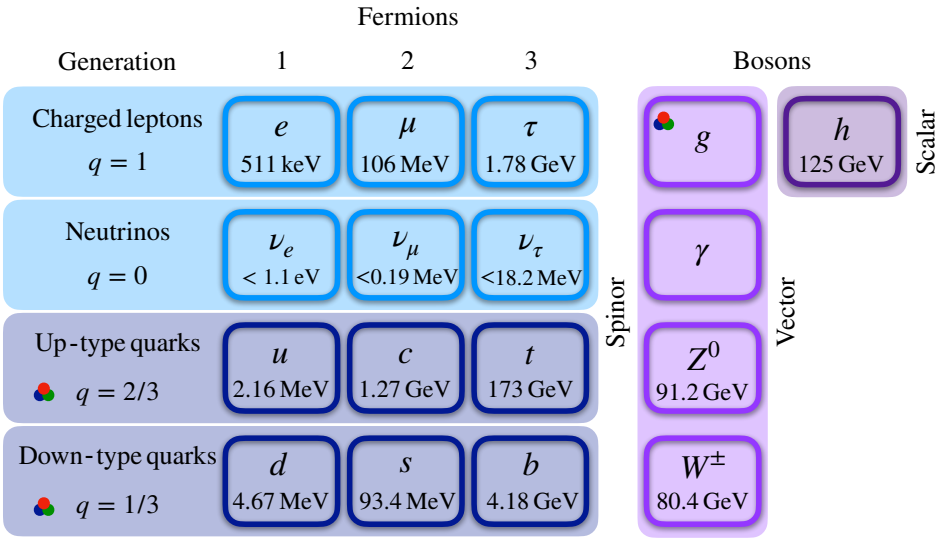


Figure 1.1: Summary of the particle content of the SM showing the three generations of fermions and the gauge bosons. Electric charges are denoted by q (or $\pm$). Particles that carry colour charge are labelled using the colored marker. All masses are given in natural units. The values used are obtained from [7].

forces. Particles interact via four types of forces, of which three[1] are described by the SM: electromagnetism, strong and weak. The photon is the mediator of the electromagnetic interaction, which is, for example, responsible for the electrostatic attraction between electrons and nuclei that holds atoms together. The nucleus of the atom is bound by the strong nuclear force, which is a manifestation of the strong interaction, mediated by gluons. The weak force, responsible for fusion reactions and nuclear β decays, is mediated by the massive $W^{\pm}$ and Z^0 bosons. Among the fermions, all quarks and the leptons that are electrically charged, i.e. the e, μ and τ, can undergo the electromagnetic interaction. In contrast to the charged leptons, which carry integer units of charge, quarks carry fractional charge. The up-type quarks (u, c and t) have charges of $+2/3$ while the down-type quarks (d, s and b) have charges of $-1/3$. Quarks also carry the charge of the strong interaction, or color, which comes in three varieties, denoted red (r), green (g) and blue (b), thus they can interact via the strong interaction. While all fermions can interact through the weak force, the nature of the interaction varies depending on their quantum mechanical chirality state. The dynamics of fermions can be described by the relativistic Dirac equation. An important implication of this is the existence of anti-particles with the same mass, but opposite charges. Anti-particles are denoted by the overline notation, i.e. $\bar{e}$, refers to an anti-electron (or a positron). A summary of the particle content of the SM is given in Figure 1.1.

The concept of symmetries – of operations under which a system is invariant – is extremely useful for deducing the underlying laws of nature. For example, symmetries and conserved quantities are connected by Noether's Theorem, which shows that for every continuous symmetry of the Lagrangian of a system there exists a conserved current, which is associated with a conserved quantity. The SM contains various types of symmetries. They include invariance under Lorentz transformations, which leads to the conservation of energy, momentum and angular momentum. This type of transformation is continuous in nature. It also contains discrete

[1]The gravitational force between interacting particles is small and can be neglected in the following discussions. However, the SM does not currently incorporate the gravitational interaction.

symmetries, namely symmetries under the actions of the parity, charge, and time reversal operators, $\hat{P}$, $\hat{C}$ and $\hat{T}$. Mathematically these are implemented by operators with eigenstates of ± 1. The parity operator, $\hat{P}$, inverts the sign of the spatial coordinates, $(t, \bar{x}) \rightarrow (t, -\bar{x})$. Charge conjugation reverses all internal quantum numbers (e.g. electric charge), changing particle to anti-particle and vice-versa. Time reversal changes the time component from $(t, \bar{x}) \rightarrow (-t, \bar{x})$. The strong and electromagnetic interactions preserve C, P and T, as well as the combination of C and P (CP), whereas the weak force violates C and P as well as CP. In particular, it differentiates between left- and right-handed fermions. Mathematically, left- and right-handed chiral states can be separated from a Dirac spinor, ψ, using the projection operators $P_L = (1/2)(I - \gamma^5)$ and $P_R = (1/2)(I + \gamma^5)$ where γ^5 is the fifth gamma matrix, i.e. $\psi_L \equiv P_L \psi$ and $\psi_R \equiv P_R \psi$. In calculations related to the weak interaction, they are often treated as two different sets of fields rather than as the two parts of a four component spinor, and are written separately as

$$\ell_L^1 = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \; \ell_L^2 = \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \; \ell_L^3 = \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}, \tag{1.2}$$

for the three flavours of charged and neutral leptons. The right-handed lepton fields only have single components, and they do not interact with the $W^{\pm}$ bosons,

$$\ell_R^1 = e_R \; , \ell_R^2 = \mu_R \; , \ell_R^3 = \tau_R \; . \tag{1.3}$$

Like the left-handed leptons, the left-handed quarks interact through the weak force and can be written in doublets consisting of an up-type and a down-type quark,

$$q_L^f = \begin{pmatrix} u_L^f \\ d_L^f \end{pmatrix} \; , \tag{1.4}$$

where $f \in 1, 2, 3$ refers to its flavour, while the right-handed quarks come in singlets,

$$u_R^f, \; d_R^f \; . \tag{1.5}$$

The SM Lagrangian also obeys local (and global) gauge symmetries, which is linked to the idea that the laws of nature should not depend on certain position dependent (independent) phase transformations of fields that do not appear in physical observables. This type of symmetry plays an important role in the formulation of the SM. Mathematically they refer to the invariance of the Lagrangian of the theory under the actions of members of certain unitary symmetry groups. The identification of a local (or global) gauge symmetry for a particular interaction, e.g. from experiment, restricts the terms that are allowed to be present in its Lagrangian. Furthermore, the preservation of local gauge symmetry is realised through the introduction of covariant derivatives containing additional field(s) with suitable properties, which are interpreted as the gauge bosons that mediate the interaction. The full gauge group of the SM is the $SU(3)_C \times SU(2)_L \times U(1)_Y$, where the $SU(3)_C$ is related to the strong interaction, and the $SU(2)_L \times U(1)_Y$ to the electroweak interaction, which is a unified description of the electromagnetic and weak forces.

1.2 Electromagnetic interaction

Electromagnetic interactions are described by the theory of Quantum Electrodynamics (QED), which has been tested to high precision experimentally and is therefore considered to be one

of the most successful theories to date. Yet it is also one of the simplest QFTs in the SM, and can be used to demonstrate the usefulness of the local gauge principle in the formulation of a consistent description of natural phenomena in the framework of QFT.

The Lagrangian of free fermions, which comes from the Dirac equation, is

$$\mathcal{L}_0 = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \ , \tag{1.6}$$

where ψ is a spinor field, γ^μ are gamma matrices, and m is the mass of the fermion. It is invariant under (global) transformations of the form

$$\psi \to e^{iQ\theta}\psi \ , \tag{1.7}$$

where the $Q\theta$ is an arbitrary real constant. This shows that the choice of the phase of ψ is arbitrary and without physical meaning. However, if the transformation is made position-dependent or local, i.e.,

$$\psi \to e^{iQ\theta(x)}\psi \ , \tag{1.8}$$

then it is no longer invariant due to the derivative term

$$\partial_\mu\psi \to e^{iQ\theta(x)}(\partial_\mu + i\partial_\mu\theta(x))\psi \ . \tag{1.9}$$

This seems unnatural, as it would imply that the physics of the fermion changes depending on the local phase choice. To enforce invariance under local gauge transformation, the gauge field, A_μ, which transforms as

$$A_\mu \to A_\mu + \tfrac{1}{e}\partial_\mu\theta(x) \ , \tag{1.10}$$

where e is the electromagnetic gauge coupling, is introduced along with the covariant derivative, defined as

$$D_\mu \equiv \partial_\mu - ieQA_\mu \ . \tag{1.11}$$

$D_\mu\psi$ then, has the desired transformation property, namely

$$D_\mu\psi \to e^{iQ\theta(x)}D_\mu\psi \ , \tag{1.12}$$

such that the modified Lagrangian,

$$\begin{aligned} \mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \ , \\ &= \mathcal{L}_0 + eQA_\mu\bar{\psi}\gamma^\mu\psi \ , \end{aligned} \tag{1.13}$$

which now contains an interaction term between the fermion and the gauge boson, here identifiable as the photon, is invariant under local phase transformation. Adding the gauge invariant photon kinetic term of $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}$ is electromagnetic field strength tensor, i.e.

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \ , \tag{1.14}$$

the full QED Lagrangian can be written as

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \ . \tag{1.15}$$

The addition of a mass term for the photon of the form of $m^2A^\mu A_\mu$ would violate local gauge invariance, therefore the photon is expected to be massless. Phase rotations of the form of $e^{iQ\theta(x)}$

are members of the unitary group $U(1)$. Therefore QED can be described as a gauge theory with symmetry group $U(1)_{\mathrm{EM}}$. There is a single generator associated with the global $U(1)$ transformation and thus one conserved current, from which a conserved charge (the electric charge Q) is obtained. By writing the interaction term explicitly in terms of this current, $J^\nu = -eQ\bar{\psi}\gamma^\nu\psi$, and using the Euler-Lagrange equation, the following expression can be derived,

$$\partial_\mu F^{\mu\nu} = J^\nu \ , \tag{1.16}$$

which is the covariant form of Maxwell's equations. Thus starting with the description of free fermions via the relativistic Dirac equation and enforcing the correct local gauge symmetry has allowed for the incorporation of the whole of electromagnetism.

Note that the $U(1)_{\mathrm{EM}}$ here is not the same as the $U(1)_{\mathrm{Y}}$ that appears in the full SM symmetry group. This underscores that fact that in the SM framework, the terms of the $\mathcal{L}_{\mathrm{QED}}$ emerge from the Lagrangian of the electroweak interaction (Section 1.4).

1.3 Strong interaction

The theory of the strong interaction — Quantum Chromodynamics (QCD) – describes quarks and gluons and thereby the nature of composite matter particles. Its origin lies in the quark model developed in the 1960s to explain a spectrum of particles, many newly discovered the time, as bound states of hypothetical fundamental particles with fractional electrical charges, named 'quarks'. Six types, or 'flavours', of quarks are found to describe known properties of mesons (e.g. pions) and baryons (e.g. protons), which are bound states of two ($q\bar{q}$), and three quarks (qqq), respectively, and to successfully predict new hadronic states. However, not all descriptions are aligned with expectations. For example, under this framework the wavefunction of the Δ^{++} baryon, a spin-2/3 particle comprised of three u quarks, remains symmetric under the interchange of quark spin and flavour quantum numbers. This contradicts the expectation that as a fermion, it should have an overall antisymmetric wavefunction. This type of discrepancy is resolved by the introduction of a new quantum number, 'color', which becomes the source of the required antisymmetry.

The structure of QCD shows similarities to that of QED. The free Lagrangian for quarks with flavour f is

$$\mathcal{L}_0 = \sum_f \bar{\psi}_f(i\gamma^\mu\partial_\mu - m_f)\psi_f \ , \tag{1.17}$$

where ψ_f is a three component vector in colour space,

$$\psi_f = \begin{pmatrix} \psi_f^r \\ \psi_f^b \\ \psi_f^g \end{pmatrix} \ . \tag{1.18}$$

Analogous to the electric charge and the symmetric group $U(1)_{\mathrm{EM}}$, this 'colour symmetry' is associated with the symmetry group of $SU(3)_C$, which correspond to transformations of the type

$$\psi \to e^{ig_s t^a \theta^a(x)}\psi \ , \tag{1.19}$$

where t^a are the eight generators of $SU(3)_C$ group, which are given by $t^a = \frac{1}{2}\lambda^a$, where λ^a, $a = 1, 2, ...8$ are the Gell-Mann matrices. The Lagrangian is invariant under global $SU(3)_C$ transformations in colour space. To make it invariant also under local transformations, the

quark derivative is replaced by a covariant derivative of the form

$$D_\mu \equiv \partial_\mu - ig_s t^a G^a_\mu \ , \tag{1.20}$$

where g_s is the gauge coupling constant of the strong interaction, and G^a_μ, which denotes the field strength tensors of eight gluons, is given by

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \ , \tag{1.21}$$

where f^{abc} is the structure constant of $SU(3)$, defined by $[t^a, t^b] = if^{abc}t^c$. Adding in the gluon kinetic term, the QCD Lagrangian can be written as

$$\mathcal{L}_{\mathrm{QCD}} \equiv -\frac{1}{4} G^{\mu\nu}_a G^a_{\mu\nu} + \sum_f \bar{\psi}_f (i\gamma^\mu D_\mu - m_f)\psi_f \ , \tag{1.22}$$

which can be expanded into

$$\begin{aligned} \mathcal{L}_{\mathrm{QCD}} = &-\frac{1}{4}(\partial^\mu G^\nu_a - \partial^\nu G^\nu_a)(\partial_\nu G^a_\nu - \partial_\nu G^a_\mu) + \sum_f \bar{\psi}_f (i\gamma^\mu D_\mu - m_f)\psi_f \\ &- \frac{g_s}{2} f^{abc}(\partial^\mu G^\nu_a - \partial^\nu G^\mu_a) G^b_\mu G^c_\nu - \frac{g_s^2}{4} f^{abc} f_{ade} G^\mu_b G^\nu_c G^d_\mu G^e_\nu \ , \end{aligned} \tag{1.23}$$

where the first line containing the kinematic terms of the gluons and quarks and the interaction term between them is fully analogous to that of QED. The gluons, which are massless like the photon, couple only to particles with the colour charge, i.e. to quarks, but not leptons. The second line, which does not appear in QED, describes interactions between gluons, which also carry the charge of its interaction. This feature is expected to be the source of two important characteristics of QCD – confinement and asymptotic freedom.

The strength of the QCD coupling depends strongly on the energy of the interaction. At low energy scales or large distances, it becomes very large, thus single quarks cannot be separated and quarks are confined to states that do not have net colour charge. The size of the coupling in this regime creates difficulties for the calculation of quantities such as hadronic matrix elements from theory. At a scale of around $\sim 1\,\mathrm{GeV/c^2}$, it is of $\mathcal{O}(1)$, thus perturbation methods generally cannot be used close to or below this limit. Conversely at high energies or short distances ($\sim 100\,\mathrm{GeV/c^2}$) it becomes smaller ($\approx 0.1$), and quarks within hadrons behave as though they are free particles. Nevertheless, it is not small per se (for comparison, in QED the value is $\sim 1/137 = 0.007$). Therefore higher order terms in the perturbation expansion are less negligible, which causes QCD calculations to be generally more complicated than that of QED.

1.4 Electroweak unification

The theories of QED and QCD show multiple structural similarities, such as the involvement of massless mediators and the conservation of C, P and CP. In contrast, the weak interaction is mediated by massive gauge bosons, and is found to violate all the aforementioned symmetries. The description of the weak interaction began with the Fermi theory in 1934 [8], which is formulated prior to the discovery of parity violation. Under this framework, the matrix element of the β-decay ($n \to pe^-\bar{\nu}_e$), which is related to its transition probability, is expressed in terms

of a contact interaction that occurs at a single point in space-time, i.e.

$$\mathcal{M}_{\text{eff}} \sim G_F(\bar{\psi}_e\gamma^\mu\psi_\nu)(\bar{\psi}_u\gamma_\mu\psi_d) \ , \tag{1.24}$$

where G_F, the Fermi constant, acts as the coupling constant of the interaction. This is an effective description in the sense that it approximates the full matrix element (featuring the W propagator) in the low energy limit, where the square of the four-momenta of the exchanged virtual particle (q^2) is much smaller than that of the mass of the real W boson ($q^2 \ll m_W^2$). In addition, the spinor combinations of $\bar{\psi}\gamma^\mu\psi$, which is of a vector nature, conserve parity. Following the discovery in 1956 of parity violation in the β-decay of polarised cobalt-60 nuclei [9], the structure of the weak interaction was modified. Instead of only including vector terms, which change sign under parity operation, axial-vector terms of the form of $\bar{\psi}\gamma^\mu\gamma^5\psi$ that do not change sign are added. Combining both in a vector minus axial vector (VA) structure, i.e. $\bar{\psi}\gamma^\mu(1-\gamma^5)\psi$ allows for the introduction of P (and C) violation.

In the 1960s, a unified description of both the weak and the electromagnetic forces started to emerge, motivated by the tantalising possibility of finding more fundamentals theories of nature that underlie the apparently diverse interactions. The resulting Glashow-Weinberg-Salam model of the electroweak interaction [10, 11, 12] is based on the combined $SU(2)_L \times U(1)_Y$ symmetry group, in which several massless gauge bosons mix to produce the photon with its parity conserving interactions, and the parity violating $W^\pm$ and Z^0 bosons. This symmetry group corresponds to the invariance of the free Lagrangian under local gauge transformations of the type

$$\psi(x) \rightarrow e^{ig_1 Y\theta_Y(x)}e^{ig_2 T^a\theta_L^a(x)}\psi(x) \ , \tag{1.25}$$

for left-handed fermion fields (Equations 1.2, 1.4), where T^a denote the three generators of $SU(2)_L$, where the L subscript refers to left-handed fields, Y is the hypercharge operator, and θ_Y and θ_L^a are real numbers. Note that the $SU(2)_L$ transformation, unlike that of the $U(1)_Y$, does not act on the right-handed singlets (Equations 1.3, 1.5). The conservation of local gauge invariance requires the introduction of three gauge bosons for $SU(2)_L$ (W^a), and one for $U(1)_Y$ (B). Their corresponding field strength tensors, written out explicitly, are

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \ , \\ W^a_{\mu\nu} &= \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2\epsilon^{abc}W^b_\mu W^c_\nu \ , \end{aligned} \tag{1.26}$$

where ϵ^{abc} is the antisymmetric structure constant of $SU(2)$. The associated covariant derivative is

$$D_\mu \equiv \partial_\mu - ig_1 Y B_\mu - ig_2 T^a W^a_\mu \ , \tag{1.27}$$

where g_1 and g_2 are the coupling strengths of the hypercharge and weak interactions, respectively. Quantities conserved by the electroweak interaction are the weak isospin, $T_a = 1, 2, 3$, and the hypercharge, for $SU(2)_L$ and $U(1)_Y$, respectively. These are related to the electric charge operator, Q, by

$$Y = Q - T^3 \ . \tag{1.28}$$

The mixing of the W^a and B bosons to give rise to the photon and the Z^0 boson can be described in terms of the weak mixing angle, θ_W, as

$$\begin{pmatrix} Z \\ A \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W^3 \\ B \end{pmatrix} \ , \tag{1.29}$$

where

$$\sin\theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \ , \ \cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \ . \tag{1.30}$$

The expressions for A and Z^0, written using θ_W, are

$$\begin{aligned} A_\mu &= \cos\theta_W B_\mu + \sin\theta_W W^3_\mu \ , \\ Z_\mu &= \cos\theta_W W^3_\mu - \sin\theta_W B_\mu \ . \end{aligned} \tag{1.31}$$

The $W^\pm$ bosons can be identified as combinations of the W^1 and W^2 bosons,

$$W^\pm_\mu = \frac{W^1_\mu \mp i W^2_\mu}{\sqrt{2}} \ . \tag{1.32}$$

The coupling of the electromagnetic interaction, e, can be related to g_1 and g_2 through the weak mixing angle via

$$e = g_1 \cos\theta_W = g_2 \sin\theta_W \ . \tag{1.33}$$

Hence the electroweak interaction is a merged, or 'unified' description of the electromagnetic and weak interactions.

1.4.1 Charged and neutral currents

The interactions between fermions and the gauge bosons of the weak and electromagnetic forces are contained within the covariant derivative terms ($\mathcal{L}_{\text{kin+int}}$ of Equation 1.1), which can be written as

$$\mathcal{L}_{\text{kin+int}} = \sum_{i,\,L,\,R} i \bar{f}_i \gamma^\mu D_\mu f_i \ , \tag{1.34}$$

where γ^μ is the gamma matrix and the f_i denote (chiral) spinor fields. The electroweak part of the covariant derivative, D_μ, given in Equation 1.27, can be written in terms of the photon and the $W^\pm$ and Z^0 fields by making the substitutions

$$\begin{aligned} B_\mu &= \cos\theta_W A_\mu - \sin\theta_W Z_\mu \ , \\ W^3_\mu &= \sin\theta_W A_\mu + \cos\theta_W Z_\mu \ , \\ W^1 T^1 + W^2 T^2 &= \tfrac{1}{\sqrt{2}}(W^+ T^+ + W^- T^-) \ , \end{aligned} \tag{1.35}$$

where $T^\pm$ are the SU(2) raising and lower operators. The resulting expression is

$$\begin{aligned} D_\mu \equiv\ & \partial_\mu \\ & - i\tfrac{g_2}{\sqrt{2}}(W^+_\mu T^+ + W^-_\mu T^-) \\ & - i Z_\mu (g_2 \cos\theta_W T^3 - g_1 \sin\theta_W Y) \\ & - i A_\mu (g_2 \sin\theta_W T^3 + g_1 \cos\theta_W Y) \ , \end{aligned} \tag{1.36}$$

where the first line is the kinetic term, and the remaining three lines, from top to bottom, describe the interactions involving the $W^\pm$ bosons ('charged currents'), the Z^0 boson and the electromagnetic interaction ('neutral currents'), respectively. Inserting the second line of Equation 1.36 into 1.34, and noting that the $W^\pm T^\pm$ term gives zero when acting on right-handed

singlets, the term describing the $W^{\pm}$ interaction is given by

$$\mathcal{L}_W = \frac{g_2}{\sqrt{2}}(\bar{\nu}_L \gamma^\mu W^+_\mu e_L + \bar{u}_L \gamma^\mu W^+_\mu d_L)) + h.c. \ , \tag{1.37}$$

for a single family of leptons and quarks as example (*h.c.* denotes Hermitian conjugate terms omitted for brevity). Using the projection operators, it can be written as

$$\mathcal{L}_W = \frac{g_2}{2\sqrt{2}}(W^+_\mu(\bar{\nu}\gamma^\mu(1-\gamma_5)e + \bar{u}\gamma^\mu(1-\gamma_5)d) + h.c. \ , \tag{1.38}$$

which only involves left-handed doublets and shows its distinctive VA structure. Note that at this point the fermions (and the $W^{\pm}$ boson) are massless and CP remains a good symmetry. The introduction of mass terms takes place through the Higgs mechanism (Section 1.4.2), and the mismatch between the interaction (flavour) basis and the mass basis of the quarks allows for the introduction of CP violation (Section 1.4.3).

The term describing the interaction between fermions and the Z^0 boson can be found using the second line of Equation 1.36 and the relationships of 1.33 and 1.28, as

$$\mathcal{L}_Z = \frac{g_2}{\cos\theta_W} \sum_{i,\,\mathrm{L,\,R}} (T^3 - Q\sin^2\theta_W)\bar{f}_i \gamma^\mu f_i Z_\mu \ , \tag{1.39}$$

which includes both left- and right-handed fermions. Again using P_L, and noting that T^3 is zero for right-handed singlets, it is possible to write it in a more standard form as

$$\mathcal{L}_Z = \frac{g_2}{2\cos\theta_W} \sum_i \bar{\psi}_i \gamma^\mu (c_{iV} - c_{iA}\gamma_5)\psi_i Z^0_\mu \ , \tag{1.40}$$

where the full spinors are used and the vector and axial vector couplings c_{iV} and c_{iA} are given by

$$\begin{aligned} c_{iV} &= T^i_3 - 4Q_i \sin^2\theta_W \ , \\ c_{iA} &= T^i_3 \ . \end{aligned} \tag{1.41}$$

In contrast to the $W^{\pm}$ bosons for which $c_{iV} = c_{iA} = 1$, the Z^0 boson does not couple purely to left-handed fermions, and its coupling strength varies for fermions of different types.

The interactions between fermions and the photon is given by the third line of Equation 1.36. Again, by using Equation 1.33 and 1.28 for e and Q and noting that the term containing T^3 cancels, it can be written in a more familiar form (c.f. Equation 1.13) as

$$\mathcal{L}_\gamma = eA_\mu \sum_i \bar{\psi}_i \gamma^\mu Q_i \psi_i \ , \tag{1.42}$$

which is consistent with the parity conserving nature of the electromagnetic interaction.

1.4.2 The Higgs mechanism

The theory of the electroweak interaction would not be complete without the Higgs mechanism [1, 2]. While the $W^{\pm}$ and Z^0 bosons are known to be massive, no corresponding mass terms can be added to its Lagrangian without breaking the $SU(2)_L \times U(1)_Y$ gauge symmetry. This also affects all the fundamental fermions, as their mass terms would comprise both

left (doublet) and right-handed (singlet) states, which transform differently under $SU(2)_L$. To resolve this issue, in the 1960s, Higgs, Brout and Englert formulated a mechanism where this symmetry is preserved at high energies, but broken spontaneously at low energies. This is realised through the introduction of a single $SU(2)_L$ doublet scalar field (Higgs field). In this way, the $W^{\pm}$ and Z^0 bosons and the fermions can be regarded as being fundamentally massless, but acquire mass through their interactions with the Higgs boson.

Gauge boson mass generation

The Higgs field, ϕ, can be written as a complex doublet of scalar fields, specifically

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} , \tag{1.43}$$

and its Lagrangian is given by,

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) , \tag{1.44}$$

where the $(D_\mu \phi)^\dagger (D^\mu \phi)$ is the kinetic term of ϕ, and the potential term, V, can be written in a general form as

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 . \tag{1.45}$$

When $-\mu^2$ is chosen to be negative and λ positive, the ground state is no longer invariant under the initial symmetry of $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ of the Lagrangian, thus it is said to be spontaneously broken. The value of the potential at the minimum, the vacuum expectation value, is

$$\nu = \frac{\mu^2}{\lambda} . \tag{1.46}$$

Expanding ϕ about this vacuum state and removing non-physical degrees of freedom by gauge choice (unitary gauge), it can be written as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h \end{pmatrix} , \tag{1.47}$$

where h is the real, massive scalar field of the Higgs boson. After spontaneous symmetry breaking, the residual symmetry obeyed by the Lagrangian is that of the $U(1)_{\text{EM}}$ of electromagnetism, which is a linear combination of the original $U(1)_Y$ and a subgroup of $SU(2)_L$. In this process, the massless electroweak gauge bosons of W^a and B mix to give rise to the massive $W^{\pm}$ and Z^0 bosons of the weak force. The relevant terms are contained within $\mathcal{L}_{\text{Higgs}}$, namely

$$\begin{aligned} (D_\mu \phi^\dagger)(D^\mu \phi) = & \tfrac{1}{2} (\partial_\mu h)(\partial^\mu h) \\ & + \tfrac{1}{8} g_2^2 (W_\mu^1 - i W_\mu^2)(W^{1\mu} + i W^{2\mu})(\nu + h) \\ & + \tfrac{1}{2} Y^2 (g_2 W_\mu^3 - g_1 B_\mu)^2 (\nu + h)^2 , \end{aligned} \tag{1.48}$$

the first line of which is the kinetic term of the Higgs boson, and the second and third lines result in the mass terms of the $W^{\pm}$ and Z^0 bosons, respectively. Using the relationship of 1.31, the second line can be written as

$$\tfrac{1}{8} g_2^2 W_\mu^+ W^{\mu -} (\nu + h) , \tag{1.49}$$

which contains terms that include h and $W^{\pm}$. The former represents the coupling between the Higgs and the $W^{\pm}$ bosons, and the latter is the mass term, $\frac{1}{2}m_W^2 W_\mu^+ W^{\mu-}$, where $m_W^2 = g_2^2\nu^2/4$, with $Y = 1/2$ (Higgs hypercharge value). Similarly, writing an expression for the Z^0 in terms of the electroweak bosons as

$$Z_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W_\mu^3 - g_1 B_\mu) \ , \tag{1.50}$$

allows for the ν^2 term of the third line to be written as $\frac{1}{2}m_Z^2 Z_\mu Z^\mu$, where $m_Z^2 = (g_1^2 + g_2^2)\nu^2/4$, which is again in the form of a mass term. Note that the photon, which is in an orthogonal state to the Z^0 boson,

$$A_\mu = \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2 W_\mu^3 + g_1 B_\mu) \ , \tag{1.51}$$

does not have a corresponding term in Equation 1.48 and therefore remains massless as required.

Lepton and quark masses

The generation of mass terms for the fundamental fermions is realised through $\mathcal{L}_{\text{Yukawa}}$, which contains the couplings between the scalar field and the fermions. It is of the form of

$$\mathcal{L}_{\text{Yukawa}} = -y_f(\bar{f}_L \phi f_R + \bar{f}_R \phi f_L) \ , \tag{1.52}$$

where y_f is the Yukawa coupling that differs for the three generations of fermions (f). For the leptons (ℓ), inserting the Higgs vacuum expectation value (Equation 1.47) into the Yukawa term leads to the following expression

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -\sum_f \frac{y_f^\ell}{\sqrt{2}}(\bar{e}_L^f(\nu + h)e_R^f + \bar{e}_R^f(\nu + h)e_L^f) \ , \tag{1.53}$$

where the summation is over the three generations ($e_{L,R}$ is used to refer to the field of a charged lepton). Using $\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R = \bar{\psi}\psi$, this expression can be simplified into

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -\sum_f (\frac{y_f^\ell \nu}{\sqrt{2}}\bar{e}^f e^f + \frac{y_f^\ell}{\sqrt{2}} h \bar{e}^f e^f) \ , \tag{1.54}$$

the first term of which is in the form of a mass term, and the second contains the couplings between the Higgs and the charged leptons. The masses of the leptons are given by

$$m_\ell^f = \frac{y_f^\ell \nu}{\sqrt{2}} \ , \tag{1.55}$$

which depend on the Yukawa couplings that increase in size from the first generation electrons to the third generation tau leptons.

The masses of the quarks are generated in a similar fashion. The Yukawa term for the down-type quarks, the fields of which are denoted by d (and that of an up or down-quark is denoted by q), is given by

$$\mathcal{L}^{q_d}_{\text{Yukawa}} = -\sum_{i=1}^{3}\sum_{j=1}^{3}(y_{ij}^d \bar{q}_L^i \phi d_R^j + (y_{ij}^d)^* \bar{d}_R^j \phi^\dagger q_L^i) \ , \tag{1.56}$$

where the couplings y^d_{ij} are now the complex entries of a 3×3 matrix. To write a term for the up-type quarks, it is possible to define the conjugate Higgs doublet,

$$\tilde{\phi} \equiv i\sigma_2 \phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} \nu + h \\ 0 \end{pmatrix} \ , \tag{1.57}$$

which has the same transformation properties as the original. The resulting expression is then analogously

$$\mathcal{L}^{q_u}_{\text{Yukawa}} = -\sum_{i=1}^{3}\sum_{j=1}^{3} (y^u_{ij} \bar{q}^i_L \tilde{\phi} u^j_R + (y^u_{ij})^* \bar{u}^j_R \tilde{\phi}^\dagger q^i_L) \ . \tag{1.58}$$

After inserting the Higgs vacuum expectation value, the quark mass terms are found to be

$$\begin{aligned} \mathcal{L}^q_{\text{Yukawa}} = &- \begin{pmatrix} \bar{u}_1 & \bar{u}_2 & \bar{u}_3 \end{pmatrix}_R M^u_{ij} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}_L \\ &- \begin{pmatrix} \bar{d}_1 & \bar{d}_2 & \bar{d}_3 \end{pmatrix}_R M^u_{ij} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}_L + h.c. \ , \end{aligned} \tag{1.59}$$

where

$$M^u_{ij} = \frac{y^u_{ij}\nu}{\sqrt{2}}, \ M^d_{ij} = \frac{y^d_{ij}\nu}{\sqrt{2}} \ , \tag{1.60}$$

are the entries of the mass matrices for up and down-type quarks. Note that these matrices are in flavour space. They are complex, and contain off-diagonal terms. To obtain quark mass eigenstates, M^d_{ij} and M^u_{ij} need to be diagonalised. This can be achieved using four unitary matrices, U_L, U_R, D_L and D_R acting on up and down-type quarks, respectively, which rotate the quark fields between the flavour (subscript $i = 1, 2, 3$) and mass basis (u, d, c, s, t, b) as

$$\begin{aligned} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}_{L,\,R} &= U_{L,\,R} \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,\,R} \ , \\ \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}_{L,\,R} &= D_{L,\,R} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,\,R} \ . \end{aligned} \tag{1.61}$$

Inserting the expressions of 1.61 into 1.59 results in

$$\begin{aligned} \mathcal{L}^q_{\text{Yukawa}} = &- \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_R U^\dagger_R M^u_{ij} U_L \begin{pmatrix} u \\ c \\ t \end{pmatrix} \\ &- \begin{pmatrix} \bar{d} & \bar{s} & \bar{b} \end{pmatrix}_R D^\dagger_R M^u_{ij} D_L \begin{pmatrix} d \\ s \\ b \end{pmatrix} + h.c. \ , \end{aligned} \tag{1.62}$$

where

$$
\begin{aligned}
U_R^\dagger M^u U_L &= \frac{\nu}{\sqrt{2}} \begin{pmatrix} y_u^u & 0 & 0 \\ 0 & y_c^u & 0 \\ 0 & 0 & y_t^u \end{pmatrix} = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \ , \\
D_R^\dagger M^u D_L &= \frac{\nu}{\sqrt{2}} \begin{pmatrix} y_d^d & 0 & 0 \\ 0 & y_s^d & 0 \\ 0 & 0 & y_b^d \end{pmatrix} = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} \ ,
\end{aligned}
\tag{1.63}
$$

are the diagonalised mass matrices for the quarks, which contains terms of the form $m_q^f \bar{q_f} q_f$, where $m_q^f = y_f \nu / \sqrt{2}$.

1.4.3 CKM matrix

In the expressions of the Lagrangians of the charged and neutral current interactions, i.e. Equations 1.39, 1.38 and 1.42, the quark fields are given in the flavour basis, and can be rotated to the mass basis using the matrices of 1.61. Doing so leaves the neutral current Lagrangians unchanged, as the terms involved are of the form of $\bar{u}\gamma^\mu u$ (and $\bar{d}\gamma^\mu d$). This correctly reflects the lack of quark flavour mixing in these interactions. The situation is different for the charged current due to terms of the form $\bar{u}^f \gamma^\mu (1-\gamma_5) d^f$, which lead to

$$
\begin{aligned}
\bar{u}^f \gamma^\mu (1-\gamma_5) d^f &= \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L U_L^\dagger \gamma^\mu D_L \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \dots \ , \\
&= \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix}_L \gamma^\mu V \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L + \dots \ ,
\end{aligned}
\tag{1.64}
$$

where the extra factor of $V = U_L^\dagger D_L$ is the Cabibbo-Kobayashi-Maskawa [13, 14] (CKM) matrix, which is a unitary matrix that contains the couplings of the $W^\pm$ to pairs of up and down-type quarks. It can be written as

$$
\begin{aligned}
V_{\text{CKM}} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\
&= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}c_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \ ,
\end{aligned}
\tag{1.65}
$$

in which $s_{ij} \equiv \sin\theta_{ij}$ and $c_{ij} \equiv \cos\theta_{ij}$, where the $i, j = 1, 2, 3$ refer to the three generations, θ_{12}, θ_{13}, θ_{23} are the (real) quark mixing angles, and δ is the complex phase responsible for CP violation. This is also the only source of CP violation in the SM. These four parameters need to be determined experimentally. The magnitudes of the measured V_{CKM} elements show clear hierarchical structure, with strong couplings within the same generations of $\mathcal{O}(1)$, reduced couplings between first and second generations of around 0.22 for $|V_{cd}|$ and $|V_{us}|$, and between second and third generations of around 0.04 for $|V_{cb}|$ and $|V_{ts}|$, and very small couplings between first and third generation quarks of < 0.01 [7].

1.5 Beyond the Standard Model

The Standard Model is a remarkably successful theory. With the discovery of the Higgs boson in 2012 [15, 16], all fundamental particles of the model have been found. One striking example of its predictive power is in the measurement of the electron anomalous magnetic moment, the result of which agrees with theoretical prediction up to the 12th digit [17, 18]:

$$\begin{aligned} a_e^{\text{experiment}} &= 1.00115965218073 \pm 0.00000000000028 \\ a_e^{\text{theory}} &= 1.00115965218178 \pm 0.00000000000077 \ . \end{aligned} \tag{1.66}$$

However, despite its successes, it leaves key observations unexplained:

- The SM incorporates three out of the four known fundamental forces – it does not include the gravitational interaction;
- The only significant source of CP violation in the SM is the complex phase that appears in the CKM matrix, which is insufficient to explain the observed asymmetry between matter and anti-matter in the universe [19];
- The SM explains around 5% of the matter content in the universe – it does not appear to provide candidates that can satisfactorily explain dark matter, the presence of which is strongly suggested by cosmological observations [20].

Furthermore, on the theoretical front, several aspects of the model are considered puzzling or unsatisfactory:

- The SM has a large number (~ 18)[2] of free parameters that are not predicted from theory and need to be measured.
- The couplings in the CKM matrix show a hierarchical structure that may hint at an underlying mechanism, but is not currently explained. Another unexplained, but intriguing structure is existence of three families of quarks and leptons that differ greatly in mass but are otherwise identical.
- Experimental evidence of neutrino oscillation, e.g. [21], strongly suggest that they are not massless[3], albeit their masses are expected to be very small. While it is possible to generate mass terms using the Higgs mechanism (e.g. through the addition of three right-handed neutrino fields, ν_{Ri}, where $i = 1, 2, 3$, and via the strategy used for the up-type quarks), their extremely small values prompt the question of whether or not a different mechanism may be responsible.
- Like the weak force, the structure of the QCD Lagrangian allows for CP violation through a CP violating phase. However, the strong force is observed to preserve this symmetry, and this phase needs to be made vanishingly small $(< 10^{-9})$ without strong theoretical justifications [22].
- The mass of the Higgs boson is very sensitive to quantum corrections. Obtaining a value consistent with existing experimental constraints requires precise cancellations between the bare mass and these corrections, which constitutes a level of fine-tuning that is often seen as 'unnatural' [23, 24, 25].

[2] This number varies in the literature depending on the choices made when counting, such as the inclusion of parameters related to neutrino masses and mixing.

[3] They have been treated as exactly massless in the previous discussions of the SM.

These shortcomings suggest that the SM could be considered as an effective description of nature valid at low energies, and that new fields and interactions may be present.

The search for 'New Physics' (NP) can take place in a direct or indirect way. Direct searches often involve the production and detection of new particles at colliders. While capable of producing unambiguous results, it suffers from the drawback of being limited by the maximum achievable energy. Indirect searches tend to focus on the detection of subtle effects through precision measurements of the properties of decays that can be modified from SM expectations by contributions from virtual NP particles. They have higher energy reach, which allows for the detection of NP effects before the technology and/or equipment necessary for direct production becomes available. Therefore, they are complementary to direct searches, and provide important constraints on NP scenarios. The study of the decays of hadrons containing b quarks is one of the most promising ways to carry it out, a particular type of which is the focus of the sections below.

2 The $B^0 \to K^{*0} e^+ e^-$ decay

The angular analysis of the $B^0 \to K^{*0} e^+ e^-$ decay makes up the core of this thesis. In the following, important theoretical aspects related to $B^0 \to K^{*0} \ell^+ \ell^-$ decays in general are introduced and relevant observables as well as experimental measurements are discussed.

2.1 Flavour changing neutral current

The decay of $B^0 \to K^{*0} \ell^+ \ell^-$ involves an underlying $b \to s \ell^+ \ell^-$ quark transition. This is a 'flavour changing neutral current' (FCNC) process, where the quark flavour changes (s to b) without a change in the electric charge. In the SM, the only type of interaction that can change flavour are charged current interactions involving the $W^\pm$ bosons. This means that FCNC processes cannot take place at leading order ('tree-level'), but must proceed through higher order ('loop-level') terms, examples of which are shown in Figure 2.1. Such decays tend to be have low branching fractions (are 'rare') as they are higher order effects that are further suppressed by the small sizes of the off-diagonal couplings in the CKM matrix. These characteristics make them useful for the indirect search of NP, as its impact may be proportionally large. Examples of mediation by NP particles are shown in Figure 2.2.

2.2 Effective field theory

In the field of flavour physics, the decays of B mesons are often studied using the formalism of effective field theory (EFT) [26, 27, 28]. This is because they involve widely separated energy scales, which makes direct calculations from the full theory intractable. The weak interaction responsible for FCNCs is governed by the electroweak scale set by mass of the $W^\pm$ boson, $m_W \sim 80\,\mathrm{GeV/c^2}$. The strong interaction responsible for hadron formation is governed by the scale of the strong interaction, $\Lambda_{\mathrm{QCD}} \sim 0.2\,\mathrm{GeV/c^2}$. The mass of the b quark, $m_b \sim 4\,\mathrm{GeV/c^2}$, defines an intermediate scale that is lower than m_W, but higher than Λ_{QCD}. The framework of EFT simplifies this problem by providing an approximation to the full theory that is valid at the relevant energy scale, in this case m_b. An important part of this process is the use of the local operator product expansion (OPE) method, which treats the weak interaction as point-like from the view of m_b (and Λ_{QCD}). The high energy, or short distance effects above the m_b scale are then encoded in the Wilson Coefficients (WCs), which are multiplied to local operators that contain only light SM fields, i.e. leptons, photons, gluons, quarks, but excluding the top quark (as $m_t \gg m_b$). This is akin to Fermi's theory of weak interactions in which the interaction vertices are treated as a single point, and the full theory is contained within the coupling constant G_F. The initial conditions of the WCs are obtained at the electroweak scale by matching the decay amplitudes of the full theory to those of the effective theory. They are then evolved down to the energy scale of m_b using the renormalisation group equations [29, 30]. The presence of NP at high energy corresponds to the existence of alternative initial conditions,

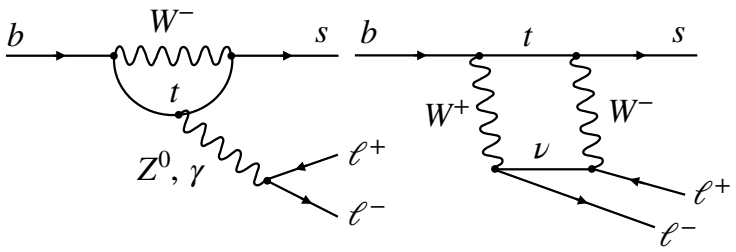


Figure 2.1: Leading order SM penguin (left) and box (right) Feynman diagrams of the $b \to s\ell^+\ell^-$ transition.

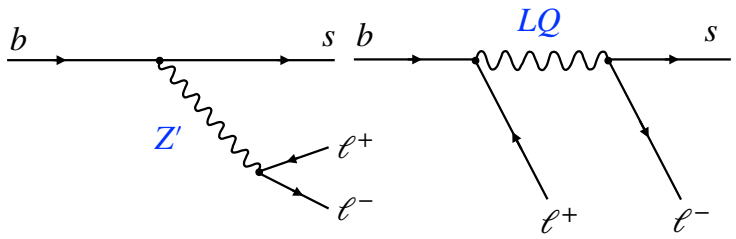


Figure 2.2: Hypothetical tree-level Feynman diagrams of the $b \to s\ell^+\ell^-$ transition involving a Z' boson (left) and a leptoquark (right).

leading to shifts from SM values, i.e.

$$\mathcal{C}_i \to \mathcal{C}_i^{\text{SM}} + \mathcal{C}_i^{\text{NP}} \ , \tag{2.1}$$

where the WCs are denoted by $\mathcal{C}_i$. Under this framework, the effective Lagrangian of the $b \to s$ transition is given by[1] [27]

$$\mathcal{L}_{\text{eff}} = \frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \frac{\alpha_e}{4\pi} \sum_i \mathcal{C}_i \mathcal{O}_i \ , \tag{2.2}$$

where G_F is the Fermi constant, V_{tb} and V_{ts} are CKM matrix elements, α_e is the fine structure constant, and the $\mathcal{O}_i$ are the local operators. While the summation extends in principle over an infinite number of operators, for a given problem and to a chosen precision, only a finite number are important. The typical choice for flavour physics is to keep ones up to dimension six. The most relevant among these include the radiative ($\mathcal{O}_7$) and electroweak penguin operators ($\mathcal{O}_9$ and $\mathcal{O}_{10}$),[2]

$$\begin{aligned} \mathcal{O}_7 &= \frac{m_b}{e} \bar{s}\sigma^{\mu\nu} P_R b F_{\mu\nu} \ , \\ \mathcal{O}_9 &= \bar{s}\gamma_\mu P_L b \bar{\ell}\gamma^\mu \ell \ , \\ \mathcal{O}_{10} &= \bar{s}\gamma_\mu P_L b \bar{\ell}\gamma^\mu \gamma_5 \ell \ , \end{aligned} \tag{2.3}$$

[1] Contributions from the u quark are suppressed by the small $V_{ub}V_{us}^*$ factor and can be neglected.

[2] These operators also have chirality flipped counterparts ($\mathcal{O}'$) that can be obtained by exchanging $P_L \leftrightarrow P_R$. Their WCs ($\mathcal{C}_i'$) are suppressed in the SM due to the left-handed nature of the charged current interaction, but this may not be the case for NP.

and the current-current operators $\mathcal{O}_1$ and $\mathcal{O}_2$,

$$\begin{aligned}\mathcal{O}_1^q &= \frac{4\pi}{\alpha_e}\bar{s}\gamma_\mu P_L b\bar{c}\gamma^\mu P_L c \ , \\ \mathcal{O}_2^q &= \frac{4\pi}{\alpha_e}\bar{s}\gamma_\mu P_L c\bar{c}\gamma^\mu P_L b \ .\end{aligned} \tag{2.4}$$

In these expressions, s, b, c and ℓ denote the quark and lepton fields. The SM values of $\mathcal{C}_7$, $\mathcal{C}_9$ and $\mathcal{C}_{10}$, calculate at next-to-next-leading order, are [31]

$$\mathcal{C}_7^{\rm SM} = -0.3 \ , \quad \mathcal{C}_9^{\rm SM} = +4.2 \ , \quad \mathcal{C}_{10}^{\rm SM} = -4.2 \ . \tag{2.5}$$

The WCs $\mathcal{C}_{1,2}^{\rm SM}$ are comparable in size to $\mathcal{C}_{7,9,10}^{\rm SM}$, with $\mathcal{C}_1^{\rm SM} \sim -0.2$ and $\mathcal{C}_2^{\rm SM} \sim 1.0$.

The decay of a B^0 meson to a $K^{*0}\ell^+\ell^-$ final state does not only proceed through the FCNC $b \to s\ell^+\ell^-$ transition ('rare mode'), but also, for example, through intermediate resonances. This is reflected by the terms of the $\mathcal{L}_{\rm eff}$, and can be illustrated using its differential decay rate as a function of the dilepton invariant mass squared, q^2, as shown in Figure 2.3. Experimental sensitivity to the WCs depends on the q^2 regime. At very low q^2 ($\lesssim 1\,{\rm GeV}^2/{\rm c}^4$), the dominate contribution comes from the $\mathcal{O}_7$ term, which corresponds to the $b \to s\gamma$ transition. Observables of this type of radiative decays such as $B \to K^*\gamma(\to \ell^+\ell^-)$ are sensitive to WC $\mathcal{C}_7$. Existing measurements of this and other modes (e.g. $B \to X_s\gamma(\to \ell^+\ell^-)$, where X_s denotes a hadron containing an s quark) [32, 33, 34, 35, 36] show good agreement with SM predictions e.g. [37], which set strong constraints on NP in this WC. The $b \to s\ell^+\ell^-$ transition, which corresponds to operators $\mathcal{O}_9$ and $\mathcal{O}_{10}$, plays a leading role in the 'central q^2' region of $1 \lesssim q^2 \lesssim 6.0\,{\rm GeV}^2/{\rm c}^4$. At higher q^2 values, contributions from $\mathcal{O}_{1,2}$ operators that describe the tree-level $b \to sc\bar{c}$ process become increasingly prominent until the virtual charm loop goes on-shell (hadronise into a real meson) around the masses of the J/ψ (and $\psi(2S)$) resonances at $\sim 9\,{\rm GeV}^2/{\rm c}^4$ ($\sim 13.5\,{\rm GeV}^2/{\rm c}^4$), and the differential branching fraction becomes dominated by charmonium decays (e.g. $B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)$). These modes constitute a dangerous source of background for studies of the rare mode, therefore the q^2 regions where they dominate are generally excluded from measurements. The high q^2 region ($\gtrsim 14\,{\rm GeV}^2/{\rm c}^4$), in particular above the energy threshold beyond which pairs of c-hadrons (e.g. $D\bar{D}$ pairs) can be produced from $c\bar{c}$ pairs (open charm threshold), is populated with broad $c\bar{c}$ resonances. While it offers potential access to WC $\mathcal{C}_9$ and $\mathcal{C}_{10}$, it is more challenging to analyse. Due to the different experimental and theoretical features present at different q^2 values, analyses are often tailored to specific regions. The relatively clean central q^2 region is generally favoured for the measurements of rare mode observables, and it is also used for the analysis discussed in this thesis.

2.2.1 Computation of observables

The calculation of observables requires values for the WCs, which can be obtained relatively precisely using perturbative methods due to the asymptotic freedom of QCD. In contrast, the operators that encode long distance (non-perturbative) effects are the leading sources of theoretical uncertainties. The amplitude of the decay of a B^0 meson to a $K^{*0}\ell^+\ell^-$ final state can be obtained by inserting the $\mathcal{L}_{\rm eff}$ between the initial and final hadronic states. It can be written illustratively as [28]

$$\langle K^{*0}\ell^+\ell^-|\mathcal{L}_{\rm eff}|B^0\rangle = \frac{G_F}{\sqrt{2}}\sum_i V_{tb}V_{ts}^* C_i \langle K^{*0}\ell^+\ell^-|\mathcal{O}_i|B^0\rangle \ , \tag{2.6}$$

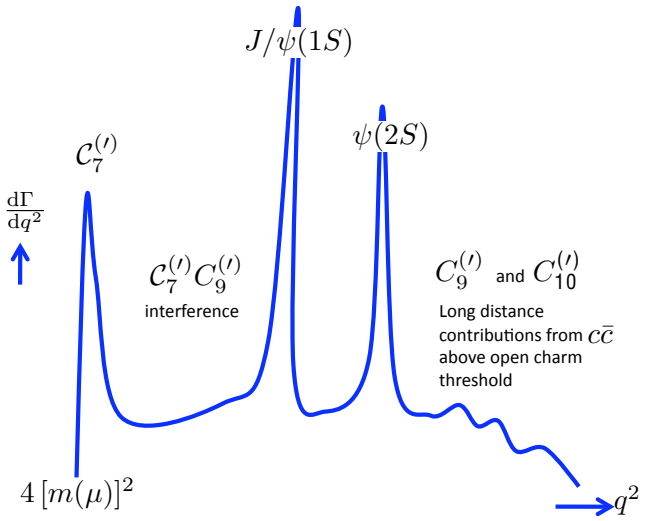


Figure 2.3: Schematic illustration of the q^2 dependence of the differential decay rate of the $B^0 \to K^{*0}\ell^+\ell^-$ decays [38].

where the $\langle K^{*0}\ell^+\ell^-|\mathcal{O}_i|B^0\rangle$ terms are the matrix elements. The calculations can be made under a 'naive factorsation' [39] scheme, where the operators are expressed as products of lepton and quark currents ($\mathcal{O}_i = J_{i\ell} \cdot J_{iq}$), and the full expression is separated into a (trivial) leptonic and a hadronic part, i.e.

$$\langle K^{*0}\ell^+\ell^-|\mathcal{O}_i|B^0\rangle \sim \langle \ell^+\ell^-|J_{i\ell}|0\rangle \cdot \langle K^{*0}|J_{iq}|B^0\rangle \,, \tag{2.7}$$

where the latter describes the $B^0 \to K^{*0}$ transition. These amplitudes can be parametrised using QCD form factors, which are non-perturbative quantities that can be calculated numerically using the framework of Lattice QCD at high q^2 or analytically via Light Cone Sum Rules (LCSR) at low q^2. Both tend to be affected by uncertainties that are sometimes relatively large. Besides form factors, another source of uncertainty arises from long distance charm loop contributions associated with the $\mathcal{O}_{1,2}$ operators, which are not fully factorisable. While this effect is at its strongest in regions dominated by charmonium resonances, its influence extends to some extent throughout the full q^2 range [40, 41]. Mathematically, this contribution has a vector structure like $\mathcal{O}_9$, and therefore enters into the decay amplitude in a similar way [42]. This can lead to difficulties in disentangling NP from SM effects in particular for WC $\mathcal{C}_9$. As a result, there is significant ongoing effort in the theory community to improve its assessment.

2.3 Angular definitions

The decay of $B^0 \to K^{*0}(\to K^+\pi^-)\ell^+\ell^-$, where the K^{*0} refers to the $K^{*0}(892)$ vector meson and its CP conjugate mode of $\bar{B}^0 \to \bar{K}^{*0}(\to K^-\pi^+)\ell^+\ell^-$, provide access to a large number of observables. In particular, the presence of the K^{*0} vector meson introduces additional degrees of freedom to the system, and allows the flavour of the B^0 ($\bar{B}^0$) meson to be unambiguously determined [43].

For an on-shell K^* meson, the distribution of its final state particles can be fully described by three angles, θ_ℓ, θ_K, ϕ and q^2. The definitions of the angles are illustrated in Figure 2.4, following the convention adopted by the LHCb collaboration [44].[3] The angle θ_ℓ is the angle

[3]The convention used in the theory community is often different e.g. [43, 45].

between the direction of the $\ell^+\,(\ell^-)$ in the dilepton rest frame, and that of the dilepton in the $B^0\,(\bar{B}^0)$ meson rest frame. The angle θ_K is the angle between the direction of the kaon in the $K^{*0}\,(\bar{K}^{*0})$ meson rest frame and that of the $K^{*0}\,(\bar{K}^{*0})$ meson in the $B^0\,(\bar{B}^0)$ meson rest frame. The angle ϕ is the angle between the planes containing the two leptons and the plane containing the kaon and the pion. Explicitly, the angles are calculated as follows for the decay of the B^0 meson,

$$\begin{aligned}
\cos\theta_K &= \left(\hat{p}_{K^+}^{(K^{*0})}\right)\cdot\left(\hat{p}_{K^{*0}}^{(B^0)}\right) = \left(\hat{p}_{K^+}^{(K^{*0})}\right)\cdot\left(-\hat{p}_{B^0}^{(K^{*0})}\right)\ , \\
\cos\theta_\ell &= \left(\hat{p}_{\ell^+}^{(\ell^+\ell^-)}\right)\cdot\left(\hat{p}_{(\ell^+\ell^-)}^{(B^0)}\right) = \left(\hat{p}_{\ell^+}^{(\ell^+\ell^-)}\right)\cdot\left(-\hat{p}_{B^0}^{(\ell^+\ell^-)}\right)\ , \\
\phi &= \arccos\left[\left(\hat{p}_{\ell^+}^{(B^0)}\times\hat{p}_{\ell^-}^{(B^0)}\right)\cdot\left(\hat{p}_{K^+}^{(B^0)}\times\hat{p}_{\pi^-}^{(B^0)}\right)\right]\ ,
\end{aligned} \tag{2.8}$$

where $\hat{p}_x^y$ refers to the unit vector giving the direction of x in the rest frame of y, and for its CP conjugate $\bar{B}^0$,

$$\begin{aligned}
\cos\theta_K &= \left(\hat{p}_{K^-}^{(\bar{K}^{*0})}\right)\cdot\left(\hat{p}_{\bar{K}^{*0}}^{(\bar{B}^0)}\right) = \left(\hat{p}_{K^-}^{(\bar{K}^{*0})}\right)\cdot\left(-\hat{p}_{\bar{B}^0}^{(\bar{K}^{*0})}\right)\ , \\
\cos\theta_\ell &= \left(\hat{p}_{\ell^-}^{(\ell^+\ell^-)}\right)\cdot\left(\hat{p}_{(\ell^+\ell^-)}^{(\bar{B}^0)}\right) = \left(\hat{p}_{\ell^-}^{(\ell^+\ell^-)}\right)\cdot\left(-\hat{p}_{\bar{B}^0}^{(\ell^+\ell^-)}\right)\ , \\
\phi &= \arccos\left[-\left(\hat{p}_{\ell^+}^{(\bar{B}^0)}\times\hat{p}_{\ell^-}^{(\bar{B}^0)}\right)\cdot\left(\hat{p}_{K^-}^{(\bar{B}^0)}\times\hat{p}_{\pi^+}^{(\bar{B}^0)}\right)\right]\ .
\end{aligned} \tag{2.9}$$

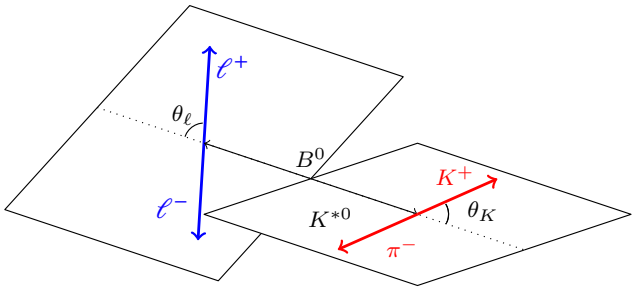
In both cases, the sign of the ϕ angle is determined using

$$\sin\phi = \left[\left(\hat{p}_{\ell^+}^{(B^0/\bar{B}^0)}\times\hat{p}_{\ell^-}^{(B^0/\bar{B}^0)}\right)\times\left(\hat{p}_{K^\pm}^{(B^0/\bar{B}^0)}\times\hat{p}_{\pi^\mp}^{(B^0/\bar{B}^0)}\right)\right]\cdot\hat{p}_{(K^{*0}/\bar{K}^{*0})}^{(B^0/\bar{B}^0)}\ . \tag{2.10}$$

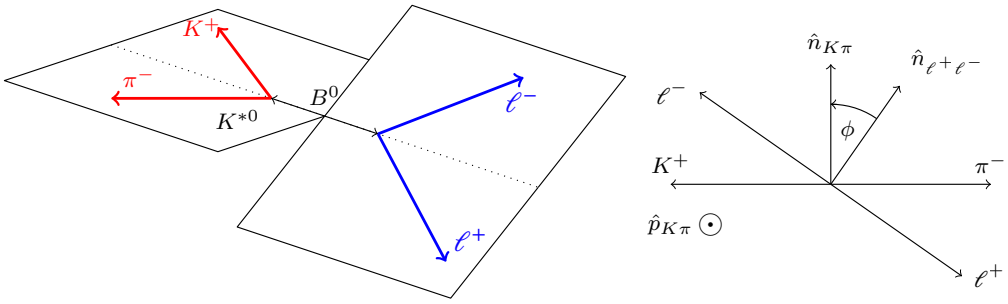
The boundaries of the angular phase space are given by $-1<\cos\theta_\ell<1$, $-1<\cos\theta_K<1$ and $-\pi<\phi<\pi$.

2.4 Differential decay rate

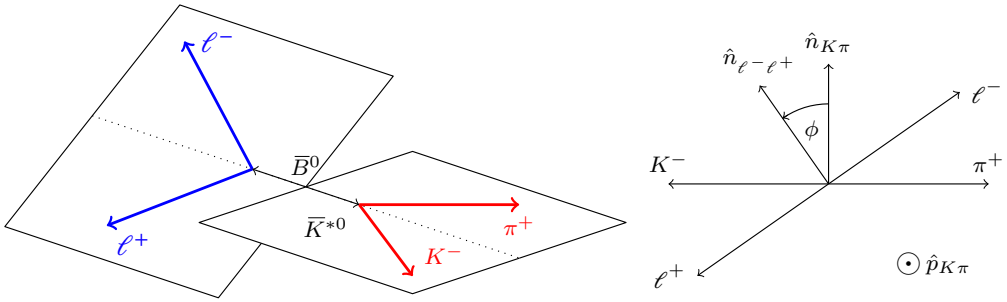
To obtain the differential decay rate of the $B^0\to K^{*0}(\to K\pi)\ell^+\ell^-$ decay, the hadronic part of its matrix element can be written in terms of $B^0\to K^{*0}$ form factors with the assumptions that the K^{*0} resonance is on-shell and has a narrow width. Squaring the matrix element, summing over the spin of the leptons, integrating over the $K\pi$ invariant mass, and making use of the



(a) θ_K and θ_ℓ definitions for the B^0 decay



(b) ϕ definition for the B^0 decay



(c) ϕ definition for the $\overline{B}^0$ decay

Figure 2.4: Illustration of the angular basis used to parametrise the $B^0 \to K^{*0}(\to K^+\pi^-)\ell^+\ell^-$ decay [44].

four-body phase-space with the angular definitions of Section 2.3, allows it to be written as [46]

$$\begin{aligned}\frac{d^4\Gamma[\bar{B}^0 \to \bar{K}^{*0}\ell^+\ell^-]}{dq^2\, d\bar{\Omega}} &= \frac{9}{32\pi}\sum_i I_i(q^2,\bar{\Omega}) \\ &= \frac{9}{32\pi}\Big[\, I_{1s}\sin^2\theta_K + I_{1c}\cos^2\theta_K + \\ &\qquad I_{2s}\sin^2\theta_K\cos 2\theta_\ell + I_{2c}\cos^2\theta_K\cos 2\theta_\ell + \\ &\qquad I_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + I_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi + \\ &\qquad I_5\sin 2\theta_K\sin\theta_\ell\cos\phi + I_6^s\sin^2\theta_K\cos\theta_\ell + \\ &\qquad I_7\sin 2\theta_K\sin\theta_\ell\sin\phi + I_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi + \\ &\qquad I_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\,\Big]\ .\end{aligned} \tag{2.11}$$

The trigonometric terms in this equation stem from the phase space of the decay, while the I_i observables depend on seven complex K^{*0} amplitudes that are denoted as $\mathcal{A}_0^{L/R}$, $\mathcal{A}_\parallel^{L/R}$, $\mathcal{A}_\perp^{L/R}$ and $\mathcal{A}_t$, where the 0, $\parallel$, $\perp$ and t subscripts refer to the transversity states of the K^{*0} meson, namely its one longitudinal, two transverse and one time-like state. The L and R superscripts correspond to the chirality of the lepton current. The I_i terms and their corresponding amplitude dependencies are summarised in Table 2.1, following the convention of [47].

In addition to $\mathcal{A}_{0,\parallel,\perp,t}^{L,R}$, which refer to the spin-1 K^{*0} meson (P-wave), the $K\pi$ system can also be in a spin-0 configuration (S-wave). Experimentally these two contributions cannot be disentangled. Sources of the latter include non-resonant $K\pi$ pairs, and $K\pi$ pairs originating from the decay of a spin-0 resonance (e.g. $K^{*0}(1430)$). The S-wave and its interference with the P-wave introduces two new complex amplitudes and additional observables. However, its contribution is relatively small.

2.5 Angular observables

The angular distributions of $B^0 \to K^{*0}\ell^+\ell^-$ decays and its CP conjugate mode of $\bar{B}^0 \to \bar{K}^{*0}\ell^+\ell^-$ can be described by Equation 2.11, specifically

$$\begin{aligned}\frac{\mathrm{d}^4\Gamma[\bar{B}^0 \to \bar{K}^{*0}\ell^+\ell^-]}{\mathrm{d}q^2\,\mathrm{d}\bar{\Omega}} &= \frac{9}{32\pi}\sum_i I_i(q^2,\bar{\Omega})\ , \\ \frac{\mathrm{d}^4\bar{\Gamma}[B^0 \to K^{*0}\ell^+\ell^-]}{\mathrm{d}q^2\,\mathrm{d}\bar{\Omega}} &= \frac{9}{32\pi}\sum_i \bar{I}_i(q^2,\bar{\Omega})\ .\end{aligned} \tag{2.12}$$

	$I_i(q^2)$	$f_i(\bar{\Omega})$
$1s$	$\frac{2+\beta^2}{4}[\lvert\mathcal{A}_\perp^L\rvert^2 + \lvert\mathcal{A}_\parallel^L\rvert^2 + (L \to R)]$	$\sin^2\theta_K$
	$+(1-\beta^2)\mathcal{R}e(A_{\perp L}A_{\perp R}^* + A_{\parallel L}A_{\parallel R}^*)$	
$1c$	$\Big[\lvert\mathcal{A}_0^L\rvert^2 + \lvert\mathcal{A}_0^R\rvert^2\Big] + (1-\beta^2)\Big[\lvert A_t\rvert^2 + 2\mathcal{R}e(A_{0L}A_{0R}^*)\Big]$	$\cos^2\theta_K$
$2s$	$\frac{\beta^2}{4}\Big[\lvert\mathcal{A}_\perp^L\rvert^2 + \lvert\mathcal{A}_\parallel^L\rvert^2 + (L \to R)\Big]$	$\sin^2\theta_K\cos 2\theta_l$
$2c$	$-\beta^2\Big[\lvert\mathcal{A}_0^L\rvert^2 + (L \to R)\Big]$	$\cos^2\theta_K\cos 2\theta_l$
3	$\frac{\beta^2}{2}\Big[\lvert\mathcal{A}_\perp^L\rvert^2 - \lvert\mathcal{A}_\parallel^L\rvert^2 + (L \to R)\Big]$	$\sin^2\theta_K\sin^2\theta_l\cos 2\phi$
4	$\frac{\beta^2}{\sqrt{2}}\Big[\mathcal{R}e(\mathcal{A}_0^L\mathcal{A}_\parallel^{L*}) + (L \to R)\Big]$	$\sin 2\theta_K\sin 2\theta_l\cos\phi$
5	$\sqrt{2}\beta\Big[\mathcal{R}e(\mathcal{A}_0^L\mathcal{A}_\perp^{L*}) - (L \to R)\Big]$	$\sin 2\theta_K\sin\theta_l\cos\phi$
$6s$	$2\beta\Big[\mathcal{R}e(\mathcal{A}_\parallel^L\mathcal{A}_\perp^{L*}) - (L \to R)\Big]$	$\sin 2\theta_K\cos\theta_l$
7	$\sqrt{2}\beta\Big[\mathcal{I}m(\mathcal{A}_0^L\mathcal{A}_\parallel^{L*}) - (L \to R)\Big]$	$\sin 2\theta_K\sin\theta_l\sin\phi$
8	$\frac{\beta^2}{\sqrt{2}}\Big[\mathcal{I}m(\mathcal{A}_0^L\mathcal{A}_\perp^{L*}) + (L \to R)\Big]$	$\sin 2\theta_K\sin 2\theta_l\sin\phi$
9	$\beta^2\Big[\mathcal{I}m(\mathcal{A}_\perp^L\mathcal{A}_\parallel^{L*}) + (L \to R)\Big]$	$\sin^2\theta_K\sin^2\theta_l\sin 2\phi$

Table 2.1: Angular coefficients, I_i, expressed in terms of $B^0 \to K^{*0}$ amplitudes with the relativistic factor $\beta = \sqrt{1-4m_\ell^2/q^2}$, where m_ℓ is the lepton mass [47].

As the flavour of the B^0 meson can be distinguished using the charges of the K^{*0} decay products, it is possible to measure both CP-averaged, S_i, and CP-asymmetric, A_i, quantities[4],

$$\begin{aligned} S_i &= \frac{I_i + \bar{I}_i}{(\mathrm{d}\Gamma + \mathrm{d}\bar{\Gamma})/q^2}\ , \\ A_i &= \frac{I_i - \bar{I}_i}{(\mathrm{d}\Gamma + \mathrm{d}\bar{\Gamma})/q^2}\ . \end{aligned} \tag{2.13}$$

In angular analyses, measurements are often made in intervals of q^2, which effectively takes the averages of the observable values over the q^2 range. For the S_i observables, the differential

[4]This thesis, however, focuses on the CP-averaged observables.

decay rate in terms of the angles can be written as

$$\begin{aligned}\frac{1}{\mathrm{d}(\Gamma+\bar{\Gamma})/\mathrm{d}q^2}\frac{\mathrm{d}^4(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2\mathrm{d}\bar{\Omega}} = \frac{9}{32\pi}\Big[\ &\tfrac{3}{4}(1-F_L)\sin^2\theta_K\ +\ F_L\cos^2\theta_K\ + \\ &\tfrac{1}{4}(1-F_L)\sin^2\theta_K\cos 2\theta_\ell\ -\ F_L\cos^2\theta_K\cos 2\theta_\ell\ + \\ &S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi + S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi\ + \\ &S_5\sin 2\theta_K\sin\theta_\ell\cos\phi + \tfrac{4}{3}A_{FB}\ \sin^2\theta_K\cos\theta_\ell\ + \\ &S_7\sin 2\theta_K\sin\theta_\ell\sin\phi + S_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi\ + \\ &S_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi\Big]\ .\end{aligned} \tag{2.14}$$

where F_L is the fraction of the longitudinal polarisation of the K^{*0}, and A_{FB} is the forward-backward asymmetry of the dilepton system. Note that this expression is obtained from the definitions of the I_i in Equation 2.11 in the limit that the q^2 is large compared to the mass of the lepton, that is, $q^2 \gg 4m_\ell^2$, such that its mass can be neglected. In this case, the following relationships hold between the CP averages of I_1^c, I_1^2, I_2^c and I_2^s, which correspond to S_1^c, S_1^s, S_2^c, S_2^s, and F_L [44]:

$$F_L = S_1^c = -S_2^c\ ,\quad 1-F_L = \tfrac{4}{3}S_1^s = 4S_2^s\ . \tag{2.15}$$

Additionally, A_{FB} is related to S_6 via

$$A_{FB} = \tfrac{3}{4}S_6\ . \tag{2.16}$$

In addition to the S-basis observables, it is also possible to define a set of optimised observables based on them where form factor uncertainties cancel at leading order, namely [48]

$$\begin{aligned}P_1 &= \frac{2S_3}{1-F_L}\ , \\ P_2 &= \frac{2A_{FB}}{3(1-F_L)}\ , \\ P_3 &= \frac{-S_9}{1-F_L}\ , \\ P'_{4,5,6,8} &= \frac{S_{4,5,7,8}}{\sqrt{F_L(1-F_L)}}\ .\end{aligned} \tag{2.17}$$

These are referred to as the P-basis observables. Note that while uncertainties related to form factors are suppressed, they are still affected by non-local contributions.

The presence of the S-wave adds six additional parameters. In this case, the modified differ-

ential decay rate can be written as

$$\begin{aligned}\frac{1}{\mathrm{d}(\Gamma+\bar{\Gamma})/\mathrm{d}q^2}\frac{\mathrm{d}^4(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2\mathrm{d}\bar{\Omega}}\Big|_{S+P} = (1-F_S)\frac{1}{\mathrm{d}(\Gamma+\bar{\Gamma})/\mathrm{d}q^2}\frac{\mathrm{d}^4(\Gamma+\bar{\Gamma})}{\mathrm{d}q^2\mathrm{d}\bar{\Omega}}\Big|_{P} + \\ \tfrac{3}{16\pi}\big[\, F_S \sin^2\theta_l + \\ S_{S1}\sin^2\theta_l\cos\theta_K + \\ S_{S2}\sin 2\theta_l\sin\theta_K\cos\phi + \\ S_{S3}\sin\theta_l\sin\theta_K\cos\phi + \\ S_{S4}\sin\theta_l\sin\theta_K\sin\phi + \\ S_{S5}\sin 2\theta_l\sin\theta_K\sin\phi\,\big]\,,\end{aligned} \tag{2.18}$$

where F_S is the S-wave fraction, and the S_{Si} are related to the interference between S- and P-wave amplitudes. This contribution is limited in data. For the muon mode in the central q^2 region, F_S is found to be around 10% in the $K\pi$ invariant mass region of $796 < m_{K\pi} < 996\,\mathrm{MeV/c^2}$ [49]. Therefore, sensitivity to these terms is only possible at relatively high statistics (Section 9.1).

2.6 Experimental status

Starting with the angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay in 2013 [50], a number of $b \to s\ell^+\ell^-$ observables have been found to deviate from SM predictions. They include differential branching fractions and angular observables of multiple muonic modes, and (until recently) ratios of branching fractions R_K and $R_{K^{*0}}$ that allow for precision tests of lepton flavour universality. While none of these results, often referred to as the 'B anomalies'[5], are statistically significant enough for a claim of discovery[6], they seemed to hint at a shared underlying pattern of deviations from the SM featuring lepton flavour universality violating NP contributions to muons. However, in late 2022, the updated simultaneous measurement of the two aforementioned branching fraction ratios made using improved experimental strategies revealed good agreement with SM predictions. These results are expected to modify the overall picture, although their full implications have yet to be explored. The observables and measurements involved, as well as details pertaining to the updated results, are discussed in the following sections.

2.6.1 Differential branching fractions

Measurements of the differential branching fractions in bins of q^2 ($d\mathcal{B}/dq^2$) have been performed by the LHCb collaboration for the decays of $B^0 \to K^{*0}\mu^+\mu^-$ [49], $B_s^0 \to \phi\mu^+\mu^-$ [52], $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^0\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ [53]. In all cases, the decay rates are found to be lower than SM predictions with varying degrees of statistical precision. Figure 2.5 shows the

[5]Measurements of decays featuring the charged current $b \to c\ell\nu$ transition, in particular that of the lepton flavour universality testing ratios involving taus (third generation leptons), also reveal tensions with the SM at the level of $\sim 3\sigma$ [51], and may be considered as part of the flavour anomalies in literature. These tensions persist at the time of writing.

[6]The threshold that is often used is five standard deviations (5σ).

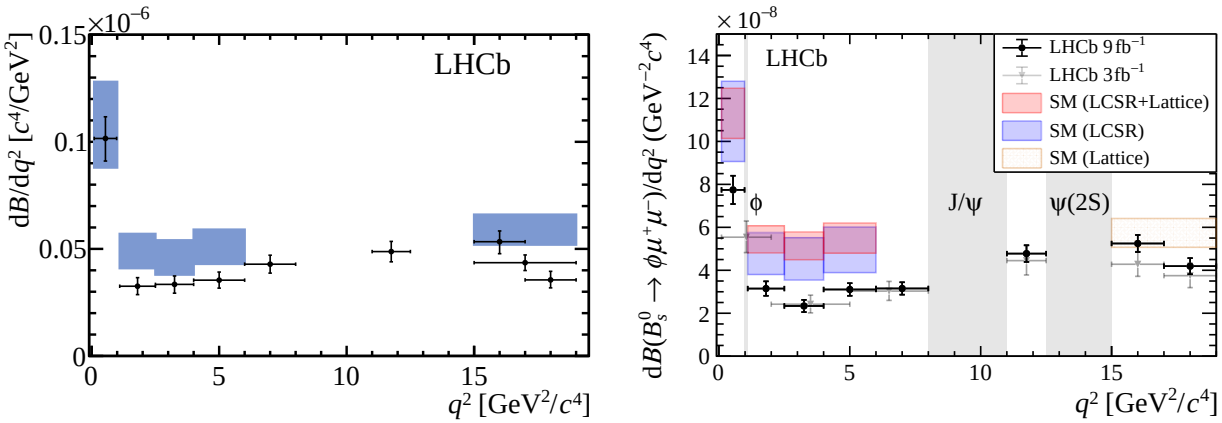


Figure 2.5: Differential branching fractions of the $B^0 \to K^{*0}\mu^+\mu^-$ (left) and $B_s^0 \to \phi\mu^+\mu^-$ (right) decays. Both measurements are carried out in bins of q^2, avoiding regions dominated by decays featuring charmonium resonances. The SM predictions and their associated uncertainties are also shown.

two most recent results where the precision of the measurements is starting to exceed that of the theoretical predictions. Results obtained by the BaBar, Belle and CMS collaborations [54, 55, 56] also hint at the same trend, albeit with larger uncertainties. In general, the interpretation of these deviations is complicated not only by the limited precision of the SM hadronic form factors but also by the treatment of non-local contributions.

2.6.2 Angular analyses

Angular analyses of the $B^0 \to K^{*0}\mu^+\mu^-$, $B_s^0 \to \phi\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ modes have been carried out by the LHCb collaboration [57, 58, 59]. The ATLAS and CMS collaborations also analysed the $B^0 \to K^{*0}\mu^+\mu^-$ decay [60, 61], and the Belle collaboration made a combined measurement of both $B^0 \to K^{*0}\mu^+\mu^-$ and $B^+ \to K^{*+}\mu^+\mu^-$ modes [62]. One observable of particular interest is P_5'. In the most recent LHCb analysis of $B^0 \to K^{*0}\mu^+\mu^-$, it shows tensions with SM predictions in the q^2 bins of $4.0 < q^2 < 6.0\,\text{GeV}^2/\text{c}^4$ and $6.0 < q^2 < 8.0\,\text{GeV}^2/\text{c}^4$ of $2.5\,\sigma$ and $2.9\,\sigma$, respectively. This further solidifies the discrepancies seen in the earlier analyses [63, 50] and is generally in good agreement with the results found by other collaborations (one exception is CMS), which nevertheless have much larger statistical uncertainties. Interestingly, the recent analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ mode reveals a similar trend in P_5', but again the statistical uncertainties are large. The $B_s^0 \to \phi\mu^+\mu^-$ angular observables are found to show good agreement with the SM. However in this case the flavour symmetric $\phi \to K^+K^-$ decay does not allow the identity of the B_s^0 ($\bar{B}_s^0$) to be determined, therefore some observables including P_5' cannot be accessed. A summary of the measured P_5' values is shown in Figure 2.6. Note that the tension seen in this observable is in some sense the 'tip of an iceberg'. Global fits made using all angular observables point to an overall tensions at the level of 2-3 σ [57, 58, 59], which is not driven solely by P_5'. Further discussions on this type of analysis can be found in Section 2.6.4.

2.6.3 Lepton flavour universality tests

In the SM, the couplings of gauge bosons to the leptons are universal. The only difference between the three flavours originate from their Yukawa couplings, which leads to their differ-

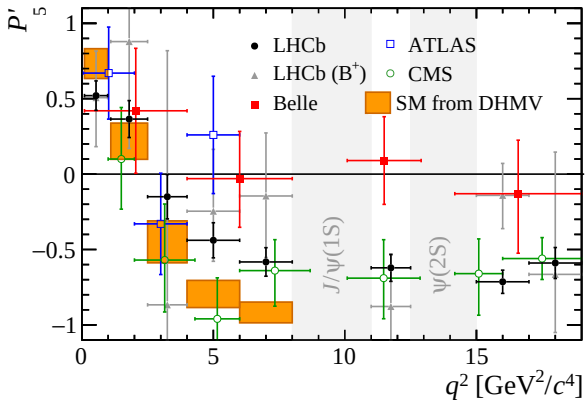


Figure 2.6: Measurements of the observable P'_5 in bins of q^2 including the most recent result obtained by the LHCb collaboration for the decay of $B^0 \to K^{*0}\mu^+\mu^-$ [57] (black) and $B^+ \to K^{*+}\mu^+\mu^-$ [59] (gray), the measurement of $B^0 \to K^{*0}\mu^+\mu^-$ by the ATLAS [60] (blue) and CMS [61] (green) collaborations, and the combined measurement of both $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*+}\mu^+\mu^-$ modes by the Belle collaboration [62] (red). The SM predictions are based on Refs. [41, 40].

ent masses. Thus the SM is said to obey 'lepton flavour universality' (LFU). This feature is frequently violated in NP models with non-trivial flavour structures. One of the most precise ways to test LFU is through the measurement of ratios of decay rates of the form

$$R_X = \frac{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(B\to X\mu^+\mu^-)}{dq^2} dq^2}{\int_{q^2_{\min}}^{q^2_{\max}} \frac{d\Gamma(B\to Xe^+e^-)}{dq^2} dq^2} \,, \tag{2.19}$$

where X is a hadron such as the $K^\pm$ or the K^{*0} meson, which allows for the cancellation of both form factor and non-local hadronic effects. Consequently the theoretical prediction for the value of $R_{K^{(*)}}$ is very precise, with uncertainties at the level of $\mathcal{O}(1\%)$ in the q^2 region of $1 < q^2 < 6\,\mathrm{GeV}^2/\mathrm{c}^4$ [64].

Several tests of LFU have been carried out by the LHCb collaboration, they include most prominantly R_K and $R_{K^{*0}}$, which were found to be [65, 66]

$$R_K = 0.846^{+0.042}_{-0.039}\,(\text{stat.})\,^{+0.013}_{-0.012}\,(\text{syst.}) \,, \tag{2.20}$$

and

$$R_{K^{*0}} = 0.69^{+0.11}_{-0.07}\,(\text{stat.}) \pm 0.05\,(\text{syst.}) \,, \tag{2.21}$$

respectively, in the q^2 region of $1.1 < q^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$. These results showed tensions with the SM at around $3.1\,\sigma$ for R_K, and $2.4\,\sigma$ for $R_{K^{*0}}$, which provided strong hints of LFU violating NP effects in muons. However, recently they have been superseded by an updated analysis, which involves the simultaneous measurement of both R_K and $R_{K^{*0}}$ using all available LHCb data (around $9\,\mathrm{fb}^{-1}$ of integrated luminosity) [67]. The new values are

$$R_K = 0.949^{+0.042}_{-0.041}\,(\text{stat.})^{+0.022}_{-0.022}\,(\text{syst.}) \,, \tag{2.22}$$

and

$$R_{K^{*0}} = 1.027^{+0.072}_{-0.068}\,(\text{stat.})^{+0.027}_{-0.026}\,(\text{syst.}) \,, \tag{2.23}$$

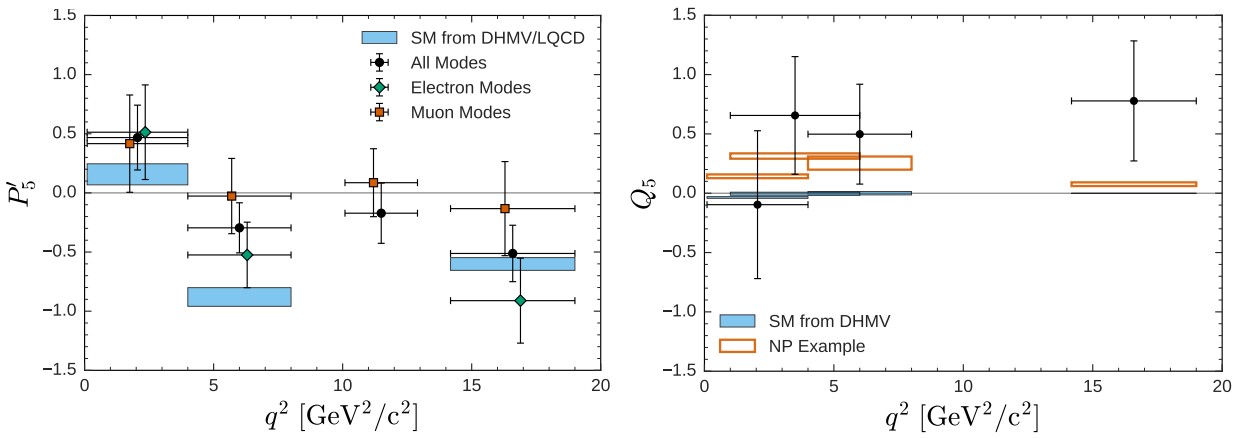


Figure 2.7: Measurements of the observable of $P_5^{\prime\ell}$ [62] (left) and Q_5 [62] (right) for the decays of $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$, $\ell = e,\ \mu$ by the Belle collaboration.

in the central q^2 region, which show, in contrast, excellent agreement with SM predictions. For R_K, the data sample used is the same as that of the previous analysis (but subject to a revised analysis strategy), and the difference is mainly systematic in nature. In particular, it is due to the use of more stringent electron identification requirements that increased signal purity as well as improvements in the background modelling strategy, both of which allow for better determination of the number of $B^+ \to K^+e^+e^-$ decays present. The same systematic effect applies to $R_{K^{*0}}$, although in this case a sizeable statistical component is also present as the previous analysis was based on a significantly smaller data sample. More detailed discussions on the implications of this systematic effect for the current analysis can be found in Section 9.2.6. Note that other collaborations, namely BaBar [54] and Belle [68, 69], also measured $R_{K^{(*0)}}$, and the results are found to agree with the SM, albeit with large uncertainties.

In addition to branching fraction ratios, the differences between the angular observables of the muon and electron modes can also be used to test LFU. One such example are the Q_i [70], which include $Q_{F_L} = F_L^\mu - F_L^e$ and $Q_i = P_i^{(\prime)\mu} - P_i^{(\prime)e}$. These are expected to have reduced sensitivity to both form factor and non-local contributions. The values of Q_4 and Q_5 have been measured by the Belle collaboration [62] in the combined analysis of the $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ modes, where $\ell = e,\ \mu$. The results generally show good agreement with the SM, although the precision of the measurement is limited.

2.6.4 Global fits

The full range of $b \to s\ell^+\ell^-$ observables can be analysed simultaneously via 'global fits' using the EFT framework introduced in Section 2.2 in order to determine whether or not they share common underlying feature(s). This can be done by parametrising the observables in terms of WCs, and performing maximum likelihood fits in which the NP contributions (the $\mathcal{C}_i^{\rm NP}$) to a selected number of them are varied [42, 71]. This work is carried out by a number of groups and updated on a regular basis, e.g. [72, 73, 71, 74, 75]. However, at the time of writing, no published results incorporating the most recent $R_{K^{(*0)}}$ measurement are available, therefore the following discussions will outline the state of the field prior to the update.

The results of global fits are often reported in two ways, both of which factor in the number of free parameters (WC(s) varied) and the degrees of freedom (observables considered). The first is the p-value, which assess whether or not the given hypothesis (SM or NP) can be

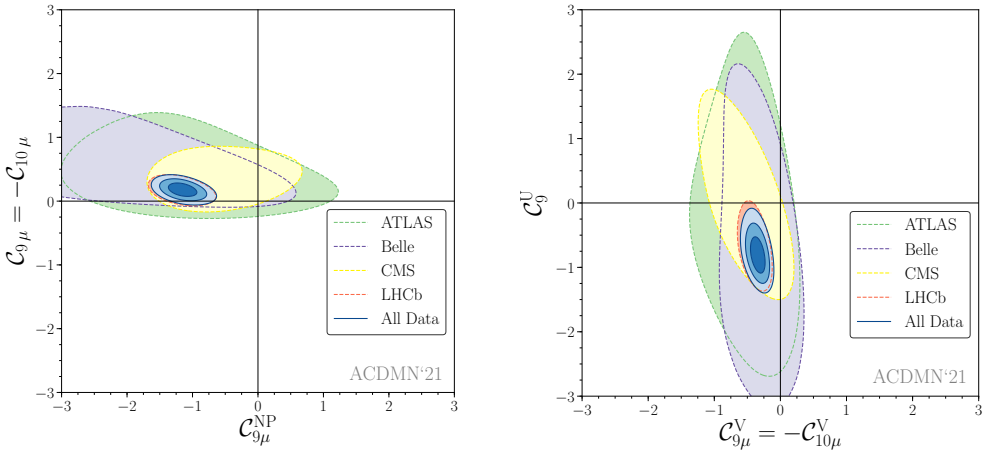


Figure 2.8: Global fits to $b \to s\ell^+\ell^-$ measurements corresponding to the two leading 2D hypotheses of $(\mathcal{C}^{\rm NP}_{9\mu}, \mathcal{C}'^{\rm NP}_{9\mu} = -\mathcal{C}'^{\rm NP}_{10\mu})$ [71] (left) and $(\mathcal{C}^U_9, \mathcal{C}^V_9 = -\mathrm{C}^V_{10})$ [71] (right). All other WCs are fixed to their SM values. The solid lines correspond to 1, 2 and 3 σ regions of the combined fit. The dashed lines correspond to the 3 σ regions found when subsets of the measurements are used.

rejected. Before the $R_{K^{(*0)}}$ update, the p-value of the SM fit is around 1.1%, which corresponds to a tension at the level of 2.5 σ [42]. The second is a quantity referred to as $\mathrm{Pull}_{\rm SM}$, which quantifies in terms of Gaussian σs the extent to which the SM hypothesis is disfavoured relative to a NP hypothesis. The $\mathrm{Pull}_{\rm SM}$ is useful in comparing alternative NP scenarios.

In the most basic case, NP contribution to a single WC is varied. These 1D fits can be helpful in determining which type of contribution, if any, is dominant in the explanation of the anomalies. Several groups [72, 73, 71, 74, 75] have performed fits of this type using almost all available measurements that involve over a hundred observables from the LHCb, ATLAS, CMS and Belle collaborations. Despite differences in methodology, in particular the choice of the form factors and the types and numbers of observables included, two LFU violating scenarios are universally favoured. One of which involves a shift in the muon WCs $\mathcal{C}^{\rm NP}_{9\mu}$ of ~ -1, and the other shifts in both $\mathcal{C}^{\rm NP}_{9\mu}$ and $\mathcal{C}^{\rm NP}_{10\mu}$ simultaneously such that $\mathcal{C}^{\rm NP}_{9\mu} = -\mathcal{C}^{\rm NP}_{10\mu} =\sim -0.4$ The $\mathrm{Pull}_{\rm SM}$ values of these two scenarios have increased over the years as more observables are added and the precision of measurements as well as theoretical predictions increased. For the $\mathcal{C}^{\rm NP}_{9\mu}$ only scenario, the change is from $\sim 5\,\sigma$ in 2016 to $\sim 7\,\sigma$ in 2021, and for $\mathcal{C}^{\rm NP}_{9\mu} = -\mathcal{C}^{\rm NP}_{10\mu}$, the increase is from $\sim 4\,\sigma$ in 2016 to $\sim 6\,\sigma$ in 2021.[7]

In addition to 1D fits, 2D fits where an additional WC (or a constrained combination) is varied, are also performed. In line with the indications provided by the former, almost all favoured scenarios involve variation in $\mathcal{C}_{9\mu}$ and/or $\mathcal{C}_{10\mu}$, including combinations with right handed currents, i.e. $\mathcal{C}'_{9\mu}$ and $\mathcal{C}'_{9\mu}$. One example is $\mathcal{C}^{\rm NP}_{9\mu}$ and $\mathcal{C}'^{\rm NP}_{9\mu} = -\mathcal{C}'^{\rm NP}_{10\mu}$, which produces a large $\mathrm{Pull}_{\rm SM}$ of $\sim 7\,\sigma$ (Figure 2.8 (left)). Note that all the aforementioned scenarios consider NP contribution to muons only. Some groups, e.g. [73, 74], also consider the possibility of a LFU NP contribution. In this case, the muon and electron WCs are rewritten as $\mathcal{C}_{i\mu} = \mathcal{C}^V_{i\mu} + \mathcal{C}^U_i$ and $\mathcal{C}_{ie} = \mathcal{C}^U_i$, respectively, where $U\,(V)$ denotes lepton flavour universal (violating) NP. The combination of $\mathcal{C}^U_9$ and $\mathcal{C}^V_9 = -\mathrm{C}^V_{10}$ also results in a sizeable $\mathrm{Pull}_{\rm SM}$ of $\sim 7\,\sigma$ (Figure 2.8 (right)).

[7] Note that while these values exceed 5 σ, they should be regarded as strong indications for possible NP rather than a discovery of NP. Experimental challenges aside, the underlying theory assumptions and the set-up of the fits (only selected WCs are allowed to vary) complicate the latter interpretation.

Leading explanations of possible LFU violation in $b \to s\ell^+\ell^-$ decays often involve models featuring new vector bosons (Z') and/or leptoquarks, which couple to leptons and quarks. Both are able to mediate the FCNC decay at tree level (Figure 2.2). The Z' boson can also explain LFU NP. However, it should be noted that unaccounted for SM hadronic contributions can mimic the presence of a universal NP contribution.

The new $R_{K^{(*0)}}$ results, which show good agreement with SM and therefore LFU, will likely affect the global picture by reducing preference for sizeable LFU violating NP contributions to muons. However, the continued presence of anomalies in the muon mode still require explanation. In this context, the exploration of hitherto underutilised observables, in addition to increasing the precision of existing measurements, may provide useful inputs. In particular, $\mathcal{C}_9$ dominated ones can help to disentangle between LFU NP and unaccounted for SM QCD effects [76]. An important observable of this type is Q_5. Therefore, the angular analysis of the electron mode can contribute to the further clarification of the B anomalies.

Part II

The LHCb detector at the LHC

3 The LHC

The Large Hadron Collider (LHC) [77, 78] at the European Center for Nuclear Research (CERN) is a circular particle accelerator and collider with a circumference of around 27 km situated about 100 meters below ground. Protons, circulating in two adjacent rings inside vacuum pipes in opposite directions are designed to collide at a large centre-of-mass (CoM) energy of 14 TeV. The trajectories of the particles are bent by 1232 main superconducting dipole magnets capable of generating a strong magnetic field of 8.3 T, and the beam is focussed and stabilised by hundreds of additional multipole magnets.

Protons are supplied to the main LHC ring through a number of smaller accelerators. They are initially separated from hydrogen molecules using a duoplasmatron before being delivered to Linac2, a linear accelerator, where they are accelerated to an energy of 50 MeV. Next, they are sent into a series of synchrotrons, starting with the Proton Synchrotron Booster (PSB), where they are accelerated up to 1.4 GeV, followed by the Proton Synchrotron (PS), where they reach energies of 25 GeV, and finally the Super Proton Synchrotron (SPS), which accelerates them up to 450 GeV before injection into the LHC for further acceleration. In the process of increasing the energies of the proton beams, the radio frequency (RF) cavities of these synchrotrons split it up into multiple packets of protons, or bunches. Their frequencies are tuned to enable the final output of the SPS to fulfil the required bunch structure of the LHC. A schematic diagram of the accelerator complex is shown on Figure 3.1. Note that the LHC also collides ions. The acceleration of ions starts with a different linear accelerator, Linac3, followed by the Low Energy Ion Ring (LEIR), before they enter the PS and follow the same path as the protons.

The layout of the LHC largely follows the geometry of the tunnels excavated for an older experiment, the Large Electron-Positron collider (LEP), and is illustrated on Figure 3.2. The LHC ring is comprised of eight arcs joined by eight straight sections, which host experimental or utility equipment. The two beams are made to collide at four locations around the ring, namely at Points 1, 2, 5 and 8, where detectors of the main experiments are stationed. Two general purpose detectors, ATLAS [79] and CMS [80], occupy points 1 and 5, respectively. The main aims of these experiments include the measurement of the properties of the Higgs boson after its discovery in 2012, and the direct search for NP. Point 2 is occupied by ALICE [81], a specialised detector constructed to investigate QCD effects at the scale of Λ_{QCD}, by studying the collisions of heavy nuclei at high energies. The LHCb detector [82], located at Point 8, is a forward spectrometer designed to carry out precision measurements of b and c hadron decays. Point 4 contains RF cavities. Operated at 400 MHz, they are designed to capture and accelerate the proton beams injected into the LHC ring close to Points 2 and 8. Beam collimation systems occupy Points 3 and 7. These systems are vital for the stable operation of the machine. Given that each proton beam stores around 300 MJ, the deposition of even a small fraction of this energy would be sufficient to disrupt superconducting magnets and damage accelerator components. Multiple types of collimators are installed around the beam pipe to scatter particles that deviate from the core of the beam away to be safely absorbed. Finally, Point 6 houses the beam abort system. When the beams need to be removed, e.g. at the end of a circulation period, a series of magnets are turned on to redirect them along two ~ 700 m long extractions lines, which lead to graphite and steel blocks that are designed to

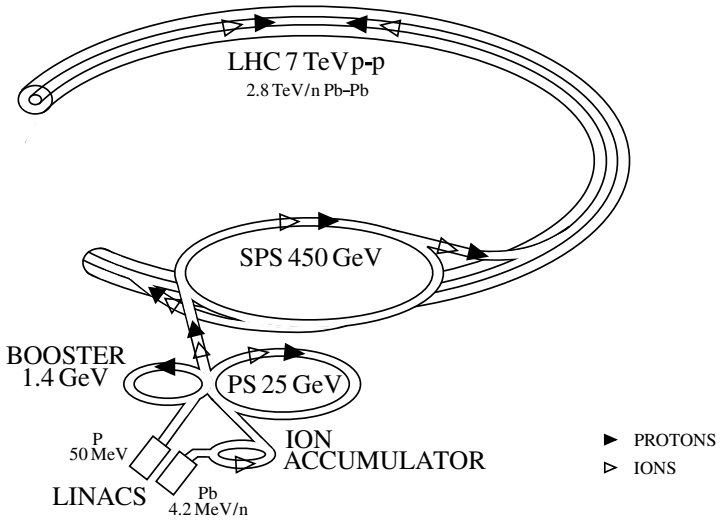


Figure 3.1: Schematic diagram of the accelerator complex that provides the LHC with protons [78].

absorb their energies.

Two important quantities related to the performance of the LHC are the beam energy, and luminosity. The former is related to the energy available for particle production and determines the types of particles that can be produced. The latter is a measure of the number of collisions, and is therefore related to the statistics available for analyses. The rate at which a given process occurs (e.g. the production of $b\bar{b}$ pairs) can be expressed as

$$\frac{dN}{dt} = L\sigma \ , \tag{3.1}$$

where L is the instantaneous luminosity of the collider, and σ is its production cross-section. The total number of events produced can be found via

$$N_{\text{total}} = \sigma \int L dt \ , \tag{3.2}$$

where $\int Ldt$ is referred to as the integrated luminosity.

Assuming a Gaussian beam profile, the luminosity can be related to accelerator and beam properties through the following expression

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon_n \beta^*} F \ , \tag{3.3}$$

where N_b and n_b are the number of particles per bunch and the number of bunches per beam, respectively, f_{rev} is the revolution frequency, γ_r the relativistic gamma factor, F is a factor that reduces the luminosity based on the crossing angle of the beams at the interaction point, ϵ_n is the normalised transverse beam emittance (a quantity related to the width of the beam), and β^* is the beta function (a measure of the beam focus) at the collision point.

The LHC is designed to reach an instantaneous luminosity of $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ and a beam energy of 7 TeV. After the first test runs in 2008 to 2009, operation began in earnest in 2010 [83]. From 2010 to 2013, also referred to as Run 1 of the LHC, beam energy values of 3.5 TeV (2010-2011) and 4 TeV (2012-2013) were reached, and the peak luminosity of around $10^{33}\,\text{cm}^{-2}\text{s}^{-1}$

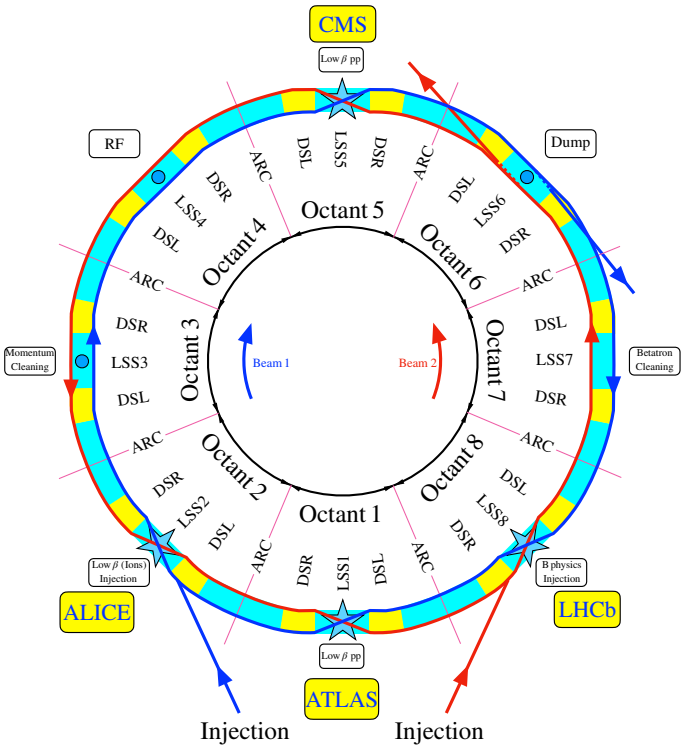


Figure 3.2: Schematic diagram showing the layout of the LHC [77]. The two counter rotating beams are shown in red and blue. The four major experiments, and the key utility and experimental components are labelled.

was achieved. After a shut down of two years for technical maintenance and consolidation, in 2015-2018 (Run 2) [84], the beam energy was increased up to 6.5 TeV, and the peak luminosity exceeded $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

The LHC is an excellent source of b hadrons [85]. Operating at a CoM energy of 14 TeV, the $b\bar{b}$ production cross section is expected to be $\sim 500\,\mu$b, leading to the production of around 10^{12} $b\bar{b}$ pairs per year. Furthermore, a large variety of b hadrons can be produced from proton-proton (pp) collisions, ranging from B mesons to Λ_b baryons. These factors allow for the study of a wide range of b hadron decays, including the rare FCNC decays that can reveal NP. However, pp collisions are complex events that can lead to the production of copious particles through quark hadronisation. The use of a detector capable of operating in this harsh hadronic environment is essential for the realisation of these physics goals.

4 The LHCb detector

The LHCb detector [82, 86] is optimised for the precision measurements of CP violation and the rare decays of b and c hadrons. The $\sigma_{b\bar{b}}$ production cross section, though large at the LHC, is small compared to the total hadronic cross-section. To enable the efficient reconstruction of events of interest in a background rich environment, the detector is built with a robust and efficient tracking system, a flexible and adaptable trigger system, and relatively thin and light components (in terms of radiation and interaction lengths) to reduce unwanted secondary interactions that negatively affect event reconstruction. Furthermore, as the $b\bar{b}$ pairs produced are strongly boosted in the forward (and backward) direction (Figure 4.1), it is constructed as a forward spectrometer – a compact detector that concentrates high precision instruments within a limited solid angle.

To make precision measurements, events of interest arising out of pp collisions need to be retained efficiently, and to be well reconstructed. In order to achieve optimal performance, the (instantaneous) luminosity at the LHCb interaction point needs to be controlled, and often kept at a level that is lower than the maximum achievable value at the LHC. While higher luminosities imply more statistics, operating under such conditions leads to increased occupancies in the detector subsystems, which degrades the quality of event reconstruction. A related quantity that is particularly important for the trigger system is the 'pileup', which refers to the number of visible[1] pp collisions per bunch crossing [89]. Large pileup adds pressure to the system, leading to long processing times. To keep both quantities under control and at values suitable for the LHCb detector, a real-time luminosity levelling system [86, 89] is introduced in 2011, where control is achieved by adjusting the transversal overlap of the proton beams[2]. This procedure allows the luminosity to be kept stable to within $\sim 5\%$, as illustrated in Figure 4.2. Throughout the years of 2011 to 2018, the detector operated at a luminosity of $\sim 4 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, which exceeded its design value of $2 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The integrated luminosity collected over the years is shown in Figure 4.3.

The LHCb detector is installed at Point 8 of the LHC. It is approximately 10 m high, 13 m wide and 21 m long. The view of the detector from the side is shown schematically in Figure 4.4. LHCb uses a right-handed coordinate system, where the z axis points along the beamline from the interaction point towards the end of the detector, and the y axis points vertically upwards. The trajectories of the particles are bent by a magnet in the xz plane ('bending plane'). The detector covers an angular acceptance of around 10 mrad to 300 mrad in the bending plane, and around 10 mrad to 250 mrad in the vertical plane. Equivalently in units of pseudorapidity[3], it covers the range of $2 < \eta < 5$. Multiple components of the detector that form the tracking and particle identification systems provide information on the traversing particles to allow their momenta, and identities to be determined. This information is also used by the trigger system to filter out the collision events that likely produced processes of interest for further analysis.

[1] A visible interaction is defined as one where at least two tracks are reconstructed in the VELO, and both point towards the collision region [88].

[2] The number of pileup events, μ, is related to the luminosity via $\mu = (L\sigma_{pp})/(N_b f_{\mathrm{rev}})$, where σ_{pp} is the cross-section of inelastic pp collisions [86, 89]. Therefore controlling the luminosity is generally equivalent to controlling the pile-up.

[3] Pseudorapidity, η, is defined as $\eta = \ln(\tan\theta/2)$, where θ is the polar angle with respect to the beam axis.

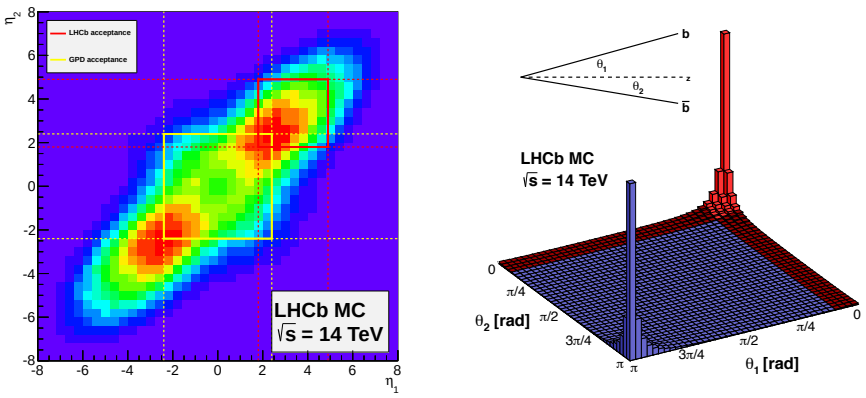


Figure 4.1: Distribution of $b\bar{b}$ pairs produced at 14 TeV CoM energy in terms of the pseudo-rapidities of the b and $\bar{b}$ (left) [87]. The region within the acceptance of the LHCb detector is marked by red lines, and the region covered by the general purpose detectors, ATLAS and CMS, is marked by yellow lines. The right plot shows the distribution of $b\bar{b}$ pairs in terms of the polar angle.

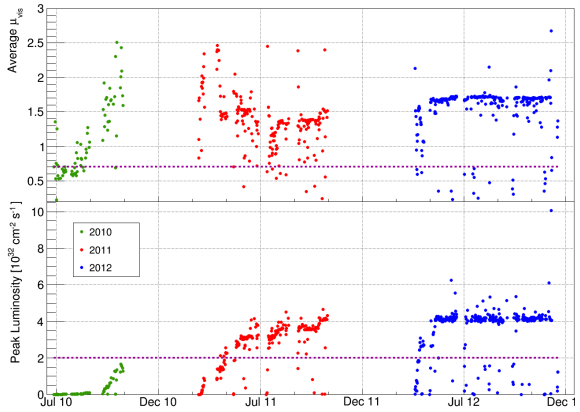


Figure 4.2: Detector pileup [86] (top) and instantaneous luminosity [86] (bottom) in 2010-2012. The design values are shown by the purple line. The detector is found to be able to maintain good performance at higher luminosity and pileup values.

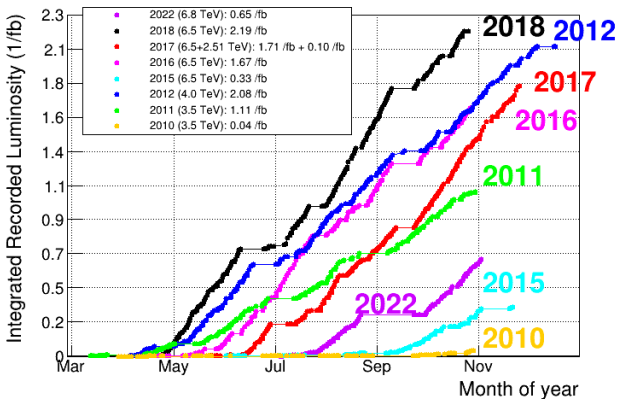


Figure 4.3: Integrated luminosity recorded by the LHCb detector at different CoM energies of the LHC [90].

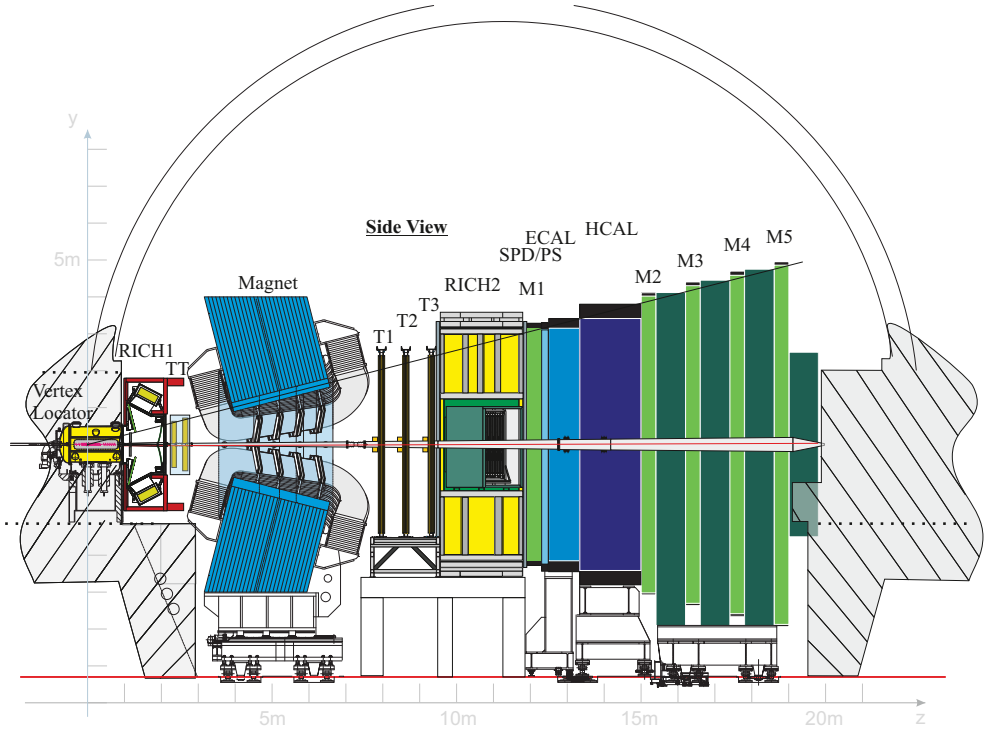


Figure 4.4: Schematic diagram of the LHCb detector viewed from the side (y-z plane) [91, 88].

These systems are discussed in detail in the sections below.

4.1 Tracking system

The tracking system [91, 86] consists of a series of sub-detectors position up and downstream of a warm (non-superconducting) dipole magnet [92] that bends the trajectories of charged particles with an integrated field strength of 4 Tm. The first of these is the vertex locator (VELO)[93], a silicon microstrip detector that surrounds the interaction region, which provides precise measurements for the identification of primary vertices (PVs) from pp collisions as well as secondary vertices, which originate from relatively long-lived particles like the B meson, which may traverse millimeters of detector material before decaying. Good reconstruction of the latter is essential for the separation of signal from background. The VELO is followed by another silicon-based detector, the Tracker Turicensis [94] (TT), and three tracking stations [94] (T1-T3), each of which is comprised of a silicon-based inner region referred to as the inner tracker (IT) [95], and an outer region covered by layers of straw tubes, or outer tracker (OT) [96]. Information on the positions of the particles is used to reconstruct their tracks before and after the magnet, the slopes of which can be used to determine their momenta.

4.1.1 Dipole magnet

The dipole magnet [92, 91, 86] is located around 5 m from the interaction point behind the TT and in front of the tracking stations. The conical saddle-shaped design of its aluminium coils is used to match the required detector acceptance. Two sets of coils are placed mirror symmetrically to each other and are held in place by a rectangular iron yoke. The magnet

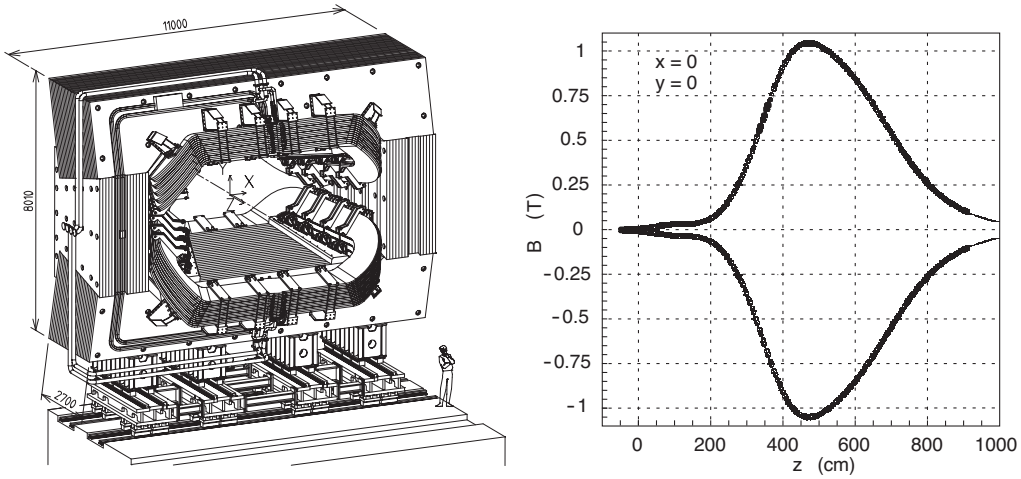


Figure 4.5: Schematic diagram of the dipole magnet [91] (left) showing the central saddle shaped coils surrounded by the rectangular yoke. The variation of its field strength along the z axis (beamline) is shown for the tracking region from $z = 0\,\text{m}$ to $z = 10\,\text{m}$ [91] (right), for both polarity orientations.

produces a magnetic field that is chiefly in the y-direction, which bends the trajectories of charged particles in the $x-z$ plane. Within the 10 m of the tracking region, which extends from the collision point to the end of the tracking stations, it maintains an integrated field strength of around 4 Tm. The polarity of the magnet is periodically reversed during data taking. This is done to reduce systematic uncertainties related to potential differences in performance between components on the left and right sides of the detector for precision CP violation measurements. A schematic diagram of the magnet, and a plot of its field strength variation along z, are shown in Figure 4.5.

4.1.2 Vertex locator

The VELO [93, 91, 86] is a silicon micro-strip detector that surrounds the pp interaction region. It is built with two retractable halves that can be moved inwards to a minimum distance of about 8 mm from the beam, but can also be pulled out to a distance of 3 cm to avoid damage when the beam is unstable (e.g. during injection). Each half is made up of 21 standard modules, which hold in place two approximately semi-circular discs containing two types of strip sensors, the R and ϕ sensors. The R strips are arranged in concentric semi-circles centred around the beampipe and divided into four 45 deg arc segments to minimise occupancy, while the straight ϕ sensors radiate outwards from the beam, and are split into two groups. These sensors provide measurements of particle positions using cylindrical polar coordinates, where the r coordinate describes the radial distance from the beam axis, and ϕ the angle around the axis. In addition, two extra modules with R sensors only are located in the most upstream positions. Information from these 'pile-up' sensors was originally intended to be used in the trigger system to reject events with more than one pp interaction, which are challenging to analyse. However, due to a combination of efficient trigger design and a lower than expected impact from such events, they are no longer used in this way [97]. Instead, they are mainly used for luminosity monitoring and measurements. The two halves of the VELO are made to overlap slightly when fully closed in order to ensure full angular coverage. The VELO operates in a vacuum separated from that of the LHC by a corrugated aluminium-alloy foil in order to protect the LHC from the out-gassing of the VELO modules, and to protect the VELO from RF-pickup from the LHC beam. To

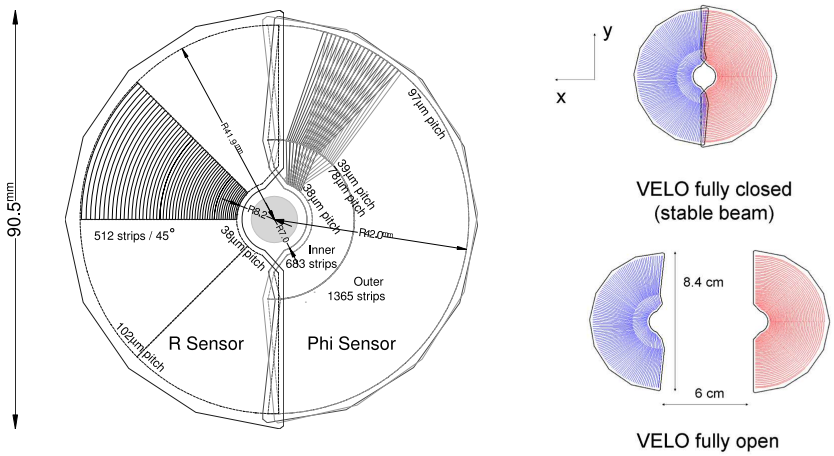


Figure 4.6: Illustration of the $r-\phi$ geometry of the VELO sensor [91] (left). For clarity, only a selected number of strips are shown. The ϕ strips of two adjacent sensors are shown to highlight the stereo angle. The two halves of the VELO can be closed during stable operating conditions, and retracted when the beam is unstable [91] (right).

reduce heat during operation and limit radiation damage, a dedicated cooling system is used to keep the sensors at a temperature of -10°C to 0°C. A schematic diagram of the VELO is shown in Figure 4.6.

The VELO is designed to achieve high vertex and impact parameter (IP) resolution, the latter of which is defined for a given track as its distance from the PV at its point of closest approach to it. Selections based on IP are used frequently to suppress background. The VELO achieves a resolution of about 13 μm for the x and y coordinates, and 71 μm for the z coordinate, for PVs with around 25 tracks. The IP resolution is found to be around 13 μm in both the x and y directions for tracks with asymptotically high momenta transverse to the beam axis.

4.1.3 Tracking stations

The tracking stations include the TT [94, 91, 86], a 150 cm wide and 130 cm high planar structure situated between RICH 1 and the magnet, and three stations, T1-T3 (T stations), positioned downstream of the magnet. Each downstream station is composed of an IT [95] covering a 120 cm wide and 40 cm high cross shaped region surrounded by an OT [96], which extends to the full detector acceptance. The TT and the IT of stations T1-T3 use silicon micro-strip sensors, while the OT is a straw tube detector. In both cases, individual detection planes only provide measurements of the x-coordinate of a traversing particle based on the strip or tube that registered the hit (in addition to the z-coordinate information given by their position along the beamline). To allow the y-coordinate to be measured, groups of four planes are arranged in an 'x-u-ν-x' geometry, where the silicon strips/gas tubes of the middle two planes, u and ν, are rotated at angles of 5° and -5°, respectively with respect to the first and last, which are aligned to the y axis.

Silicon tracker

The TT [94, 91, 86] and the IT [95, 91, 86] are collectively referred to as the Silicon Tracker. The TT is composed of four layers divided into two pairs. They are arranged such that the the first two planes, x and u, are located close to each other, and the last two planes, ν

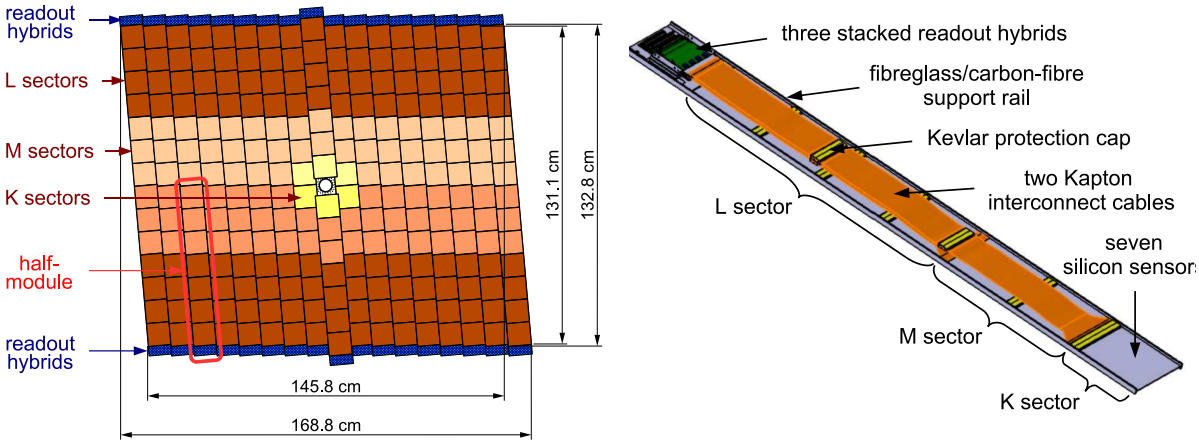


Figure 4.7: Schematic diagram showing the layout of the third TT layer [91] (left), with a (4-3) half module marked in red, and the illustration of a 4-2-1 type half module with seven silicon sensors [91] (right).

and x, are situated around 27 cm further downstream. All layers are housed within a single light tight, thermally and electrically insulated container that is kept at a low temperature to suppress radiation damage. Each layer is constructed in two separable halves such that they can be retracted for beampipe maintenance. These halves are made up of multiple vertical modules, each of which consists of a row of seven silicon sensors, organised into three readout sectors to keep detector occupancy in all regions manageable. Due to the high particle flux near the beampipe, the seven sensors of the half modules closest to it are organised based on a 4-2-1 scheme, while the modules further away use a two-sector 4-3 scheme. The individual silicon sensors are 500 μm thick, 9.64 cm wide, 9.44 cm long and each carry 512 readout strips with a strip pitch of 183 μm, designed to achieve a single hit resolution of about 50 μm. The performance of the TT is found to be in line with expectations [86]. A schematic diagram of a single detection plane is shown in Figure 4.7 along with an illustration of a half module.

The IT is constructed out of four individual boxes arranged around the beam pipe, which are insulated and cooled in the same way as that of the TT. Each box contains four detection layers, and each layer is composed of seven modules, the designs of which vary depending on their location. The modules in the boxes situated to the left and right of the beampipe ($\pm z$) consist of pairs of silicon sensors, which are 410 μm thick, 7.6 cm wide, 11 cm long, and carry 384 readout strips with strip pitches of 198 μm. The boxes positioned above and below the beampipe ($\pm y$) contain modules made up of single sensors, which are thinner (320 μm), but are otherwise of the same dimensions. The thickness of the sensors is chosen to provide sufficiently high signal-to-noise ratio for each module type. The IT was found to achieve a hit resolution of around 50 μm, which is compatible with its design requirement [86]. The layout of this sub-detector is shown in Figure 4.8.

Outer Tracker

The OT [96, 91, 86] is a gaseous straw tube detector. This type of technology is chosen to limit the material budget, and to cover a large area economically, whilst enabling good tracking efficiency and high momentum resolution. In order to deliver good performance in a high multiplicity environment, a fast drift time is important. This is achieved by using a gas mixture composed of Argon (70%) and CO_2 (30%), which allows for a drift time below 50 ns.

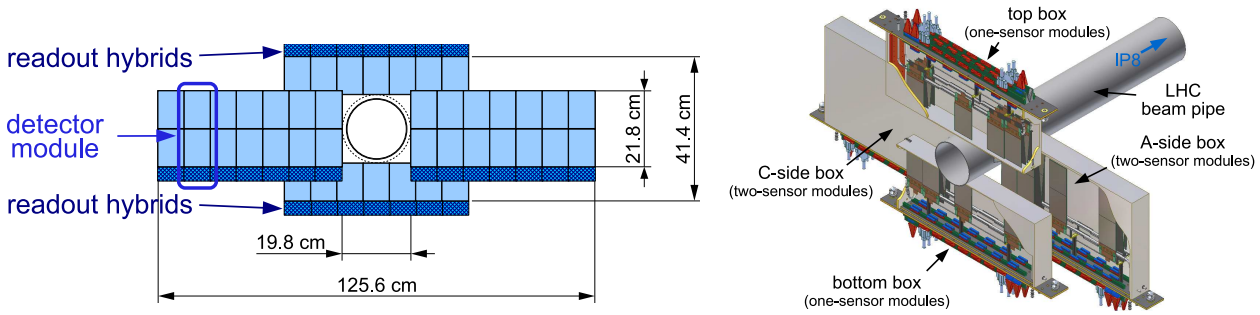


Figure 4.8: Illustrations showing the layout of a single IT layer [91] (left), and the four detector boxes of the IT mounted on the sides of the beam pipe [91] (right).

The three OT stations are split into two retractable halves, each of which is built from two types of modules, namely seven long F-modules with active lengths of 4850 mm containing a total of 256 tubes, and shorter S-modules located above and below the beam pipe, which have half the length of the F-modules, and contains 128 tubes. In both cases, the tubes, with inner diameters of 4.9 mm, are laid down in two staggered layers.

The OT achieves a single hit efficiency of more than 99%, and a spatial resolution of around 200 μm for tracks with momenta greater than 10 GeV/c [98, 99].

4.1.4 Track reconstruction

Track reconstruction at the LHCb relies on information from the VELO and tracking stations in the form of electronic signals, or hits, produced when particles traverse sensors. By identifying a set of hits produced by a particle, a fit can be performed to obtain the best estimate of its trajectory. Several types of tracks can be reconstructed [86, 100, 101], these include:

- **VELO tracks** – short tracks that pass only through the VELO. They can have large angles with respect to the z axis or point in the backward direction. These tracks are useful for primary vertex reconstruction;
- **Upstream tracks** – tracks traversing the VELO and TT before being bent out of the acceptance of the detector. These tracks tend to be associated with low momenta particles (and ones with significant radiative energy losses);
- **T tracks** – tracks based on hits in the T stations only. They are often due to secondary particles produced from material interactions;
- **Downstream tracks** – tracks based on hits in the TT and T stations only. Particles associated with these tracks may have originated from the decays of long lived particles (e.g. K_s^0 and Λ), which may have decay vertices that are significantly displaced from the primary vertex (e.g. outside the VELO region);
- **Long tracks** – tracks that extend through the full tracking system and are reconstructed based on hits in the VELO and the T stations. Information from the TT may also be incorporated. These tracks are favoured for physics analyses, as they allow for the best possible estimations of particle momenta. The analysis described in this thesis uses long tracks exclusively.

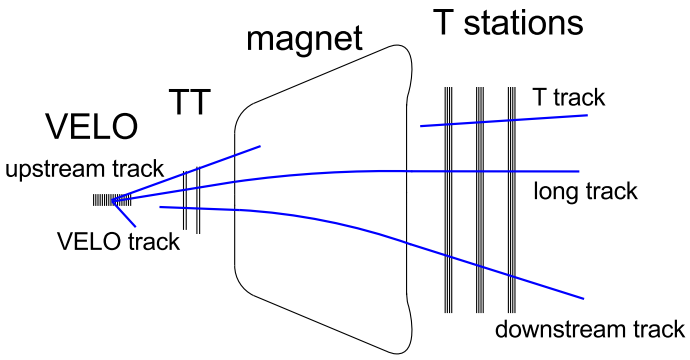


Figure 4.9: Schematic digram showing the types of reconstructible tracks at the LHCb and the sub-detectors involved [102].

An illustration of these types of tracks is given in Figure 4.9.

The reconstruction of a track proceeds via three main steps. First, pattern recognition algorithms are used to search for groups of potentially related hits. Next, track fits are carried out to obtain best estimations of particle trajectories. Finally, the sample of track candidates is cleaned by, for example, removing duplicates reconstructed by the different algorithms.

The reconstruction of long tracks uses the outputs of two standalone algorithms, VELO tracking (FastVeloTracking) [103] and T-track seeding (PatSeeding) [104]. As the effect of the magnetic field is insignificant in the VELO, tracks in this region are expected to be straight lines to a good approximation. Tracks that cross at least four sensors, i.e. leave four hits ('quadruplets'), are better determined than those with fewer hits, and are therefore searched for first. The tracking algorithm starts near the end of the sub-detector by searching for quadruplets in the last four R sensors. This is done as track separation is at its greatest at large z. Search windows are then opened to find hits in the stations located at progressively lower z values, which are added to the existing quadruplets. After all quadruplets have been processed, searches are also made in a similar way for groups of three hits using only previously unused hits. Next, for each of the track candidates found, searches are made for ϕ hits that are consistent with forming a straight line. Finally, a track fit is carried out using χ^2 minimisation to produce VELO tracks candidates. Note that tracks in the backward ($-z$) direction are also searched for in an analogous way after the identification of forward going tracks. While they are of little use to analyses directly due to the lack of instrumentation in this region, they can be useful for PV reconstruction.

Unlike the VELO, the magnetic field present in the region of the T stations causes tracks to be bent. The T-track algorithm first searches for hits in the x layers of the T1 and T3 stations. Next, a search window is opened in the x layers of station T2. If a compatible hit is found, a parabola is built from these three hits in the $x-z$ plane. After this, hits from all other x layers are added to this set if they are sufficiently close to the parabola, and $x-z$ track candidates are constructed if additional quality requirements are passed, and a least-squares fit to the hits converges and produces a sufficiently low χ^2 value. For each $x-z$ candidate, searches are made for compatible hits in the stereo layers using a Hough Transform-based approach. More specifically, after potentially related hits are selected, their y coordinates are projected linearly to a fixed z plane (straight lines passing through the origin and the hits in question are created). The y positions of these projections are then calculated at the z plane. Hits belonging to a common track are expected to cluster around a central y value, whereas unrelated ones will be distributed more randomly. Stereo hits belonging to clusters that pass additional selection criteria are added to the x hits, and track fits are made and used to perform further selections.

Track candidates that remain after this process are sorted based on the number of hits they contain and their fit χ^2. T track candidates are taken progressively from this list starting from the highest quality tracks. Candidates are only retained if they do not share a sizeable fraction of hits with ones that have already been selected.

VELO tracks are used as inputs to the forward tracking algorithm (PatForward) [105], which is one of the two independent algorithms used to reconstruct long tracks. The VELO tracks are first extrapolated linearly into the T stations, and potentially compatible clusters of hits are searched for in the x planes of each station using a Hough transformation, where each hit is projected to a reference plane at a fixed z value ($z = 8520\,\text{mm}$) using the direction of the VELO track and knowledge of the magnetic field. Potential clusters are identified, and the hits therein are fitted using a third order polynomial function. In this process, hits that might not belong to the cluster are identified based on their contributions to the fit χ^2 and are removed. Track candidates are further selected based on quality criteria related to the minimum number of hits or the maximum χ^2/ndf, after which another cluster search and fit process is similarly performed for the hits in the stereo layers. Finally, the best tracks corresponding to each VELO track are selected using a quality variable based on its momentum, fit χ^2/ndf, number of hits, and the compatibility in the y coordinate of the VELO and the T track segments. At this point, VELO tracks are extrapolated into the z location of the TT, and hits are searched for within its layers. The momentum information of the tracks are used to constrain the size of the search window, and a similar Hough Transform approach is used to associated the most consistent clusters with the track candidates. Note that the inclusion of TT hits is optional for long tracks.

In addition to forward tracking, the track matching algorithm (PatMatch) [106, 107] is also used to reconstruct long tracks. In this case, both VELO tracks and T tracks are used as inputs. Good quality candidates of both types are extrapolated to $z = 830\,\text{mm}$ (approximately the end of the T stations). The angles of the VELO tracks and the momenta associated with the T tracks are used to calculate the momenta of the long track candidates in the direction transverse to the beamline. If a minimum threshold is exceeded, a χ^2 value, which quantifies the goodness of the match, is calculated. As multiple combinations using the same track segments are possible, only the ones that result in the lowest χ^2 values are kept. Finally, the VELO tracks of identified candidates are extrapolated into the TT region, and consistent hits, if found, are also added.

Long track candidates produced by the two algorithms are fitted using a Kalman Filter [108, 109] to obtain best estimations of track parameters, taking the detector geometry (and the magnetic field) as well as uncertainties due to sources of noise (e.g. multiple scattering) into consideration. These are further processed by removing artefacts of reconstruction, one type of which originates from duplicated tracks, or 'clones'. A pair of tracks are considered clones if they share at least 70% of the hits in the VELO and part of the T stations. This type of background is reduced by a dedicated 'Clone Killer' [110, 111] algorithm, which takes all tracks in an event into consideration (excluding those used in the reconstruction of other tracks, e.g. some VELO/T tracks), and if clones are found, it identifies the best track with the largest number of hits (and lowest χ^2/ndf if the hits equal) and flags the other as a clone. To further suppress background, namely contributions from fake or 'ghost' tracks, which are tracks that do not correspond to the trajectories of particles, but are instead reconstructed from mismatched hits or detector noise, a Neural Network ('ghost probability') [112] trained using simulation is used. This algorithm assigns a value ranging from zero to one to each track, which can be interpreted as the probability of it being a ghost (`GhostProb`).

The design of the tracking system hardware and the reconstruction algorithms allow for

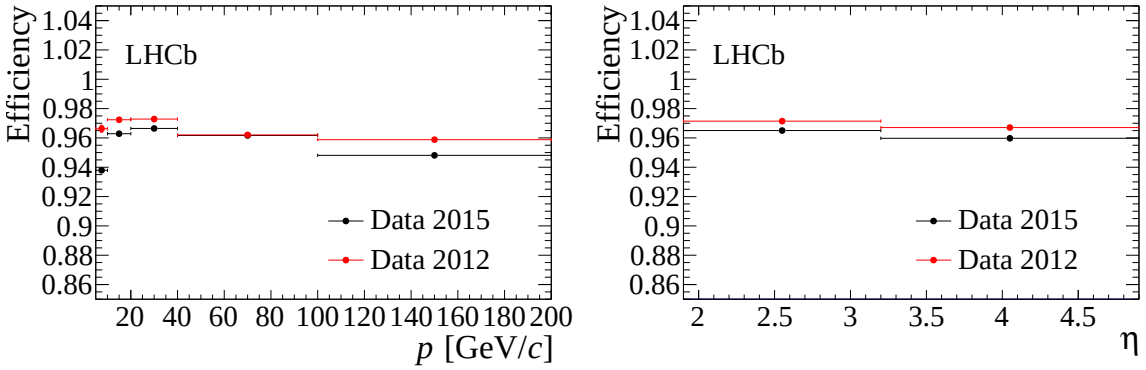


Figure 4.10: Track reconstruction efficiency as a function of the momentum [113] (left) and pseudorapidity [113] (right) determined using Run 1 (2012) and Run 2 (2015) data. The slight reduction in Run 2 is related to spillover effects in the OT, which has a readout window longer than 25 ns, when the bunch spacing is reduced from 50 ns to 25 ns.

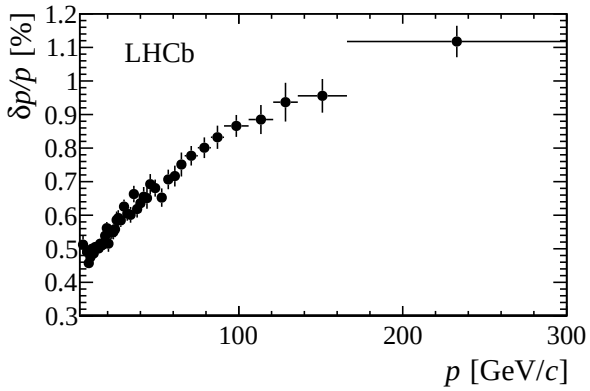


Figure 4.11: Relative momentum resolution as a function of track momentum for long tracks in data determined using J/ψ decays [86].

excellent track finding efficiency and momentum resolution. The tracking efficiency can be estimated in data by applying a tag-and-probe approach to $J/\psi \to \mu^+\mu^-$ decays. One of the final state muons (the tag) is fully reconstructed, while the other one (the probe) is partially reconstructed using only information from selected sub-systems, which are chosen to be as unrelated as possible to the tracking sub-systems for which the efficiency is to be determined. The tracking efficiency is then found by calculating the fraction of fully reconstructed long tracks that can be matched to probe tracks. Figure 4.10 shows the track reconstruction efficiency in Run 1 (2012) and Run 2 (2015) as functions of momentum (p) and pseudorapidity (η). In both cases, the efficiency is high ($\sim 96\%$) within the range of $20\,\text{GeV/c} < p < 200\,\text{GeV/c}$ and $2 < \eta < 5$, which provides good coverage of the phase space of the detector. The momentum resolution for long tracks determined using $J/\psi \to \mu^+\mu^-$ candidates is found to range from 0.5% for tracks with $p \sim 20\,\text{GeV/c}$, to 1.0% for $p \sim 200\,\text{GeV/c}$, as shown in Figure 4.11.

4.1.5 Primary vertex reconstruction

The algorithm used to reconstruct primary vertices (PatPV3D) [114, 100] makes use of VELO tracks that have been passed through the Kalman Filter, and proceeds in two steps. In the first,

PV candidates are found, and in the second they are fitted. The search for candidates involves finding locations where sufficient numbers of tracks pass close to each other. This is done through an iterative procedure that runs through all VELO tracks. For each track (base track), additional close tracks are found based on the distance of closest approach (DOCA) – tracks are considered close if the DOCA is less than 1 mm. A base track is skipped if the number of close tracks is less than four. The x, y and z coordinates of each group of close tracks are calculated via a weighted average, and are added to the collection of PV candidates. Tracks used in the construction of a PV candidate are marked as used, and the process continues until all tracks have been used. The list of candidates are sorted according to their track multiplicity, and the ones with the highest values are reconstructed first. This reduces the incorrect reconstruction of secondary vertices (e.g. from $b\bar{b}$ event), which typically have lower multiplicity, as PVs. A fit is performed for every candidate using an adaptive weighted least squares method. An important point of this approach is the assignment of weights to tracks based on the values of their impact parameter χ^2 ($\chi^2_{\text{IP}}\,(\text{PV})$), which is a measure of the significance of the minimum distance between the trajectory of the track and the PV. Tracks consistent with originating directly from a PV will generally have lower $\chi^2_{\text{IP}}\,(\text{PV})$ values than ones that originate from secondary vertices. The fit makes use of VELO tracks, as well as long and upstream tracks, which contain VELO segments. The position of a PV is determined iteratively. In every iteration, a new PV position is determined, and the $\chi^2_{\text{PV}}\,(\text{PV})$ is recalculated accordingly. This iteration stops when sufficient numbers of tracks have been associated to the PV, and when the shift in the z position of the PV in successive iterations is less than a small set threshold value. Successfully fitted vertices are also required to be spatially separated from previously reconstructed ones.

The design specifications of the VELO allow for the precise determination of PV position and track IP. The PV resolution can be measured in data by randomly splitting the set of tracks present in events into two, and running the reconstruction algorithm on both [115]. Vertices in these sets are matched to each another if the differences in their z positions are small. Finally, the differences between the x, y and z positions of matching PVs are calculated if the numbers of tracks used in the two reconstructions are the same. This is done as the PV resolution is strongly correlated to the number of tracks used. The distributions of the differences are fitted with Gaussian functions, and the resolution for each track multiplicity is obtained from its width divided by $\sqrt{2}$. Figure 4.12 shows the resolution of the x position (the results for y are similar) and that of the z as functions of the number of tracks for Run 1 and Run 2. For a PV with 50 tracks, the x coordinate resolution is around 10 μm, and that of the z is around 50 μm. The IP resolution can be determined using good quality tracks that are consistent with originating from a PV [115]. In this case, the difference of the IP from zero is due to measurement resolution. Only events with one reconstructed PV are used, and the PV is required to be based on a minimum of 25 tracks to reduce the impact of PV resolution on the IP. The IP of each track is calculated with respect to the PV refitted without its contribution. The resulting values for the x and y directions can be binned in quantities upon which it shows dependence and fitted with Gaussian functions, the widths of which are taken as the resolution. Figure 4.13 shows the resolution as a function of the inverse transverse momentum (p_T) of the tracks involved.[4] For particles with high p_T, the IP resolution is found to be around 13 μm in both the x and y directions.

[4]The transverse momentum, p_T, refers to the momentum of a particle transverse to the beampipe, i.e. $p_T = \sqrt{p_x^2 + p_y^2}$.

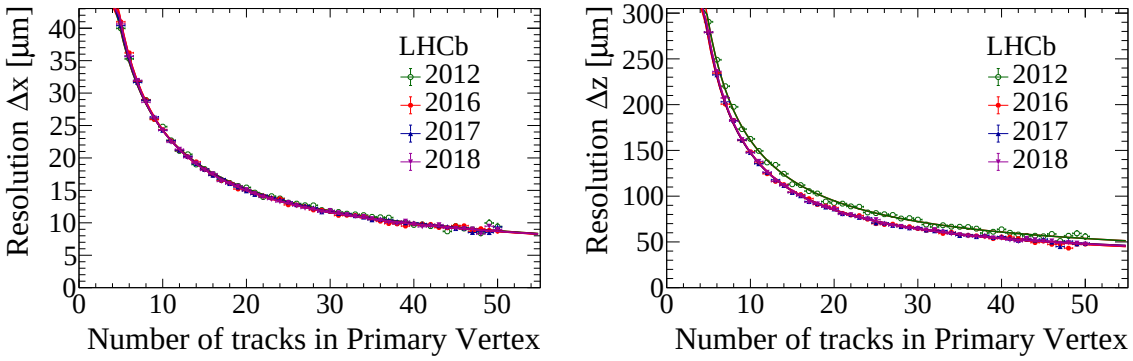


Figure 4.12: Resolution of the x and z positions of PVs versus the number of tracks used in its reconstruction. The values for the y coordinate are similar to that of the x.

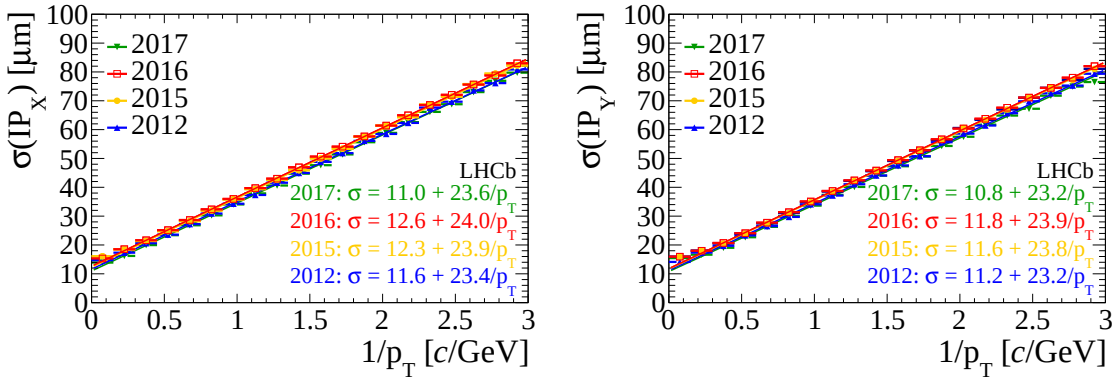


Figure 4.13: Resolution of the IP in the x and y directions as functions of the inverse track transverse momentum.

4.2 Particle identification system

The particle identification system [91, 86] is made up of a diverse range of sub-detectors employing different technologies. They include two ring imaging Cherenkov detectors (RICH 1 and 2) [116], an electromagnetic calorimeter (ECAL) preceded by the scintillating pad detector (SPD) and the pre-shower detector (PS) [117], a hadronic calorimeter (HCAL) [117], and five muon stations (M1-M5) [118, 119, 120], one of which (M1) is positioned between RICH 2 and the PS. The RICH detectors carry out particle identification using Cherenkov radiation emitted when charged particles traverse dielectric materials faster than the (phase) velocity of light in these materials. They provide information for the identification of charged hadrons (π, K, p), and also contribute to the identification of charged leptons (e, μ). The calorimeters, constructed in a layered structure with high density materials interspersed by plastic scintillators, measure the energy deposited by traversing particles. Together with the PS and the SPD, they allow for the identification of neutral (photons and neutral pions) as well as charge particles, in particular electrons. Finally, the muon stations, which are gas based detectors sandwiched by thick iron absorbers, are situated towards the end of the detector as muons from the decays of interest tend to be minimum ionizing particles [7].

4.2.1 Ring imaging Cherenkov detectors

Two RICH detectors with similar designs, RICH 1 and RICH 2 [116, 91, 86], are position up and downstream of the magnet, respectively. These detectors make use of the following relationship between the angle of emission of Cherenkov photons, and the velocity of the particle traversing the dielectric medium,

$$\cos\theta_c = \frac{1}{n\beta} \ , \tag{4.1}$$

where θ_c is the angle of emission, n is the refractive index of the medium, and β is given by the velocity of the particle over the speed of light (in vacuum). The combination of the θ_c information, and the knowledge of the particle momentum from the tracking system allows for its invariant mass and hence its identity to be determined.

In the forward region within the acceptance of the LHCb, there is a strong correlation between the polar angle of the trajectory of a particle and its momentum, such that high momentum particles tend to be produced at small angles close to the beampipe. This feature motivates the use of two RICH detectors filled with materials with different n values suitable for the separation of particle species within different momentum ranges [121]. RICH 1 is situated between the VELO and the TT, and covers the low to intermediate range of $2-40\,\mathrm{GeV/c}$ over the full detector acceptance. RICH 2, positioned between the T stations and the first muon station, covers the higher momentum region of around $15\,\mathrm{GeV/c}$ to over $100\,\mathrm{GeV/c}$ with a more limited acceptance of $15-120\,\mathrm{mrad}$ where most high momentum particles are produced. During Run 1, RICH 1 operated with two different radiators kept at room temperature and pressure – fluorobutane (C_4H_{10}) gas with $n = 1.00014$ as well as plates of aerogel[5] with $n = 1.03$ [121]. RICH 2 uses carbon tetraflouride (CF_4) gas with $n = 1.0005$ [121]. Schematic diagrams of both detectors are shown in Figure 4.14.

Both RICH 1 and 2 have similar optical systems that are composed of sets of planar and spherical mirrors, distributed symmetrically around the beampipe, which deflect Cherenkov photons produced inside the radiators towards arrays of Hybrid Photodetectors (HPDs) positioned on either sides of the detector. These HPDs measure the spatial positions of the photons to allow for the reconstruction, for each particle, of a ring of Cherenkov light, the radius of which is proportional to θ_c. In Run 2, RICH 1 and 2 achieved an average angular resolution of 1.67 mrad and 0.67 mrad, respectively, which are comparable to Run 1 and simulation values [122]. Figure 4.15 shows an example of reconstructed Cherenkov rings from an event, and the relationship between θ_c and particle momentum.

4.2.2 The calorimeter system

The calorimeter system [117, 91, 86] is comprised of two main calorimeters, ECAL and HCAL, preceded by two scintillator pads, the SPD and the PS, which are separated by a lead sheet. The ECAL and HCAL make use of the same working principle, namely, the measurement of particle energy and position by allowing it to deposit its energy in the detector volume, and converting a portion of this energy into electronic signal. In both cases, this is achieved by using layers of high density materials, which increase the probability of particle-material interactions, interspersed by plastic scintillators that fluoresce when exposed to ionising radiation from high energy photons. This light can then be detected and related to the amount of energy deposited. For high energy electrons, the dominant process of energy loss is bremsstrahlung [7]. High energy

[5]The aerogel plates are removed for Run 2, as their ability to provide separation power for very low momentum particles is compromised by the large number of photons in RICH 1 due to the increase in luminosity [122].

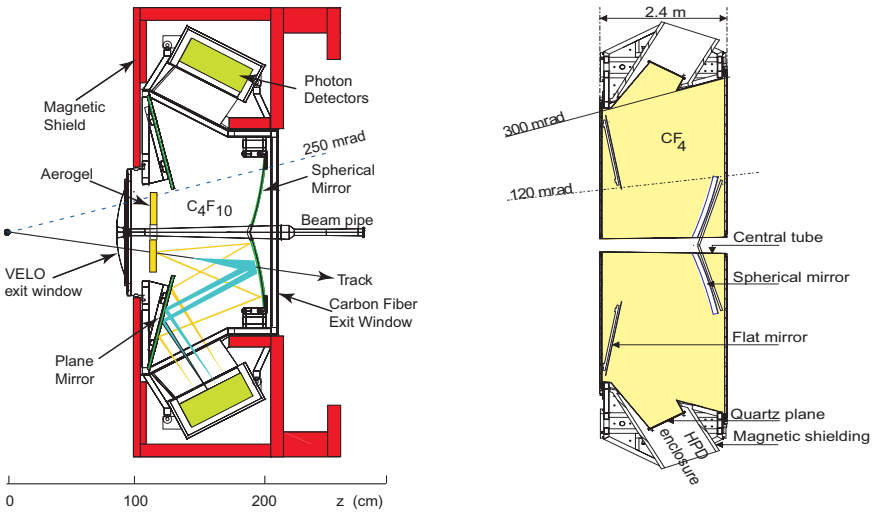


Figure 4.14: Schematic diagram of RICH 1 [91] (left) as seen from the side, and RICH 2 [91] (right) as seen from the top.

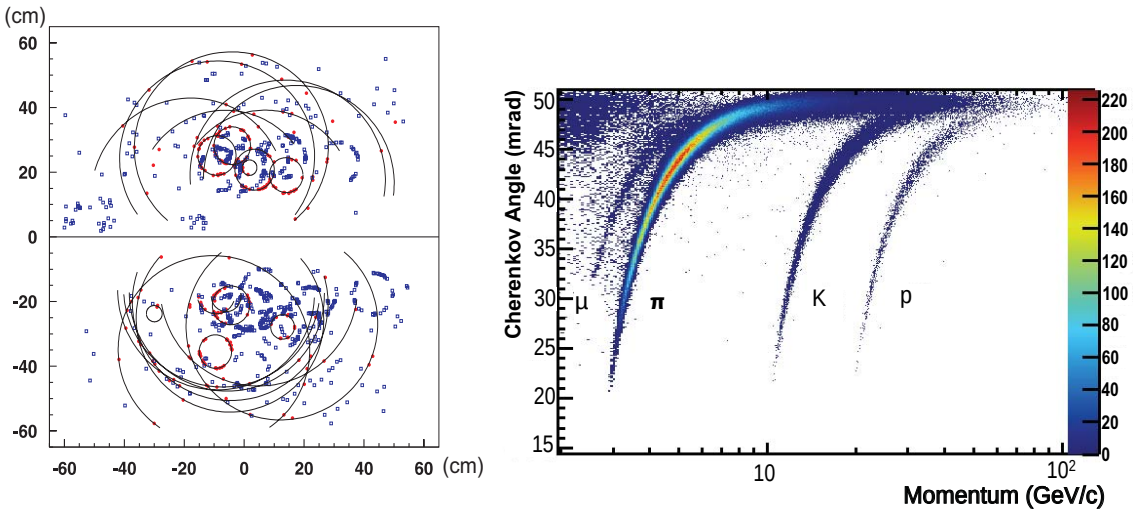


Figure 4.15: Simulation of a single event in RICH 1 [91] (left). The smaller (larger) rings correspond to the C_4F_{10} gas (aerogel) radiators. The relationship between reconstructed Cherenkov angle and momentum for different particle species is shown on the right for the C_4F_{10} radiator [121].

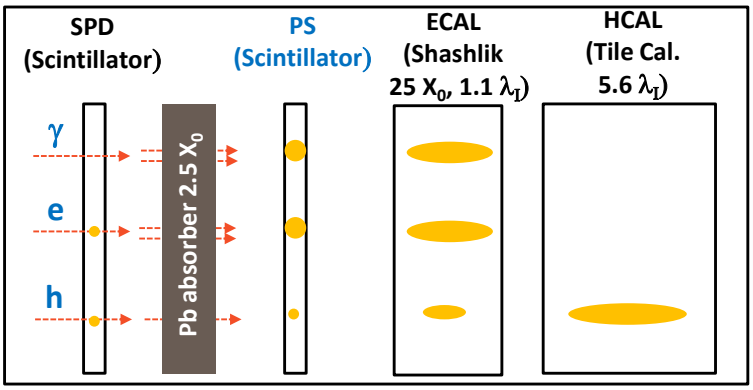


Figure 4.16: Schematic diagram showing the role played by the SPD and the PS in the separation of photons, electrons and hadrons [123].

photons tend to lose energy by e^+e^- pair production. This causes electrons and photons entering the block materials to generate showers of additional electrons and photons ('electromagnetic showers'). The mean distance over which the energy of an electron is reduced to $1/e$ of its initial value by bremsstrahlung is referred to as the radiation length, X_0. The ECAL is designed to extend to $25\,X_0$ in order to achieve good containment of electromagnetic showers, and therefore good energy resolution. Hadrons can also collide inelastically with detector materials to produce secondary particles, which may decay and/or further interact with the nuclei of materials, leading to complex cascades of particles ('hadronic shower'). A useful quantity in this case is the hadronic interaction length, λ_I, which is the mean free path of a hadron between inelastic collisions. The HCAL is built to cover only $5.6\,\lambda_I$ due to space limitations and less stringent requirements. The three-layered SPD/PS system does not constitute a fully fledged calorimeter. Its purpose is rather to provide additional information for the trigger system based on the location and characteristics of energy depositions.

The SPD and PS

The SPD and the PS [117, 91, 86] are plastic scintillator pads that are almost identical in size. The PS has a sensitive area that is about 7.6 m wide and 6.2 m high, while the dimensions of the SPD are around 0.45% smaller, as it is located further upstream. The lead layer in between the two planes is 15 mm thick, which corresponds to $2.5\,X_0$. The SPD-PS system provides important information that, together with ECAL and HCAL measurements, helps to determine electron, photon and hadron candidates for the trigger system. In contrast to electrons, which readily leave signals in the SPD, photons (and other neutral particles) do not interact directly with the scintillator material to produce signals. However, within the lead layer, both photons and electrons begin to initiate electromagnetic showers, which results in them leaving signals in the PS. In the case of the electron, this second signal is typically larger than the first. Hadrons (mostly pions) tend to leave weak signals (if any) in the SPD and the PS due to their large interaction lengths in the materials involved. These characteristics aid the separation of electrons, photons and hadrons, as illustrated by Figure 4.16. In addition, information on the number of hits registered by the SPD (n_{SPD}), which is positively correlated with the number of tracks in the event, is used by the trigger system to veto complex events that would be difficult to analyse.

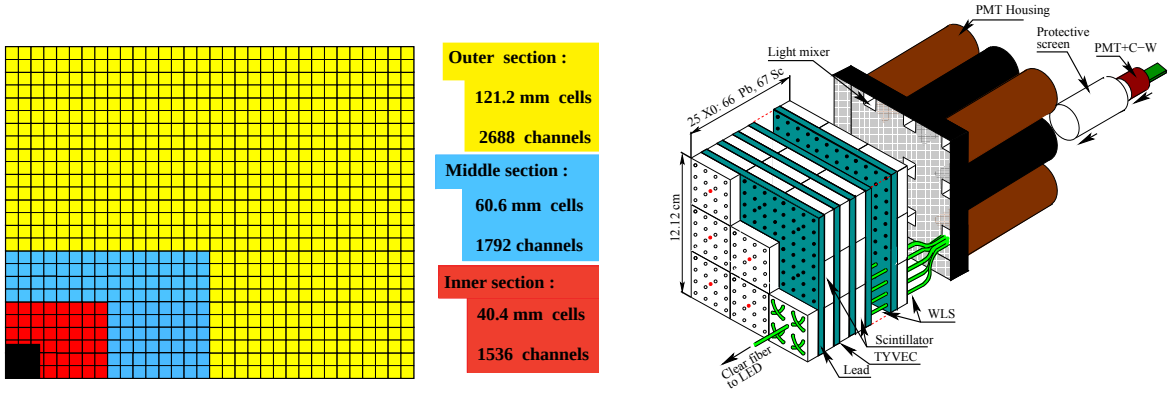


Figure 4.17: Diagram showing the segmentation used in the layers of the ECAL and SPD/PS [124] (left), and a single ECAL cell [124] (right).

The ECAL

The ECAL [117, 91, 86, 124] is a calorimeter designed to contain electromagnetic showers in order to provide optimal energy resolution. The entire sub-detector is around 6.3 m wide and 7.8 m high, and extends 835 mm in the z direction, which corresponds to 25 X_0 or 1.1 λ_I. Constructed as a sampling calorimeter, it is composed of 66 alternating layers of 2 mm thick lead and 4 mm thick polystyrene. Scintillation light generated within the polystyrene are transmitted by wavelength shifting fibres to photomultiplier tubes (PMT), where it is converted into electronic signals that are read out by Front-End Boards (FEBs). The FEBs converts the signals into energy values, E. As the hits density varies over two orders of magnitude as a function of the distance from the beam pipe, the ECAL is divided into inner, middle and outer sections, as shown in Figure 4.17. Each of these regions is populated by cells of different sizes, with the inner region, which is subject to the highest density of hits, featuring the smallest cells. An example of an ECAL cell is shown on Figure 4.17.

ECAL information is used for the trigger system and offline analysis. An important quantity is the transverse energy, E_T, which is defined as

$$E_T = E \sin\theta \ , \tag{4.2}$$

where E is the energy registered by one cell, and θ is the angle between the z axis and the line passing through the centre of that cell from the interaction point. As one of the main signatures of b hadron decays is the production of particles with high transverse momenta, the deposition of large E_T in clusters of cells is used together with other information from the SPD/PS and HCAL to detect events of interest. The energy resolution of the ECAL is around $\sigma_\mathrm{E}/\mathrm{E} = 10\%/\sqrt{\mathrm{E}} \oplus 1\%$, where E has units of GeV [91].

The HCAL

The HCAL [117, 91, 86] is a sampling calorimeter made up of iron and scintillating tiles arranged in planes that are bundled together parallel to the z axis. It is 6.8 m wide, 8.4 m high, and extends 1.65 m in the z direction, which corresponds to 5.6 λ_I. The dimensions of the HCAL are limited by spatial constraints, and as such its thickness does not allow for good containment of hadronic showers. Compared to the ECAL, it is found to have a more limited energy resolution of around $\sigma_\mathrm{E}/\mathrm{E} = (69 \pm 5)\%/\sqrt{\mathrm{E}} \oplus (9 \pm 2)\%$ (E is the deposited energy in

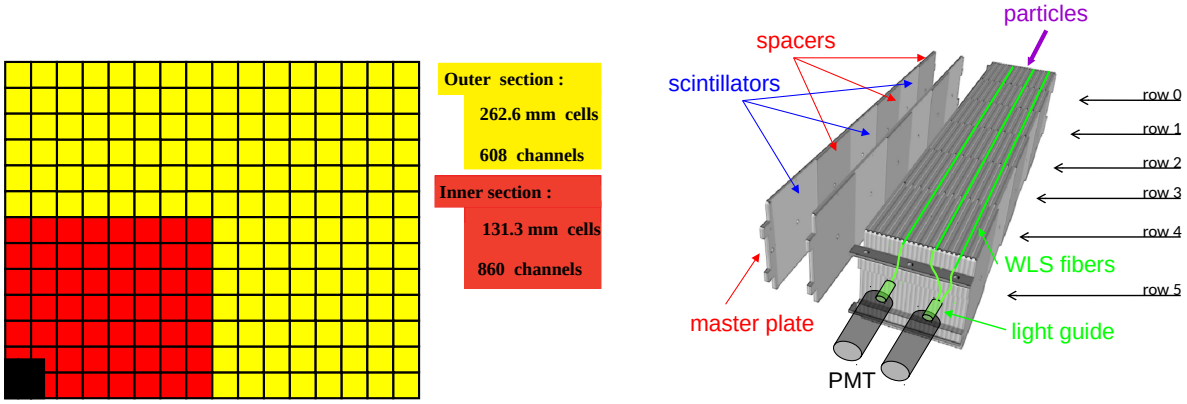


Figure 4.18: Diagram showing the segmentation used in layers of the HCAL [124] (left) and a single HCAL cell [124] (right). Note that the orientation of the HCAL is different from the ECAL, such that traversing particles pass through the sides, rather than faces of the material layers.

GeV) [91]. Nevertheless, as its main purpose is to provide measurements of the E_T of hadrons for the trigger system, moderate energy resolution is sufficient. Like the ECAL, it is divided into regions. Given the increased dimensions of hadronic showers, only two regions with larger cell sizes are used. The HCAL segmentation and an example of a HCAL cell are shown in Figure 4.18.

4.2.3 Muon system

The muon system [118, 119, 120, 91, 86] is composed of five rectangular stations, M1-M5, one of which (M1) is positioned in front of the calorimeter system while the other four (M2-M5) are located at the end of the detector. Stations M2-M5 are interleaved with 80 cm thick iron absorbers, while the calorimeters between M1 and M2 play the role of an absorber. Each station is divided into four concentric regions, R1 to R4, with R1 closest to the beampipe, the dimensions of which are scaled according to the ratio of 1:2:4:8. Multi-wire proportional chambers are used in these regions, with the exceptions of the innermost region R1 of M1, which uses triple gas electron multiplier based detectors due to its exposure to high levels of radiation. The chambers are partitioned into physical channels, which are combined into logical pads with variable dimensions that return binary outputs corresponding to the detection, or not, of a particle. The sizes of these pads scale with the same ratio as the sizes of the regions, such that the occupancy is kept roughly constant throughout a given station. The pads are made to be thinner in the bending plane to allow for more accurate measurements of transverse momenta. A schematic diagram illustrating the design of a station quadrant is given in Figure 4.19.

The muon stations can function as a tracking system, and it is used in this way with a standalone algorithm to reconstruct muon tracks for the trigger system. The calorimeter system together with the iron layers act as an absorber with around 20 λ_I. This allows for the effective isolation of penetrating muons from most other particles. Stations M1-M3 are used to reconstruct the direction of the muon track and to determine its p_T due to their high granularity (small logical pads). The resolution of last two stations M4 and M5 is limited, and they are mainly used for muon identification.

Due to the use of hits from all five stations in the trigger system, the detection efficiency of each station needs to be very high to guarantee high trigger efficiency. Tests show that all five

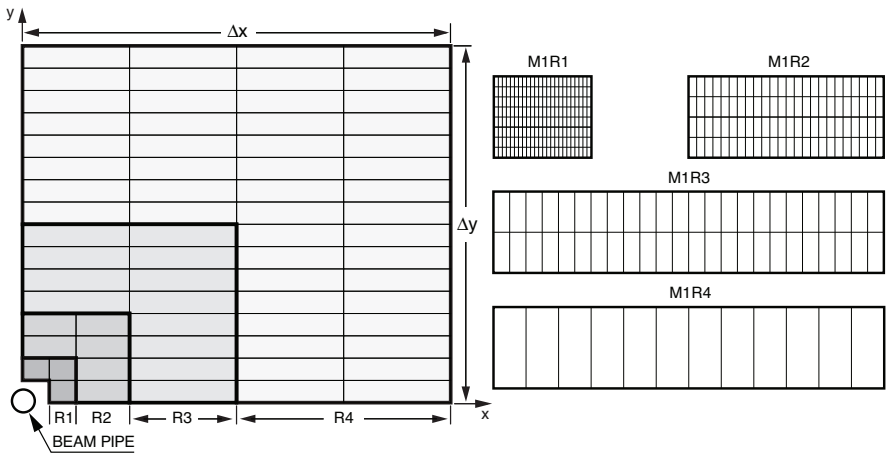


Figure 4.19: Diagram showing one quadrant of a single layer of the M1 station, with rectangular blocks illustrating the sizes of the chambers in the four different regions [91] (left). These chambers are further partitioned into logical pads [91] (right).

stations fulfilled or exceeded the design efficiency requirement of 99% [125].

4.2.4 Particle identification methods

Information from the calorimeter system [117], RICH detectors [116] and the muon stations [118, 119, 120] are used to construct quantities, such as the difference in log-likelihoods between one particle hypothesis and another, which can be used to separate hadrons, electrons, photons, and muons [121, 91, 86, 126]. Other quantities such as the binary IsMuon variable are also made to aid the identification of specific particle species.

The RICH detector is mainly used to separate charged pions, kaons and protons, and provide supplementary information for the separation of charged leptons. Conceptually, particle identification can be carried out for a track that traverse at least one RICH active volume by first finding its associated Cherenkov photons, and then fitting the resulting ring to obtain θ_C. However, in practise, it is more convenient to take an alternative approach [127]. Rather than assigning photon hits to a given track, the pattern of expected hits is determined for the track under the assumption that it belongs to different particles (proton, kaon, pion, muon, electron and deuteron). This is then compared to the observed pattern to assess the likelihoods of these mass hypotheses. One further complication is that the RICH sub-detectors operate in a high occupancy condition, which causes a significant number of Cherenkov rings to overlap. This motivates an inclusive approach. Instead of treating each track separately, a likelihood is constructed based on the probability of seeing photons in each pixel of the detector for a given set of tracks and mass hypotheses. All tracks in the whole event from both RICH sub-detectors are taken into consideration. The likelihood minimisation starts with the assumption that all tracks belong to pions as they are the most abundantly produced particle species in pp collisions. The algorithm then calculates the overall event likelihood. The mass hypothesis of a single track is changed, leaving that of all others unchanged, and the likelihood is recomputed. This iterative procedure is carried out for all individual tracks and mass hypotheses. The change that leads to the largest improvement is identified, and the mass hypothesis of that track is set to this preferred value. The procedure is then repeated to find the next largest increase. The minimisation stops when no further improvements can be found. The results of this process can be formulated, for each track, as the difference in the overall event likelihood

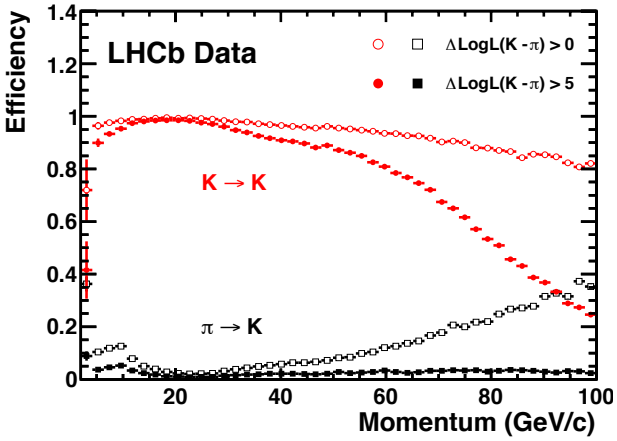


Figure 4.20: The kaon identification efficiency and the pion to kaon misidentification rate as a function of particle momenta measured in data [86]. Values obtained using a looser and a tighter threshold of the $\Delta\log\mathcal{L}^{\mathrm{RICH}}(K-\pi)$ of greater than 0 and 5, respectively, are shown.

when its mass hypothesis is changed from that of a pion to particle x, i.e. $\Delta\log\mathcal{L}^{\mathrm{RICH}}(x-\pi)$, which can be combined with similar information from other sub-detectors.

The RICH particle identification performance can be studied using high purity data samples. These can be obtained from exclusive decay modes with large branching fractions that can be selected without using RICH information, such as the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ decay for kaons and pions. Figure 4.20 shows the performance of kaon identification for two different thresholds of the $\Delta\log\mathcal{L}^{\mathrm{RICH}}(K-\pi)$, showing the effect of a looser threshold ($\Delta\log\mathcal{L}^{\mathrm{RICH}}(K-\pi) > 0$) that favours high signal efficiency, which constitutes the requirement that the kaon mass hypothesis is more likely compared to that of the pion, and that of a tighter threshold leading to stronger background suppression. For the looser cut, averaging over the momentum range of $2-100\,\mathrm{GeV/c}$, the kaon efficiency is found to be around 95%, with a pion misidentification rate of around 10%. The more stringent threshold lowers the misidentification rate to about 3%, at the cost of an approximately 10% reduction in signal efficiency.

The calorimeter system contributes mainly to the identification of electrons, photons and neutral pions [91, 86, 128, 129]. The identification of electrons relies mostly on information from the ECAL, with additional inputs from the PS and the HCAL. For the ECAL, the quantities with the most distinguishing power are related to the balance of track momentum and energy of the charged cluster in the ECAL that it is associated with (Figure 4.21), and the quality of the match between the track and the cluster. These inputs are used to compute the difference in the log-likelihood between the electron and hadron hypotheses, $\Delta\log\mathcal{L}^{\mathrm{ECAL}}(e-h)$ for each track.

Another contribution to $\Delta\log\mathcal{L}^{\mathrm{ECAL}}(e-h)$ comes from presence (or absence) of bremsstrahlung emission. Electrons tend to lose more energy than any other relevant charged particles through this process, therefore the detection of bremsstrahlung photons can be used for its identification. The detector is designed with very little material in the magnet region, which allows for the broad categorisation of bremsstrahlung photons into those emitted upstream of the magnet, and those emitted downstream. Photons belonging to the latter category generally end up in the same calorimeter cells as the electron and are indistinguishable from it. However, those emitted before the magnet are generally well separated. The position of the potential bremsstrahlung photon for a given electron can be predicted by extrapolating its track segment

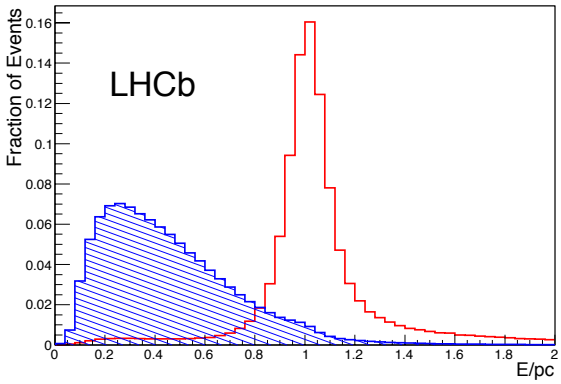


Figure 4.21: Distribution of the ratio between the energy of the cluster in the ECAL and the track momentum (E/pc) in data for electrons and hadrons [86].

before the magnet linearly to the ECAL plane, as illustrated on Figure 4.22. The extent of the agreement between the predicted photon position and the centre of reconstructed photon cluster provides additional information for the separation of electrons from hadrons.

Note that the concept of matching bremsstrahlung photons to electrons is also used to improve momentum resolution. As electron momentum is measured using the curvature of its track, energy loss due to photon emission upstream of the magnet will lead to a reduction in the measured values. This can be compensated for by adding the four-vector(s) of the photon(s) found back to that of the electron. This type of 'bremsstrahlung correction' is made by a dedicated algorithm, `BremAdder` [130], which is ran at a separate stage using a more refined, albeit time-consuming, photon search strategy for improved accuracy. A notable difference is that in this case an extrapolation of the track is also made from its origin point close to the beam into the plane of the ECAL, in addition to the extrapolation from the region before the magnet (end of the TT). Photons are then searched for within a window between these points, the size of which is allowed to vary based on the uncertainties of the extrapolation and that of the photon position.

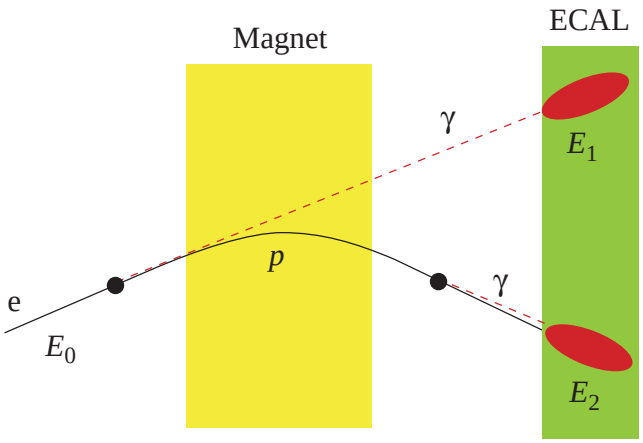


Figure 4.22: Illustration of the bremstrahlung recovery method used to provide information for the identification of electrons [91].

Information from the PS and HCAL also provide distinguishing power. Electrons tend to

deposit more energy in the PS than hadrons. This energy deposition, along with the track momentum, is used to produce values for the difference in the log-likelihood between the electron and hadron hypotheses, $\Delta\log\mathcal{L}^{\mathrm{PS}}(e-h)$. The leakage of electromagnetic showers into the HCAL is expected to be small for true electrons, whereas true hadrons mostly generate showers within the HCAL, therefore $\Delta\log\mathcal{L}^{\mathrm{HCAL}}(e-h)$ is obtained using energy deposition in the HCAL and track momenta. The individual $\Delta\log\mathcal{L}(e-h)$ values are summed to produce the combined $\Delta\log\mathcal{L}^{\mathrm{Calo}}(e-h)$ estimator, which allows for good electron identification performance. As an example, the average identification efficiency of electrons from the decay of $B^+ \to K^+ J/\psi(\to e^+e^-)$ is found to be $(91.9 \pm 1.3)\%$, with a pion misidentification rate of $(4.54 \pm 0.02)\%$ after imposing the requirement of $\Delta\log\mathcal{L}(e-h) > 2$ [86, 131].

Photon identification mainly relies on information from the PS and ECAL. The difference in the log-likelihood between the photon and background hypotheses is obtained using energy deposition in the PS, the quality of the track-cluster match, and the ratio of the energy in the centre cell of a cluster to that of the whole cluster. A Multi-Layer Perceptron (MLP) algorithm is also used to help separate photons from hadrons and electrons. The separation between photons and neutral pions with high E_T, which can decay into a pair of photons that do not leave separable clusters in the ECAL ('merged pions') is aided by another MLP. This algorithm is trained using simulation to distinguish between the two based on the shape of the ECAL clusters, as those left by merged pions tend to be broader and more elliptical than those left by single photons.

The muon stations provide important inputs for muon identification [126, 86]. It is responsible for two main variables, namely IsMuon, which is a binary quantity that offers efficient selection of muon candidates based on their penetration of the muon stations, and the difference in the log-likelihood between the muon and background hypotheses for each track, $\Delta\log\mathcal{L}(\mu)$. For the former, optimised search windows are opened around reconstructed tracks to search for hits in the muon stations. Depending on the momentum of the track, different numbers of muon stations need to register hits for the track to pass the IsMuon criteria. Tracks with momenta in the range of $3 < p < 6\,\mathrm{GeV/c}$, are required to have associated hits in stations M2 and M3. Those in the range of $6 < p < 10\,\mathrm{GeV/c}$ are required to have hits in stations M4 or M5 (or both) in addition to hits in M2 and M3. Very high momenta tracks ($p > 10\,\mathrm{GeV/c}$) are required to be associated with hits in stations M2, M3, M4 and M5. For the latter, the cumulative probability distribution of the averaged squared distance significance (D^2) of hits with respect to the linear extrapolation of the track into the muon stations is used to calculate likelihoods for the muon and non-muon hypotheses. True muons are more likely to leave hits in the muon chambers, resulting in D^2 values closer to zero.

The muon identification efficiency can be evaluated in data by applying a tag-and-probe method to $J/\psi \to \mu^+\mu^-$ decays, and the misidentification rate of hadrons as muons can be quantified using high purity pion and kaon tracks from exclusive decays. In the case of the IsMuon selection, for particles with $p > 3\,\mathrm{GeV/c}$ and $p_T > 0.8\,\mathrm{GeV/c}$, the signal efficiency is around 98%. For $p > 3\,\mathrm{GeV/c}$ and the full p_T range, the proton, kaon and pion to muon misidentification rates are all around 1%. For the $\Delta\log\mathcal{L}(\mu)$, at a moderate cut of $\Delta\log\mathcal{L}(\mu) > 1.74$ with a signal efficiency of around 93% (for tracks with $p > 3\,\mathrm{GeV/c}$ and $p_T > 0.8\,\mathrm{GeV/c}$), the three types of misidentification rates are reduced to around 0.21%, 0.78% and 0.52%, respectively.

The various likelihoods from the different sub-detectors are added linearly to produce a set of powerful combined variables for analyses (DLL), which quantifies, for each track, how likely

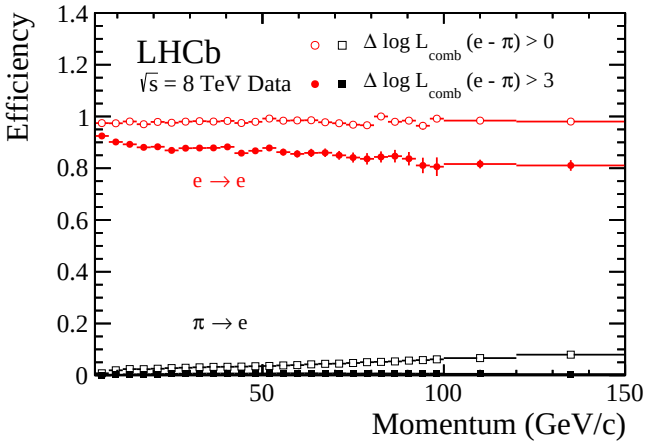


Figure 4.23: The electron identification efficiency and pion to electron misidentification rate as a function of track momentum for a looser and a tighter threshold of the combined $\mathrm{DLL}_{e\pi}$ [86].

a given mass hypothesis is relative to that of the pion, i.e.

$$\mathrm{DLL}_{x\pi} = \log\mathcal{L}_x - \log\mathcal{L}_\pi \ , \tag{4.3}$$

where $x = p$, K, e, μ. This results in improved performance. For example, using a threshold of $\mathrm{DLL}_{e\pi} > 4$, it is possible to achieve an electron identification efficiency of $(92.8 \pm 1.2)\%$ for electrons from the decay of $B^+ \to K^+ J/\psi(\to e^+e^-)$ with a misidentification rate of $(1.01 \pm 0.01)\%$, which is lower compared to the value achieved using calorimeter information only [131]. The electron identification efficiency versus pion to electron misidentification rate in data as a function of track momentum is shown in Figure 4.23.

Alternatively, information from the sub-detectors can be combined using artificial neural network (ANN) algorithms resulting in a set of `ProbNN` variables, which are normalised to be within the range of zero to one, and can be interpreted as a probability [132, 86]. Separate ANNs are trained using simulation to distinguish between particle x and background, resulting in outputs labelled as `ProbNN`x. The advantage of this method is that it can take into account correlations between sub-detector outputs, and allow for the inclusion of additional information, such as PID related variables that cannot be expressed as likelihoods and track quantities that are not directly related to PID. The ANNs are trained using simulation of inclusive B events. Their outputs are designed to quantify, for each track, the probability that it belongs to different particle species (electron, muon, pion, kaon and proton), as well as the probability of it being a ghost. `ProbNN` variables show comparable and often superior performance compared to the DLL.

4.3 Trigger system

The trigger system [86, 91, 86] is comprised of a hardware based L0 trigger, and two software based stages, the High Level Triggers 1 and 2 (HLT1 and HLT2). The L0 trigger uses information from the calorimeter and the muon systems to reduce the event rate from 40 mHz, the rate of bunch crossings, to a more manageable 1 mHz, which allows partial event reconstruction to be made at the HLT level. HLT1 and HLT2 reduce the event rate down to about 5 kHz (Run

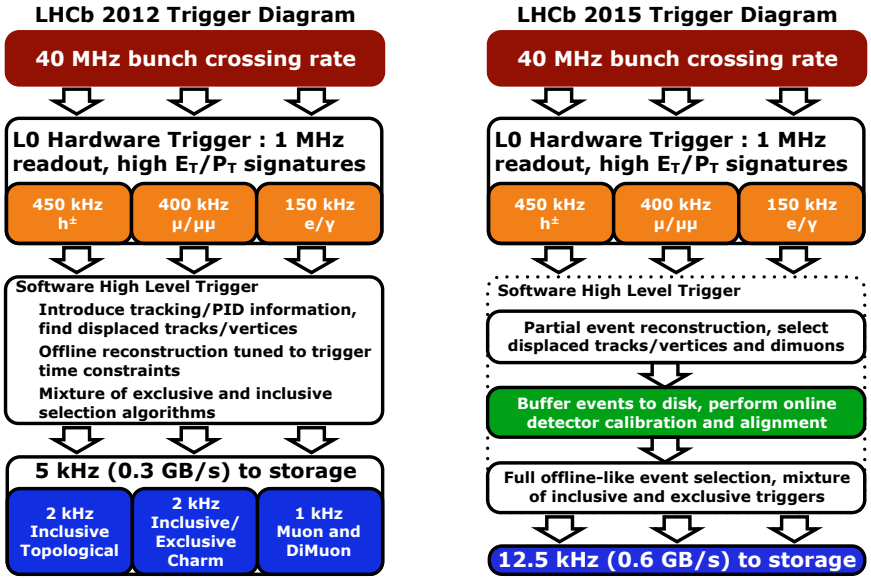


Figure 4.24: The trigger scheme used in 2012 [133] (left) and 2015 [133] (right).

1) and 12.5 kHz (Run 2) for storage. The trigger schemes used in 2012 and 2015 are shown in Figure 4.24.

4.3.1 Hardware trigger

The L0 hardware trigger [86, 134, 113] is implemented on custom electronics boards and ran synchronously with the full bunch crossing rate of 40 mHz. Its purpose is to filter out events based on the typical signatures of B decays (high p_T tracks and large E_T depositions) in order to reduce the event rate down to about 1 mHz. The L0 trigger can be divided into three independent units: L0-Muon, L0-Calorimeter and L0-Pileup. The latter, based on the pile-up stations in the VELO, is only used for the determination of luminosity, and will not be discussed further. The L0-Muon trigger relies on information from the muon system, which is used to fit track segments to determine the p_T of muon candidates. The L0-Calorimeter trigger uses information from the calorimeter system to search for hadron, photon and electron candidates. These candidates are defined as follows:

- **L0Muon (L0DiMuon)** – muon track segment with the highest p_T (product of the p_T of the muon track with the highest p_T and that of the track with the second highest p_T);
- **L0Hadron** – HCAL cluster with the highest E_T. If there is an ECAL cluster in front of the HCAL cluster, the E_T measurements of both calorimeters are summed;
- **L0Photon** – highest E_T cluster in the ECAL that also has 1 or 2 PS hits in front of it but no hits in the corresponding SPD cells. If the cluster is located in the inner region of the ECAL, it is also accepted if there are 3 or 4 PS hits in front of it;
- **L0Electron** – highest E_T cluster in the ECAL with the same requirements as L0Photon. In addition, at least one SPD cell in front of the PS cells needs to be hit.

Events containing one or more of the above are retained if they pass the p_T or E_T thresholds that are chosen based on the operating conditions of the detector. A threshold on the n_{SPD} is

also applied to veto high multiplicity events that would take up a disproportionate amount of processing time in the software stage.

The total output of the L0 trigger of 1 mHz consists of approximately 400 kHz from the muon triggers, 450 kHz from the hadron trigger, and 150 kHz from the combined electron and photon triggers.

4.3.2 Software trigger

The software trigger [86, 113, 88, 101] is implemented in the Event Filter Farm (EFF), a computing farm which contains around 1700 nodes, 800 of which is added for Run 2, with around 27000 processing cores. It is based on the same software framework that is used in offline[6] data processing (simulation and reconstruction), which enables event reconstruction to be performed at the trigger level. This allows the HLT to access more complex quantities. However, the time available for the processing of an event is much shorter online than offline. While the latter uses ~ 2 s per event, at the L0 output rate of ~ 1 MHz, and given the computing power, the available time is ~ 20 ms per event in Run 1 and ~ 35 ms in Run 2. The HLT is therefore divided into two stages. During HLT1, a fast partial event reconstruction is made, which aims to filter out ones that contain potential b hadron decays to reduce the event rate to a point where full reconstruction using offline-like configurations becomes possible.

In Run 1, the event is processed using simplified versions of the offline tracking and PID algorithms at the HLT1 level. In particular, relatively high momentum thresholds are used for track reconstruction. The presence of a single detached high-momentum track of interest is used to trigger the event [137] (`Hlt1Track`). In Run 2, following the addition of more computing resources and the optimisation of the data processing framework, the minimum track momentum thresholds are significantly reduced, and a line based on the detection of two-tracks that form a secondary vertex can be introduced [138]. Multivariate classifiers are used to increase efficiency and reduce background (`Hlt1TrackMVA` and `HLT1TwoTrackMVA`).

HLT1 reduces the event rate to about 80 kHz (110 kHz) in Run 1 (Run 2), which allows the HLT2 to have a time budget of ~ 150 ms per event in Run 1 and ~ 350 ms in Run 2. In Run 1, a simplified version of the offline reconstruction chain is again used, while in Run 2 modifications are made to allow for very good alignment between the online and offline processing strategies. The information available at this stage allows for large numbers of trigger lines (~ 500) to be developed. Among them, the topological trigger lines [139, 140, 113] (`Hlt2Topo`), which take up around 40% of the HLT2 output rate, are particularly relevant. They are designed to trigger on generic b hadron decays (to be 'inclusive') with at least two charged decay products and a displaced secondary vertex. The strategy used is to first select for tracks of interest based on p_T, χ^2_{IP} with respect to the PV, and the fit χ^2/ndof. The latter two quantities are used to select for tracks that are less likely to originate from particles produced directly from pp interactions (rather than from the decays of e.g. b hadrons), or be ghost tracks. Then two, three and four body candidates are made by progressively combining together tracks that are consistent with sharing the same vertex, are significantly displaced from the PV, and have topologies compatible with the decays of b hadrons. Multivariate classifiers are used to select events in both Run 1 and Run 2. The inputs to these classifiers are in the form of the secondary vertices that they contain. They are trained to separate b hadron decays that can be fully reconstructed within the acceptance of the detector from ones that cannot be,

[6] At the LHCb, the data can be processed 'online' in the EFF during (or close to the time of) data taking before it is sent to permanent storage, and 'offline' in the LHC Computing Grid using stored events that passed the HLT [135, 136].

and to suppress backgrounds from the displaced vertices of c hadrons originating from the PV. An event is kept if at least one vertex passes the classifier decision. In Run 1, a type of boosted decision tree (bonsai boosted decision tree), is used. In Run 2, the MVA choice is revised and a more efficient algorithm (MatrixNet) is used along with a different selection of variables for training. Thus these lines are referred to as the `Hlt2TopoNBodyBBDT` (Run 1) and `Hlt2TopoNBody` (Run 2), where $N = 2, 3, 4$. The additional requirement that one of the tracks of interest (or both) passes the electron identification criteria results in the more specific lines of `Hlt2TopoENBodyBBDT` (Run 1) and `Hlt2TopoENBody` (Run 2) for which the MVA threshold can be slightly reduced with respect to the general `NBody` lines.

Part III

Angular analysis of $B^0 \to K^{*0} e^+ e^-$ decays

5 Analysis strategy

Analyses carried out by experiments at the LHC have repeatedly confirmed the remarkable success of the SM, but revealed as of yet no definitive signs of NP. In this context, the emergence of a seemingly coherent set of deviations in rare FCNC decays in the last decade comprised of anomalies in the differential branching fractions and angular observables of several $b \to s\mu^+\mu^-$ decays together with hints of LFU violation in ratios of decay rates between muon and electron modes has led to considerable interest. Both sets of results are amenable to relatively simple explanations featuring shifts in a small number of WCs (Section 2.6.4), in particular LFU violating shifts in $C_{9\mu}^{\mathrm{NP}}$. The updated simultaneous measurement of R_K and $R_{K^{*0}}$, however, show good agreement with the SM. Nevertheless, the differences with respect to the previous results are due almost entirely to the treatment of decays with electrons in the final state. Therefore, anomalies of the muon mode persist, and require explanation. Measurements of the angular observables of the $B^0 \to K^{*0}e^+e^-$ decay can help clarify the nature of these tensions [76], which may be due to unaccounted for QCD effects, such as the long distance $c\bar{c}$ contributions, or NP that obeys LFU. In addition, the angular analysis of this decay in the central q^2 region will be the first of its kind in a hadronic machine, demonstrating the viability of the measurement in this challenging experimental environment.

Within the flavour anomalies, single electron mode measurements are rare due to the challenges brought about by the low mass of the electron, which leads to significant energy loss through photon emission. This includes bremsstrahlung that occurs due to interactions between the electron and the electrostatic field of the nuclei in the parts of the detector that it travels through, but also intrinsic final state radiation (FSR) – decays with charged particles in the final state will unavoidably take place with photon emission. Nevertheless, material bremsstrahlung and FSR are experimentally inseparable. Both are often referred to as 'bremsstrahlung', and both contribute to resolution degradation. At the LHCb, a dedicated algorithm (`BremAdder`) is used to correct for this effect (Section 4.2.4), which, nevertheless, has limitations. Firstly, not all emitted photons can be recovered, as only those within the ECAL acceptance and have transverse momenta of $p_T > 75\,\mathrm{MeV/c}$ can be reconstructed. Secondly, due to the occupancy of the ECAL, reconstructible photons may be missed if they land in the same cells as other particles, and in general it is not always possible to match the right photons to their associated electrons. These complications motivate and sometimes necessitate the use of methods to control or reduce the impact of bremsstrahlung, which at times comes at the cost of introducing additional features and sources of systematic uncertainties.

One of these methods is the use of an alternative q^2 definition. The q^2 quantity referred to in the definition of the angular observables does not include FSR or bremsstrahlung effects and can be called the 'true' q^2. The reconstructed q^2 values calculated directly from the measured four-vectors of the electron-position pair have limited resolution, therefore defining a q^2 window based on it would lead to the inclusion of large numbers of signal events, which have very different true q^2 values. Backgrounds, in particular charmonium candidates, i.e. $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ decays, can also 'leak' into the signal region. To improve resolution and reduce the migration of events in q^2, the dielectron four-momentum is obtained through a fit where the reconstructed B^0 meson is constrained to originate from its associated PV, and a

second constraint is also used to force the invariant mass calculated from the final state particle tracks to equal the nominal mass of the B^0 meson. This alternative variable, hereafter known as q^2_c, is more closely aligned to the true q^2 for signal candidates, which simplifies the correction of remaining FSR and bremsstrahlung effects, and enables superior signal-background separation.[1] Its use allows the analysis to be performed more safely in the standard central q^2 region of $1.1 - 6.0\,\text{GeV}^2/\text{c}^4$. Furthermore, it allows the range to be extended up to $7\,\text{GeV}^2/\text{c}^4$, which increases the signal yield by about 20%.

Like the q^2, the invariant mass of the B^0 candidates, which is used to separate signal from backgrounds, can be calculated using the PV constraint to improve resolution, resulting in the variable denoted as $m(K\pi ee)_{\text{PV}}$. For the control mode of $B^0 \to K^{*0}J/\psi(\to e^+e^-)$, which is used mainly for validation purposes and to determine nuisance parameters of the signal mass distribution, it is also possible to constrain the invariant mass of the dielectron pair to match the nominal mass of the J/ψ meson ($m(K\pi ee)_{\text{PV},\,J/\psi}$). However, as this variable shows non-negligible correlations with $\cos\theta_\ell$, it is only used for background suppression.

The angular analysis of $B^0 \to K^{*0}(\to K^+\pi^-)e^+e^-$ decays [141] is performed using data collected by the LHCb detector in the years of 2011-2012 (Run 1) at CoM energies of $\sqrt{s} = 7\,\text{TeV}$ (2011) and $\sqrt{s} = 8\,\text{TeV}$ (2012), 2015-2016 (Run 2p1) and 2017-2018 (Run 2p2) at a CoM energy of $\sqrt{s} = 13\,\text{TeV}$, corresponding to a total of around $9\,\text{fb}^{-1}$ of integrated luminosity. Measurements of CP averaged angular observables are made in the central q^2 regions of $1.1 < q^2 < 6.0\,\text{GeV}^2/\text{c}^4$ and $1.1 < q^2 < 7.0\,\text{GeV}^2/\text{c}^4$, the results of which can be compared to that of the muonic channel. The key components of this analysis, which will be discussed in detail in the following chapters, are outlined below[2]:

- **Samples and corrections** – the analysis uses data collected by the LHCb detector and simulation of the rare and control modes, as well as that of backgrounds. The simulation needs to be a good proxy for the data, especially since the correction of distortions to the signal q^2 and angular distributions relies on good agreement between the two. Corrections are made to address known simulation-data differences in the modelling of PID variables, trigger efficiencies, and kinematic and event multiplicity variables. Simulation and data samples are discussed in Chapter 6, where a brief overview of the process of simulation generation at LHCb is also provided.

- **Candidate selection** – the selection of signal candidates involves cut-based selection

[1] The constrained q^2_c is found in simulation to reduce the net migration of truth-matched *Signal* (*Low-mass background*) candidates with $q^2_{\text{true (MC)}} = |p_{B^0} - p_{K^{*0}}|^2$ (calculated from the difference of the true B^0 and K^{*0} four vectors to mostly avoid FSR effects, which is only possible in simulation) values beyond the range of $1.1 < q^2_{\text{true (MC)}} < 7.0\,\text{GeV}^2/\text{c}^4$ from around 14% (26%) in the case of the unconstrained q^2, to around 2.4% (3.8%), and similar values are found for the smaller $q^2_{\text{true (MC)}}$ range. Overall, the percentage of events that migrated greater than $|1\,\text{GeV}^2/\text{c}^4|$ is reduced from around 20% (34%) to 2.6% (5.5%).

[2] The work discussed in this part is carried out by the author with inputs from other members of the LHCb collaboration (mainly the other proponents of the $B^0 \to K^{*0}e^+e^-$ angular analysis [141]). The production of simulation and data samples, as well as PID corrections (Section 6.5.1), are the results of work done by the other proponents. Trigger efficiency corrections and corrections to kinematic and multiplicity variables (Sections 6.5.3 and 6.5.4) are made using the work of the other proponents with additional inputs from the author. In both cases, errors are the responsibility of the author. These sections are included for completeness. The contents of Chapters 7 to 9 are mainly the work of the author, although it should be noted that aspects of the analysis strategy, such as the preselection (Section 7.1) and the general approach used for acceptance correction (Section 8.1.1), are based heavily upon (and sometimes intentionally aligned to) that of similar analyses [142, 67, 57]. They are also influenced by (and in some cases based upon) the results of previous studies [143], in particular the set-up of the MVA (Section 7.3). Additional references are provided where necessary.

to reduce generic backgrounds, targeted vetoes to reduce contamination from specific sources, and a cut on the output of a multivariate classifier to suppress remaining background that originates from the random combinations of tracks. This selection strategy is based on that of existing analyses [66, 67] and the work carried out in [143]. Details on the selection criteria and the set-up of the multivariate classifier can be found in Chapter 7.

- **Effective acceptance** – the angular distribution of the final state particles of the signal decay can only be described by Equation 2.14 in the absence of experimental distortions, bremsstrahlung (and FSR). The strategy used to correct for these effects, which involves the parametrisation of 'effective acceptance functions', is discussed in Section 8.1.

- **Components modelling** – significant sources of backgrounds that remain after applying the selection criteria are included in the angular fit. The procedures used to obtain models for the mass and angular distributions of backgrounds, as well as the mass distribution of the signal, are detailed in Section 8.2.

- **Angular fit** – the weighted unbinned maximum likelihood function that is used to extract the observable values in data is introduced in Section 8.3. A blinded data fit is made to obtain estimations of the number of signal and background events. Based on these yields, pseudoexperiments are generated to ascertain the expected sensitivity to the observables of interest, and to study the behaviour of the fit. The set-up of the pseudoexperiments and the results of these studies are discussed in Section 8.4.

- **Control mode validation** – the high statistics control mode is used to validate aspects of the analysis strategy. The method used to extract observable values for the signal is applied to control mode candidates with as few modifications as possible. As LFU is well established for this type of decay, the resulting observable values are compared against the external values of the $B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ mode. This check and its results are discussed in Section 9.1.

- **Systematic uncertainties** – choices made in parts of the analysis, such as the simulation correction strategy, the parametrisation of the effective acceptance functions, and the modelling strategies used for the background components, can affect the values of the observables obtained. Significant sources of systematic uncertainties are evaluated, and the results are reported in Section 9.2.

- **Blinded results** – due to the complexity of the analysis, the strategy needs to be thoroughly cross-checked (and potentially adjusted) before the results can be revealed ('unblinded') to reduce the risk of bias. Nevertheless, aspects of the current results are shown in Section 9.3, and the expected sensitivity of the measurement is also presented.

6 Samples

The samples used for the analysis include data collected by the LHCb detector during its Run 1 and Run 2 operations. As the total amount of stored data is large, only filtered subsamples, which contain reconstructions of pp collision events that are likely to contain the decay of interest, are used. In addition, other subsamples of data are used to study backgrounds. Simulations are relied upon extensively in many parts of the analysis. As such, corrections need to be made to address known sources of simulation-data differences. The strategies used in the initial data processing and corrections are informed by that of similar analyses [142, 67, 57], as well as the results of the initial studies documented in [143]. In the following sections, the production of simulation, and the processing of both real and simulated data are briefly outlined, and the relevant samples are discussed.

6.1 Simulation generation and data processing

The production of simulation and the processing of simulated and real data are carried out at the LHCb via a software system [144] that consists of custom applications implemented within the `Gaudi` framework [145, 146], a platform designed to facilitate the interconnections between, and usage of, data processing and analysis applications for high energy physics experiments. The main applications therein include those used to simulate pp collision events as well as the detector response, and those used to carry out event reconstruction.

Event generation and detector simulation are organised by the `Gauss` framework [147], which calls upon multiple external packages, the most relevant of which are `Pythia` [148], `EvtGen` [149] and `PHOTOS` [150] for the former, and `Geant4` [151] and `Boole` [152], for the latter. `Pythia` is an event generator that simulates pp collisions and quark hadronisation. It is responsible for the generation of the particle content arising from collisions, and the simulation of the kinematics of the resulting hadrons. Particles with specified decay models are then handled by `EvtGen`, which simulates the stages of their subsequent decay chains using the corresponding amplitudes. In addition, `EvtGen` calls the `PHOTOS` package for relevant decays to modify existing events in order to introduce FSR effects. At this point, the generated, or 'generator level' events are stored in a specific format before being passed on to the `Geant4` software, which simulates the propagation of particles through the detector and interactions with detector materials. The hits in the virtual detector are digitised by the `Boole` software, which also simulates the detector response as well as that of the L0 trigger. The output of this final stage is digitised data that mimics real data, and can be processed in the same way.

The processing of both simulated and real data starts with the trigger software `Moore` [153], which implements the algorithms of the HLT (Section 4.3). Events passing the trigger are handled by the offline reconstruction programme `Brunel` [154] before being stored and made available to the `DaVinci` [155] analysis software, which is ran to filter out subsamples of manageable sizes based on user requirements ('stripping lines'), which are in the form of a series of initial cuts on a selected number of PID, kinematic and topological variables with signal-background separation power, the choice of which is optimised for the decay modes of interest. The main stripping line used for this analysis is the `Bu2LLKee` line, which is commonly used

for the $b \to s e^+ e^-$ modes in LFU tests. As the processing of large quantities of data is computationally demanding, central 'stripping campaigns' are ran infrequently, and would involve hundreds of user created lines. Data after stripping are stored and can be accessed by individual users through `DaVinci` to produce samples for analyses.

6.2 Data

The data samples used in this analysis contain around $1\,\text{fb}^{-1}$ of integrated luminosity collected by the LHCb detector at a CoM energy of $\sqrt{s} = 7\,\text{TeV}$ in 2011, $2\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$ in 2012, $2\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$ in 2015 and 2016, and $4\,\text{fb}^{-1}$ at $\sqrt{s} = 13\,\text{TeV}$ in 2017 and 2018.

All samples are passed through the requirements of the same stripping line (`Bu2LLKee`), which are given in Table 6.1. An important feature of this line is that it uses the `DiElectronMaker` tool to create electron-position pairs, which ensures that if the same bremsstrahlung photon candidate can be associated with both candidates, it is only added to one, selected at random. This reduces instances of over-correction. This stripping line includes a series of requirements on PID, track and vertex quality, topological and kinematic, and global event variables to perform an initial selection. Events with too many tracks that are difficult to reconstruct properly are removed using a cut on the n_{SPD}. Decaying particles are required to have good quality decay vertices (low $\chi^2_{\text{DV}}/\text{ndf}$ values), and to be distinctly separated from their vertices of origin (minimum threshold set for PVχ^2 separation). An upper limit is set on the significance of the distance between the track of the reconstructed B^0 meson and the PV ($\chi^2_{\text{IP}}\,(\text{PV})$), as its track is expected to point towards the PV, while those of the final state particles are required to exceed a minimum threshold. The cosine of the angle between the momentum vector of the B^0 meson track and the line joining its associated PV and decay vertex (`DIRA`) is required to be close to one (angle close to zero), such that only candidates where the two are well aligned are kept. Requirements are also imposed to reduce misidentification (DLL cuts), and sanity cuts are made to only retain candidates which have reconstructed B^0 and K^{*0} meson masses close to their nominal values.

For background studies and modelling, background dominated data samples are used. These include ones that contain charge conservation violating $B^0 \to K^\pm \pi^\mp e^\pm e^\pm$ ('same-sign') candidates, which can arise when misidentified, or correctly identified final state kaon, pion, and electron tracks from unrelated processes are combined and retained when they happen to fulfil signal requirements. Thus, they are expected to resemble backgrounds that arise from the random combination of tracks in the main data samples ('combinatorial' background). This, as well as another type of background that involves the partial reconstruction of decays, can also be studied with (charged) lepton flavour violating[1] (LFV) $B^0 \to K^+ \pi^- e^+ \mu^-$ candidates. The stripping criteria applied to both samples are kept as similar as possible to that of the `Bu2LLKee` line.

[1]The SM gauge bosons do not couple to leptons of different generations (e.g. e and μ), such that the lepton number, L_i for generation i, which is $+1$ (-1) for lepton (anti-lepton), is effectively conserved. This is an 'accidental symmetry' in contrast to the gauge symmetries that are imposed in the construction of the model. While neutrino oscillations allow for its violation, the associated branching fractions are vanishingly small ($\mathcal{O}(10^{-50})$) e.g. [156].

Table 6.1: Summary of the requirements of the main stripping line (`Bu2LLKee`).

	Requirement
Event	$n_{\mathrm{SPD}} \leq 600$
B^0	$\lvert m - m^{\mathrm{PDG}}_{B^0} \rvert < 1500\,\mathrm{MeV/c^2}$
	`DIRA` > 0.9995
	$\chi^2_{\mathrm{IP}}\,(\mathrm{PV}) < 25$
	$\chi^2_{\mathrm{DV}}/\mathrm{ndf} < 9$
	PV χ^2 separation > 100
K^{*0}	$\lvert m - m^{\mathrm{PDG}}_{K^{*0}} \rvert < 300\,\mathrm{MeV/c^2}$
	$p_T > 500\,\mathrm{MeV/c}$
	$\chi^2_{\mathrm{DV}}/\mathrm{ndf} < 25$
K	$\mathrm{DLL}_{K\pi} > -5$
	$\chi^2_{\mathrm{IP}}\,(\mathrm{PV}) > 9$
π	$\chi^2_{\mathrm{IP}}\,(\mathrm{PV}) > 9$
e^+e^-	$m < 5500\,\mathrm{MeV/c^2}$
	$\chi^2_{\mathrm{DV}}/\mathrm{ndf} < 9$
	PV χ^2 separation > 16
e	$\mathrm{DLL}_{e\pi} > 0$
	$p_T > 300\,\mathrm{MeV/c}$
	$\chi^2_{\mathrm{IP}}\,(\mathrm{PV}) > 9$

6.3 Simulation

The main simulation samples used in this analysis are those of the signal decay, which are generated using WCs from [157, 158] and LCSR form factors from [159]. In addition, corresponding generator level samples without selection cuts, detector simulation or FSR are produced[2], which allows for all selection and resolution effects to be parametrised. As the control mode is used for validation purposes, full simulation of the $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ decay is also used along with its corresponding generator level samples.

Simulations of other modes are used to study or model specific backgrounds. One important type comes from multibody decays, which have been reconstructed with one or more final state particles missing. They include decays such as $B^+ \to K^+\pi^+\pi^- e^+e^-$ where the kaon and pions may originate from intermediate resonances. The complexity of this type of 'partially reconstructed background' motivates the use of a data-driven approach, rather than one that is based entirely on simulation. To this end, a sample of $B^+ \to K_1(\to K^+\pi^+\pi^-)e^+e^-$ events is generated with the `LSFLAT` model, which produces a flat distribution in the $K^\pm\pi^\pm\pi^\mp$ invariant mass. This distribution is later reweighted to resemble background-subtracted data of the $B^+ \to K^+\pi^+\pi^- J/\psi$ mode. Alternative samples are also produced to study the impact of this choice leading to the quantification of associated systematic uncertainties. They include simulation of the $B^+ \to K_1^+(1270)e^+e^-$ mode, with around 40% of decays proceeding via an intermediate $K^+\rho^0(770)$ pair, 20% via $K^{*0}(892)\pi^+$, $< 1\%$ via $K^+\omega$, and the rest taking place nonresonantly (around 40%), and that of the $B^+ \to K_2^{*+}(1430)e^+e^-$ mode, with a similar mixture of intermediate states. Another type of background consists of double semi-leptonic decays such as $B^0 \to D^-(\to K^{*0}e^-\bar{\nu}_e)e^+\nu_e$, where the neutrinos cannot be reconstructed. Likewise it is difficult to simulate the composition and characteristics of this type of background in data, and it is modelled using LFV data samples, although the simulation of the dominant mode is used to study systematic uncertainties. For the control mode, samples of the $\Lambda_b^0 \to pK^- J/\psi(\to e^+e^-)$ decay are generated for background modelling in the main analysis, and simulations of partially reconstructed backgrounds originating from the decays of three types of B mesons, B^+, B^0 and B_s^0, are used to study the systematic uncertainties of the control mode angular fit made for the validation of the main fit strategy.

6.3.1 Truth matching

Candidates reconstructed from simulated events can be categorised using the generation information [160]. They are classified as *Signal* if all the final state particles are correctly identified (none missing) and are matched to the same true decaying particle, which undergoes the specified signal decay. Classification as *Signal* also requires all intermediate resonances belonging to the requested decay chain to be correctly reconstructed, otherwise candidates are classified as *Quasi-signal*. True signal decays can also be classified as *Low-mass background* if they have associated unreconstructed photons that exceed a given momentum threshold (300 MeV/c) and are therefore counted as a 'missing particle'. Another important category is *Ghost*, which takes in candidates comprised of at least one fake track that is reconstructed from a group of largely unrelated hits. In most parts of the analysis, including the parametrisation of effective acceptance functions and the modelling of signal mass distributions, all the aforementioned categories are considered effectively as signal. The inclusion of *Low-mass background* events is motivated by the leakage of true signal decays with the emission of hard photons into this category, and

[2] In contrast to the full simulations, which are centrally produced, these samples are generated locally using standard LHCb software.

the inclusion of *Ghost* candidates is due to their tendency to peak in the signal mass region. Note that simulated candidates can also be partially reconstructed, as in the case where the simulated decay of $B^+ \to K^+\pi^+\pi^-e^+e^-$ is reconstructed as the signal, and they can be combinatorial in nature (e.g. when the true origin of one or more of the 'final state particles' is the PV). Due to the complexity of data events, simulated combinatorial background is generally not expected to reproduce all of its features. Nevertheless, in some cases they can still replicate its main characteristics.

6.4 Trigger selection

The L0 and HLT triggers filter out potential events of interest for storage. Thus, all data samples are composed of events that passed one or more trigger lines. As different lines make use of different criteria, only the outputs of the ones that are the most relevant for the decay of interest (and are suitable for analysis[3]) are used. For the L0 they include:

- **L0TIS** – events triggered by candidates that are not part of the signal candidate[4] ('Trigger Independent of Signal' or TIS), i.e. hadron, muon and photon candidates that passed the L0Hadron and/or L0Muon and/or L0Photon triggers;
- **L0En** – events triggered by a signal candidate ('Trigger on Signal' or TOS), i.e. one or both electrons of the signal candidate passed the L0Electron trigger, which do not simultaneously belong to the L0TIS category.[5]

These two categories are designed to be mutually exclusive, and to contain approximately equal statistics. They are expected to show subtly distinct features. For example, the L0En sample will generally show slightly improved resolution compared to the L0TIS due to the more stringent requirement on the E_T of at least one electron belonging to the signal candidate, but the L0TIS sample is expected to be less sensitive[6] to potential trigger-related differences between muon and electron modes. Differences in background composition, as well as signal efficiency across the phase space of interest can also be expected. Nevertheless, due to limited available statistics, the two categories are combined in most parts of this analysis save for certain simulation-data corrections. The impact of using this combined (or averaged) strategy is limited, especially since the ratio of L0TIS and L0En events match that of the data to a good extent after trigger efficiency corrections.

The HLT1 lines used include the `Hlt1TrackAllL0` (Run 1) and `Hlt1TrackMVA`[7] (Run 2), which retain events based on the presence of a good quality track after partial reconstruction, and the HLT2 lines include the topological triggers based on two and three-track combinations with or without additional electron identification requirements, i.e. the `Hlt2Topo(E)[2,3]` (Section 4.3.2).

[3]Lines that e.g. admit too few events and thus do not contribute enough statistics to justify necessary corrections (and the systematic uncertainties they may introduce) are neglected.

[4]The particles triggering the event may have originated from the decay of the other b hadron produced from the original $b\bar{b}$ pair.

[5]L0En, or L0E *exclusive*, does not include candidates, which also belong to the L0TIS category. In parts of the analysis, the inclusive L0E category is also used.

[6]The L0TIS is chosen as the primary category for this reason for the $R_{K^{*0}}$ analysis [67].

[7]The `Hlt2TrackMVA` line is not used as no dedicated alignment procedure is available (Section 6.5.2).

Table 6.2: Summary of the trigger lines used for this analysis, which are required to be TOS unless stated otherwise.

Run 1	**Run 2 (2015)**	**Run 2 (2016-18)**
L0Electron or L0Hadron(TIS) or L0Muon(TIS) or L0Electron(TIS)		
Hlt1TrackAllL0	Hlt1TrackMVA	
Hlt2Topo(E)[2,3]BodyBBDT	Hlt2Topo[2,3]Body	Hlt2Topo(E)[2,3]Body

6.5 Corrections to the simulation

Simulation is used in multiple aspects of the analysis, including the modelling of the signal mass distribution, the training of the multivariate classifier against combinatorial background, and the parametrisation of the effective acceptance. Good agreement between simulation and data is therefore essential. However, several factors lead to differences between the two. They include limitations in the simulation of the B^0 hadron kinematics, imperfect description of detector materials, and changes in operating conditions during data taking that are not reflected in the simulation settings. These factors lead to discrepancies in a number of important variables related to kinematics, particle identification (PID), and event characteristics, as well as the trigger efficiency.

Several data-driven correction strategies are used to reduce differences through the creation of adjusted variables or correction weights. Trigger alignments are also made by emulating the effects of data cuts. PID variables are corrected first, as they are used for background suppression for the subsequent corrections. Alignment in the HLT trigger requirements is then made, after which the simulated efficiencies of the L0 and HLT triggers are corrected to match those of the data. Finally, kinematic and event multiplicity variables are corrected using simulation and data samples passing all preselection requirements, taking previous corrections into account. The choice of the ordering of the last two steps is made based on the nature of the ratio-based trigger efficiency correction method, which is not well suited to taking kinematic corrections into account. Attempting to correct trigger efficiency after kinematic (and multiplicity) correction leads to reduced agreement in the these variables (Appendix A). The aforementioned stages are discussed in the following sections.

6.5.1 Particle identification variables

PID variables used in the analysis are obtained by combining, either directly or through a multivariate analysis, the outputs of many subdetectors (Section 4.2.4). This makes them challenging to reproduce, as they require good simulation of the kinematics of traversing particles together with the response of the detector, which is influenced by many factors including occupancy, instrumental alignment, temperature and gas pressure that are subject to fluctuations. The common approach to address this problem is through data-driven methods based on 'calibration samples' – decay modes that can be isolated in whole or in part without the use of PID variables, which, after the statistical subtraction of background, can provide unbiased distributions of signal PID variables. These samples are used in the `PIDCalib` software package [161]

to perform corrections according to a variety of methods implemented as separate tools, the most relevant of which are `PIDGen` and `PIDCorr` [162, 143].

The `PIDGen` (for PID 'generation') approach involves the replacement of simulated PID variables by new ones that are generated through the inverse transform sampling of their corresponding distributions in the calibration samples. More specifically, the correction starts with the determination of the PID distribution in the calibration samples as functions of a few kinematic and event variables, usually combinations of p_T, η and the number of tracks (n_{Tracks}), which are correlated to PID variables, i.e. $p_{\text{calib}}(x|p_T,\, \eta,\, n_{\text{Tracks}})$, where x refers to the PID variable in question. The cumulative distribution function (cdf) of this pdf is given by,

$$P_{\text{calib}}(x|p_T,\, \eta,\, n_{\text{Tracks}}) = \int_{-\infty}^{x} p_{\text{calib}}(x'|p_T,\, \eta,\, n\text{Tracks})\, dx' \ . \tag{6.1}$$

The new, or corrected PID variables can then be obtained from

$$x_{\text{corr}} = P^{-1}_{\text{calib}}(u|p_T,\, \eta,\, n_{\text{Tracks}}) \ , \tag{6.2}$$

where u, the sampling variable (effectively the cumulative probability value), can be drawn from a uniform distribution with values ranging from 0 to 1. The resulting x_{corr} will be distributed according to p_{calib}, as desired. However, one important shortcoming of this approach is that the generated variables are fully decoupled from their original versions, such that their correlations with variables other than p_T, η and n_{Tracks}, are lost. Such correlations are nevertheless expected to be present in data due to the dependency of PID variables, in particular the neural network based `ProbNN`, on multiple input parameters. Furthermore, correlations also exist between different PID variables, e.g. the more likely a given track is to belong to a kaon, the less likely it is to belong to a pion. The strategy of simply adding more variables tends to suffer from the drawbacks of increased dimensionality, which, for a fixed calibration sample size, means a reduction in the precision of the resampling.

The preservation of correlations motivates the modification of Equation 6.1. Instead of taking values for u from a uniform distribution, it is taken from the cdf of the simulated PID distribution, that is,

$$u = P_{\text{MC}}(x_{\text{MC}}|p_T,\, \eta,\, n_{\text{Tracks}}) = \int_{-\infty}^{x_{\text{MC}}} p_{\text{MC}}(x'|p_T,\, \eta,\, n_{\text{Tracks}})\, dx'. \tag{6.3}$$

The new distribution is then obtained from

$$x_{\text{corr}} = P^{-1}_{\text{calib}}(P_{\text{MC}}(x_{\text{MC}}|p_T,\, \eta,\, n_{\text{Tracks}})|p_T,\, \eta,\, n_{\text{Tracks}}), \tag{6.4}$$

which is effectively x_{corr} written as a function of x_{MC}. Thus, when the difference between the distributions of p_{calib} and p_{MC} are small, x_{MC} and x_{calib} are highly correlated. This allows for the preservation of correlations present between the original variable (x_{MC}) and other PID variables as well as event and track parameters, and is the approach implemented by `PIDCorr` (for PID 'correction').

Both `PIDCorr` and `PIDGen` use unbinned descriptions of the calibration pdfs in four dimensions (x, p_T, η and n_{Tracks}), realised through a modified kernel density estimation (KDE) technique using the Meerkat library [163], which offers improved descriptions of boundaries and narrow features for statistically limited samples compared to more traditional KDE methods.

The calibration samples used include the decays of $D^0 \to K^-\pi^+$ (from $D^{*+} \to D^0\pi^+$) for the kaon and pion and $J/\psi \to e^+e^-$ (from $B^+ \to K^+J/\psi$) for the electrons. Where possible, the `PIDCorr` approach is used to correct the distributions of all DLL and `ProbNN` variables used to select signal candidates. They include $K\,(\pi)\,\mathrm{DLL}_{K\pi}$, $e\,\mathrm{DLL}_{e\pi}$, $K\,(\pi)$ `ProbNNk`, π `ProbNNpi`, $K\,(\pi)$ `ProbNNp` and e `ProbNNe`. The effects of these corrections are illustrated in Figure 6.1. In general, the corrected distributions are found to show improved agreement with those of the background subtracted control mode candidates.

6.5.2 Trigger cuts alignment

The settings of the L0 and HLT are configured by Trigger Configuration Keys (TCKs), which set the sequence of algorithms ran and the thresholds used in cuts. TCKs also correspond to specific versions of the `Moore` software, which is responsible for running the HLT triggers in both data and simulation. This feature is designed for reproducibility, as any changes to the code would need to take place with the creation of a new TCK. As such it is possible to switch fully to a particular setting by using the right TCK (and `Moore` version).

During data taking, the trigger settings are changed relatively frequently, for example, to adjust trigger rates. This flexibility is one of the key benefits of running the trigger in software, however, it needs to be taken into account in the corresponding simulation. While the data are processed using multiple TCKs, the simulation samples are produced with a single TCK per year, which is often the configuration used for the largest percentage of data events. Nevertheless, remaining mismatches lead to simulation-data differences. This necessitates the alignment of the trigger requirements, in particular that of the HLT1 line, which is most strongly affected, for the years of 2012 and 2016.

The approach used closely follows that of the $R_{K^{(*0)}}$ analysis [67]. For 2012 samples, the HLT1 requirements of the `HLT1AllTrackL0` line varied during data taking, in particular the thresholds on track momenta and transverse momenta. However, only the loosest requirements are used for the simulation. To emulate the effects of tighter cuts that are applied in the collection of a fraction of the data, the same percentage of randomly selected simulated events that passed the original loose requirements are selected with a tighter threshold, and the `HLT1AllTrackL0` decision is updated accordingly. As bremsstrahlung correction is not carried out prior to the HLT1 decision, uncorrected track momenta values are used for the electrons.

A similar approach is used for the 2016 samples to correct differences in the simulation of the `HLT1TrackMVA` line. The cut applied to each track, which differs depending on the TCK setting, is given by

$$\begin{aligned} &(p_T(\mathrm{GeV}) > 25 \text{ and } \log(\chi^2_{\mathrm{IP}}) > 7.4) \\ &\qquad\qquad \text{or} \\ (\log(\chi^2_{\mathrm{IP}}) > &\frac{1}{(p_T(\mathrm{GeV})-1)^2} + \frac{b}{25}\cdot(25 - p_T(\mathrm{GeV})) + \log(7.4)), \end{aligned} \tag{6.5}$$

where p_T and $\log\chi^2_{\mathrm{IP}}$ are the transverse momenta and the logarithm of the significance of the impact parameter, respectively, and b is the tunable parameter. A single value of $b = 1.1$ is used in the processing of 2016 simulation samples, while three different values are used during

[8] Note that while attempts have been made to suppress residual backgrounds in data without biasing the variables in question, some events remain, leading to features such as the tail at large $\pi\,\mathrm{DLL}_{K\pi}$ values. Therefore these plots should been seen as a visual check rather than a rigorous assessment of the level of agreement between simulation and data.

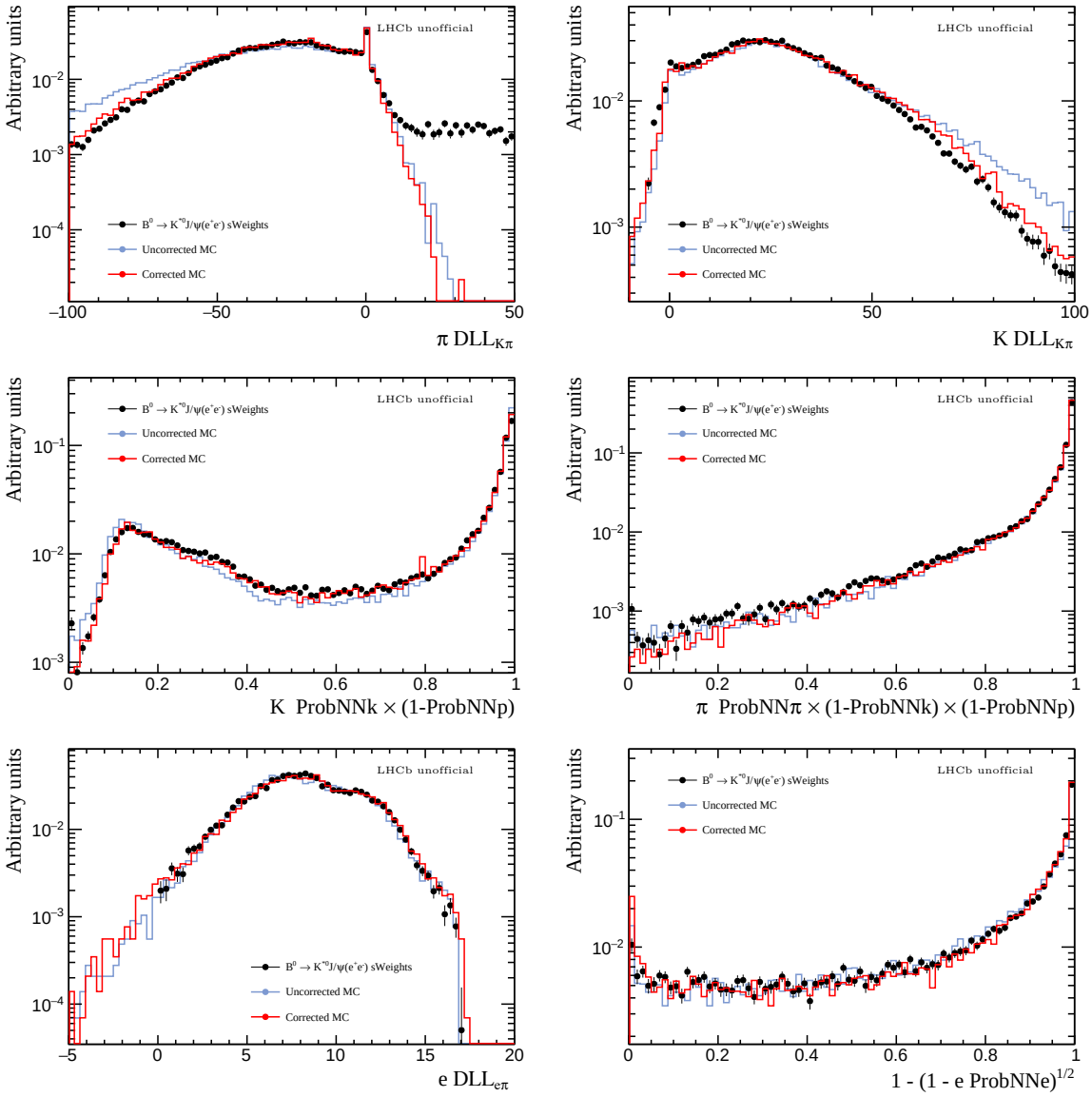


Figure 6.1: Distributions of relevant PID variables in simulation (blue and red) and background subtracted data containing $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ candidates (black) for data taken in 2018 and simulation samples produced with the corresponding conditions [141]. Both single resampled variables and combinations of variables show improved agreement with data compared to the original distributions.[8]

Table 6.3: The fractions of the 2016 data (taken with the two magnet polarity configurations) that are obtained with different values of the parameter b, which varies depending on the TCK setting.

Polarity	$b = 1.1\,[\%]$	$b = 1.6\,[\%]$	$b = 2.3\,[\%]$
MagDown	89.8	0.0	10.2
MagUp	30.0	15.5	54.5

Table 6.4: Thresholds used in cuts applied to both simulation and data samples belonging to the L0E category for the different years of data taking.

Year	E_T^{L0} [MeV]
2011	>2500
2012	>3000
2015	>3000
2016	>2700
2017	>2700
2018	>2400

data collection, as shown in Table 6.3. To emulate the effects of the tighter requirements, simulated events passing the original `HLT1TrackMVA` line are randomly split into subsamples, the sizes of which reflect the fractions of data taken with the different TCK settings. These samples are cut using higher values of the b parameter, and the decision of the `HLT1TrackMVA` line is updated accordingly. As before, transverse momenta used for the electrons are values calculated without bremsstrahlung correction.

In addition to the alignment of the HLT1 trigger settings, the ageing of the ECAL causes the meaning of the E_T thresholds of the L0 trigger to change with time. For example, damage to the PMTs due to high currents during operation leads to a reduction in gain and therefore signal strength. Moreover, radiation damage to the scintillator tiles and optical fibres leads to decreased performance. Both these factors can alter the relationship between detected and actual energy deposition. However, these effects, which mainly affect the electron L0 trigger output, are not reproduced by simulation. A way to reduce differences between simulation and data is to apply cuts to the electron transverse energy measurements that served as inputs to the L0 trigger in order to remove regions where the disagreements are most pronounced [67]. As the original quantities are not retrievable offline, a proxy quantity (E_T^{L0}) calculated using a special tool (`L0CaloTool`) in `DaVinci` is used. A common cut that is tighter than the TCK requirements of both simulation and data is applied to candidates belonging to the L0E category. It requires the E_T^{L0} value of the electron and/or positron that passed the L0Electron trigger to exceed given thresholds, which are shown in Table 6.4.

6.5.3 Trigger efficiency corrections

The efficiency of triggering on signal candidates is not well reproduced by simulation due to its dependency on the detector occupancy, and correlations between the signal and the

underlying event. In principle the correction can be made by simply calculating the efficiencies in simulation and data using the number of signal events before and after the trigger in question, i.e. $\epsilon^{\rm MC}_{\rm trig} = N^{\rm MC}_{\rm trig}/N^{\rm MC}_{\rm all}$ and $\epsilon^{\rm data}_{\rm trig} = N^{\rm data}_{\rm trig}/N^{\rm data}_{\rm all}$, and then use the ratio of $\epsilon^{\rm data}_{\rm trig}/\epsilon^{\rm MC}_{\rm trig}$ as weights. However in practise this strategy cannot be applied directly as data collection always involves the use of the trigger system ($N^{\rm all}_{\rm data}$ inaccessible). Instead, the 'TISTOS' approach [164] is used. This method involves the use of a reference sample that is obtained without using the trigger of interest (and can be assumed to be independent from it) and taking the number of events that pass both this reference trigger and the trigger of interest divided by the total number of events that pass the reference trigger as the efficiency. For the two L0 categories of L0TIS and L0E (inclusive), the simulation (and data) efficiencies, $\epsilon^{\rm MC\,(data)}$, can be found via

$$\epsilon^{\rm MC\,(data)}_{\rm L0TIS} = \frac{N^{\rm MC\,(data)}_{\rm L0E\,and\,L0TIS}}{N^{\rm MC\,(data)}_{\rm L0E}} \ , \tag{6.6}$$

and

$$\epsilon^{\rm MC\,(data)}_{\rm L0E} = \frac{N^{\rm MC\,(data)}_{\rm L0E\,and\,L0TIS}}{N^{\rm MC\,(data)}_{\rm L0TIS}} \ , \tag{6.7}$$

where $N^{\rm MC\,(data)}_{\rm L0E\,(L0TIS)}$ is the number of events in simulation (or data) that pass the L0E (or L0TIS) requirement, and $N^{\rm MC\,(data)}_{\rm L0E\,and\,L0TIS}$ is the number of events in simulation (or data) that pass both. Note that the inclusive L0E category needs to be used, as the overlap of the exclusive L0En with L0TIS is zero due to their mutually exclusive definitions.

For the HLT triggers, which are always required to be TOS, the TIS lines can be used as the reference trigger, as these two sets of requirements are not mutually exclusive. The efficiency calculations are made separately for events in the L0TIS and L0E categories. Conceptually, this is given by

$$\epsilon^{\rm L0TIS\,MC\,(data)}_{\rm HLT} = \frac{N^{\rm L0TIS\,MC\,(data)}_{\rm TIS\,and\,TOS}}{N^{\rm L0TIS\,MC\,(data)}_{\rm TIS}} \ , \tag{6.8}$$

for the former, and

$$\epsilon^{\rm L0E\,MC\,(data)}_{\rm HLT} = \frac{N^{\rm L0E\,MC\,(data)}_{\rm TIS\,and\,TOS}}{N^{\rm L0E\,MC\,(data)}_{\rm TIS}} \ , \tag{6.9}$$

for the latter.

The correction weights are calculated on the basis of the PID corrections and after performing trigger alignment. They are obtained in bins of the quantities upon which the efficiency depends, namely the transverse momenta of the B^0 candidates in the case of the L0TIS category and for the HLT trigger, and in regions of the ECAL, each of which is split into bins of E_T^{L0}, for the L0E category. Two sets of weights are used – one for the L0 and one for the HLT – and in each case they are obtained separately for the two L0 categories and for each year of data taking. More details are given in the sections below.

L0 trigger correction

The data used in this analysis are comprised of events that pass two sets of L0 trigger requirements (Section 6.4). In both cases, obtaining per-event correction weights, $\omega_{\rm L0}$, necessitates the isolation of high purity signal samples. This is done using preselected control mode candidates to which L0 trigger requirements are not applied. An additional cut of $\pm 60\,{\rm MeV/c^2}$ around the constrained B^0 mass ($m(K\pi ee)_{{\rm PV},\,J/\psi}$) is used to strongly suppress remaining background.

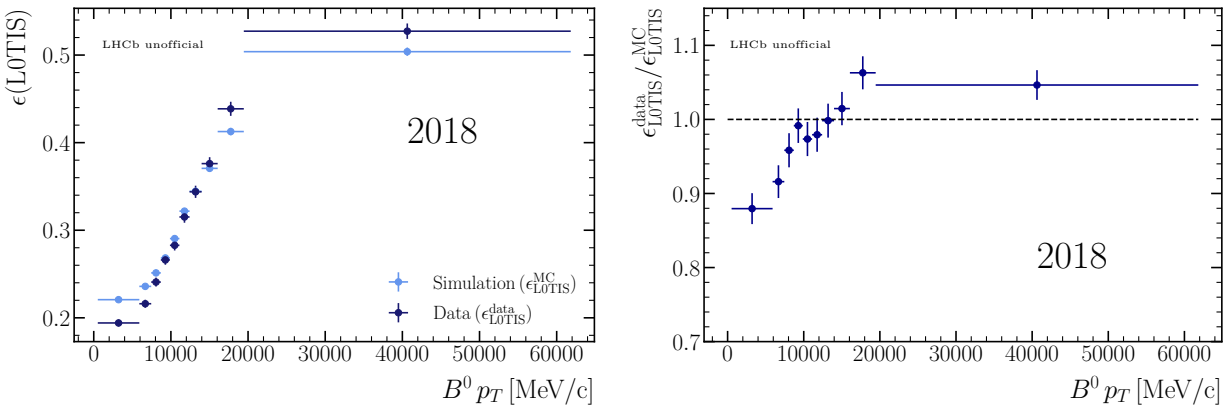


Figure 6.2: Signal efficiencies for the L0TIS trigger requirement (left) estimated using the L0E as the reference trigger for the simulation (light blue) and data (dark blue) for 2018, and correction weights calculated from their ratios in bins of $B^0\,p_T$ (right).

Corresponding simulation samples are selected in the same way and are also required to pass truth-matching requirements (*Signal*, *Quasi-signal* and *Low-mass background*).

The correction weights for the L0TIS sample are obtained using the L0E sample. The efficiency of $\epsilon^{\text{data}}_{\text{L0TIS}}$ is calculated via equation 6.6 using the number of control mode candidates remaining after the selections. The same is done for the corresponding simulation sample ($\epsilon^{\text{MC}}_{\text{L0TIS}}$). These values are obtained in bins of the p_T of the B^0 candidate, which is correlated with the p_T of the decay products of the other b hadron (of the initial $b\bar{b}$ pair) that may be responsible for triggering the event. For each bin, the weight is given by

$$\omega^{\text{L0TIS}}_{\text{L0}} = \frac{\epsilon^{\text{data}}_{\text{L0TIS}}}{\epsilon^{\text{MC}}_{\text{L0TIS}}} \,, \tag{6.10}$$

such that all events belonging to a given bin receive the same weights. Histograms showing the values of the L0TIS efficiencies in data and simulation, and the resulting corrections for each bin are shown in Figure 6.2 for the year of 2018 as an example.

The corrections for the L0E category can be obtained similarly using estimated efficiencies from simulation and data samples that passed the L0TIS trigger selection (Equation 6.7). In this case, as the probability of an electron triggering the event is strongly dependent upon the region of the ECAL that it passes through and the amount of transverse energy deposited, E^{L0}_T, the correction weights are obtained in bins of the latter for the three regions of the ECAL (inner, middle and outer). The weights for the inclusive L0E category can be calculated analogously to that of the L0TIS. However, the actual category used is the exclusive L0En. In this case, the efficiency to not trigger on the L0TIS requirement needs to be multiplied to the inclusive values,

$$\omega^{\text{L0En}}_{\text{L0}} = \frac{\epsilon^{\text{data}}_{\text{L0E}}}{\epsilon^{\text{MC}}_{\text{L0E}}} \cdot \frac{1-\epsilon^{\text{data}}_{\text{L0TIS}}}{1-\epsilon^{\text{MC}}_{\text{L0TIS}}} = \omega^{\text{L0E}}_{\text{L0}} \cdot \frac{1-\epsilon^{\text{data}}_{\text{L0TIS}}}{1-\epsilon^{\text{MC}}_{\text{L0TIS}}} \,. \tag{6.11}$$

Histograms showing the L0E efficiencies in data and simulation are given in Figure 6.3 for 2018 samples.

[9]Note that the differences between simulation and data are more pronounced near the E^{L0}_T thresholds in particular for 2017 and 2018 samples. While it is possible to tighten the threshold at the cost of reducing signal efficiency, its impact on the analysis (in terms of increased systematic uncertainty) is expected to be much lower than both the expected statistical uncertainty of the measurement and the dominant sources of

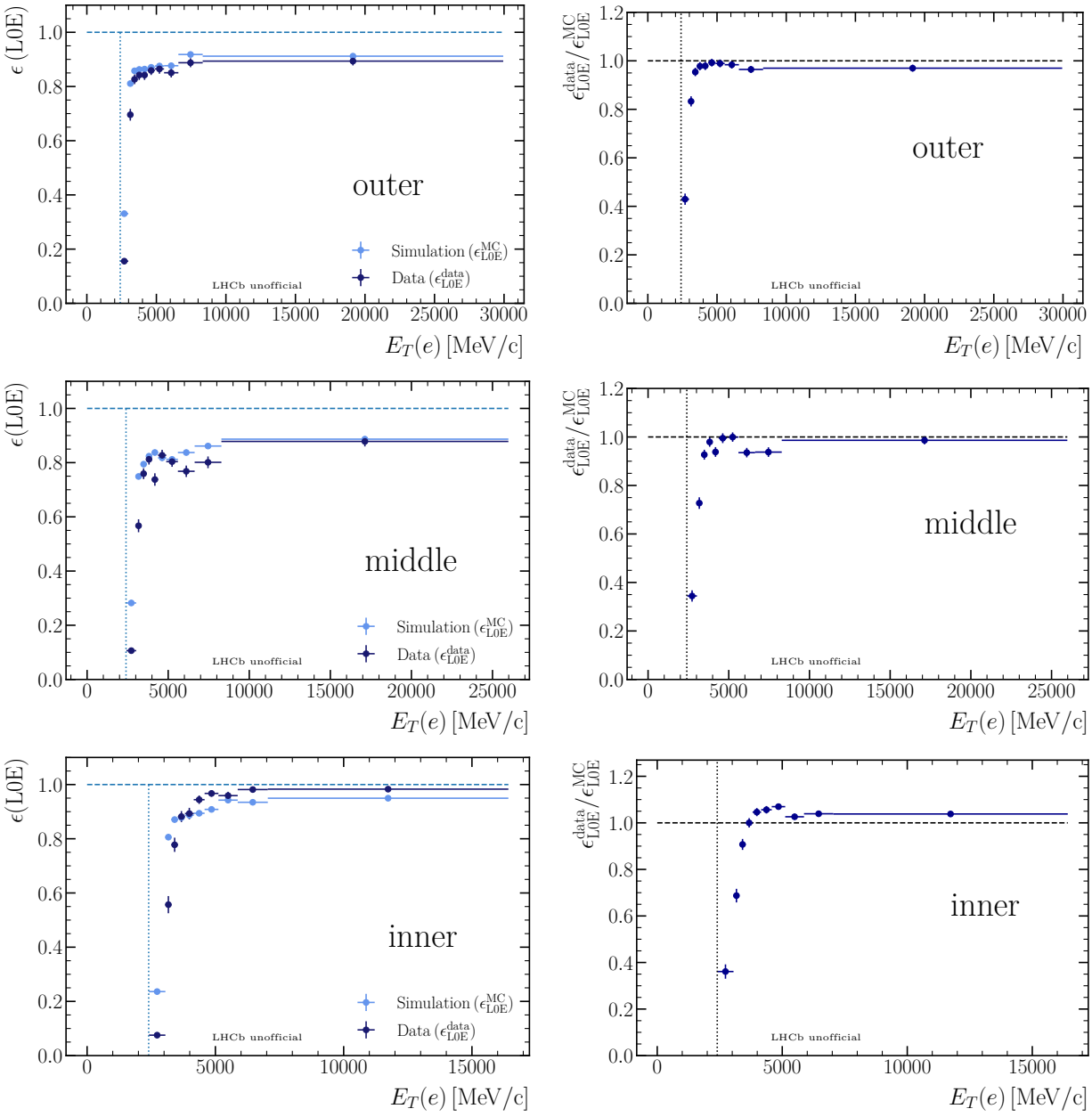


Figure 6.3: Signal efficiencies for the L0E trigger requirement (left) estimated using the L0TIS as the reference trigger for the simulation (light blue) and data (dark blue) for 2018 in bins of E_T^{L0} in the three regions of the ECAL and their ratios (right). The final correction weights for the exclusive L0En category also include ϵ_{L0TIS}.[9]

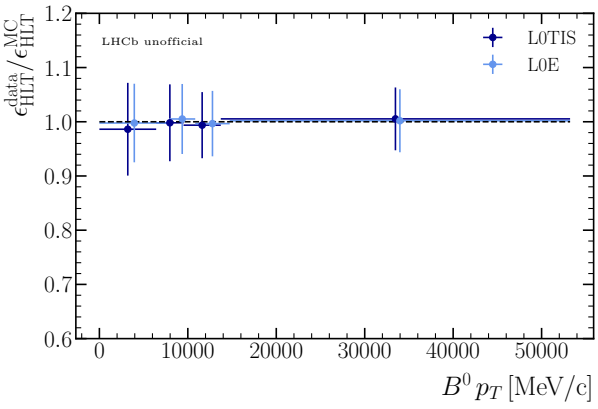


Figure 6.4: Correction weights for the HLT trigger efficiency in bins of $B^0\,p_T$ for the L0 categories of L0TIS and L0E.

HLT trigger correction

The HLT trigger correction is made via per-event weights separately for the L0E and L0TIS categories, for each year of data taking and in bins of $B^0\,p_T$, with the inclusion of the L0 correction weights. Signal-rich samples used to estimate the HLT efficiencies are obtained by selecting control mode candidates in simulation and data using a set of HLT TIS lines.

The HLT efficiency in simulation is obtained using control mode candidates in the TIS sample that are also TOS with respect to the lines of interest, and the number of candidates passing TIS in total (Equations 6.8 and 6.9). Note that to take L0 correction weights into account, for the simulation, the ratio is calculated using the sum of the L0 weights rather than the number of events. Fits are made to separate signal from background in data, and the corresponding ratios are calculated using the signal yields found. This procedure is repeated for all bins of B^0 p_T, the edges of which are defined to allow them to be populated approximately equally. The resulting correction weights, ω_{HLT} for L0TIS events in the simulation are

$$\omega^{\text{L0TIS}}_{\text{HLT}} = \frac{\epsilon^{\text{L0TIS data}}_{\text{HLT}}}{\epsilon^{\text{L0TIS MC}}_{\text{HLT}}} \ , \tag{6.12}$$

and those for the L0En are

$$\omega^{\text{L0E}}_{\text{HLT}} = \frac{\epsilon^{\text{L0E data}}_{\text{HLT}}}{\epsilon^{\text{L0E MC}}_{\text{HLT}}} \ . \tag{6.13}$$

Note that the inclusive L0E category is used to determine HLT weights for the exclusive L0En due to its larger statistics, and because differences between the weights obtained using L0E and L0En are small. Histograms showing the estimated data and simulation efficiencies, and the sizes of the correction weights in bins of $B^0\,p_T$ are shown in Figure 6.4 for 2018 samples.

6.5.4 Reconstruction, kinematic and multiplicity corrections

The simulation of the kinematics of the B^0 mesons, and of the particle content of events, are known to show difference with respect to data. As the detector response is dependent upon quantities such as p_T and the event multiplicity, they should be corrected to resemble the data as much as it is possible to do so. These corrections are often made using a data-driven approach

systematic uncertainties, therefore the present thresholds are kept unchanged.

based on a sample of relatively pure signal candidates belonging to a control mode that closely resembles the characteristics of the decay of interest. This sample is selected from data, and its mass distribution is fitted to allow statistical background subtraction to be performed. The variables to be corrected are then binned in the same way for both simulation and data, and correction weights are extracted through simple division, taking weights used for background subtraction (and previous corrections) into account, i.e. for simulated events belonging to a given bin, the correction weight, $\omega_{\text{Kin}}^{\text{bin}}$, can be calculated as

$$\omega_{\text{Kin}}^{\text{bin}} = \frac{N_{\omega\,\text{bin}}^{\text{data}}}{N_{\omega\,\text{bin}}^{\text{MC}}} \,, \tag{6.14}$$

where $N_{\omega\,\text{bin}}^{\text{data}}$ and $N_{\omega\,\text{bin}}^{\text{MC}}$ refer to the weighted sum of data and simulated events in the bin, respectively. This histogram reweighting strategy has several disadvantages. If the corrections are made by binning the variables in one dimension, then important correlations will be lost. The use of multiple dimensions can lead to sparsely populated bins and unreliable weights if the simulation, or data samples are insufficiently large.

In this analysis, a multidimensional reweighting strategy based on a Boosted Decision Tree (BDT) is used [165]. This method is conceptually similar to the aforementioned histogram reweighting in that it similarly involves obtaining ratios between background subtracted data and simulation, but it enables the use of multiple variables whilst limiting the impact caused by increased dimensionality. This is achieved by using decision trees (DTs) to split the multidimensional space into several regions, the optimal boundaries of which are identified by maximising a symmetrised χ^2 quantity, defined as,

$$\chi^2 = \sum_{\text{region}} \frac{(N_{\omega\,\text{region}}^{\text{MC}} - N_{\omega\,\text{region}}^{\text{data}})^2}{N_{\omega\,\text{region}}^{\text{MC}} + N_{\omega\,\text{region}}^{\text{data}}} \,, \tag{6.15}$$

where $N_{\omega\,\text{region}}^{\text{MC}}$ and $N_{\omega\,\text{region}}^{\text{data}}$ are the sum of the weights of the simulated and data events in a given region. The value of the χ^2 is higher for regions where the differences between the sums are large. Hence, maximising this quantity effectively involves isolating regions where the simulation-data differences are the largest. In this way, the multidimensional space is not split up unnecessarily.

The BDT reweighting process uses an iterative approach. An iteration starts with the construction of a DT to maximise the χ^2. Next, correction weights are calculated for all regions and assigned to simulated events. Then, a new DT is introduced that attempts to maximise the χ^2 taking the previous weights into account. This procedure is repeated such that each additional DT in the sequence is optimised to reduce the remaining discrepancies in the previous step. The final BDT reweighter includes multiple DTs.

The inputs to the reweighter consist of control mode simulation and data candidates that pass all preselection requirements. Likelihood fits are performed for the constrained B^0 mass, $m(K\pi ee)_{\text{PV},\,J/\psi}$, separately for each year and bremsstrahlung category, the results of which are used to obtain weights for the statistical subtraction of backgrounds using the $s\mathcal{P}$lot [166] technique.[10] Corresponding control mode simulation samples are selected in the same way. The variables used include n_{Tracks}, and quantities related to the B^0 meson, specifically its p_T,

[10]The $s\mathcal{P}$lot approach is a statistical method that can be used to unfold signal and background contributions in data for variables of interest based on the result of a maximum likelihood fit to a discriminating variable (typically the invariant mass distribution), for which their distributions are known. An important requirement for the validity of this procedure is that the discriminating variable is uncorrelated with the variables of interest.

η, the logarithm of the significance of its flight distance[11], $\log(\chi^2_{\mathrm{FD}}\,(\mathrm{PV}))$, the logarithm of the significance of its impact parameter with respect to its associated PV, $\log(\chi^2_{\mathrm{IP}}\,(\mathrm{PV}))$, and the fit quality of its decay vertex, χ^2_{DV}. One dimensional projections of the corrected simulation and background subtracted data distributions are shown in Figure 6.5 for 2018 samples, together with the original, unweighted distributions. The quantities that underwent the largest changes include n_{Tracks} and $B^0\,p_T$, while the changes for the others are relatively small. Note that the distributions of the corrected simulation are modified by both trigger and kinematic correction weights.

[11]The flight distance (FD) refers to the distance between the origin and decay vertices of a particle.

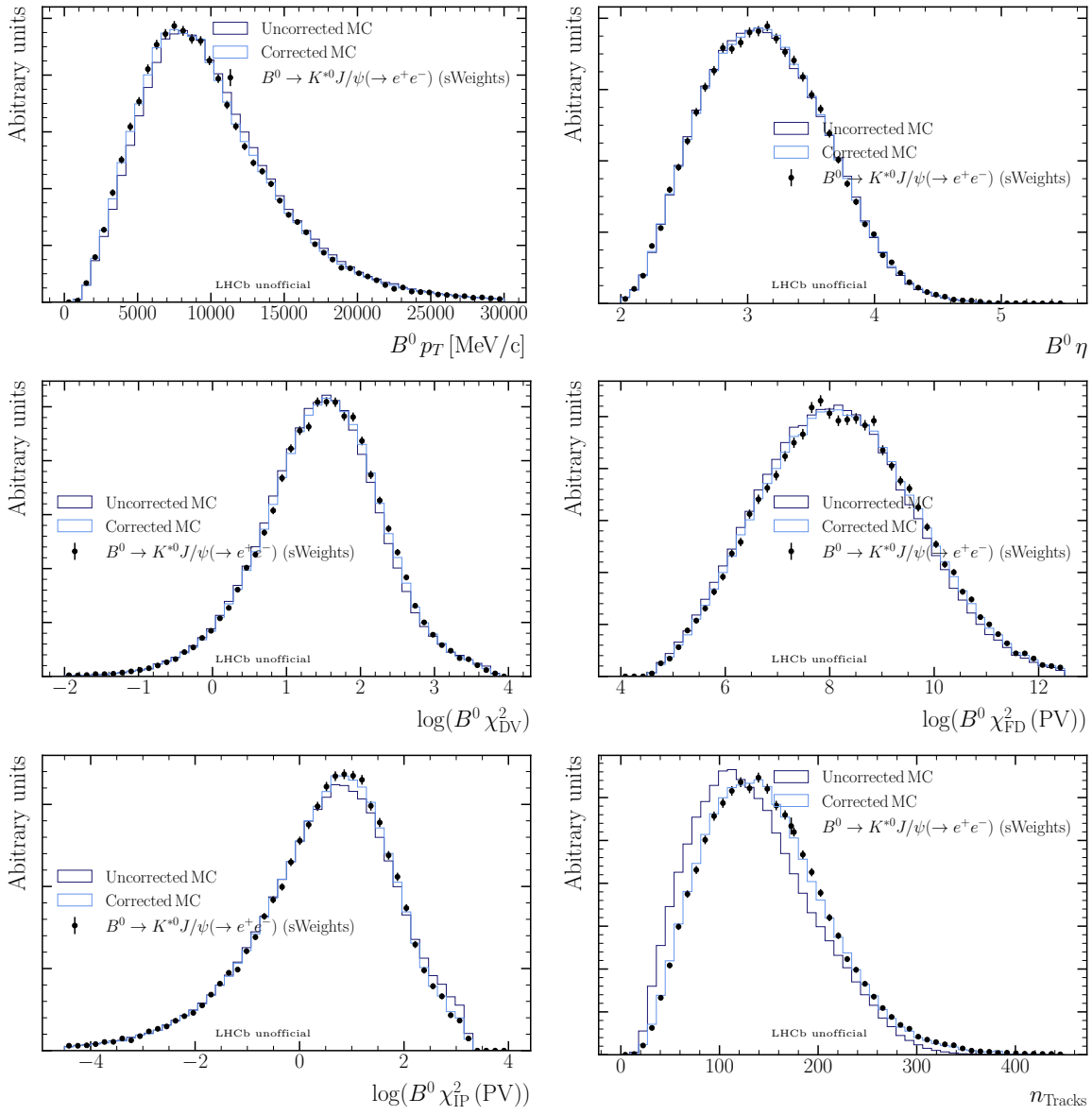


Figure 6.5: Distributions of variables involved in the BDT reweighting for 2018 samples. The distributions of background subtracted $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ candidates are shown in black, and the uncorrected and corrected simulation distributions are shown in dark and light blue, respectively.

7 Candidates selection

To determine the angular observables of the signal decay with good sensitivity, it is important to formulate a selection that is capable of separating signal from backgrounds effectively and efficiently. In addition, it needs to take detector limitations into account by removing candidates that are known to be poorly measured and improve alignment between samples whenever it is necessary to do so.

The isolation of signal candidates begins with the operation of the trigger system during data taking, and continues through the central offline stripping campaigns, which also divide larger samples into manageable sub-samples. Nevertheless, at this point they remain background dominated. To prepare them for analysis, a series of requirements ('preselection') are made based on variables related to track and event characteristics, PID, kinematics and decay topology, followed by the use of a dedicated multivariate classifier (MVA) to suppress combinatorial background. Selections against specific backgrounds (vetoes) are also applied. The choice of the selection requirements is based on that of existing analyses, in particular the $R_{K^{*0}}$ analysis [67], and previous studies of this mode [143], with modifications to tailor them to the current measurement.

The selections applied to post-stripping simulation and data samples are discussed in the following sections, starting with the preselection requirements in Section 7.1, followed by strategies used to control specific backgrounds in Section 7.2, the MVA and the determination of its optimal threshold in Section 7.3, the phase space cut to remove low efficiency regions in Section 7.4, and finally the removal of multiple candidates as a last step in Section 7.5.

7.1 Preselection

The preselection requirements are a generic set of cuts designed to reduce background, and to ensure good alignment between the main and auxiliary samples, such as the ones used to correct the PID response in simulation (Section 6.5). They can be broadly separated into three categories, which include cuts related to (sub-)detector acceptance and reconstruction quality, fiducial cuts for sample alignment, and cuts to suppress background.

High multiplicity events, i.e. events containing too many particles and interactions to be well reconstructed, are removed via cuts on the n_{SPD}. All tracks are required to pass basic quality cuts on the track χ^2/ndf variable, which is a measure of the goodness of the track fit. In addition, the output of a neural network [112], `GhostProb`, which can be interpreted as the probability of a track to be an artefact of reconstruction rather than one that is associated with a real particle, is used to suppress background. To improve the alignment between the phase space coverage of the calibration samples used for PID correction and that of the main data samples, cuts are made on the momenta and transverse momenta of all final state particles. In the case of the electrons, requirements are also imposed on the p_T of the dielectron pair, as well as the p_T of the individual electrons calculated without taking bremsstrahlung corrections into account ($p_{T,\,\text{track}}$). All tracks are required to have associated hits in the RICH detectors and electron tracks are also required to be within the acceptance of the ECAL and to have associated clusters in at least one of the calorimeters or the PS. As some cells in the inner region

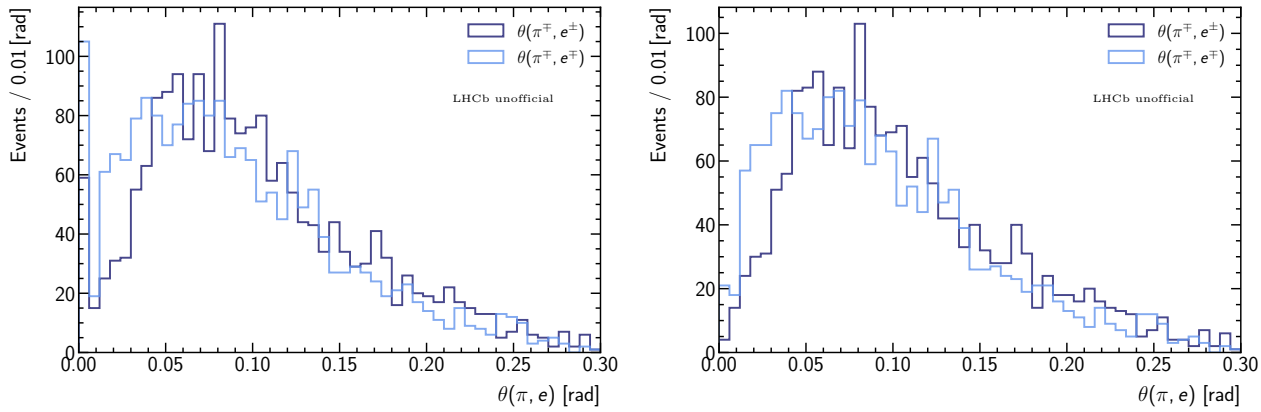


Figure 7.1: Distribution of the angle between the tracks of the pions and electrons of signal candidates in data that passed all preselection requirements with the exception of cuts to remove clone tracks (left), where a pathological peak can be seen at very low values of $\theta(\pi, e)$. This peak is removed following the application of the cut (right).

of the ECAL have not been read out properly during data taking [66], this region is removed from all samples to reduce simulation-data differences via the geometrical cut of

$$|\texttt{xProjection}^{\texttt{L0Calo}}_{\text{ECAL}}| < 363.6\,\text{mm} \;\text{ and }\; |\texttt{yProjection}^{\texttt{L0Calo}}_{\text{ECAL}}| < 282.6\,\text{mm}, \tag{7.1}$$

where `xProjection` and `yProjection` refer to the x and y dimensions of the ECAL planes.

Backgrounds related to particle misidentification are suppressed using both DLL and `ProbNN` variables (Section 4.2.4). Requirements on single DLL variables are imposed to reduce kaon-pion and electron-pion misidentification ($K\,\text{DLL}_{K\pi} > 0$ and $e\,\text{DLL}_{e\pi} > 2$), and the cut of $e\,\texttt{ProbNN}e > 0.2$ is made for electron tracks to reduce electron misidentification in general. In addition, combinations of DLL variables are also used to reduce backgrounds from events where the kaon and pion have been misidentified as each other ($K\,\text{DLL}_{K\pi} - \pi\,\text{DLL}_{K\pi} > 0$), i.e. $B^0 \to K_{\to\pi}\pi_{\to K}e^+e^-$, and combined `ProbNN` cuts are made to suppress pion-hadron (kaon and proton) misidentification ($\pi\,\texttt{ProbNNpi}\cdot(1-\pi\,\texttt{ProbNNk})\cdot(1-\pi\,\texttt{ProbNNp}) > 0.1$) and kaon-proton misidentification ($K\,\texttt{ProbNNk}\cdot(1-K\,\texttt{ProbNNp}) > 0.05$). Background suppression is also aided by a cut on the reconstructed mass of the K^{*0} candidate of $|m(K\pi) - m^{PDG}_{K^{*0}}| < 100\,\text{MeV/c}^2$, where $m^{PDG}_{K^{*0}}$ refers to the nominal mass of the neutral $K^{*0}(892)$ meson [7].

One specific type of background related to the reconstruction process is composed of 'clone tracks', which are tracks that share more than 70% of their total hits. These tracks may or may not carry the same charge, and may have different momenta values and PID hypotheses. The tendency of electrons to emit bremsstrahlung can lead to changes in the track direction, which causes additional complications for track reconstitution, and makes the creation of clone more likely. This background is reduced by the Clone Killer algorithm [110], which is ran at the end of the track reconstruction sequence. One important feature of these pairs is that the angular separation between them is typically close to zero. As such, it is possible to check for remaining contributions by calculating the angles between the tracks of all final state particles. As shown on Figure 7.1, the spikes at very small angles between some pion and electron tracks strongly suggests that clone tracks remain in the sample. Therefore, as a precaution, a cut of $\theta > 0.005\,\text{mrad}$ is applied to all track combinations.

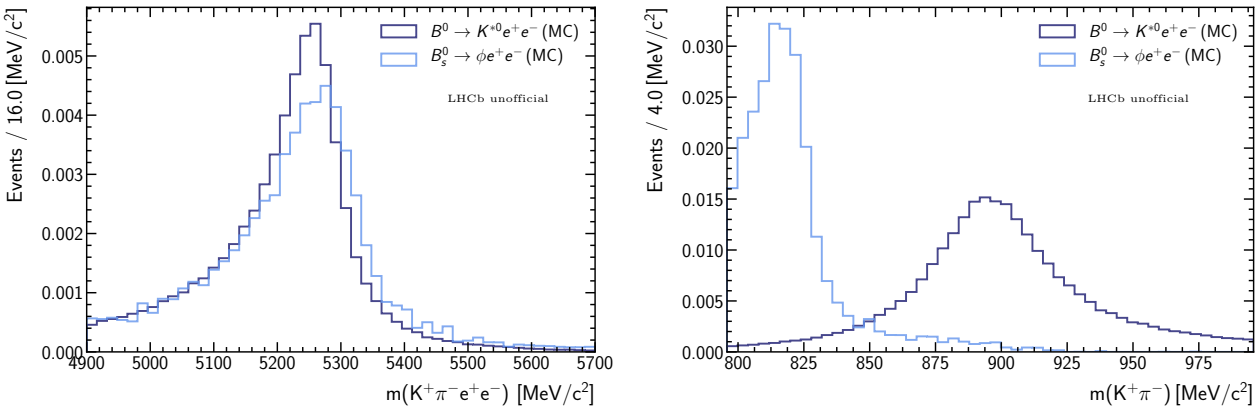


Figure 7.2: The $m(K\pi ee)_{\rm PV}$ and $m(K^+\pi^-)$ distributions of simulated $B_s^0 \to \phi e^+e^-$ candidates reconstructed as the signal, and simulated signal candidates that passed all preselection requirements with the exception of the dedicated veto against this background.

7.2 Specific backgrounds

After the preselection stage, dedicated cuts or 'vetoes' are applied to suppress backgrounds originating from specific decay modes. These include the decays of $B_s^0 \to \phi e^+e^-$, $B^+ \to K^+e^+e^-$, and semileptonic cascade decays such as $B^0 \to \bar{D}^0(\to K^+\pi^-)\pi^-e^+\nu_e$, $B^0 \to D^-(\to K^{*0}\pi^-)e^+\nu_e$, and $B_s^0 \to D_s^-(\to K^{*0}K^-)e^+\nu_e$. Although the use of the constrained q_c^2 variable strongly limits contributions from the control mode decay of $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (and the analogous decay involving a $\psi(2S)$ meson), events where the electron is misidentified as a hadron and vice versa may pass the usual requirements. A veto based on masses calculated with constraint on the J/ψ (or $\psi(2S)$) meson mass and changes in particle hypotheses is used to reduce its contribution. However, not all known backgrounds are controlled through vetoes. In some cases, no dedicated cuts are applied to backgrounds that are known to contribute significantly to the final data sample in order to model them in the angular fit. This applies to partially reconstructed and double semileptonic decays.

7.2.1 $B_s^0 \to \phi e^+e^-$ background

Candidates from the $B_s^0 \to \phi e^+e^-$ decay can be mistaken for signal if one of the kaons from the dominant $\phi \to K^+K^-$ mode is misidentified as a pion. The resulting kaon-pion combinations can be included within the K^{*0} mass window cut, and peak in the signal region, as shown in Figure 7.2. This background can be reduced by requiring that the invariant mass of the kaon-pion pair, calculated with the mass hypothesis of the pion changed to that of the kaon ($m(K\pi_{\to K})$), is greater than the nominal mass of the ϕ meson of 1040 MeV/c^2 [7]. To further improve signal efficiency, instead of rejecting all events with $m(K\pi_{\to K}) < 1040\,\text{MeV/c}^2$, only events belonging to this region, which also have pions with low probability of being pions, π `ProbNN`$\pi < 0.8$, are removed. The resulting veto has a signal efficiency of almost 100%, and rejects around $80 - 85\%$ of background events. The region removed is shown in red on Figure 7.3.

7.2.2 $B^+ \to K^+e^+e^-$ background

The decays of $B^+ \to K^+e^+e^-$ constitute a source of background if a random pion is combined with its decay products to form a signal candidate, or if the kaon is misidentified as a pion,

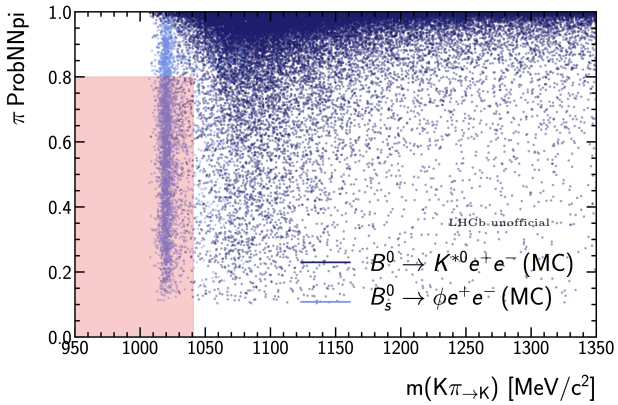


Figure 7.3: Distribution of simulated $B_s^0 \to \phi e^+ e^-$ and signal candidates passing all preselection requirements without the dedicated veto in π `ProbNNpi` and $m(K\pi_{\to K})$. The region removed by the veto is shown in red.

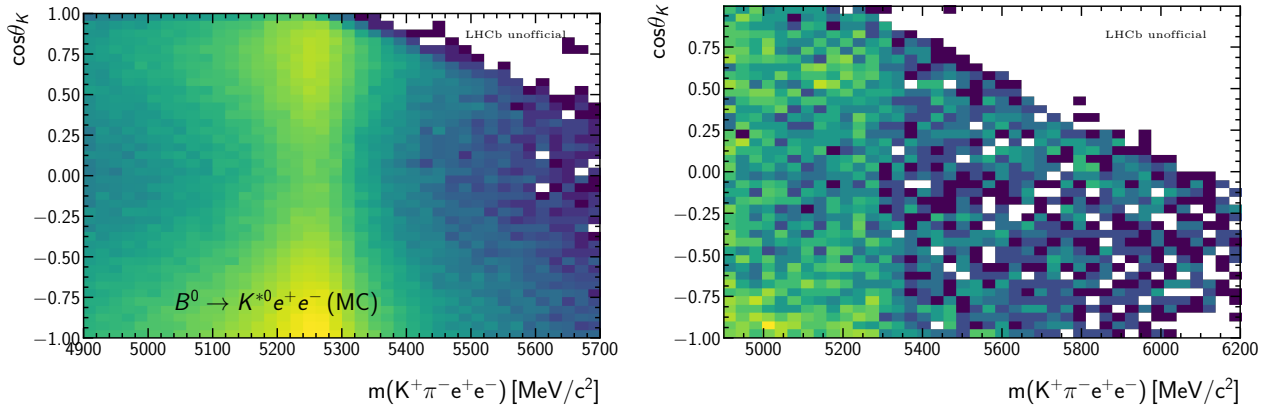


Figure 7.4: The impact of the veto on simulated and truth-matched signal candidates (left), and signal candidates in data with invariant masses well above the nominal invariant mass of the B^0 meson, which are mainly combinatorial in nature (right). In both cases, events within a triangular region at high mass and $\cos\theta_K$ near 1 are almost fully removed.

and a random kaon from the rest of the event is picked up. In either case, a fake K^{*0} is reconstructed. Due to this over-reconstruction, these events tend to form a peaking structure that is much broader than that of the signal in the upper mass region. Although the bulk of its distribution is far from the invariant mass of the B^0 meson, it can affect the determination of the shape and yield of the combinatorial background, and therefore affect the signal indirectly. To suppress this contribution, a veto based on the three-body invariant masses of

$$\max(m(K^+e^+e^-),\, m(e^+e^-\pi_{\to K})) < 5100\,\mathrm{MeV/c^2} \tag{7.2}$$

is imposed, which has a signal efficiency of around 98%. This background suppression strategy comes at the cost of introducing correlations between the reconstructed B^0 mass and $\cos\theta_K$, as illustrated in Figure 7.4, which is considered as a source of systematic uncertainty (Section 9.2.5).

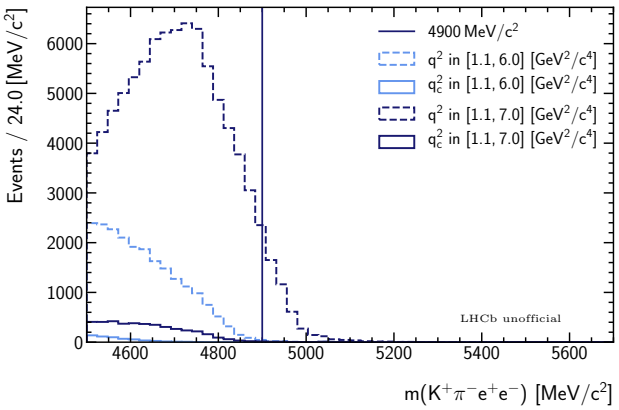


Figure 7.5: Distribution of preselected simulated $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ events with cuts on the unconstrained and constrained q^2 variables.

7.2.3 Charmonium contributions

Decays featuring charmonium resonances, in particular $B^0 \to K^{*0} J/\psi(\to e^+e^-)$, which is used as the control mode due to its comparable topology and kinematics as well as its large branching fraction, can be a significant source of background for these very reasons. While the signal q^2 window is chosen to avoid regions where they dominate, resolution effects can broaden their distributions and lead to leakage into the measurement region. The use of the q^2_c strongly suppresses this contribution, as does the choice of a restricted mass window. However, the constraints cause background events that contain true J/ψ mesons but are otherwise of combinatorial nature to be shifted into the signal region. Additionally, charmonium decays with misidentification, especially where an electron is identified as a hadron and a hadron is identified as an electron ('electron-hadron swaps') generally escape selection based on the q^2_c. Although they are strongly suppressed by PID requirements, they can be further reduced using cuts based on constrained masses. More details on charmonium backgrounds and their treatments are given in the sections below.

Charmonium decays

Control mode candidates have non-trivial angular distributions, and may therefore give rise to large systematic uncertainties if left uncontrolled. While they are mainly located close to $m^2_{J/\psi} = 9.6\,\mathrm{GeV}^2/\mathrm{c}^4$ [7], they tend to leak into the signal q^2 window. This 'J/ψ leakage' background is strongly suppressed by using the constrained rather than the unconstrained q^2 variable, as illustrated in Figure 7.5. The signal mass window can also be chosen to start at $4900\,\mathrm{MeV}/\mathrm{c}^2$ in order to avoid most of the remaining events. In addition, the phase space cut to remove low-efficiency regions (Section 7.4) removes the edges of $\cos\theta_\ell$, where this distribution tends to peak, leading to further reduction. Due to the low expected yield of this background of around 12 (4) events in the larger (smaller) q^2_c ranges, it is not included as a component in the angular fit. The impact of this choice is assessed as a source of systematic uncertainty in Section 9.2.3.

Combinatorial-like components

Using the q^2_c tends to shift combinatorial background in the low mass region away from the signal q^2_c window, but in the high mass region they can be shifted into it. This is in general not a

problem for contributions without distinctive q^2 features, as the resulting distribution typically remains smooth and exponential to a good approximation. However, this is not the case for combinatorial backgrounds that contain true J/ψ candidates. These events are concentrated within a relatively narrow q^2 region. The application of the constraint forces them to enter the signal q_c^2 region at high B^0 invariant masses, resulting in the formation of a broad peaking structure. This effect is illustrated in Figure 7.6 and 7.7 using simulation candidates that are truth-matched to be combinatorial in nature. This background (along with the analogous contribution from $\psi(2S)$ decays) is also clearly visible in the data sample when the nominal MVA cut is loosened (Figure 7.8). However, this feature cannot be seen in the LFV $K^+\pi^-e^+\mu^-$ or the leptonic same-sign sample ($K^+\pi^-e^\pm e^\pm$), as the combinations of $e^\pm e^\pm$ and $e^+\mu^-$ cannot come from true J/ψ decays. A comparison between their q_c^2 distributions in the upper mass region is shown in Figure 7.9.

To obtain an approximate estimation of the number of residual events, a partially data-driven approach is used. First, a fit is made to the q_c^2 distribution of the data in the upper mass sideband (above 5700 MeV/c^2) using a model for the standard combinatorial component obtained from same-sign data and an exponential distribution that is allowed to vary. The resulting yield is scaled to provide the expected value within the signal mass window based on the ratio taken from simulation. This is found to be 54±8 events. Note that this value is approximate due to the limited capability of the simulation to replicate the characteristics of combinatorial background. Nevertheless, as its impact may not be negligible, an attempt is made in Section 9.2.4 to provide a conservative estimate of the associated systematic uncertainties.

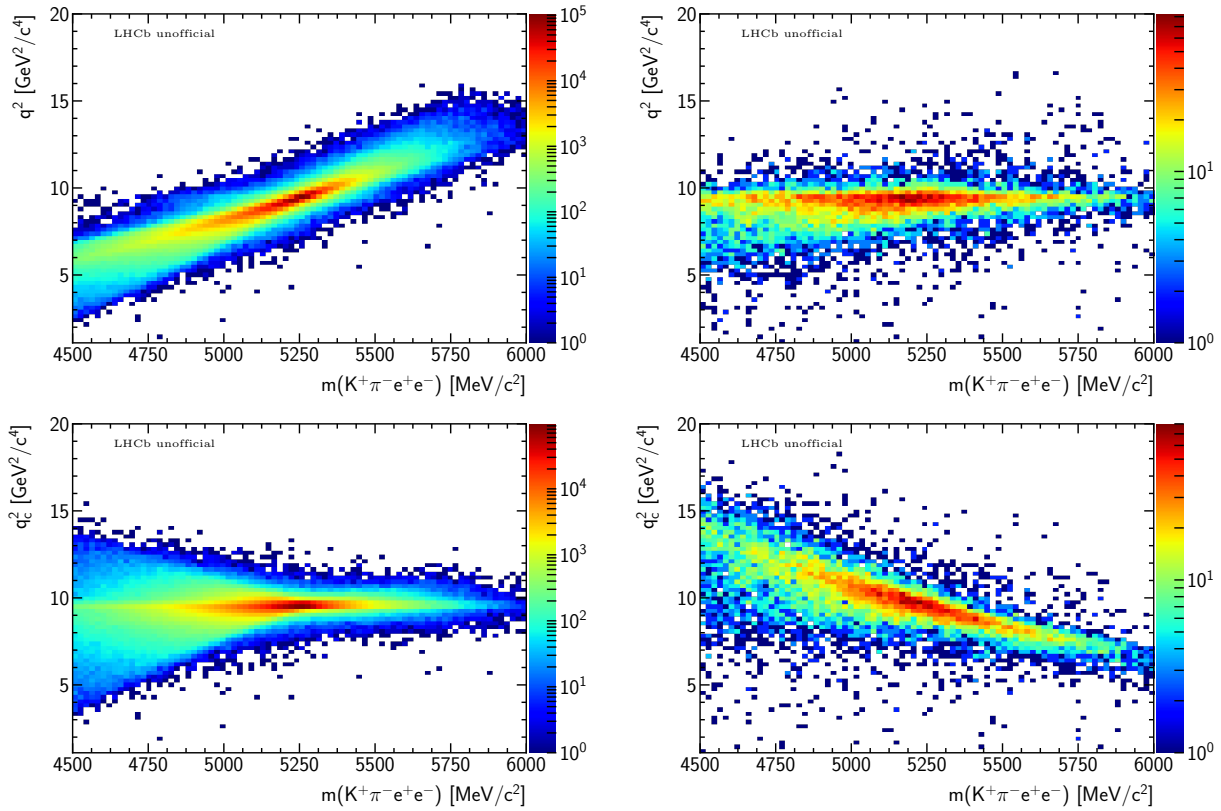


Figure 7.6: Distribution of simulated control mode candidates in $m(K\pi ee)_{\rm PV}$ and unconstrained q^2 (top row), and q_c^2 (bottom row). Correctly reconstructed signal candidates are shown on the left hand side, while incorrectly reconstructed combinatorial candidates are shown on the right.

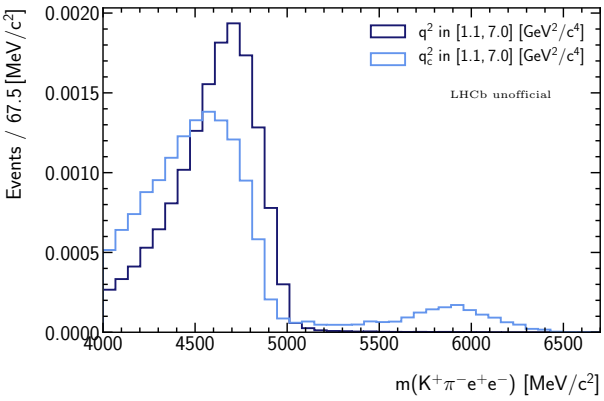


Figure 7.7: Distributions of $m(K\pi ee)_{\mathrm{PV}}$ for correctly and incorrectly reconstructed simulated $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ candidates after the application of a cut on either the unconstrained q^2, or the constrained q_c^2. In the case of the latter, the combinatorial candidates that contain true J/ψ candidates form a broad peak in the upper mass region. Note that the nominal MVA cut has been removed to make this contribution visible.

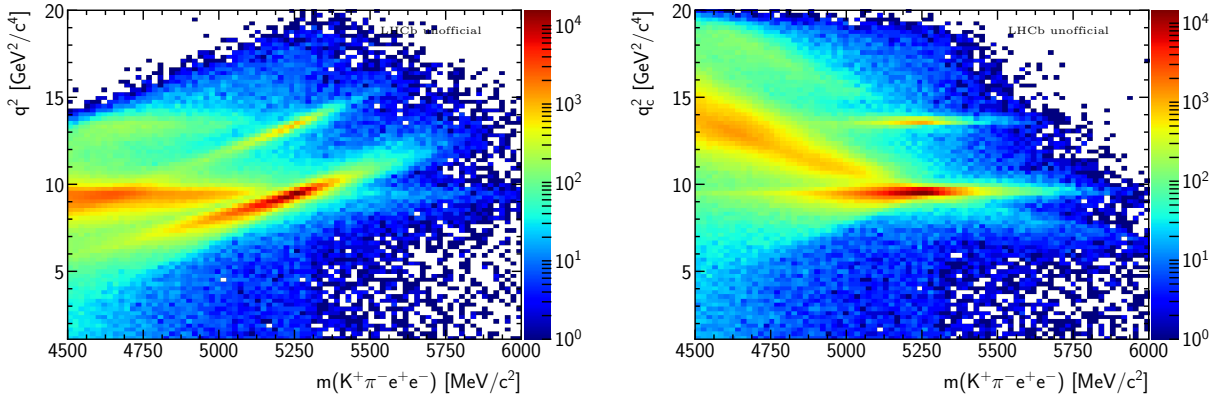


Figure 7.8: Distribution of signal candidates in data in $m(K\pi ee)_{\mathrm{PV}}$ and unconstrained q^2 (left), and q_c^2 (right). The additional band at $\sim 13.5\,\mathrm{GeV}^2/\mathrm{c}^4$ is due to contributions from the $\psi(2S)$ meson, which is expected to have negligible impact in the measurement region.

Electron-hadron swaps

Control mode candidates that feature two instances of misidentification ('swap'), e.g. with both $e \to h$ and $h \to e$, where $h = \pi$, K, are relatively rare. However, due to the large branching fraction of the control mode, such contributions are not entirely negligible. Furthermore the misidentification can allow them to escape the otherwise very effective cut on the constrained q_c^2. While cuts can be made on the B^0 masses calculated with changes in mass hypotheses, i.e. $m(K_{\to e}\pi e_{\to K} e)$ and $m(K\pi_{\to e} e e_{\to \pi})$, these vetoes tend to be inefficient due to the limited resolution. Instead, alternative variables are calculated by applying J/ψ (and $\psi(2S)$) mass constraints to the $K_{\to e} e_{\to K}$ and $\pi_{\to e} e_{\to \pi}$ pairs using the method of Lagrange multipliers [167, 67]. This improves the resolution of true swap candidates, allowing them to be vetoed more efficiently. To further increase signal efficiency, a PID requirement is added, and candidates with

$$|m(K_{\to e}\pi e_{\to K} e)_{J/\psi,\ \psi(2S)} - m^{PDG}(B^0)| < 60\,\mathrm{MeV/c^2} \text{ and } e\,\texttt{ProbNNe} < 0.8 \qquad (7.3)$$

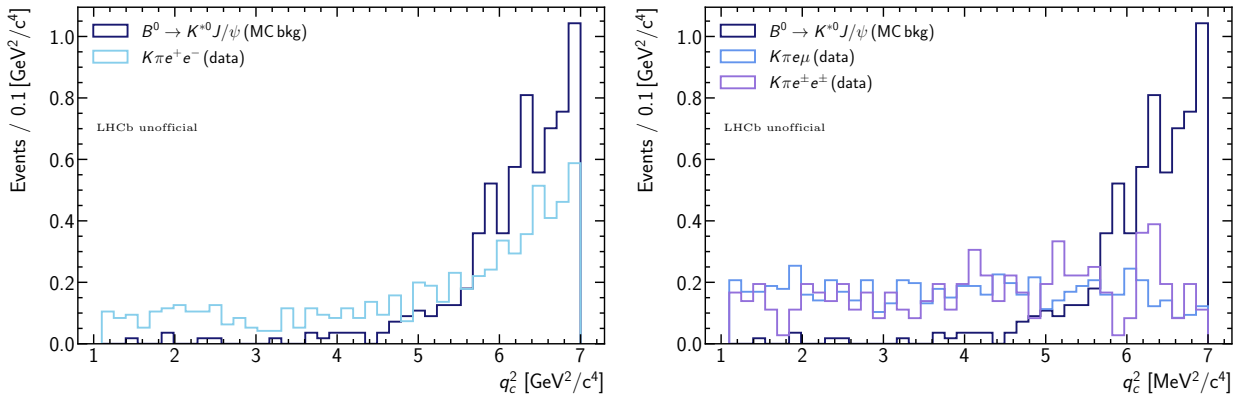


Figure 7.9: The q_c^2 distribution of combinatorial candidates in the simulated $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ samples after preselection cuts and with a loosened cut on the multivariate classifier and that of the (opposite-sign) data (left), and $K\pi e\mu$ and same-sign ($K^+\pi^- e^\pm e^\pm$) data (right). The peaking structure at high q_c^2 is only seen in cases where the two leptons can come from the decay of a real J/ψ.

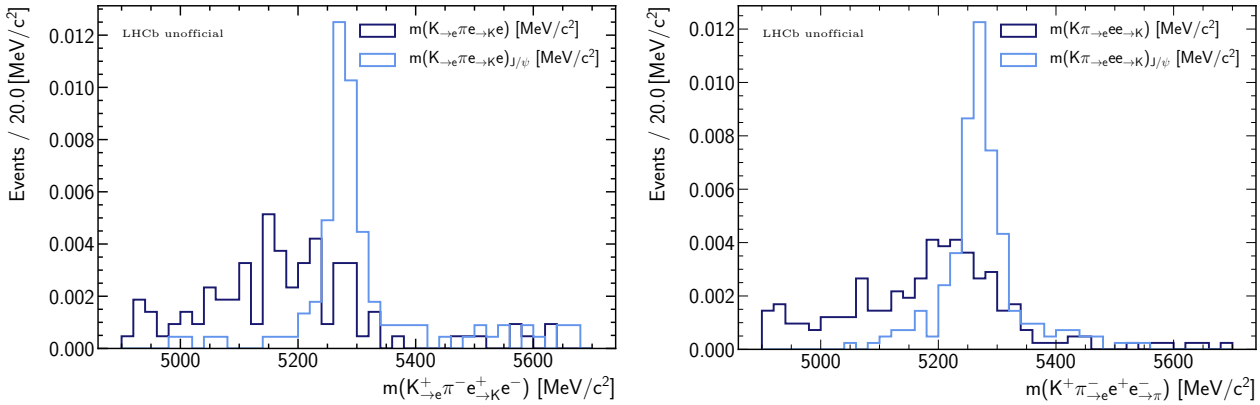


Figure 7.10: Mass distributions of simulated control mode candidates with kaon-electron swap (left) and pion-electron swap (right) calculated with the corrected mass hypotheses with and without constraints on the J/ψ mass.

and

$$|m(K\pi_{\to e}ee_{\to\pi})_{J/\psi,\ \psi(2S)} - m^{PDG}(B^0)| < 60\,\text{MeV/c}^2 \text{ and } e\,\texttt{ProbNNe} < 0.8\ , \qquad (7.4)$$

are removed. Together, these vetoes have a signal efficiency of around 98%, as estimated using truth-matched signal events that passed all other selection criteria. They are able to reject around 70% of control mode background events with double misidentification, as estimated using simulation samples for which PID requirements have been loosened for statistics.

7.2.4 Signal kaon-pion swaps

The mass distribution of signal candidates with kaon-pion misidentification resembles that of the correctly reconstructed ones, such that the two cannot be separated effectively. It is the largest source of misidentified signal decays in the simulation after preselection, and tends to have non-trivial angular distributions. To veto these events, the $\text{DLL}_{K\pi}$ of the kaon is required

to be larger than the $\mathrm{DLL}_{K\pi}$ of the pion,

$$K\,\mathrm{DLL}_{K\pi} - \pi\,\mathrm{DLL}_{K\pi} > 0\ . \tag{7.5}$$

This veto has a signal efficiency of around 99% for events passing the preselection requirements, and is able to reject around 70% of this background.[1] Control mode candidates with hadron swap are strongly suppressed by the q^2_c cut as the dielectron system is unaffected. Therefore, background from this source can be considered negligible for the rare mode.

7.2.5 Semileptonic decays with $h \to e$ misidentification

Semileptonic decays such as $B^0 \to (D^{*-} \to (\bar{D}^0 \to K^+\pi^-)\pi^-)e^+\nu_e$, and $B^0 \to (\bar{D}^0 \to K^+\pi^-)\pi^- e^+\nu_e$ can be reconstructed as signal when a pion is misidentified as an electron. Kaon-electron misidentification can also lead to the reconstruction of decays such as $B^0_s \to D^-_s(\to K^{*0}K^-)e^+\nu_e$ as signal. To reduce contributions from these modes, which have large branching fractions compared to the signal, vetoes are designed based on the removal of candidates that lie within a window of $30\,\mathrm{MeV/c^2}$ around the nominal masses of the D^0 and $D^\pm$ mesons. The invariant mass of the $K\pi$ and $K\pi e$ systems, calculated with changes in mass hypotheses, are used. To improve signal efficiency, a PID requirement is added such that candidates within the designated regions are only removed if the electrons involved also have low probabilities of being true electrons. More precisely, the veto corresponds to the removal of candidates that satisfy any of the following conditions

$$\begin{aligned}
|m(Ke_{\to\pi}) - m^{PDG}(D^0)| < 30\,\mathrm{MeV/c^2} \text{ and } e\,\texttt{ProbNNe} < 0.8\ ,\\
|m(K\pi e_{\to\pi}) - m^{PDG}(D^-)| < 30\,\mathrm{MeV/c^2} \text{ and } e\,\texttt{ProbNNe} < 0.8\ ,\\
|m(K\pi e_{\to K}) - m^{PDG}(D^-_s)| < 30\,\mathrm{MeV/c^2} \text{ and } e\,\texttt{ProbNNe} < 0.8\ ,
\end{aligned} \tag{7.6}$$

which has a high signal efficiency of around 99%.

7.2.6 Partially reconstructed decays

The partial reconstruction of multibody decays gives rise to a background that populates the lower mass region. Examples include decays such as $B^+ \to K^+_J e^+e^-$, where various excited kaon states ($K^+_J = K_1(1270)^+, K^*_2(1430)^+, ...$) can decay to the $K^+\pi^+\pi^-$ final state through a number of different intermediate resonances. For the control mode, this type of background is strongly suppressed by removing events for which the constrained $m(K\pi ee)_{\mathrm{PV,\,J/\psi}}$ is less than $5150\,\mathrm{MeV/c^2}$. While it is also possible to reduce it for the rare mode by cutting on the mass of the B^0 candidate calculated with a correction factor ('HOP mass'[2]) [66], or via multivariate techniques [143], the approach chosen for this analysis is to model it using a data-driven approach (Section 8.2.1). This choice is motivated by the limited effectiveness of

[1] This veto is not tightened to avoid reducing signal efficiency (the expected contribution of this background to the data sample after the existing cut is negligibly small).

[2] The HOP approach makes use of the fact that, barring bremsstrahlung and measurement effects, the sum of the momenta of the final state particles is not expected to produce a component that is orthogonal to the momentum vector of the B^0 candidate. This does not necessarily hold for partially reconstructed decays (or incorrect reconstructions in general), and can therefore be exploited to produce a variable with signal-background separation power. The HOP mass is calculated for the B^0 candidate by applying a correction factor of $\alpha_{\mathrm{HOP}} = p_T(K^{*0})/p_T(e^+e^-)$ to the dielectron momentum, *i.e.* $p_{\mathrm{corr}}(e^+e^-) = \alpha_{\mathrm{HOP}}\,p(e^+e^-)$.

both approaches in improving the signal-to-background ratio in the signal region (the reduction occurs primarily in the low mass region, which has limited impact on the signal sensitivity). In addition, both methods introduce distortions, which complicate the separation and modelling of the double semileptonic and combinatorial components.

7.2.7 Double semileptonic background

Double semileptonic (DSL) decays such as $B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)e^-\bar{\nu}_e)e^+\nu_e$ can be problematic as they have the same visible final state particles as the signal and large branching fractions of $\mathcal{O}(10^{-4})$ [7]. Due to energy loss from the neutrinos, their reconstructed masses peak in the low mass region, such that in the signal mass window of $4900 - 5700\,\mathrm{MeV/c^2}$, they tend be approximately exponentially distributed like the combinatorial background. However, they generally have a distinctive, asymmetric $\cos\theta_\ell$ distribution peaking at $\cos\theta_\ell = 1$. This is due to the imbalance in momentum between the electron from the decay of the D meson and the one from the B^0, and the definition of θ_ℓ (Section 2.3), which means that only the higher momentum positron (electron) from the B^0 ($\bar{B}^0$) meson is used. However, if the B^0 meson is incorrectly identified as its CP conjugate and vice versa due to, for example, kaon-pion swap, incorrect charge assignments or the incorrect combination of leptons from a true DSL decay with a random K^{*0}, a peak can form near $\cos\theta_\ell = -1$ instead. Decay modes where the K^{*0} ($\bar{K}^{*0}$) are associated with the $\bar{B}^0$ (B^0), which is the opposite of most DSL modes, an example of which is $B_s^0 \to D_s^-(\to \bar{K}^{*0}e^-\bar{\nu}_e)e^+\nu_e$, will also lead to a peak near $\cos\theta_\ell = -1$, although these tend to have low branching fractions. While in principle this feature can be used to formulate a veto, and indeed a cut of this type, namely $|\cos\theta_\ell| < 0.8$ is used in the older $R_{K^{*0}}$ analysis [66], the DSL simulation is found to only approximately describe this component in data, and non-negligible tails may still leak into the measurement region and affect observables that are sensitive to $\cos\theta_\ell$ asymmetry. In an attempt to gain greater control, instead of applying a veto, it is modelled using a data-driven approach, which is described in detail in Section 8.2.1.

7.3 Multivariate analysis

Besides backgrounds from specific decays, preselected data samples contain large numbers of combinatorial candidates, which do not have distinctive features that can be used to formulate simple vetoes. Therefore, the strategy to reduce it involves the use of a machine learning algorithm to classify candidates as signal or background based on multiple variables with distinguishing power. To this end, an optimised gradient boosting library, xgboost [168], is used through the Reproducible Experiment Platform [169], with additional functionality and tools provided by the Rare Decays package [165] and Scikit-Learn [170]. The classifier can be trained using samples that are representative of the expected characteristics of the signal and combinatorial background in data, such that it will be able to produce an output for each candidate that ranges from zero to one, with values close to zero (one) indicating that the candidate is likely to be background (signal). An optimisation procedure that involves repeated trainings using varied configurations is ran to determine the optimal set of input features and classifier settings ('hyperparameters') [143]. A cut is made on the output of the final classifier to best separate signal from background, and as full separation is not possible, the choice of this threshold is determined through another optimisation procedure using sensitivity to P_5' directly as the figure of merit. In the following sections, relevant information on the classifier algorithm is provided, followed by a summary of the optimisation procedure to determine the best input

Table 7.1: Summary of the input features used in the training of the multivariate classifier [143].

Particle	Variables
B^0	p_T, $\chi^2_{\text{IP}}\,(\text{PV})$, $\chi^2_{\text{FD}}\,(\text{PV})$, χ^2_{DV}, χ^2_{DTF}, DIRA
K^{*0}, J/ψ	χ^2_{DV}
h	$\min(K\,p_{\text{T}}, \pi\,p_{\text{T}})$, $\min(K\,\chi^2_{\text{IP}}\,(\text{PV}), \pi\,\chi^2_{\text{IP}}\,(\text{PV}))$
ℓ	min, $\max(e^+\,p_{\text{T}}, e^-\,p_{\text{T}})$, min, $\max(e^+\,\chi^2_{\text{IP}}\,(\text{PV}), e^-\,\chi^2_{\text{IP}}\,(\text{PV}))$

variables and hyperparameters, and a more detailed description of the strategy used to find the optimal cut on the classifier output.

7.3.1 Input features

The set of input features used by the multivariate classifier is selected from a list of kinematic, topological, and PID variables that are expected to provide signal-background separation power using a search procedure detailed in [143], which involves the training of multiple classifiers using subsets of the considered variables to find the combination that results in the best performance. The optimal set of fourteen features is shown in Table 7.1. It includes kinematic and topological variables of the B^0 candidate, namely its p_T, $\chi^2_{\text{IP}}\,(\text{PV})$, $\chi^2_{\text{FD}}\,(\text{PV})$, χ^2_{DV}, DIRA, and the χ^2_{DTF} of the PV constrained kinematic fit of the decay chain, which provides information on the fit quality. The χ^2_{DV} of the intermediate K^{*0} and that of the two electrons are also used. The lowest p_T value of either the kaon or the pion is included, as is the lowest $\chi^2_{\text{IP}}\,(\text{PV})$ of the two. For the electron pair, both the lowest and the highest p_T value of the two is used, along with both the lowest and the highest $\chi^2_{\text{IP}}\,(\text{PV})$. These features make use of characteristics of the signal decay – either its kinematics or the way it takes place – to separate it from the background based on random combinations. The distributions of these input features of the signal and background samples are shown in Figure 7.11.

7.3.2 Signal and background samples

The signal sample used in the training of the multivariate classifier is taken from fully corrected $B^0 \to K^{*0}e^+e^-$ simulation, and the background sample comes from the upper mass side-band in data, i.e. data candidates with reconstructed B^0 masses greater than $5600\,\text{MeV/c}^2$. Candidates in both samples must satisfy the preselection criteria, with a few exceptions made to increase the statistics of the background sample. These include modifications to PID requirements, namely the loosening of the ProbNNe requirements for electrons from greater than 0.2 to 0.05, the loosening of the kaon $\text{DLL}_{K\pi}$ requirement from greater than 0 to -5, and the loosening of the electron $\text{DLL}_{e\pi}$ cut from greater than 2 to 0. The $K\pi$ invariant mass window is enlarged from $100\,\text{MeV/c}^2$ to $200\,\text{MeV/c}^2$ around the nominal K^{*0} mass, and the lower bound of the q^2_c range is decreased to $0.1\,\text{GeV}^2/\text{c}^4$. Vetoes against semileptonic decays with $h \to e$ misidentification, cuts against $h \to e$ swap background and clones are also removed, together with additional quality cuts that tend to have little impact. In addition, the simulation candidates are required to fulfil the most stringent truth-matching criteria possible $(\textit{Signal})^3$. Separate sets of signal

[3]This choice is made due to the large sizes of these samples, and the straightforward appeal of maximising the selection efficiency for correctly reconstructed signal events (that did not emit sizeable bremsstrahlung),

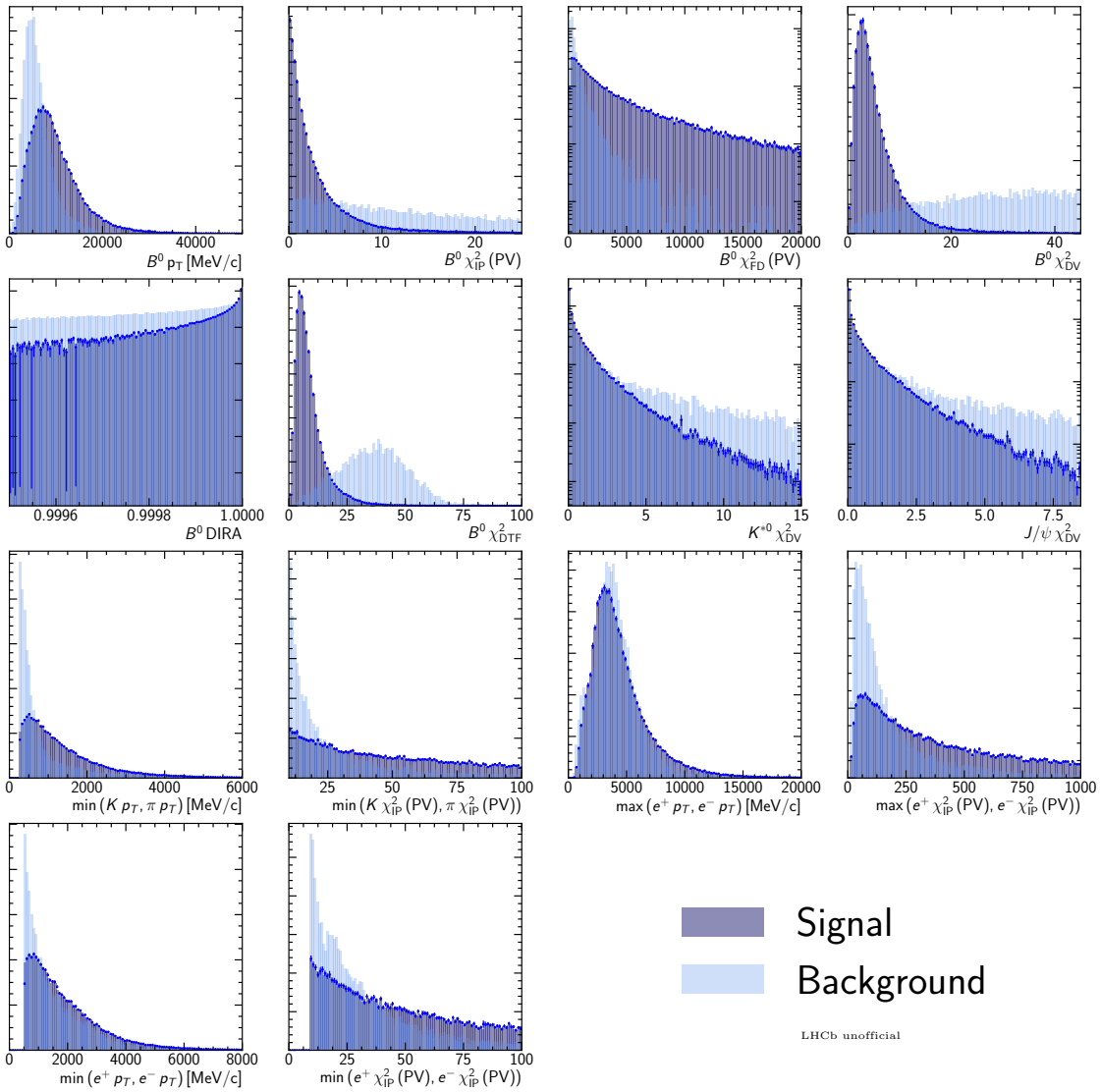


Figure 7.11: Distributions of the variables used in the MVA for the Run 2p2 signal and background training samples.

and background samples are produced to train different classifiers for the three Run periods. This is done to address potential differences due to the changes in the operating conditions of the detector.[4] The statistics of these samples are given in Table 7.2.

7.3.3 Training and results

The classifiers are trained using the k-folding approach, which allows for the full use of the training samples. First, signal and background samples are split randomly into a given number (N) of subsamples of approximately equal size. Next, a sub-classifier is trained using $N-1$

which is the contribution that is most closely described by Equation 2.18. In any case, the impact of including the other categories (*Low-mass background* and *Ghost*) is marginal.

[4] In principle, a more optimal strategy would be to also consider the L0 categories separately, as the method of triggering the event affects background characteristics. However, this would require splitting the already low-statistics background training samples further, which risks increased overtraining (and reduced performance), and is therefore not implemented.

Table 7.2: Statistics of the signal and background samples used in training.

Sample	Run 1	Run 2p1	Run 2p2
Signal	57518	122655	99308
Background	3418	6359	11910

subsamples and used to make predictions on the subsample that has been left out. This process is repeated until N sub-classifiers have been trained. The resulting group of sub-classifiers constitutes the classifier for a particular Run period. To make predictions for events that have not been used in training, the result of one sub-classifier is taken at random (and when the event has been used, the prediction from the sub-classifier that did not use it in training is taken). The number of folds, N, is chosen to be 10 based on studies documented in [143].

The performance of each sub-classifier can be checked based on its ability to separate signal from background in the subsample that has not been used in its training. A common method is to use the Receiver Operator Curve (ROC), which is a plot of the true positive rate, i.e. the fraction of true signal correctly classified as signal, against the false positive rate, or the fraction of background incorrectly classified as signal, for different cut-off values of the classifier output. In general, larger areas under the curve (ROCAUC) are indicative of better performance. The ROC curves of the Run 2p2 sub-classifiers shown in Figure 7.12 suggest that good signal-background separation can be achieved using the chosen configuration.

Another check that is commonly used involves comparing the distributions of the classifier outputs for the samples that have been used in training, and ones that have not been (test samples). Often referred to as an 'overtraining' check, its main purpose is to provide indications for the presence or absence of overtraining, which occurs when a classifier adapts itself to the specific characteristics (e.g. statistical fluctuations) of its training samples, and becomes less capable of generalising to an unrelated sample as a result. Over-training can occur due to a combination of hyperparameter choice and training sample statistics, in an analogous fashion to the overfitting of a sample by using a set of free parameters that is too large relative to the statistics available. The comparison between the train and test distributions of the Run 2p2 classifier is shown on Figure 7.13 as an example. A small but significant difference can be seen between the two especially near the score of unity, which is indicative of slight overtraining. This is not unexpected, as the background samples are statistically limited, and the line between achieving optimal performance and overtraining can be a fine one. As the classifier performance (evaluated on test samples) is good, and additional checks do not indicate any noteworthy pathological effects, its configuration is not changed from the optimal setting found based on the work in [143].

7.3.4 Response uniformity

While care has been taken to avoid input variables that are strongly correlated with key quantities of the analysis, namely the reconstructed mass of the B^0 meson, the angles and q^2, it is still possible for the classifier to 'learn' about them indirectly, such that its output may not be uniform over these quantities. Uniformity is in general desirable as the sculpting of the background distributions, in particular the mass, can complicate signal-background separation. For example, a classifier that has learned that the signal is more likely to be peak in a particular mass region may sculpt the distribution of the background in a non-uniform way (e.g. create an artificial peaking structure). Limited shaping of the angular and q^2 distributions of the signal

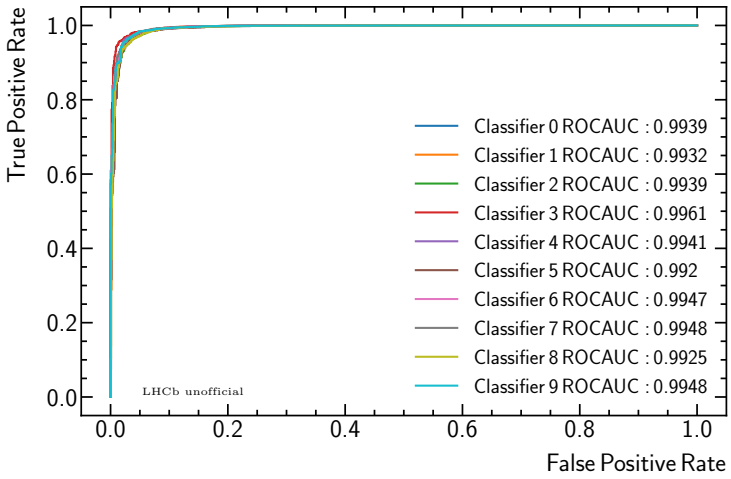


Figure 7.12: ROC curves for the ten sub-classifiers trained using Run 2p2 samples. In all cases the ROCAUC values are high, which are indicative of good separation power.

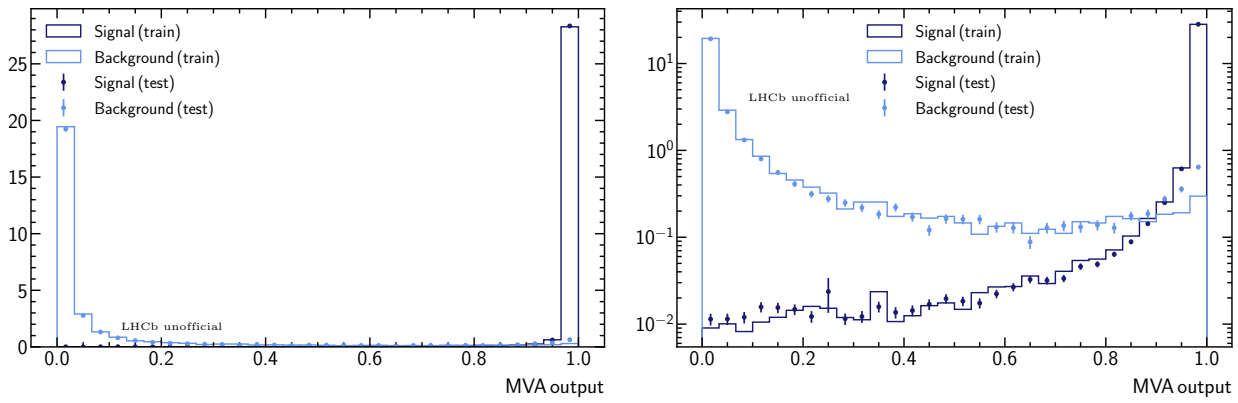


Figure 7.13: Distributions of the MVA scores of the Run 2p2 signal and background samples (left), and the same distributions shown with a logarithmic scale (right). The 'train' scores refer to the values obtained from sub-classifiers that used the event in training, and the 'test' scores are obtained from the sub-classifier that did not use the event in training.

pose fewer problems as it can be corrected for by the effective acceptance functions (Section 8.1). Its impact on the signal mass distribution can also be taken into account in the models used. Nevertheless large effects can be associated with increased systematic uncertainties.

The uniformity of the classifier output is checked by calculating the fractions of signal and background events that pass a set of thresholds that lie at the 20th, 40th, 60th and 80th percentiles of the output score distributions for the corresponding samples. In the case of signal sample, simulation-data correction weights used in training are included, although their impact is small. The calculations are performed in bins of the variables under consideration, where the bin edges are chosen to include approximately equal numbers of events in each bin. For the background sample, a sanity cut of greater than 0.1 is applied to remove very background-like events that do not represent the bulk of the combinatorial background relevant to the analysis. The results are shown for the Run 2p2 classifier as a representative example in Figure 7.15. In general, while shaping effects are visible for the signal $\cos\theta_\ell$, and $\cos\theta_K$ distributions, they remain limited. However, clear non-linear behaviour can be seen in the signal and background mass distributions. For the signal, this is not a significant problem, as its mass models are obtained after the classifier output cut and therefore take this effect into account. The shaping of the background distribution is generally smooth, with no indications of sizeable peaking effects. This is further confirmed by the same-sign sample in the nominal mass region as shown in Figure 7.16. Therefore, standard exponential distributions (Section 8.2.1) are expected to be able to accommodate these effects.

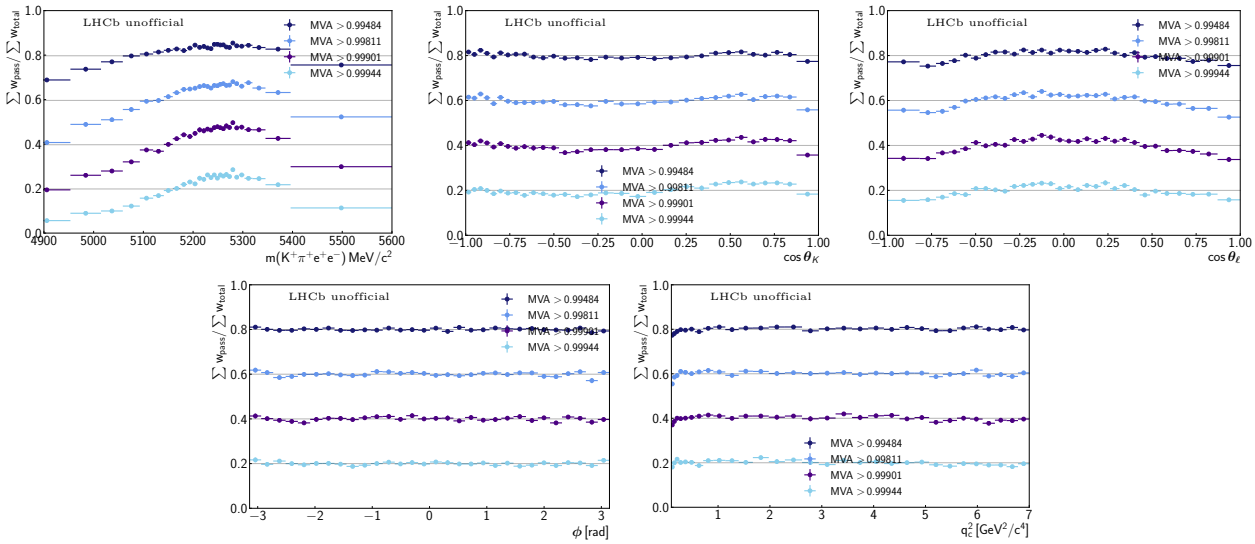


Figure 7.14: Efficiency as a function of $m(K\pi ee)_{\rm PV}$, angles and q_c^2 for the Run 2p2 classifier evaluated using the signal samples.

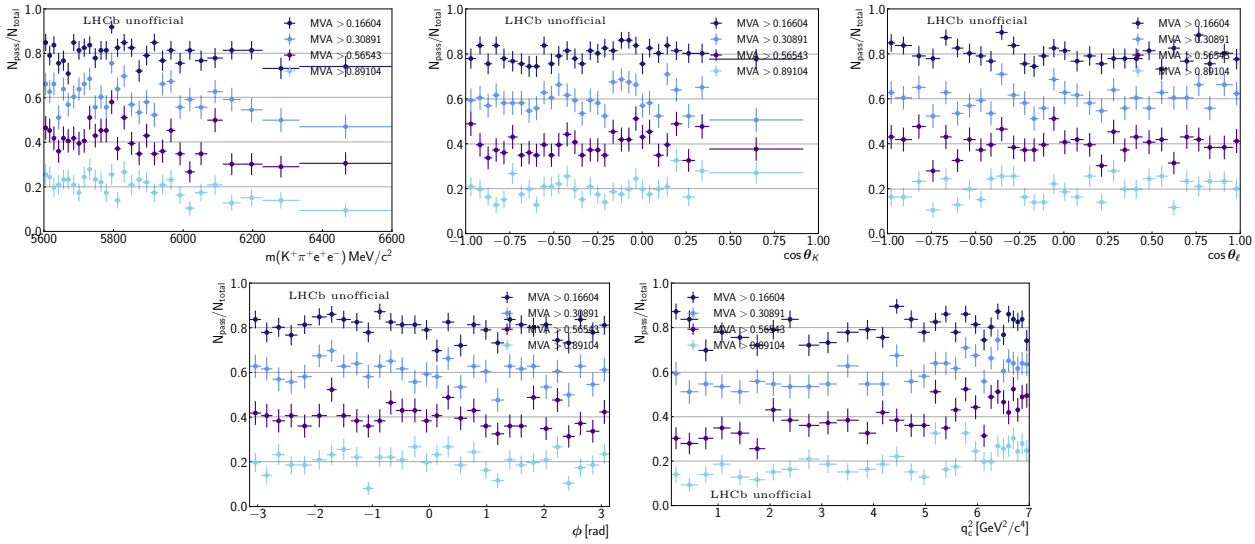


Figure 7.15: Efficiency as a function of $m(K\pi ee)_{\rm PV}$, angles and q_c^2 for the Run 2p2 classifier evaluated using the background samples.

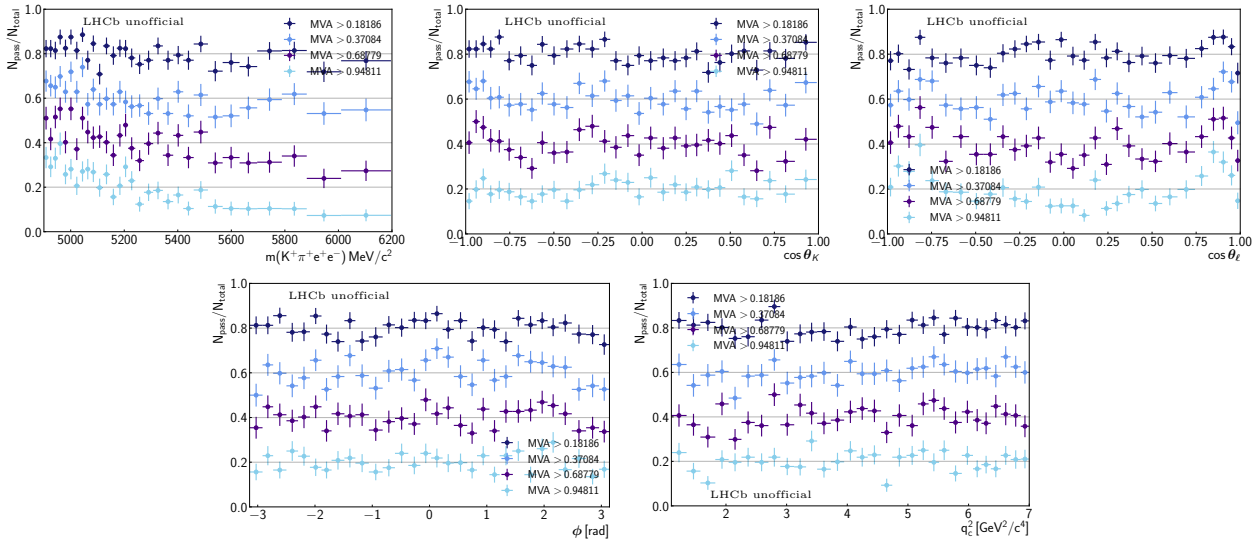


Figure 7.16: Efficiency as a function of $m(K\pi ee)_{\rm PV}$, angles and q_c^2 for the Run 2p2 classifier evaluated using same-sign data samples.

7.3.5 Threshold optimisation

The choice of the optimal cut on the classifier output can be made based on various criteria, quantified by different Figure of Merits (FoMs). The choice of which to use depends on the requirements of the problem. A commonly used FoM is based on the significance of signal with respect to background, and is given by

$$\text{FoM} = \frac{S}{\sqrt{S+B}} \ , \tag{7.7}$$

where S and B refer to the signal and background yields, respectively. This is a reasonable choice in general as sensitivity to observables of the signal tends to scale with the signal yield. For this analysis, a more explicit approach is used, where the expected statistical uncertainty on the angular observable P_5' is taken directly as the FoM. The standard significance based FoM is used to check the consistency of the results.

Note that the main optimisation procedure described below is carried out at an earlier stage of the analysis using an alternative fit strategy. Important differences include the use of only Run 1 and Run 2p1 samples, and the separate treatment of the L0 categories. Nevertheless, consistency checks show that the optimal working point is not substantially different for the current configuration, therefore this computationally expensive process is not repeated. Additional details on the set-up used can be found in Appendix B.

Optimisation procedure

The optimisation procedure involves the determination of signal, partially reconstructed and combinatorial background yields for each classifier threshold, followed by the generation of pseudoexperiments from which the expected statistical uncertainty of P_5' can be determined. More specifically it proceeds via the following steps:

- **Reference point** – a mass fit is made to a sample of signal candidates in data that passed all selection requirements as well as a tight classifier cut of MVA > 0.999 ('reference point'), which lies well beyond the optimal threshold due to its low signal efficiency. The purpose of this fit is to obtain yields for the three components in a regime where the partially reconstructed and combinatorial backgrounds can be separately relatively easily (at looser cuts the combinatorial dominates, such that the determination of the partially reconstructed background becomes challenging). These values are then scaled using simulation ratios to obtain expected values for other classifier thresholds in the following step. Note that this 'reference point fit' is only ran once at the beginning of the optimisation.

- **Thresholds (iterate)** – for each classifier threshold under consideration, the yields of the signal, partially reconstructed and combinatorial components, as well as the slope of the combinatorial background, which are necessary for pseudoexperiment generation, are determined:
  - **Signal yield** – the expected signal yield is calculated based on the yield obtained from the reference point mass fit and a scaling factor from simulation via

$$N^{\text{expected}}_{\text{MVA>x}}(x) = \frac{\epsilon^{\text{MC}}_{\text{MVA>x}}}{\epsilon^{\text{MC}}_{\text{MVA>reference}}} \times N^{\text{data}}_{\text{MVA>reference}} \ , \tag{7.8}$$

where $N^{\text{expected}}_{\text{MVA>x}}$ and $N^{\text{data}}_{\text{MVA>x}}$ refer to the expected yield for a given cut of x, and the value found from the reference point fit, respectively, and $\epsilon^{\text{MC}}_{\text{MVA>x}}$ and $\epsilon^{\text{MC}}_{\text{MVA>reference}}$ refer to the efficiency of the cut MVA > x, and that of the reference point cut (MVA > 0.999), which are determined from simulation. Obtaining an estimation for the signal yield in this way rather than extracting it directly form the fit reduces the impact of statistical fluctuations on the optimisation result.

  - **Background yields** – the yields of the partially reconstructed and combinatorial backgrounds are extracted from data by performing a mass fit. To improve fit stability, the former is Gaussian constrained based on an expected value calculated analogously to the signal using Equation 7.8, with a standard deviation of half the expected yield.[5]

- **Sensitivity** – using the yields of the signal, combinatorial and partially reconstructed components obtained in the previous steps, pseudoexperiments are generated and fitted to determine the expected sensitivity to P'_5. The same information can also be used to evaluate the standard FoM.

A schematic diagram illustrating the optimisation procedure is shown in Figure 7.17. Additional information on the mass fits as well as the pseudoexperiment configuration can be found in Appendix B.1 and B.2.

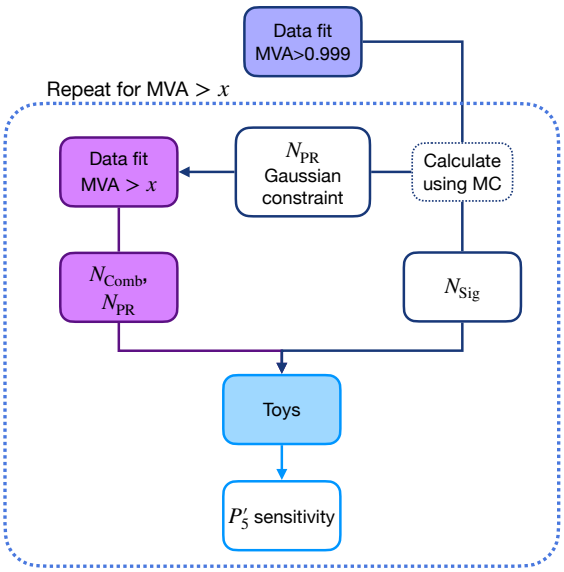


Figure 7.17: Schematic diagram showing the procedure used to determine the optimal classifier threshold using sensitivity to P'_5 as the FoM.

[5]Note that in an ideal scenario, to avoid biasing the choice of the MVA threshold, the optimisation procedure should avoid using information from the signal region (e.g. perform extrapolation based on fits to the upper/lower mass side-bands). However, in this case, it is difficult to achieve stable (and reliable) background determination without information from the nominal mass window. Nevertheless, the possibility of choosing a slightly sub-optimal threshold is not expected to lead to noteworthy complications for the analysis. Note also that the FoM obtained (along with other information) is used to inform the choice of the threshold rather than to directly determine its numerical value.

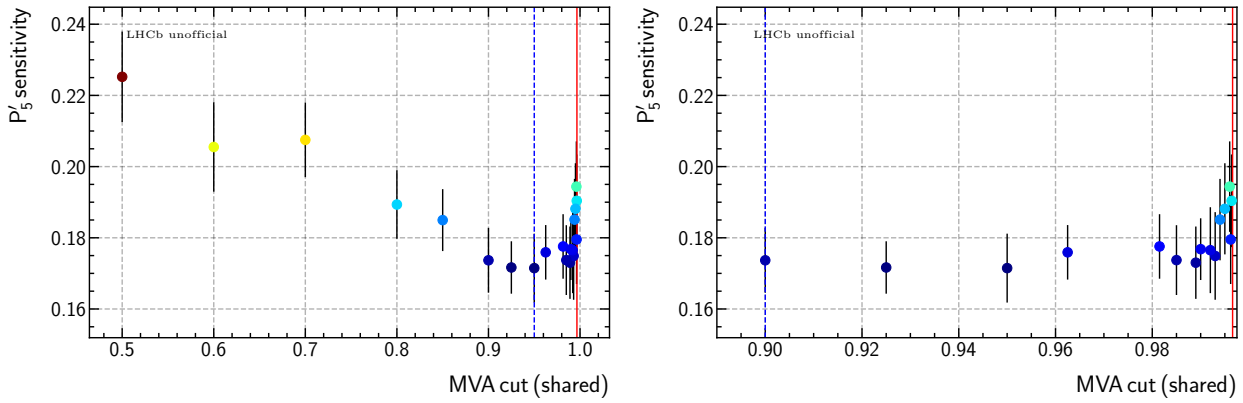


Figure 7.18: Results of the one dimensional scan (expected sensitivities to P_5') using same classifier thresholds for the L0En and L0TIS categories.

Optimisation results

Four subsamples are considered in the optimisation, namely the two L0 categories of Run 1 and Run 2p1. As it is computationally expensive to perform a grid search in four dimensions, some simplifications are made to make the problem more tractable. Firstly, the classifier thresholds are shared between the two data taking periods, and allowed to differ only for the L0 categories. This is motivated by the good compatibility of the performance of the Run 1 and Run 2p1 classifiers, and the observation that the different trigger categories can be characterised by different signal to background ratios, which may lead to preference for different thresholds. This reduces the dimensionality of the problem to two. Secondly, instead of scanning all reasonable combinations of values in the 2d grid, values along the diagonal of the two dimensional plane, which contains shared cuts for the L0En and L0TIS categories, are checked first to determine a region of interest before a finer scan is performed. This region is found to extend from approximately 0.85 to 1.0. Subsequently it is divided into bins of various sizes, with narrower bins closer to the limit at around 0.9964, beyond which pseudoexperiment fits become unstable, and points in the centres of the bins are evaluated. These results, shown in Figure 7.19, do not suggest that using different cuts for the two trigger categories is necessary. They also favour tighter values, although the sensitivity FoM seems to reach a plateau at around 0.9, before increasing again close to the boundary. The failure rate of the pseudoexperiments, shown in Figure 7.20, starts to increase around 0.9. Considering these factors, the working point can be chosen within the relatively stable region of around 0.95 to 0.99. In order to reduce background instabilities that can show up when the level of combinatorial background is too low, a relatively loose cut of 0.96 is chosen.

Standard Figure of Merit

The yields obtained for the P_5' optimisation procedure also allow for the calculation of the standard FoM. When this is done for Run 1 and Run 2p1 samples, considering the values of the shared cut between the two L0 categories, the results, shown in Figure 7.21, are compatible with the P_5' based FoM. Note that in this case while the signal yield, S, is taken directly from the estimated yield obtained using the reference point fit and Equation 7.8, the background yield, B, is taken to be the yield of the combinatorial background within the $m(K\pi ee)_{\mathrm{PV}}$ region of $5100-5400\,\mathrm{MeV/c^2}$, which is below the signal peak, and therefore most relevant to the signal

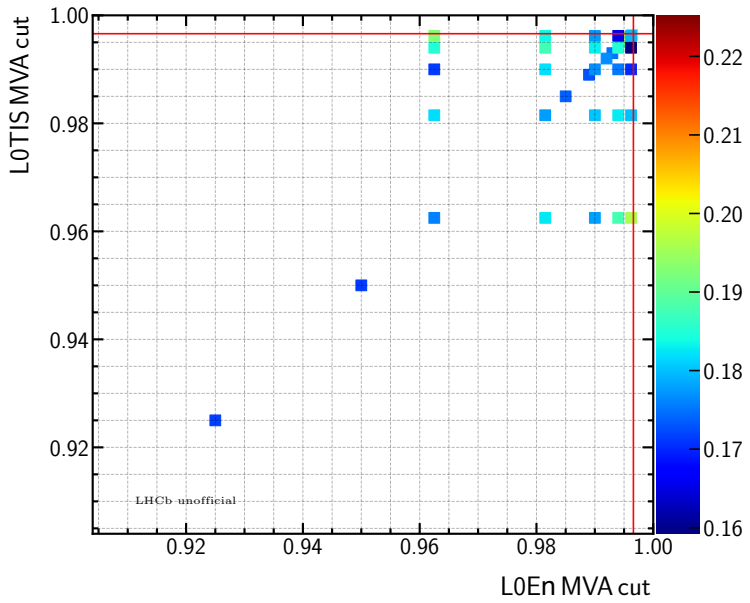


Figure 7.19: Results of the two dimensional scans in the region of interest.

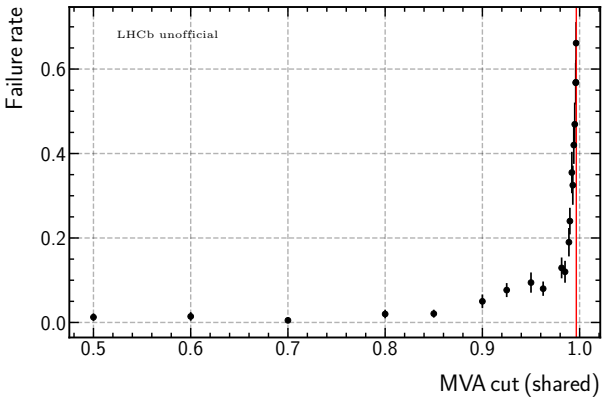


Figure 7.20: The fraction of failed pseudoexperiments over the total generated.

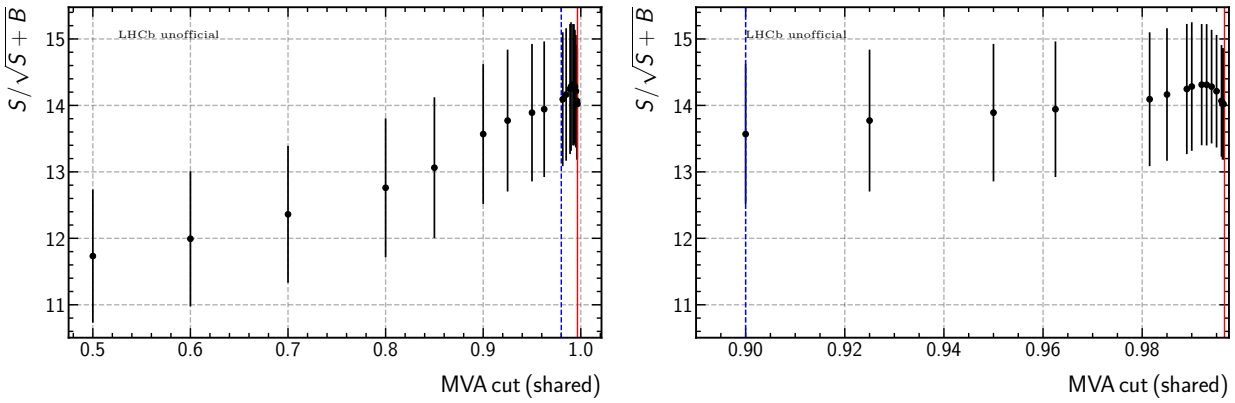


Figure 7.21: Plot of the values of the standard FoM against shared classifier thresholds between the L0TIS and L0En categories. Note that the uncertainties do not take correlations into account.

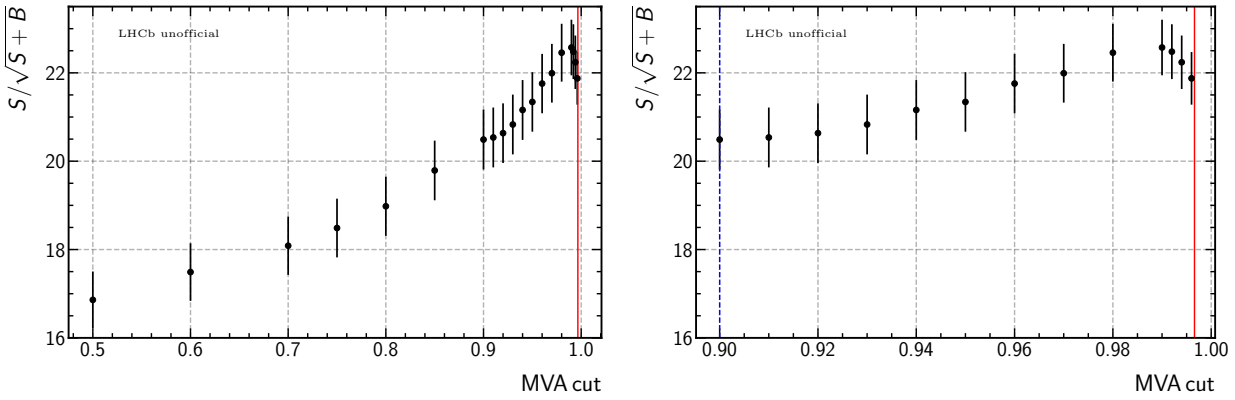


Figure 7.22: Plot of the values of the standard FoM against classifier thresholds for the current fit configuration and samples. Note that the uncertainties do not take correlations into account.

sensitivity.

Considering the good compatibility between the two FoMs, only the standard FoM is re-evaluated for the current configuration and samples to check the validity of the previous results. In this case angular fits are made for MVA thresholds ranging from 0.5 to 0.996 using the nominal configuration (Section 8.3). The resulting trend, shown in Figure 7.22, is found to be largely consistent with the previous results, which suggests that the benefit of re-performing the main optimisation is limited.

7.4 Low efficiency region cut

The angular and q_c^2 distributions of the signal candidates are corrected using per-event weights obtained from the inverse of 'effective acceptance functions' parametrised using $B^0 \to K^{*0}e^+e^-$ simulation (Section 8.1). However, due to the scarcity of simulated candidates in certain regions of the phase space and/or low efficiency, e.g. at the edges of $\cos\theta_\ell$ and near $\cos\theta_K = 1$, the resulting functions can take on very small values or even become negative. This gives rise to pathologically large, or undefined correction weights, which can lead to problematic fit behaviour. A cut of

$$|\cos\theta_\ell| < 0.9 \text{ and } \cos\theta_K < 0.9 \ , \tag{7.9}$$

is therefore applied to remove the regions where the effective acceptance effect is not well described. It has a low signal efficiency of around 93%, but it is capable of removing almost all events with negative weights, as estimated using a large statistics uniform sample, and strongly reduces the fraction of events with large weights. In addition, it has the added (unintended) benefit of reducing backgrounds from DSL decays and J/ψ leakage.

7.5 Multiple candidates

After applying the full set of selections, 0.5% of the simulated $B^0 \to K^{*0}e^+e^-$ candidates and 0.3% of the simulated $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ candidates are found to not originate from distinct events of interest. In data, the corresponding values for the rare (in the large q^2_c range) and control mode samples, which include both signal and background, are 0.5% and 0.7%, respectively. While in principle there can be more than one candidate per event, this is unlikely given the low branching fraction of the signal decay. Cases of multiple candidates are more often than not related to errors in the reconstruction process [171]. For example, one candidate can be a partial overlap of another (a true candidate), with one or more tracks taken from the rest of the event or reconstructed out of unrelated hits. Therefore, after applying all other selections, if multiple candidates are present, one is kept at the random, while the other one is discarded.

7.6 Summary of selections

To summarise, the full list of selections discussed in this chapter are shown explicitly in Table 7.3, with the exceptions of the reconstructed B^0 mass and q^2_c cuts that define the regions of interest for the rare and control modes, which are

$$1.1 < q^2_c < (6.0)7.0\,\text{GeV}^2/\text{c}^4 \text{ and } 4900 < m(K\pi ee)_{\text{PV}} < 5700\,\text{MeV}/\text{c}^2, \tag{7.10}$$

for the former and

$$7.0 < q^2_c < 11.0\,\text{GeV}^2/\text{c}^4 \text{ and } 4500 < m(K\pi ee)_{\text{PV}} < 6200\,\text{MeV}/\text{c}^2, \tag{7.11}$$

for the latter. The random removal of multiple candidates is carried out on the final samples after all selections have been applied.

Table 7.3: Summary of the selection requirements discussed in this section (with the exceptions of the cuts used to define the signal and control mode measurement regions).

Type		Requirement
Quality	all tracks	$\chi^2/\text{ndf} < 3$
		$\texttt{GhostProb} < 0.4$
	e	$\texttt{region}^{\texttt{L0Calo}}_{\text{ECAL}} \geq 0$
		not ($\lvert\texttt{xProjection}^{\texttt{L0Calo}}_{\text{ECAL}}\rvert < 363.6\,\text{mm}$
		and $\lvert\texttt{yProjection}^{\texttt{L0Calo}}_{\text{ECAL}}\rvert < 282.6\,\text{mm}$)
		$\texttt{InAccEcal}$
		$\sqrt{p^2_{x,\,\text{track}} + p^2_{y,\,\text{track}}} > 200$
Clone	K, π, e	$\theta(\pi, e) > 0.0005,\ \theta(K, e) > 0.0005,\ \theta(K, \pi) > 0.0005$
ID	K^{*0}	$\lvert m(K\pi) - m^{PDG}_{K^{*0}}\rvert < 100\,\text{MeV/c}^2$
PID	all	$\texttt{hasRich}$
	e	$\texttt{hasCalo}$
	K, π	$p_T > 250\,\text{MeV/c}$
	e	$p_T > 500\,\text{MeV/c}$, $p > 3000\,\text{MeV/c}$
	K	$\texttt{ProbNNk} \cdot (1 - \texttt{ProbNNp}) > 0.05$
	π	$\texttt{ProbNNpi} \cdot (1 - \texttt{ProbNNk}) \cdot (1 - \texttt{ProbNNp}) > 0.1$
	e	$\texttt{ProbNNe} > 0.2$
	K	$\texttt{DLL}_{K\pi} > 0$
	e	$\texttt{DLL}_{e\pi} > 2$
BKG	charmonium $e-h$ swap	not ($\lvert m_{J/\psi,\psi(2S)}(K_{\to e}\pi e_{\to K}e) - m^{PDG}(B^0)\rvert < 60\,\text{MeV/c}^2$ and $\texttt{eProbNNe} < 0.8$
		or $\lvert m_{J/\psi,\psi(2S)}(K\pi_{\to e}ee_{\to\pi}) - m^{PDG}(B^0)\rvert < 60\,\text{MeV/c}^2$ and $\texttt{eProbNNe} < 0.8$)
	$B^+ \to K^+e^+e^-$	$\max(m(Ke^+e^-), m(\pi_{\to K}e^+e^-)) < 5100\,\text{MeV/c}^2$
	$B^0_s \to \phi e^+e^-$	not ($m(K(\pi_{\to K})) < 1040\,\text{MeV/c}^2$ and $\pi\texttt{ProbNN}\pi < 0.8$)
	$B^0 \to \bar{D}^0(\to K^+\pi^-)\pi^-e^+\nu_e$	not ($\lvert m(Ke_{\to\pi}) - m^{PDG}(D^0)\rvert < 30\,\text{MeV/c}^2$ and $\texttt{eProbNNe} < 0.8$)
	$B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)\pi^-)e^+\nu_e$	not ($\lvert m(K\pi e_{\to\pi}) - m^{PDG}(D^-)\rvert < 30\,\text{MeV/c}^2$ and $\texttt{eProbNNe} < 0.8$)
	$B^0_s \to D^-_s(\to \bar{K}^{*0}(\to K^-\pi^+)K^-)e^+\nu_e$	not ($\lvert m(K\pi e_{\to K}) - m^{PDG}(D^-_s)\rvert < 30\,\text{MeV/c}^2$ and $\texttt{eProbNNe} < 0.8$)
	K $-$ π swap	$K\,\texttt{DLL}_{K\pi} - \pi\,\texttt{DLL}_{K\pi} > 0$
	part-reco (J/ψ only)	$m(K\pi ee)_{\text{PV},\,J/\psi} > 5150\,\text{MeV/c}^2$
	comb	$\texttt{MVA} > 0.96$
Low efficiency	phase space	$\lvert\cos\theta_\ell\rvert < 0.9$, $\cos\theta_K < 0.9$

8 Elements of the angular analysis

To carry out the angular analysis, effective acceptance functions are parametrised and used to correct distortions to the signal distribution using per-event weights. Models are created to describe the angular and mass distributions of backgrounds, and the mass distribution of the signal. The extraction of observable values is made by minimising a weighted negative log-likelihood function. The behavior of the fit is studied using realistic pseudoexperiments generated with an amplitude model [172, 142], and the strategy is further checked by applying it to the high statistics control mode of $B^0 \to K^{*0} J/\psi(\to e^+e^-)$. These elements are discussed in detail in the sections below.

8.1 Effective acceptance

The angular distribution of signal events in data cannot be described directly by Equation 2.14. This is due to two main effects: acceptance and resolution. The impact of parts of the reconstruction and selection procedure, such as the initial exclusion of candidates due to the limited geometrical acceptance of the detector, triggering, stripping, and preselection, is referred to collectively as the 'acceptance' effect, which is in general non-uniform across the phase space. The precision by which the momenta of particle tracks can be measured directly affects the q^2 and angular values. Uncertainties lead to shifts around the true values (smearing effect). Electrons also lose significant amounts of their energy through bremsstrahlung and FSR. While pure acceptance effects can be corrected for relatively simply by obtaining a map of the efficiency across the phase space of interest, resolution corrections (e.g. deconvolution) tend to be more complicated, and are disfavoured by the limited statistics available. Another complication is that these two effects are not fully factorisable. Therefore, an effective approach is used, where both types of distortions are accounted for simultaneously using an 'effective acceptance function'[1], ϵ_{eff}, which describes the ratio of

$$\epsilon_{\text{eff}} = P_r(\bar{\Omega}_r, q_r^2)/\mathcal{P}(\bar{\Omega}_t, q_t^2) \ . \tag{8.1}$$

In this expression, $P_r(\bar{\Omega}_r, q_r^2)$ refers to the distribution of the reconstructed angles and q^2 (the q_c^2), which can be extracted from full simulation. The $\mathcal{P}(\bar{\Omega}_t, q_t^2)$ in the denominator refers to the true angular and q^2 distributions without detector, bremsstrahlung and FSR effects, which can be obtained from generator level samples (produced without running `PHOTOS`). In this way, if the simulation replicates the distortions correctly, per-event weights from $1/\epsilon_{\text{eff}}$ would allow for the true angular and q^2 distributions in data to be retrieved.

8.1.1 Parametrisation

The effective acceptance function is parametrised as a function of the angles and q_c^2 using a sum of Legendre polynomials and Fourier terms without assuming factorisation [63, 143]. Explicitly

[1] As opposed to the 'acceptance function', which is often used in muon mode angular analyses, e.g. [57], to correct mainly for the acceptance effect as the impact of bremsstrahlung (and FSR) is negligible.

it is given by

$$\epsilon(\cos\theta_\ell, \cos\theta_K, \phi, q_c^2) = \sum_{klmn} c_{klmn} L_k(\cos\theta_K) L_l(\cos\theta_\ell) F_m(\phi) L_n(q_c^{2\prime}) \ , \tag{8.2}$$

where $L_a(x)$ is the Legendre polynomial of order a in the variable x, where $x \in [-1, 1]$. Note that in the case of q_c^2, which does not naturally lie within this range, a change of variables is made as follows

$$q_c^{2\prime} = \frac{2q_c^2 - q_{c\,\mathrm{min}}^2 - q_{c\,\mathrm{max}}^2}{q_{c\,\mathrm{max}}^2 - q_{c\,\mathrm{min}}^2} \ , \tag{8.3}$$

where $q_{c\,\mathrm{max}}^2$ and $q_{c\,\mathrm{min}}^2$ are the limits of the parametrisation range. This is done to ensure that the orthogonality property of the Legendre polynomials within -1 to 1 can be used in later calculations. As the ϕ angle is periodic, it is more appropriately modelled using Fourier terms, $F_b(x)$, which are given by

$$F_b(x) = \begin{cases} \cos\dfrac{bx}{2} & \text{if} \quad b \text{ is even} \\ \sin\dfrac{(b+1)x}{2} & \text{if} \quad b \text{ is odd} \end{cases} \ , \tag{8.4}$$

where b is a non-negative integer.

The coefficients of the function, c_{klmn}, are obtained using the method of moments by equating the analytically calculated expectation value (first moment) of random variables, $M = \epsilon(\bar{\Omega}, q_c^2)$, to its sample mean. Making use of the orthogonality of the Legendre polynomials,

$$\int_{-1}^{1} L_a(x) L_{a'}(x)\, dx = \frac{2}{2a+1}\delta_{aa'} \ , \tag{8.5}$$

where δ denotes the Kronecker delta function, and that of the Fourier terms,

$$\int_{-\pi}^{\pi} F_b(x) F_{b'}(x)\, dx = f_{bb'} \ , \tag{8.6}$$

where

$$f_{bb'} = \begin{cases} 0 & \text{if} \quad b \neq b' \\ \pi & \text{if} \quad b = b' \neq 0 \\ 2\pi & \text{if} \quad b = b' = 0 \end{cases} \ , \tag{8.7}$$

it is possible to write, for a specific set of $k'l'm'n'$ values, the expectation value of

$$\langle\, M_{k'l'm'n'} \,\rangle = (\frac{2}{2k'+1})(\frac{2}{2l'+1})(f_{m'm'})(\frac{2}{2n'+1}) c_{k'l'm'n'} \ . \tag{8.8}$$

Equating it to the sample mean of

$$\overline{M}_{k'l'm'n'} = \frac{1}{\sum_i^N \omega_i} \sum_{i=1}^{N} (\omega_i) L_{k'}(\cos\theta_{li}) L_{l'}(\cos\theta_{Ki}) F_{m'}(\phi_i) L_{n'}(q_{ci}^2), \tag{8.9}$$

which can be calculated from the sample to be parametrised with N events in total and per-event weights of ω_i, allows the coefficient of $c_{k'l'm'n'}$ to be determined. In general, c_{klmn} can be obtained via

$$c_{klmn} = \frac{1}{\sum_i^N \omega_i} \sum_{i=1}^{N} (\omega_i) \Big[L_k(\cos\theta_{li}) L_l(\cos\theta_{ki}) F_m(\phi_i) L_n(q_{ci}^2) \times (\frac{2k+1}{2})(\frac{2l+1}{2})(\frac{1}{f_{mm'}})(\frac{2n+1}{2}) \Big] , \tag{8.10}$$

where the ω_i include the weights used for simulation-data corrections as well as ones that carry out the effective division of the generator level distribution, which can be obtained from a parametrisation of the denominator of Equation 8.1.

8.1.2 Sample choice

Acceptance functions can (and often are) parametrised using simulation samples of the signal decay that are generated to be uniform (to a good approximation) across the phase space of interest, in contrast to standard ('physics') simulation samples. A key advantage of this approach is that it allows all regions of the phase space to be populated approximately evenly, which is beneficial for the estimation of the relative efficiency. For samples of the same size, using a uniform simulation results in acceptance functions with smaller pathological regions. These are regions where the function takes on very small or negative values, as illustrated in Figure 8.2. This leads to large (or undefined) weights when the inverse is taken. However, as the effective correction strategy uses an essentially shape-based method to correct for a convolution effect, which is illustrated for a simple example in Figure 8.1, it depends on the underlying physics model. Therefore, only the standard simulations generated with a realistic physics model are used. To reduce the impact of underpopulated regions, which are well confined to the edges of $\cos\theta_\ell$ and one edge of $\cos\theta_K$, as shown on Figure 8.3, a cut is made to remove this region (Section 7.4).

8.1.3 Effective acceptance function

Three main effective acceptance functions are used for this analysis. They are obtained separately from fully corrected $B^0 \to K^{*0} e^+ e^-$ physics simulation generated with Run 1, Run 2p1 and Run 2p2 conditions. To validate the analysis strategy, additional functions are made using the control mode $B^0 \to K^{*0} J/\psi (\to e^+ e^-)$ simulation. In both cases, the corresponding generator level distributions are parametrised first. This corresponds to the denominator $\mathcal{P}(\bar{\Omega}_t, q_t^2)$ in Equation 8.1. Then, per-event weights are obtained from the inverse of the generator level function using the reconstructed angles and q_c^2 of the physics simulation. This constitutes an effective division. Finally, the resulting weighted distributions are parametrised.

The lowest orders of the Legendre and Fourier polynomials that can provide good description of the distributions are used. The former values are based on those chosen for the muon mode analysis [57], while the order of the latter is motivated in part by the lack of terms beyond $\cos 2\phi$ and $\sin 2\phi$ in the signal pdf. Lower orders are favoured as the use of higher orders (at the same level of statistics) enlarges pathological regions. Events within these regions would not receive proper corrections. The ones that receive negative weights would be removed, which is undesired, as the effect of such vetoes cannot be easily accounted for in the normalisation of the signal pdf. Large weights arising from the inverse of very small numbers can introduce

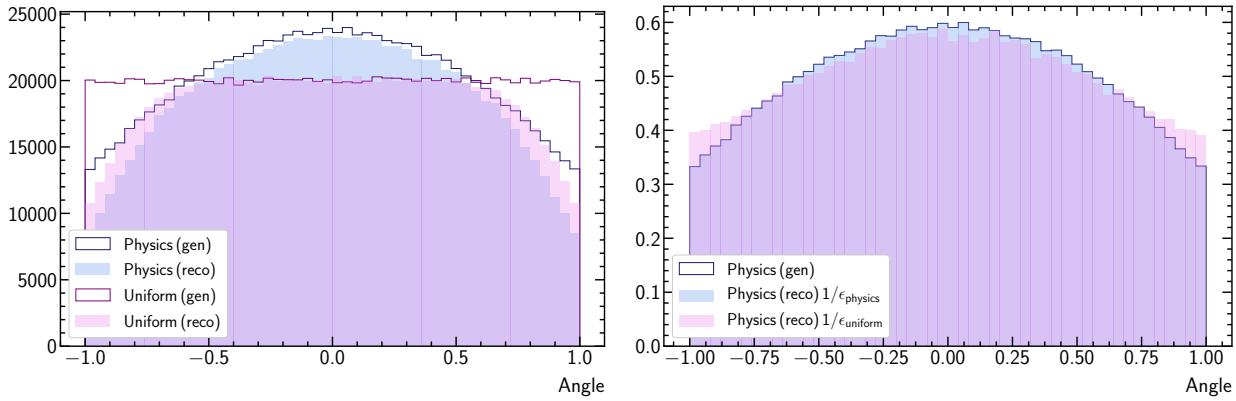


Figure 8.1: Illustration showing the model-dependency of the effective (shape-based) resolution correction. Two samples are generated, one of which is uniformly distributed in a generic 'angle', while the other is non-uniform (left). The 'reconstructed' angle is obtained by adding values drawn at random from a Gaussian distribution. Correction weights can be obtained via the effective strategy by performing a histogram division of the post-reconstruction distribution over that of the generator level. However, while both samples have been blurred using the same function, the correction weights are different. The weights obtained from the uniform distribution performs poorly when used to correct the sample that is non-uniform at generator level (right).

effects that are not necessarily straightforward to mitigate or quantify. The chosen orders for both the rare and the control modes are given in Table 8.1, and the statistics available are shown in Table 8.2. The smallest sample (Run 1) leads to an acceptance function with the largest fraction of negative weights, and the largest sample (Run 2p1) the smallest fraction (Figure 8.3).

To avoid edge effects, a q^2 range of $0.5 < q^2_{(c)} < 10.0\,\text{GeV}^2/\text{c}^4$, which is larger than the nominal q^2 window, is used for the generator level and final effective acceptance functions. Note that in the case of the control mode, the q^2 is not parametrised, as the J/ψ resonance is not suitable for the strategy of effective (shape) correction.[2] This is one of the unavoidable differences between the rare and the control modes. The one dimensional projections of the generator level parametrisations and that of the effective acceptance functions of the rare and control mode are shown in Figures 8.4 and 8.5, respectively.

The q^2 projection of the rare mode generator level sample shows a 'kink' like structure near $5.6\,\text{GeV}^2/\text{c}^4$ that is not fully described by the Legendre polynomial. This is a known feature that originates from the discontinuity of the effective WCs $\mathcal{C}_7$ and $\mathcal{C}_9$ used in the `EvtGen` model [158, 142]. While it is possible to describe using higher order polynomials, this is not done due to the need to reduce low efficiency regions and its limited impact on the observables of interest. The projections of the weighted post-reconstruction samples generally indicate that the descriptions are sufficient good. However, larger pulls can be seen near $\cos\theta_K = 1$ in some cases. This effect has been seen previously in the angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ decays [57]. Again, improved description can be achieved using higher orders, which comes at a cost. This bias towards using lower orders is assessed as a source of systematic uncertainty in Section 9.2.2.

[2]The effective correction approach relies on the presence of events in both the numerator and denominator of Equation 8.1, which is not the case for the q^2 distribution of the narrow J/ψ resonance.

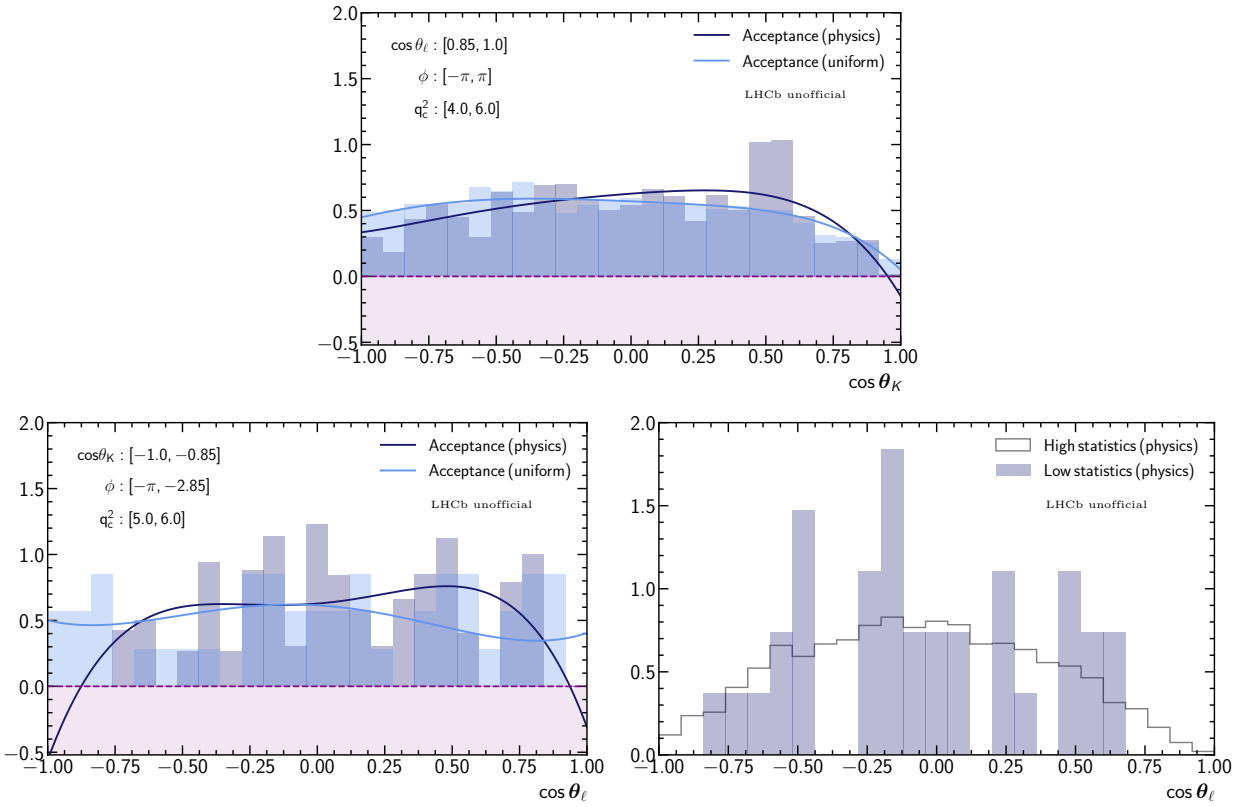


Figure 8.2: Illustration of two ways by which pathological regions can arise. The first scenario (top row) involves a region of genuinely low efficiency. Few events survive the selection in the case of both physics and uniform samples near $\cos\theta_K = 1$, and both functions can take on very small values or become negative. The second scenario (bottom row) only affects the physics samples. The true efficiency at the edges of $\cos\theta_\ell$ is not necessarily low. However, due to the shape of the generator level distribution (bottom right), when the sample statistics is low, fluctuations can easily lead to no events appearing at the edges, which can cause the calculated function to become negative.

Table 8.1: Maximum orders of Legendre polynomials and Fourier terms used for the parametrisation of generator level distributions or effective acceptance functions. In all cases, all terms of order less than or equal to the maximum order are included.

Type	$\cos\theta_K$	$\cos\theta_\ell$	ϕ	q_c^2	Total
$B^0 \to K^{*0}e^+e^-$ generator	5	4	4	9	1500
$B^0 \to K^{*0}e^+e^-$ reconstruction	5	4	4	3	600
$B^0 \to K^{*0}J/\psi(\to e^+e^-)$ generator	7	6	4	-	280
$B^0 \to K^{*0}J/\psi(\to e^+e^-)$ reconstruction	5	4	4	-	150

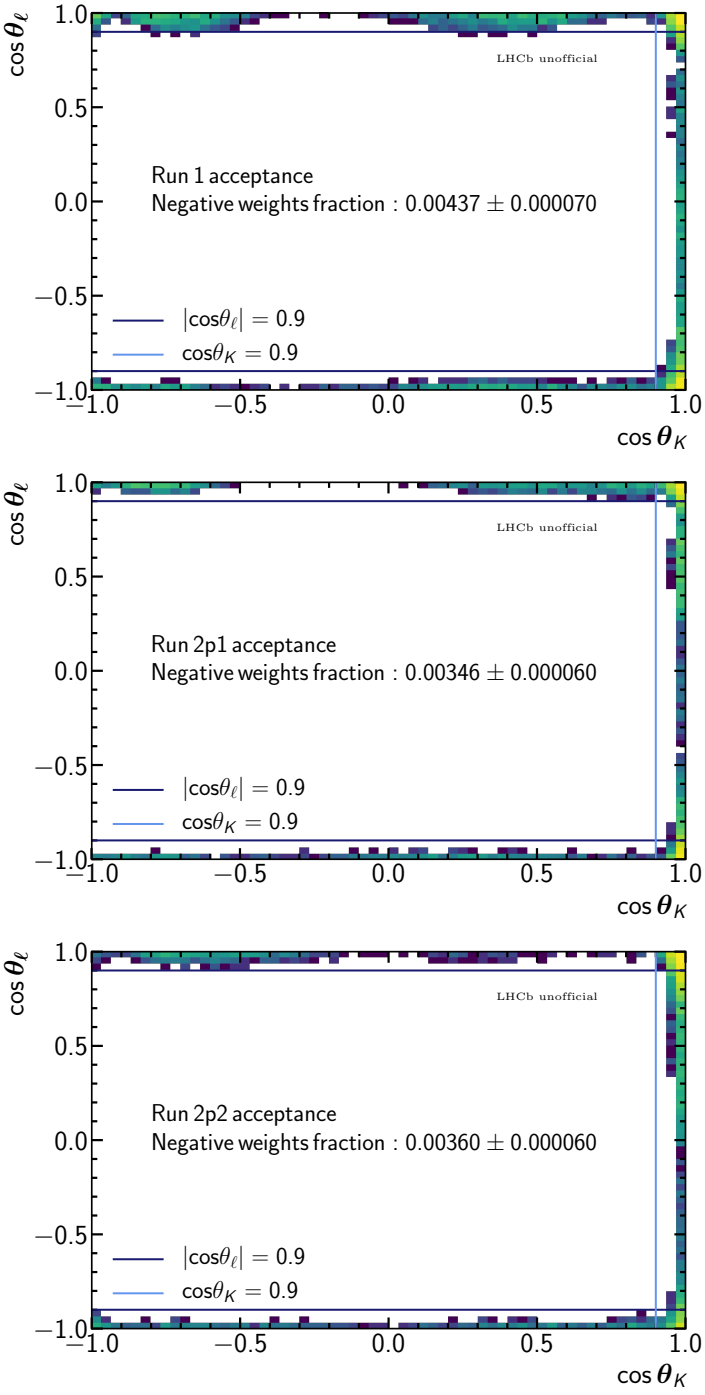


Figure 8.3: The distribution of 10^6 test events from a uniform distribution that received negative weights from the nominal Run 1 (top), Run 2p1 and Run 2p2 (bottom) effective acceptance functions and the phase space cut used to reduce them (blue lines). More than 99% of the negative weights are removed by the cut.

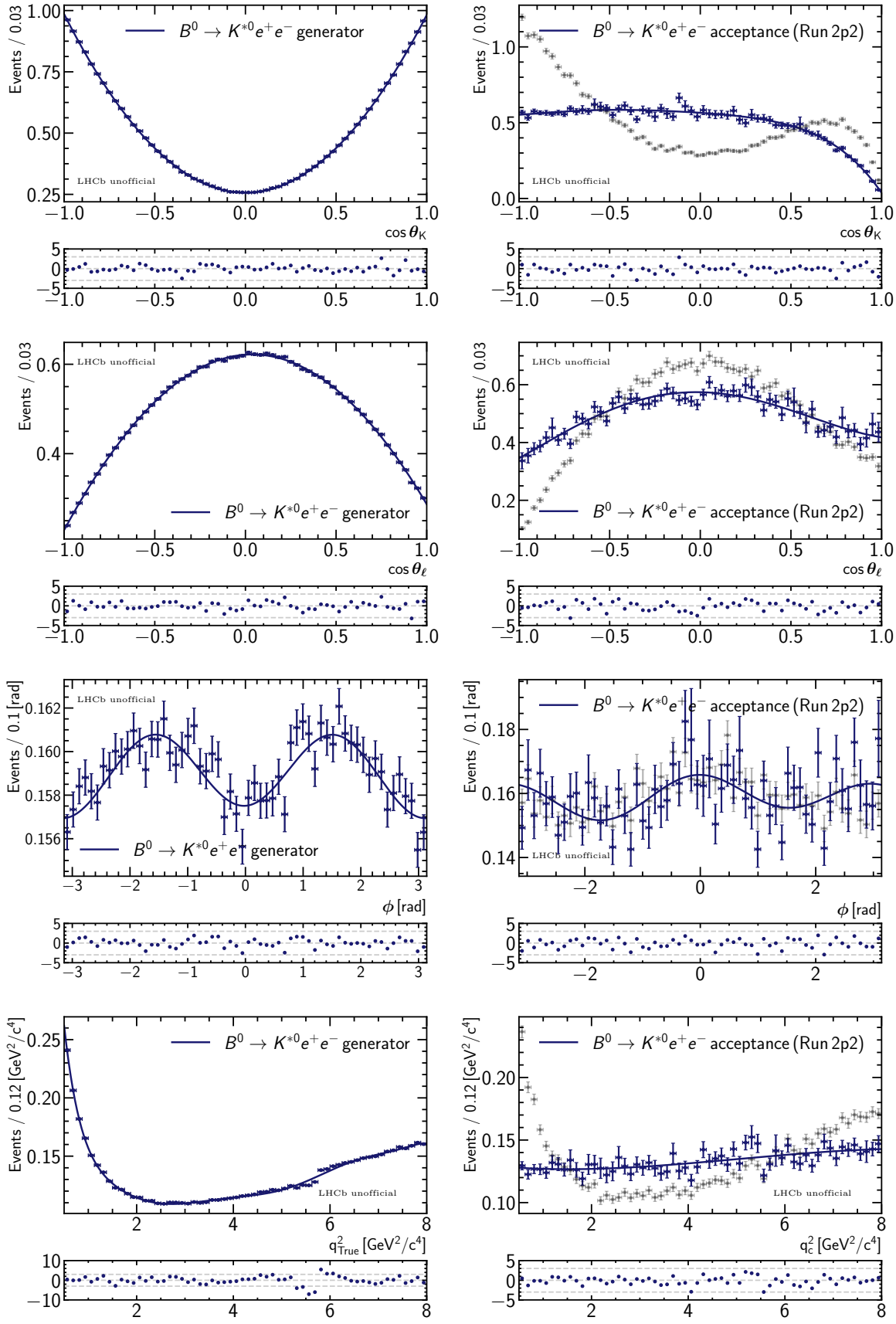


Figure 8.4: One dimensional projections of the generator level distributions (left column) and the effective acceptance functions (right column) of the rare mode. For the latter, the Run 2p2 subsample is shown as an example.

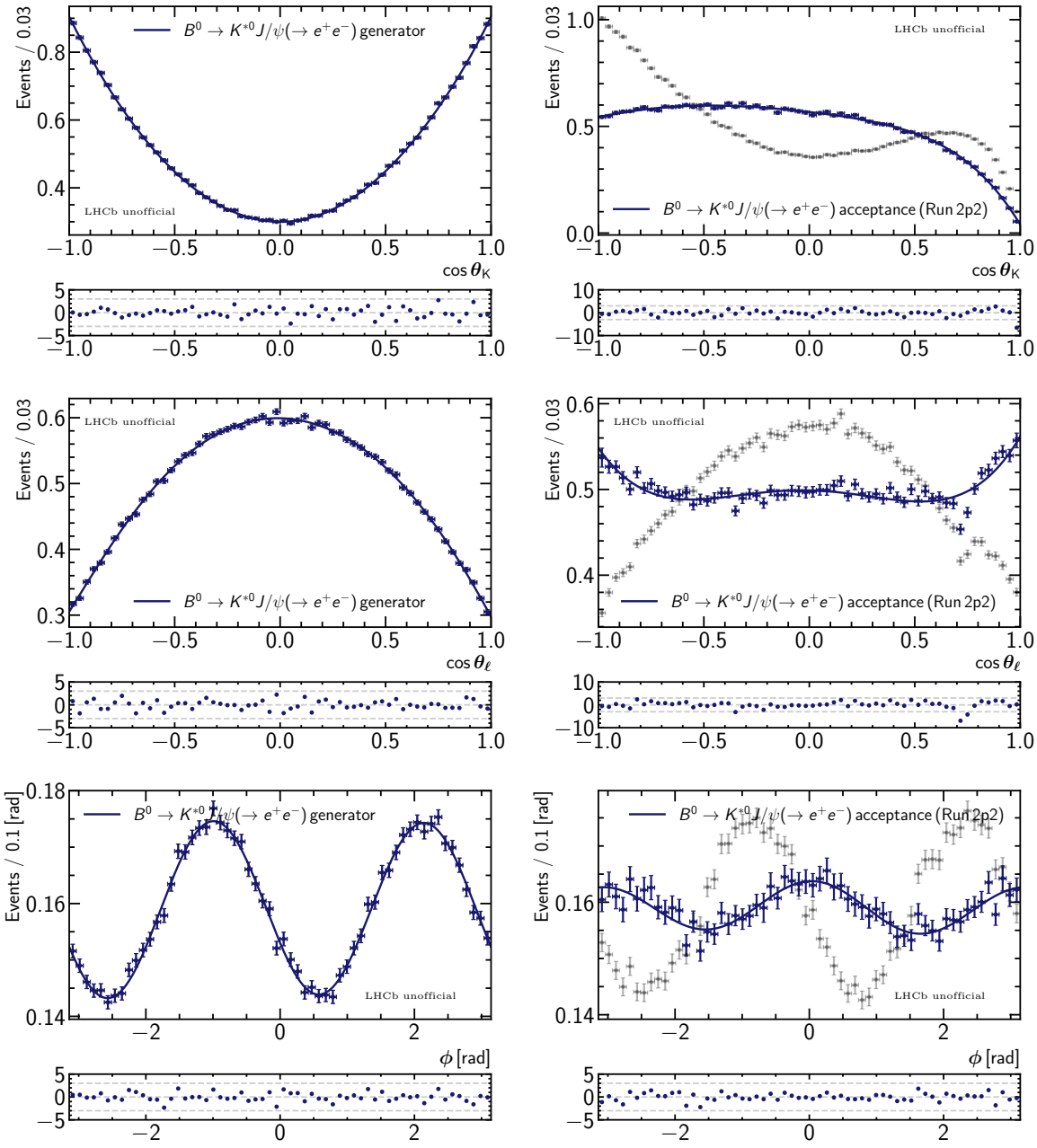


Figure 8.5: One dimensional projections of the generator level distributions (left column) and the effective acceptance functions (right column) of the control mode. For the latter, the Run 2p2 subsample is shown as an example.

Table 8.2: The statistics of the generator level and post-reconstruction simulation available for the parametrisation of rare and control mode effective acceptance functions.

Generator	All
$B^0 \to K^{*0} e^+ e^-$	2328375
$B^0 \to K^{*0} J/\psi(\to e^+ e^-)$	1000017

Reconstructed	Run 1	Run 2p1	Run 2p2
$B^0 \to K^{*0} e^+ e^-$	52941	105858	82895
$B^0 \to K^{*0} J/\psi(\to e^+ e^-)$	509066	830124	547970

8.1.4 Validation

To validate the parametrisation strategy, a 'k-folding' approach is used. In analogy to the method of the same name used in the training of the multivariate classifier, each simulation sample is split into N subsamples. A function is parametrised using all but one of the subsamples $(N-1)$, and used to assigned weights to events in the unused subsample. This procedure continues until N acceptance functions have been made, and all subsamples received unbiased weights. Considering the limited statistics of the rare mode simulation, twelve folds are used $(N = 12)$. A weighted angular fit is then made to the full sample to extract the angular observables, which are compared to values found by fitting the generator level distribution. As inverse weights from the latter is used in the parametrisation, weighting the post-reconstruction sample by the effective acceptance weights will allow for the retrieval of the generator level distribution if the parametrisation is effective.[3] This check is ran separately for the rare and the control modes of each Run subsample. Due to the inclusion of simulation-data correction weights in the parametrisation, these weights are similarly included in the fit in addition to the acceptance weights. In the case of the control mode, all angular fits are made with S-wave and interference terms (Equation 2.18) in order to check the potential impact of mismodelling on these parameters. This component is not included in the simulation, and all related parameters of the generator level fit are found to be compatible with zero, as expected.

The projections of the fits to the Run 2p2 samples are shown in Figure 8.6, and the results for all subsamples are summarised in Figures 8.7 and 8.8 for the rare and control modes, respectively. In all cases, the differences in the central values lie within two standard deviations of the generator level result. For the rare mode, all shifts are much smaller than the expected statistical uncertainty (Section 8.4), while they can be comparable to (although still smaller than) the statistical uncertainties of the control mode fit.

Two interesting features can be seen, the first of which is the slight upward shift in F_L, which is present in both modes. F_L is one of the observables that is visible in the $\cos\theta_K$ projection (the others being A_{FB}, F_S, and S_{S1}) as it is not 'integrated away' due to some symmetry of an angular term.[4] It is therefore possible gain some understanding of the origin of this effect simply

[3] More specifically, the parametrised generator level distribution is retrieved. The results of the check will therefore also include the impact on the observables due to the failure to parametrise the generator level distributions exactly.

[4] In contrast, the P_5' value, for example, has no impact on the $\cos\theta_K$ projection (or that of any other angle), as both the $\sin\theta_\ell$ and $\cos\phi$ terms that it is multiplied to are zero when integrated over their respective phase space ranges.

by looking at the $\cos\theta_K$ projections of the effective acceptance function. As shown in Figure 8.5, for example, the acceptance function undershoots slightly in the region of $\cos\theta_K \sim 0.85 - 0.90$, before overshooting $\gtrsim 0.90$. Due to the phase space cut of $\cos\theta_K < 0.9$, F_L becomes over-corrected close to the boundary[5], leading to slight upward shifts. Another interesting feature is the downward shift in A_{FB} that is prominent for the control mode, but insignificant for the rare mode. This is related to the vetoes against semileptonic decays, which creates a 'kink' in the $\cos\theta_\ell$ distribution that is not well described by the parametrisation model. These vetoes are found to predominantly affect the control mode.

The differences between the observable values found with respect to the generator level results can be thought of as biases introduced by the parametrisation strategy with their associated uncertainties. For each observable, the weighted mean of the shifts is used to correct its central value, and the uncertainty is taken into account as a systematic uncertainty (Section 9.2.2).

[5]The relative efficiencies are lower due to the undershoot, resulting in correction weights that are larger than necessary.

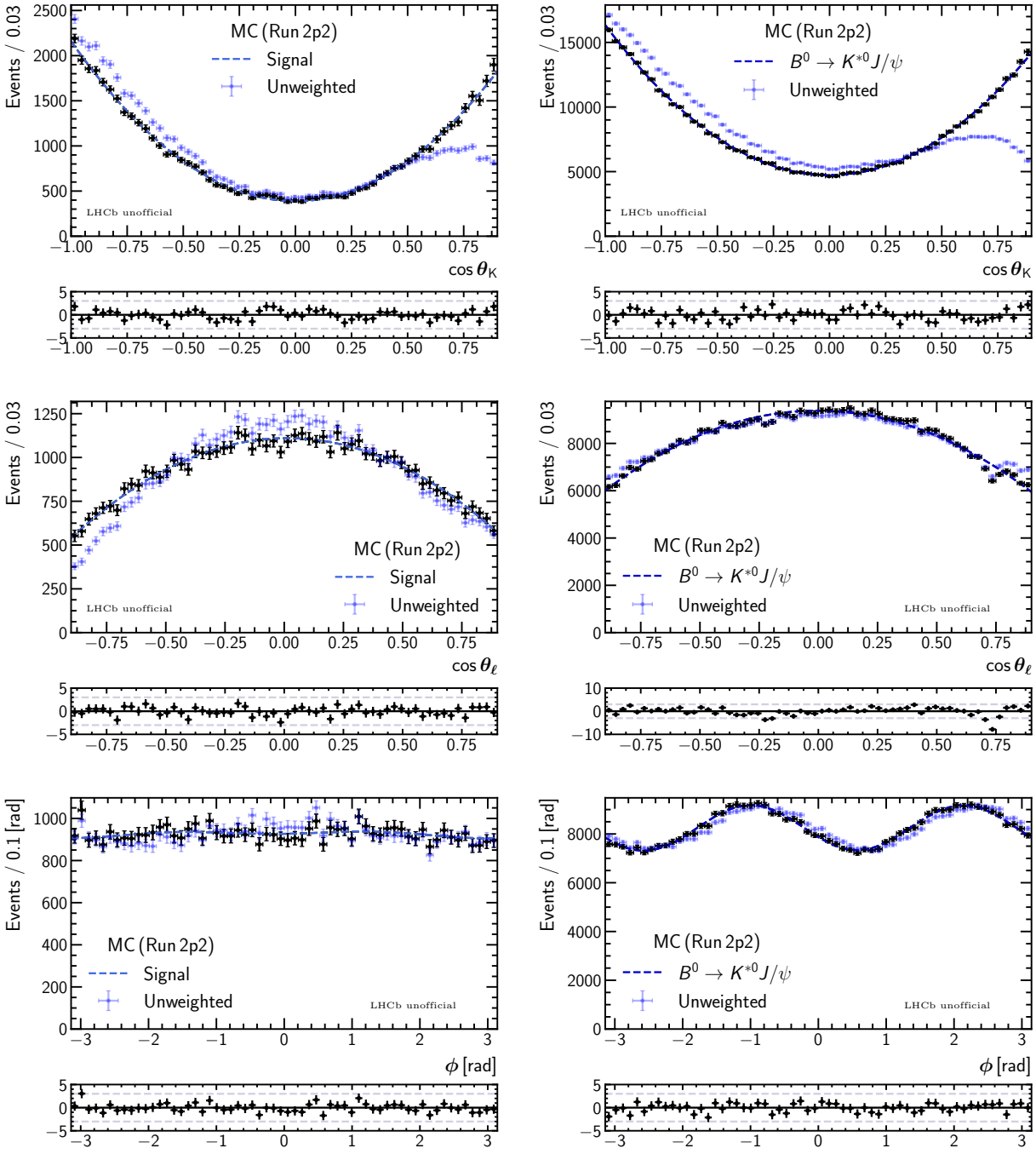


Figure 8.6: Projections of the fits made to the Run 2p2 rare mode (left column) and control mode (right column) simulation sample using per-event effective acceptance weights.

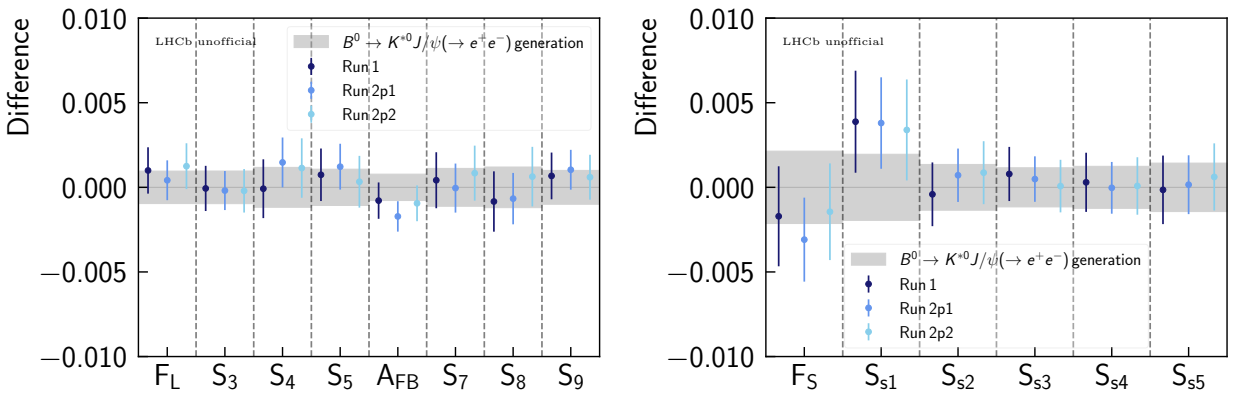


Figure 8.7: Differences between the results of the angular fits to the acceptance corrected control mode simulation samples and the generator level distribution for the three Run subsamples.

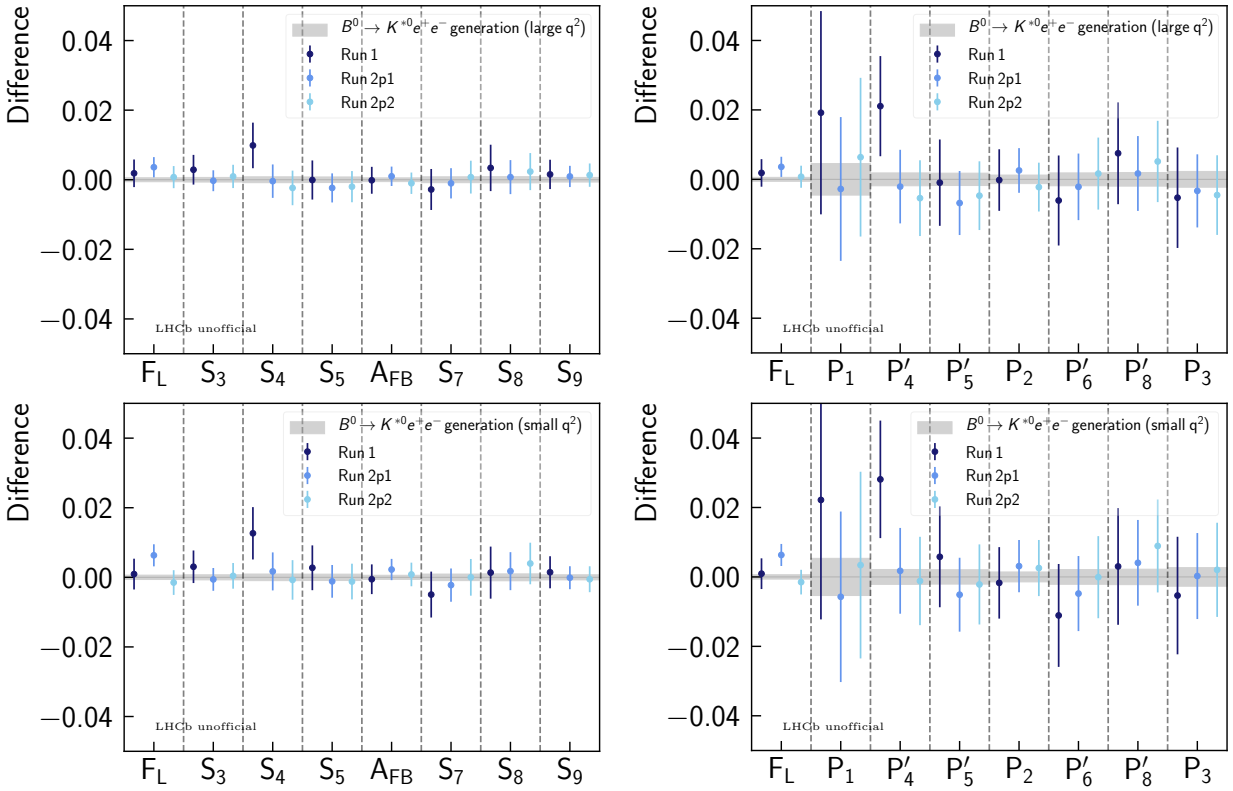


Figure 8.8: Differences between the results of the angular fits to the acceptance corrected rare mode simulation samples and the generator level distribution for the three Run subsamples. The results for the larger (smaller) q^2 ranges are shown in the top (bottom) row.

8.2 Components modelling

To separate signal from backgrounds, the mass distribution of the signal and the mass and angular distributions of the main background components need to be modelled. Effective acceptance weights (Section 8.1) are always included, as they are applied to all candidates passing the selection requirements. The signal mass distribution is described using piece-wise functions composed of Gaussian cores and power law tails. Background mass distributions are modelled using either exponential functions or kernel density estimation (KDE). Chebyshev polynomials up to second order are used for their angular distributions. Factorisation between mass and angles is assumed in all cases, and for the backgrounds, the three angles are also assumed to factorise. For the generation of pseudoexperiments, the background q_c^2 distributions are modelled in a factorised manner using Chebyshev polynomials up to second order.

8.2.1 Rare mode

The main background components present after the full selection include partially reconstructed decays, DSL decays, and combinatorial background, as illustrated in Figure 8.9. The strategies used to model these contributions are discussed in the sections below.

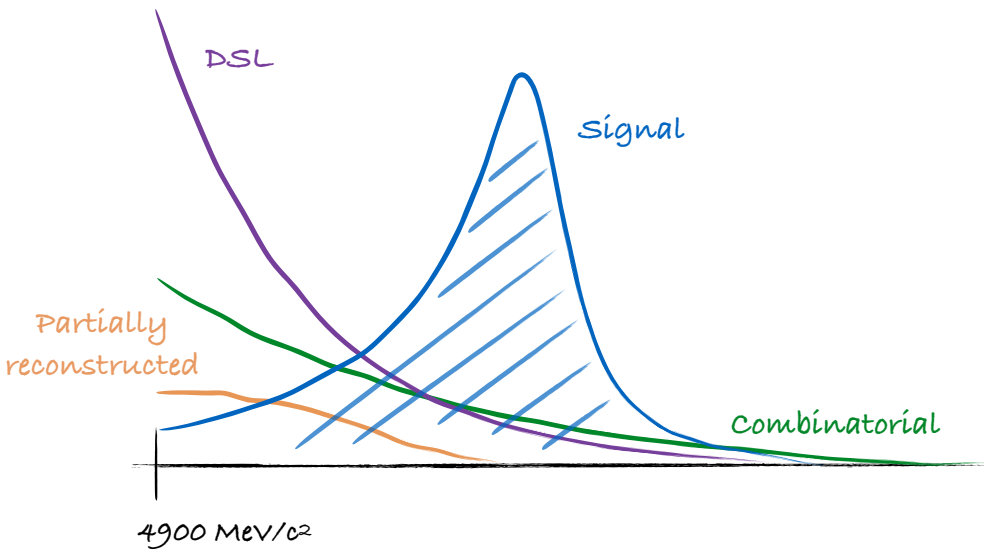


Figure 8.9: Illustration showing the mass distributions of the most prominent components present in the rare mode q_c^2 region after applying all selection requirements.

Signal mass distribution

The mass distribution of the signal is broadened by resolution effects. An empirical function that has been found to model this type of distribution well is the Crystal Ball (CB) function [173, 174], which consists of a power law tail joined to a Gaussian distribution. Here a modified version of this function, a double Crystal Ball (DCB) is used [175], where two power law tails

are joined to one shared Gaussian core, i.e.

$$\mathcal{P}_{\text{DCB}}(m|\mu,\sigma,\alpha_L,\alpha_R,n_L,n_R) = \begin{cases} \dfrac{a_L}{\left(\left|b_L - \frac{(m-\mu)}{\sigma}\right|\right)^{n_L}} & \text{if} \quad \frac{(m-\mu)}{\sigma} < -|\alpha_L| \\ e^{\frac{-(m-\mu)^2}{2\sigma^2}} & \text{if} \quad -|\alpha_L| < \frac{(m-\mu)}{\sigma} < |\alpha_R| \\ \dfrac{a_R}{\left(\left|b_R + \frac{(m-\mu)}{\sigma}\right|\right)^{n_R}} & \text{if} \quad \frac{(m-\mu)}{\sigma} > |\alpha_R| \end{cases} \tag{8.11}$$

with

$$\begin{aligned} a_{L,R} =& e^{-\frac{1}{2}|\alpha_{L,R}|^2} , \\ b_{L,R} =& \frac{|\alpha_{L,R}|}{n_{L,R}} \left(\frac{n_{L,R}}{|\alpha_{L,R}|} - |\alpha_{L,R}| \right) , \end{aligned} \tag{8.12}$$

where μ and σ are the mean and width of the Gaussian part of the function, the $n_{L\,(R)}$ governs the shape of the power-law tail on the left (right) hand side of the peak, and the $\alpha_{L,R}$ values set the locations at which the two tails are joined to the Gaussian function. The power law function on the left (lower mass) side of the mean is used to describe the tail due to bremsstrahlung energy loss, while the one on the right (higher mass) side is used to account for bremsstrahlung over-correction and stochastic tracking effects.

The $B^0 \to K^{*0}e^+e^-$ simulation is used to model the signal mass distribution. As its shape depends on the extent of the bremsstrahlung recovery, the samples are split into three categories labelled 0γ, 1γ, and 2γ, which are defined as

- 0γ – no bremsstrahlung recovery;
- 1γ – one electron received momentum correction from one or more photon clusters;
- 2γ – both electrons received momentum corrections from one or more photon clusters,

and fits using separate DCB functions are made to each, resulting in three pdfs denoted by $S_{0\gamma}(m)$, $S_{1\gamma}(m)$, and $S_{2\gamma}(m)$. The full mass model is then

$$S(m) = f_{0\gamma}S_{0\gamma}(m) + f_{1\gamma}S_{1\gamma}(m) + (1 - f_{0\gamma} - f_{1\gamma})S_{2\gamma}(m) , \tag{8.13}$$

where $f_{0\gamma}$ and $f_{1\gamma}$ are the fractions of candidates belonging to the 0γ and 1γ categories, respectively. The combined fit result is shown in Figure 8.10 for the Run 2p2 subsample.

While the fully corrected simulation is used, some residual simulation-data differences are expected to remain. To allow the mass model to accommodate these effects to some extent, the mean and width of the mass peak are allowed to vary via a shift parameter, δ_μ, which is added to the μ parameters of all three pdfs, and a scaling parameter, δ_σ, which is multiplied to their σ values.[6]

[6]In contrast to the modelling of the mass distribution, the effective acceptance functions are not parametrised separately for each bremsstrahlung category. This choice is largely driven by the statistics of the simulation – while the mass models are determined using one-dimensional fits, the full effective acceptance parametrisation involves four dimensions (without assuming factorisation). Further splitting would enlarge the low efficiency region. In addition, due to the reasonably good agreement between the fractions of the bremsstrahlung

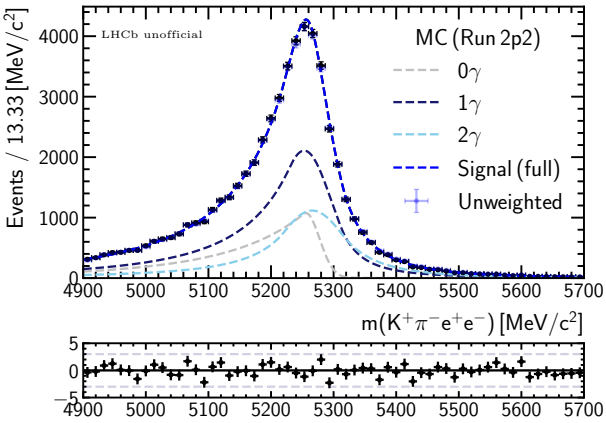


Figure 8.10: Results of fits to the mass distributions of simulated signal candidates of the Run 2p2 subsample. The full sample that includes all three bremsstrahlung categories is shown. The Run 2p2 effective acceptance function is used to produce per-event weights that are included in the fit, and the original (unweighted) distribution is also shown.

Combinatorial and double semileptonic backgrounds

The combinatorial and DSL backgrounds are obtained using a data-driven approach. While it is a common practise to use data for the former, as it contains multiple contributions and effects that are difficult to reproduce reliably in simulation, the choice of using it for the latter is motivated by the observed disagreement between simulation and data. Important causes of this include the presence of multiple modes, candidates that are partly combinatorial in nature (i.e. where one (or more) track(s) originate from unrelated source(s), such as the PV, but the combination still retains enough characteristics of the decay to distinguishes it from combinatorial background), and candidates with misidentification. Separating between combinatorial, DSL, and partly combinatorial and/or misidentified DSL-like backgrounds is inherently challenging, and further complicated by the potential presence of contributions from other sources with similar characteristics, such as residual misidentified semileptonic decays with or without combinatorial tracks. Therefore, a procedure to simultaneously obtain effective models for the DSL and combinatorial contributions is developed using data samples containing LFV $B^0 \to K^{*0}(\to K^+\pi^-)e^+\mu^-$ candidates. Details on the two backgrounds as well as the combined modelling procedure are given below.

Combinatorial background

In an ideal situation, the combinatorial background can be expected to be well described by a falling exponential distribution in the reconstructed B^0 mass, and to be distributed approximately uniformly or with mild curvature in the angles due to its random nature. Furthermore, full factorisation between mass and angles can be assumed. In practise, backgrounds of combinatorial nature that survive in data do not only contain the aforementioned 'pure' or standard combinatorial candidates, but also partly combinatorial contributions where enough particles from a decay remain correctly reconstructed such that some of its characteristics is retained. Examples include candidates where the hadron or lepton pairs originate from real K^{*0} or J/ψ resonances, and those where some (or most) of the particles come from a genuine DSL decay.

categories in simulation and data, the systematic uncertainties associated with this strategy is expected to be low compared to the sources considered.

The latter is one of the motivations for the combined parametrisation approach.

The modelling of this type of background is also complicated by two additional factors, the first of which is the veto against $B^+ \to K^+e^+e^-$ decays (Section 7.2.2), which introduces some correlation between mass and the $\cos\theta_K$ angle. Nevertheless, the impact on the measurement is limited due to the low levels of signal events in this background dominated region, and it is assessed as a source of systematic uncertainty (Section 9.2.5). The second is the cut on the q^2_c, which improves the resolution of true signal candidates, but has the side effect of distorting the distribution of combinatorial background. The extent of the impact is investigated using the $K\pi e\mu$ data sample with the DSL veto ($|\cos\theta_\ell| < 0.8$) applied. For the $m(K\pi ee)_{\rm PV}$, the distortion is found to be limited for the rare mode, but slightly more pronounced for the control mode as shown[7] in Figures 8.11 and 8.12 (top row). However, in both cases, the exponential function remains a good approximation. The (leptonic) same-sign data sample, $K^\pm\pi^\mp e^\pm e^\pm$, which does not contain true DSL decays, can also be used to check the combinatorial shape, although in this case a cut of $|m(K\pi ee)_{{\rm PV},J/\psi} - m(B^0)| > 100\,{\rm MeV/c^2}$ is applied to reduce contributions from control mode decays with incorrect electron charge assignments [67]. However, as the $m(K\pi ee)_{{\rm PV},J/\psi}$ obtained using the J/ψ mass constraint is correlated with the constrained q^2_c, this causes the unconstrained q^2 distribution to become distorted, especially in the control mode q^2 region. These results are shown in Figures 8.11 and 8.12 (bottom row).

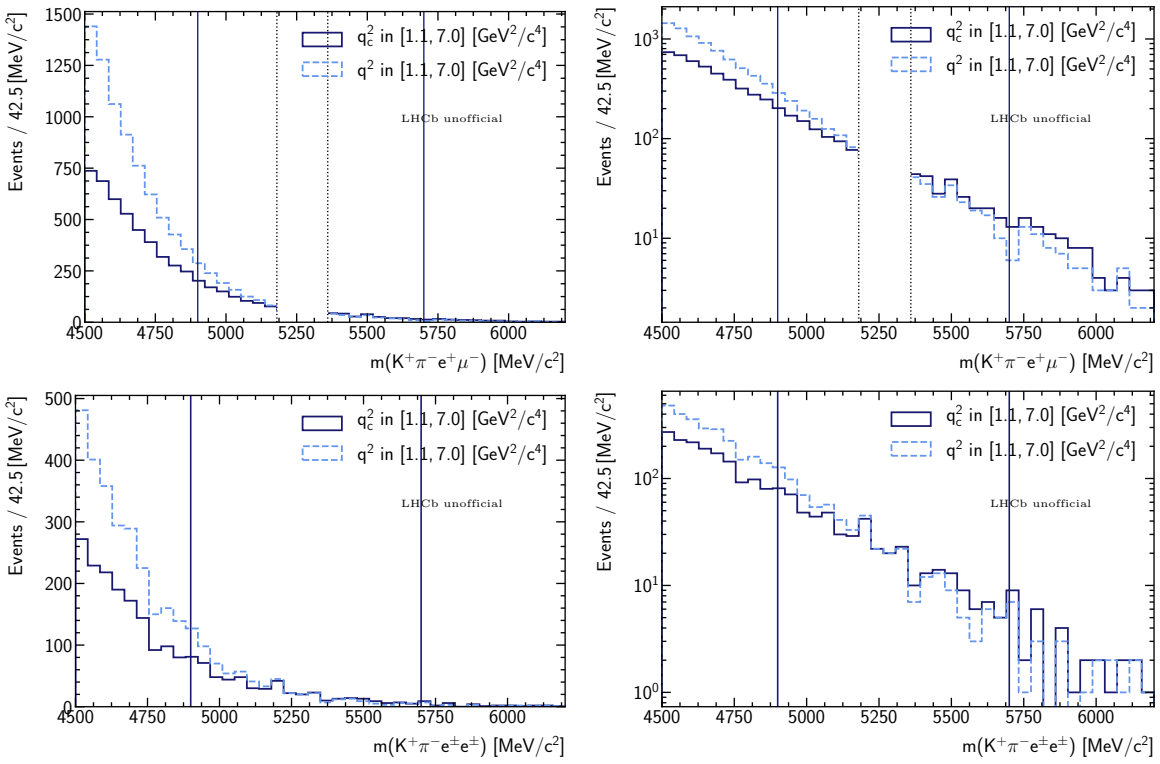


Figure 8.11: Constrained B^0 mass distributions of $K\pi e\mu$ (top) and same-sign (bottom) candidates passing signal selection and additional background vetoes. The distributions are displayed using a normal (left) and log-scale (right). The boundaries of the signal mass window are demarcated by the dark blue lines.

[7]Note that whenever the mass distribution of the $K\pi e\mu$ data is shown, the region surrounding the nominal B^0 mass is blinded. This is done due to the sensitivity of this region to LFV decays that may be observable in certain NP scenarios. Nevertheless, no contribution is expected as dedicated searches for the $B^0 \to K^{*0}e^+\mu^-$ decay using full Run 1 and Run 2 data did not reveal significant signal [176].

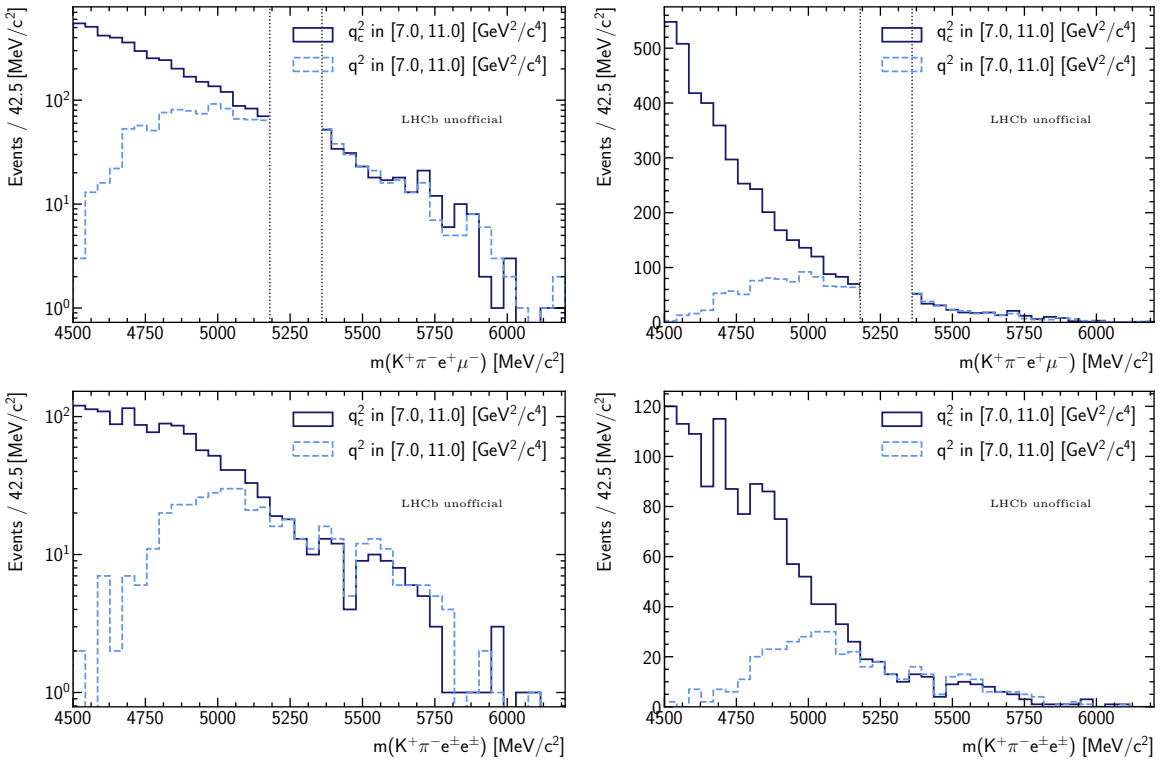


Figure 8.12: Constrained B^0 mass distributions of $K\pi e\mu$ (top) and same-sign (bottom) candidates passing control mode selection and additional background vetoes. The distributions are displayed using a normal (left) and log-scale (right). The boundaries of plot corresponds to the mass window used for the control mode.

Double semileptonic decays

In the context of this analysis, DSL decays refers to a group of decays where a B meson decays to a D meson, a lepton and an undetectable neutrino, and the D decays to a K^{*0}, a lepton, and another neutrino. Leading contributions of this type tend to have much larger branching fractions than the signal, and significant numbers remain in the data sample after preselection.

The dominant mode of $B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)e^-\bar{\nu})e^+\nu$ reconstructed as the signal, has a $\cos\theta_\ell$ distribution that peaks near $\cos\theta_\ell = 1$. Nevertheless, some modes can also contribute to the $\cos\theta_\ell = -1$ region, although these tend to be present at low levels. One example is the $B_s^0 \to D_s^-(\to \bar{K}^{*0}(\to K^-\pi^+)e^-\bar{\nu})e^+\nu$ decay, which is suppressed with respect to the dominant mode by both the ratio of hadronisation fractions ($f_s/f_d \sim 0.244$ [177]) and branching fractions (~ 0.2 [7]). However, when misidentified and/or combinatorial particles are used in the calculation of $\cos\theta_\ell$, peaks can form at both ends of its range. The central region around $\cos\theta_\ell = 0$ can also become populated if the imbalance in momentum between the two electrons disappears, because, for example, they no longer come from decaying particles with large differences in mass. Examples of these effects are illustrated qualitatively using simulation in Figure 8.13. While the single mode simulation cannot be expected to replicate the full spectrum of contributions in data, it can provide useful indications on their characteristics.

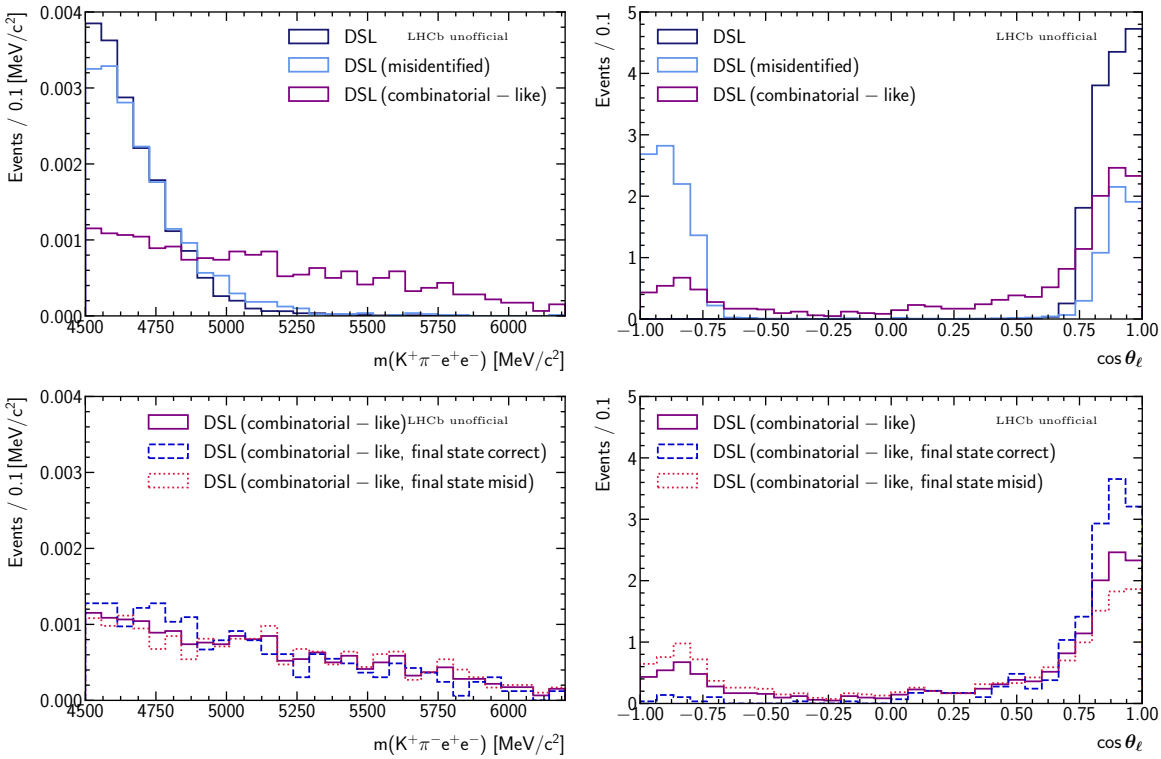


Figure 8.13: Mass (left) and $\cos\theta_\ell$ distributions of different types of 'DSL' contributions illustrated using simulation (with the nominal PID and MVA cuts removed). Properly reconstructed DSL signal is shown in dark blue (top row), together with candidates where misidentification has taken place, and candidates that are combinatorial-like. The last category is broken down further (bottom row) into combinatorial-like candidates with and without the misidentification of final state particles.

In general, this type of background can be modelled using Chebyshev polynomials up to second order for the angles of $\cos\theta_K$, and ϕ (also q_c^2), and KDE for $\cos\theta_\ell$. While in principle a cocktail of DSL modes can be generated to try and reproduce the data distribution, in practise this is challenging as the composition of what appears to be 'DSL' is not well known and may contain non-negligible contributions from sources other than the standard DSL decays, and candidates reconstructed using combinatorial track(s). This motivates a data-driven approach based on the LFV samples. Using the data bypasses, to some extent, the need to disentangle and understand each separate contribution, as it can be expected to naturally contain almost all relevant ones. For example, the $K\pi e\mu$ final state would include not only true DSL decays such as $B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)\mu^-\bar{\nu})e^+\nu$, but also candidates with combinatorial tracks and misidentification, as well as possible semileptonic decays. However, an important drawback is that experimental differences between having a muon-electron pair versus two electrons in the final state (e.g. triggering, resolution, different misidentification rates) then constitutes a source of systematic uncertainty (Section 9.2.1). Note that the same-sign sample does not contain true DSL decays and can only be used to study backgrounds that are mainly of combinatorial nature.

Data-driven effective models

The $K\pi e\mu$ sample contains combinatorial and DSL events, which includes additional contributions with DSL-like distributions. Separation between the two can be achieved to some extent using the $\cos\theta_\ell$ distribution. However, further splitting to isolate the different types of DSL-like contributions is expected to be difficult, therefore a strategy is devised to obtain

effective models for the 'DSL' component, which includes all DSL-like contributions, and for the 'combinatorial', which would take in all backgrounds that are not included in the former. This is done using a 'two-step' procedure, where, in Step 1, the angular shape of the DSL is first obtained with a tight cut on the MVA to strongly suppress combinatorial background. Then, in Step 2, a fit is made to the data sample with the nominal MVA cut. This is used to determine the angular distributions of the effective combinatorial background, as well as the mass slopes of both components. More specifically, this procedure is carried out as follows:

- **Step 1** – the angular model of the DSL component is determined using candidates passing a very tight cut on the multivariate classifier of MVA > 0.9985. A KDE lineshape is used to model the $\cos\theta_\ell$ distribution, while Chebyshev polynomials up to second order are used for the other angles. To increase statistics for parametrisation, a mass range that is larger than the nominal rare mode mass window of $4500 < m(K\pi ee)_{\rm PV} < 5700\,{\rm MeV/c^2}$ is used. This step is necessary to ensure fit stability at low sample statistics where it is unfeasible to determine the shapes of both components together. As the MVA is known to affect the mass distribution much more strongly than the angles, the slopes of the mass distributions are taken from the Step 2 fit.

- **Step 2** – a fit is made to the mass and angles of candidates passing the nominal selection criteria, with the data used in Step 1 removed. In this fit, the angular shape of the DSL component is fixed, while that of the combinatorial is allowed to vary. The slopes of the exponential functions used to describe the mass distributions of both components are also allowed to vary. The nominal mass range of $4900 < m(K\pi ee)_{\rm PV} < 5700\,{\rm MeV/c^2}$ is used.

The nominal models obtained from this two-step procedure are shown in Figures 8.14 (Step 1) and 8.15 (Step 2) for the Run 2p2 subsample as an example.

Partially reconstructed background

Backgrounds from partially reconstructed decays can come from many sources, the most prominent of which include modes featuring excited kaon states, such as $B^+ \to K_1^+ e^+ e^-$ and $B^+ \to K_2^+ e^+ e^-$, where both the K_1^+ and K_2^+ can decay to $K\pi X$ final states, where X refers to one or more pions (or kaons). These contributions show up as broad peaking structures in the region below the B^0 invariant mass due to the unreconstructed particles, and cannot be easily separated. Like the DSL background, while it is possible to emulate a mixture of modes using a combination of different simulations, the large number of possible decays and the present state of knowledge of the $K\pi\pi$ system in data makes this challenging to accomplish. Instead, a data-driven approach is used to correct the generator level distributions of the phase-space simulation of $B^+ \to K^+\pi^+\pi^- e^+ e^-$ decays [178]. This is achieved by training a BDT reweighter to modify the generator level distributions of $m(K^+\pi^+\pi^-)$, $m(K^+\pi^-)$ and $m(\pi^+\pi^-)$ using per-event weights in order to replicate the $K\pi\pi$ spectrum of the background-subtracted and efficiency-corrected $B^+ \to K^+\pi^+\pi^- J/\psi(\to \mu^+\mu^-)$ decays in data. This simulation sample is reconstructed under the $B^0 \to K^{*0} e^+ e^-$ hypothesis, and the post-reconstruction events are matched to their corresponding generator level ones to obtain the associated weights. Note that the reweighting is not performed explicitly for the angles, and it does not take all possible degrees of freedom of the system into account. This constitutes a source of systematic uncertainty that is studied using physics simulations of the dominant modes of $B^+ \to K_1^+(1270) e^+ e^-$ and $B^+ \to K_2^{*+}(1430) e^+ e^-$ in Section 9.2.1.

The modelling of this background is carried out using a KDE for the mass distribution, and Chebyshev polynomials up to second order for the angles. The result of the fit to the Run 2p2

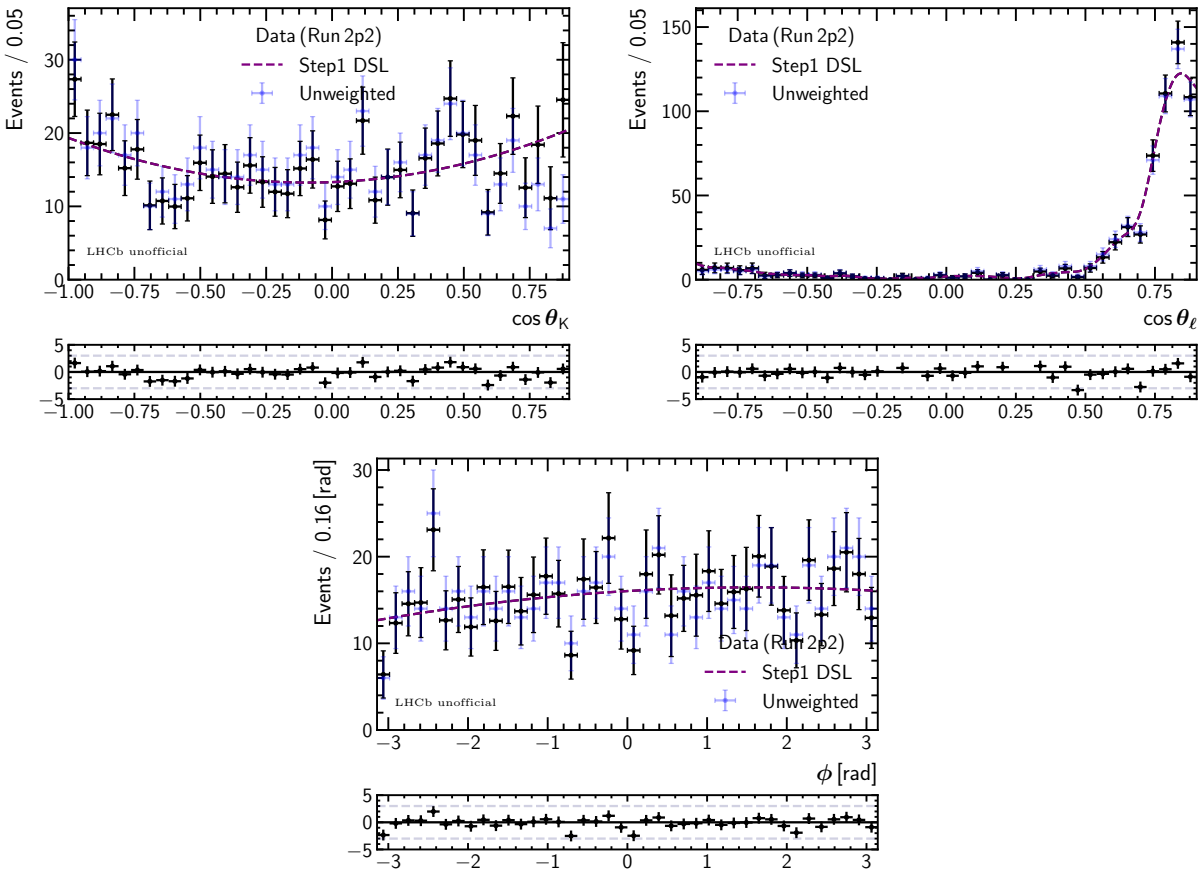


Figure 8.14: Results of the Step 1 fit to the Run 2p2 $K\pi e\mu$ data sample to extract the angular shape of the DSL component. A tight MVA cut is applied to strongly suppress combinatorial background. The fit includes per-event effective acceptance weights.

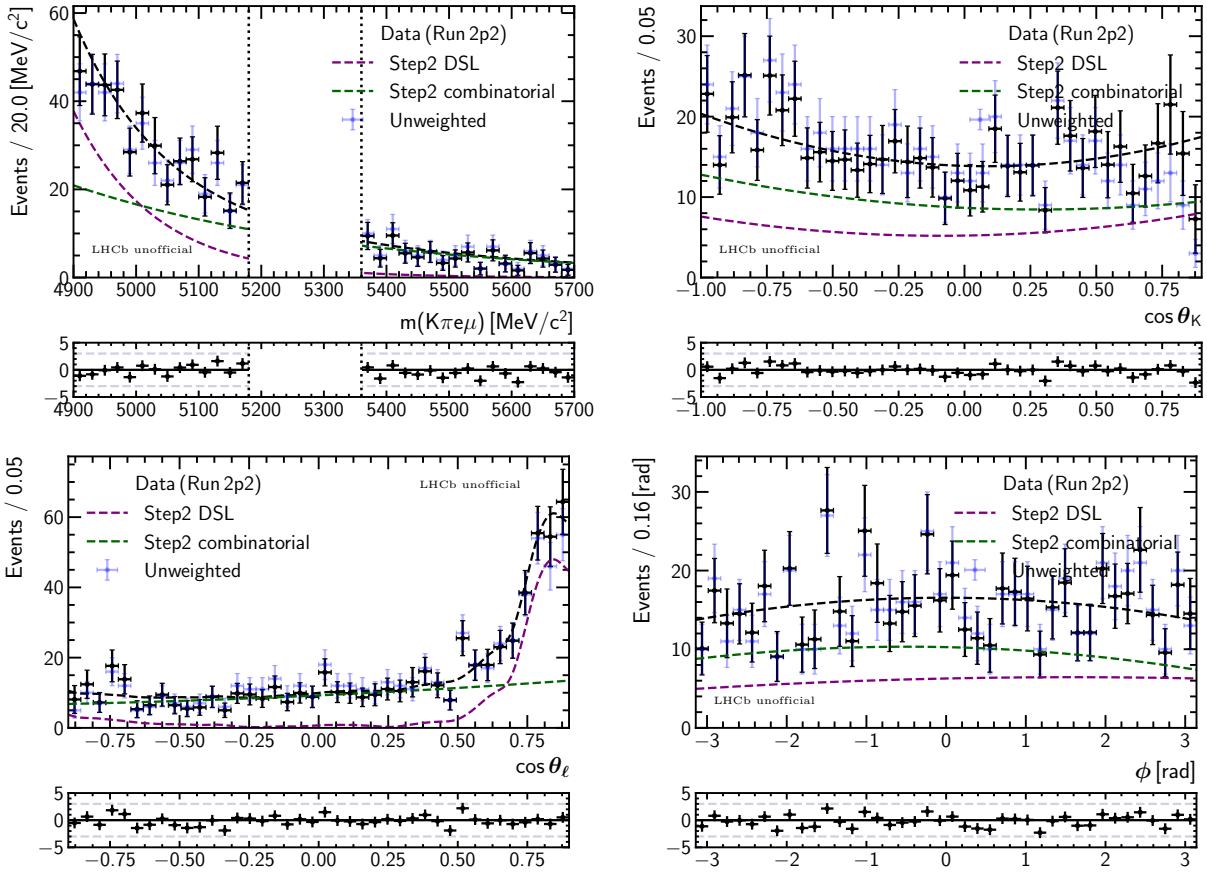


Figure 8.15: Results of the Step 2 fit to the Run 2p2 $K\pi e\mu$ data sample at the nominal MVA cut to determine the angular shape of the combinatorial component, and the slopes of both combinatorial and DSL backgrounds. The fit includes per-event effective acceptance weights.

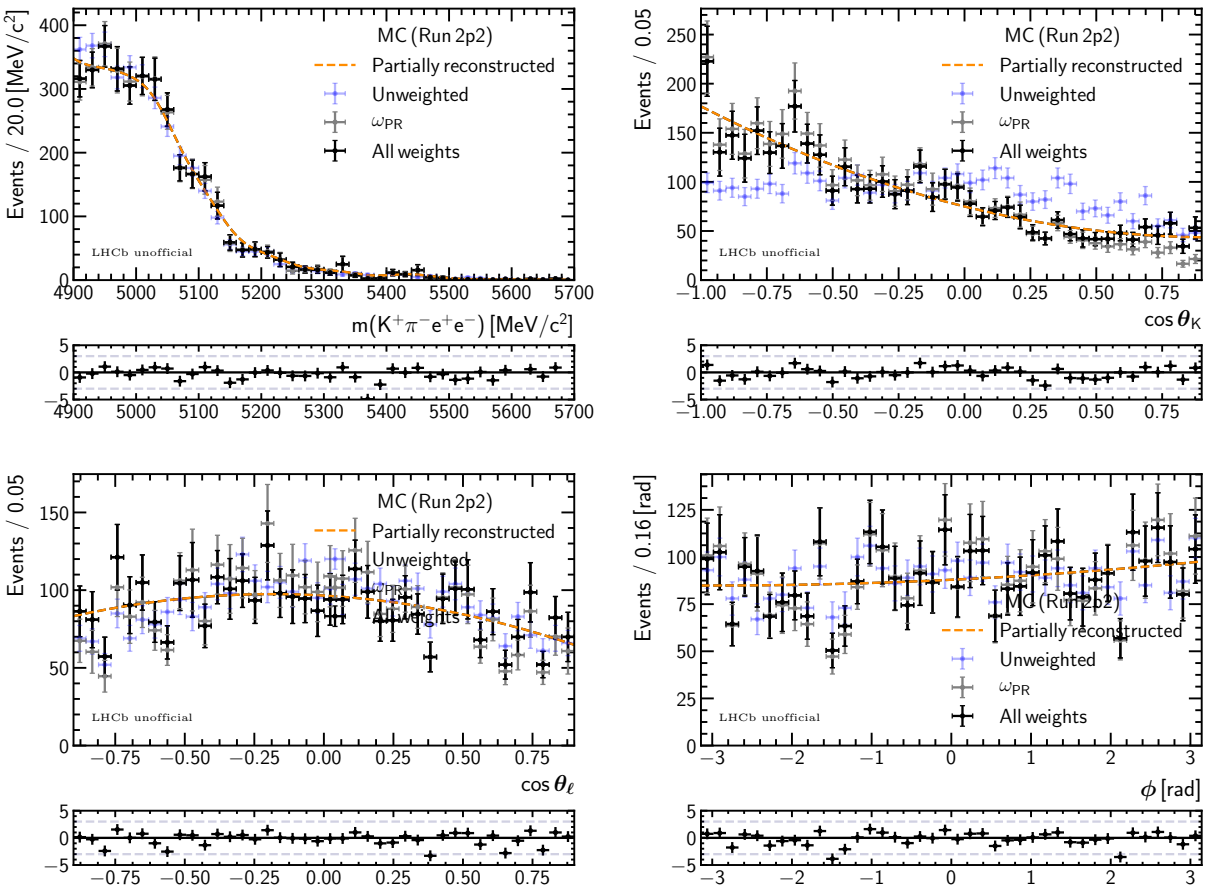


Figure 8.16: Result of the fit to the mass and angular distributions of the Run 2p2 phase-space $B^+ \to K^+\pi^+\pi^-e^+e^-$ simulation reconstructed as the signal. The unweighted sample is shown in blue, the distribution after including the data-driven correction weights ($\omega_{\rm PR}$) is shown in gray, and the final distribution used for modelling that includes both $\omega_{\rm PR}$ and acceptance weights is shown in black.

subsample is shown in Figure 8.16.

8.2.2 Control mode

The control mode of $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ is used for two main reasons, the first of which is to enable the stable determination of shift and scaling parameters of the signal mass peak, and the second is to validate the angular fit strategy. A mass-only fit is done for the former, while an angular fit is made for the latter. The configuration of the nominal mass fit is the focus of this section. Details concerning the validation fit, which does not contribute to the nominal result, are discussed further in Section 9.1.

Due to the larger branching fraction of the control mode decay, and to a lesser extent its higher selection efficiency, both its signal yield and the signal-to-background ratio exceeds that of the latter. The main components considered in the mass fit, illustrated in Figure 8.17, include misidentified $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$ decays, combinatorial background and $B_s^0 \to K^{*0}J/\psi(\to e^+e^-)$ decays.

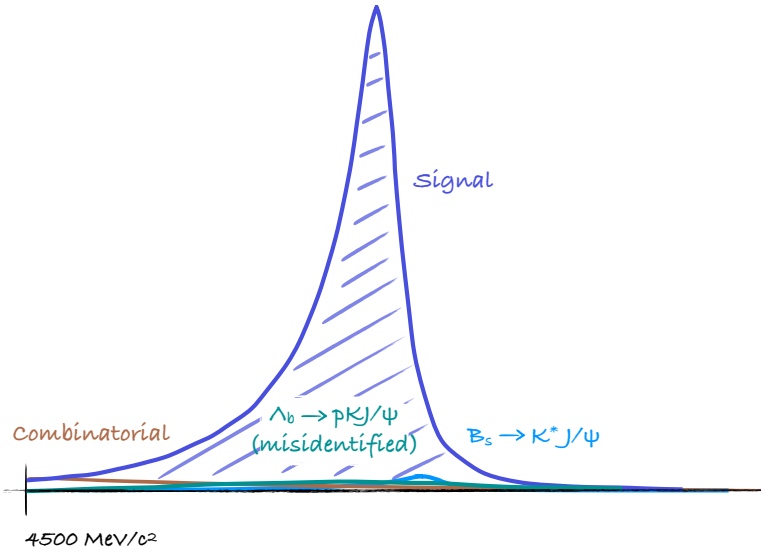


Figure 8.17: Illustration showing the mass distributions of the most prominent components present in the control mode q_c^2 region after applying all selection requirements.

Signal mass distribution

The mass window used for the control mode is enlarged ($4500 - 6200\,\mathrm{MeV/c^2}$) compared to that of the rare mode ($4900 - 5700\,\mathrm{MeV/c^2}$) in order to improve signal-background separation. Access to larger numbers of background events in the lower mass region, and the use of the high mass region, which mainly contains combinatorial candidates, adds stability to the fit. However, the tails of the signal mass distribution are slightly distorted. The high mass region, in particular, is affected by the veto against over-reconstructed B^+ decays. To improve the description of the tails, an extra Gaussian pdf is used for the bremsstrahlung categories of 1γ and 2γ. The mass model is otherwise identical to that of the rare mode.

The results of the fits to the fully corrected simulation with the inclusion of acceptance weights are shown in Figure 8.18 for the Run 2p2 subsample. Note that the larger pulls visible here, though not in the rare mode (Figure 8.10), is partly due to the imperfect modelling of the tails, and partly to the limitations of the models used, which becomes striking at the high statistics of the simulation. This effect has been seen in studies performed for similar analyses. While in principle using alternative modelling strategies may lead to improvements, the benefit is limited, as its impact on the measurement is marginal compared to the sources of systematic uncertainties considered.

Backgrounds

Backgrounds present in the control mode data sample include three main components that are relatively smoothly distributed in $m(K\pi ee)_{\mathrm{PV}}$, namely the combinatorial, DSL, and residual partially reconstructed candidates that survived the $m(K\pi ee)_{\mathrm{PV,J/\psi}} < 5150\,\mathrm{MeV/c^2}$ cut. Sources of backgrounds that peak near the nominal B^0 mass include signal with $K - \pi$ swap ($B^0 \to K^{*0}(\to K_{\to\pi}\pi_{\to K})J/\psi$), $B^0 \to \phi(\to K^+K^-)J/\psi$ decays with $K \to \pi$ misidentification, and $\Lambda_b^0 \to pKJ/\psi$ decays with the misidentification of the proton as a pion, or the misidentification of the proton as a kaon and the kaon as a pion. One additional peaking component comes from the decay of B_s^0 mesons to the same final state as the signal, i.e. $B_s^0 \to K^{*0}J/\psi(\to e^+e^-)$. Given that this mode is strongly suppressed by its lower hadronisation and branching fractions,

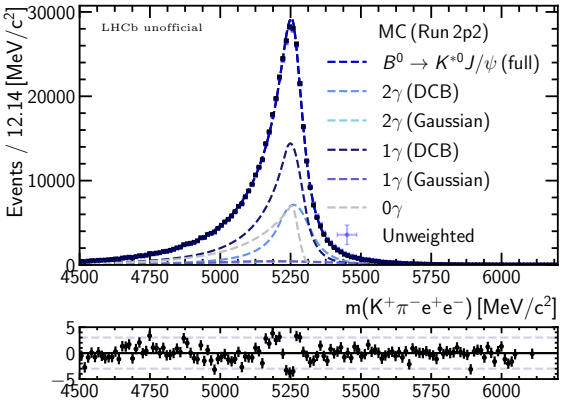


Figure 8.18: Results of fits to the mass distributions of simulated control mode signal candidates of the Run 2p2 subsample. The full sample that includes all three bremsstrahlung categories is shown together.

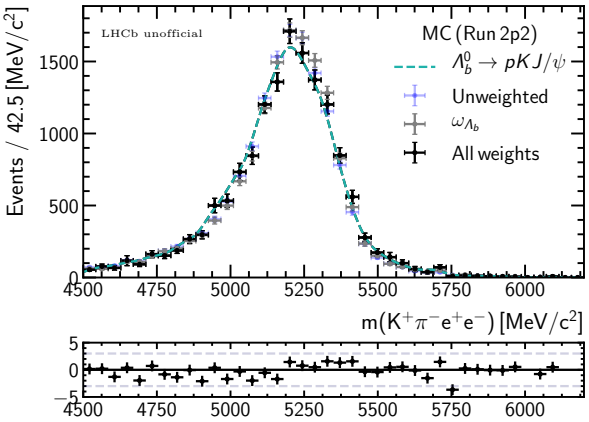


Figure 8.19: Non-parametric (KDE) model of the mass distribution of simulated misidentified $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$ decays from the Run 2p2 samples. The effect of the data-driven correction weights (ω_{Λ_b}) is shown separately in gray.

and is expected to have similar angular and mass distributions as that of the signal (with the mean of the latter shifted to the nominal mass of the B_s^0), it can be described using the same model.

Due to the relatively low levels of background after selections, the non-peaking components cannot be separated from each another reliably. They are therefore all incorporated into an effective 'combinatorial' component. Among the peaking backgrounds, the dominant source is comprised of misidentified Λ_b^0 decays. This contribution is modelled using phase-space $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$ simulation reconstructed as $B^0 \to K^{*0}J/\psi(\to e^+e^-)$. Due to the presence of multiple Λ_b resonances in data, as well as exotic (pentaquark) states [179], data-driven corrections weights [67] obtained from the $\Lambda_b^0 \to pKJ/\psi(\to \mu^+\mu^-)$ mode are used to reweigh the mass distributions of $m(K\pi)$ and $m(J/\psi p)$. This method is analogous to the one used for the partially reconstructed background of the rare mode. A KDE is used to model the weighted mass distributions, taking both the correction and the effective acceptance weights into account. The result of this is shown in Figure 8.19.

8.3 Angular fit

To measure the angular observables, a maximum likelihood fit is made to the (constrained) reconstructed invariant mass of the B^0 candidates and the three angles with the inclusion of per-event correction weights from the effective acceptance functions. The fit configuration, and the strategies used to ensure stability at low statistics and in the presence of multiple backgrounds, are discussed below. In addition, comments are made on relevant features of weighted fits.

8.3.1 Weighted unbinned maximum likelihood fit

A weighted unbinned maximum likelihood fit is made to extract the values of the angular observables from data. Analytically normalised pdfs are used for the PV constrained B^0 invariant mass ($m(K\pi ee)_{\rm PV}$) and the angular distributions of the signal and backgrounds ($\bar{\Omega} = \cos\theta_K, \cos\theta_\ell, \phi$). The full pdf that includes models for the signal and n background components can be written as

$$\begin{aligned}\mathcal{P}(m(K\pi ee)_{\rm PV}, \bar{\Omega}) = f_S P_S(m(K\pi ee)_{\rm PV}, \bar{\Omega}) \\ + \sum_i^{n-1} f_{B_i} P_{B_i}(m(K\pi ee)_{\rm PV}, \bar{\Omega}) \\ + (1 - f_S - \sum_i^{n-1} f_{B_i}) P_{B_n}(m(K\pi ee)_{\rm PV}, \bar{\Omega}) \ ,\end{aligned} \tag{8.14}$$

where f_S refers to the signal fraction and f_{B_i} is the fraction of background component i. To preserve the normalisation of the full pdf, $\mathcal{P}$, the fraction of one background component (the combinatorial), denoted by n, is taken as the remainder after subtracting all other fractions from unity. Factorisation between mass and angles is assumed for all components, and between the three angles for all background components, that is,

$$\begin{aligned}P_S(m(K\pi ee)_{\rm PV}, \bar{\Omega}) &= P_S(m(K\pi ee)_{\rm PV}) \cdot P_S(\bar{\Omega}) \ , \\ P_B(m(K\pi ee)_{\rm PV}, \bar{\Omega}) &= P_B(m(K\pi ee)_{\rm PV}) \cdot P_B(\cos\theta_K) \cdot P_B(\cos\theta_\ell) \cdot P_B(\phi).\end{aligned} \tag{8.15}$$

The weighted negative log-likelihood to be minimised with respect to the observables of interest, denoted by S_i, is given by

$$-\ln\mathcal{L} = -\sum_{e=1}^{N} \omega_e \ln\mathcal{P}(S_i, \theta; m(K\pi ee)_{\rm PV}, \bar{\Omega}) \ , \tag{8.16}$$

where θ refers to the other (nuisance) parameters that are allowed to vary in the fit, such as the background fractions, ω_e are the per-event acceptance weights, and the summation extends over all N events of the data sample. The ω_e are calculated for each event using the effective acceptance function via,

$$\omega_e = 1/\epsilon(\bar{\Omega}_e, q^2_{ce}) \tag{8.17}$$

where $\bar{\Omega}_e$ and q^2_{ce} are the values of the angles and constrained q^2_c of event e.

8.3.2 Fit strategy

The rare mode angular fit is carried out for low statistics samples and in the presence of three significant backgrounds that are challenging to separate. To achieve good stability, most parameters describing the background distributions are fixed to values found from simulation or background-rich data samples. Only the slope of the exponential function describing the mass distribution of the combinatorial background is allowed to vary. In addition, as it is difficult to determine the amount of partially reconstructed decays in the presence of significant levels of combinatorial background (and this type of physics background can be expected to scale approximately with the signal yield), its fraction with respect to the signal is shared for all Run subsamples. That is, $f_{\rm PR}$ is reformulated as $f^{\rm S}_{\rm PR} \cdot f_{\rm S}$, and a single $f^{\rm S}_{\rm PR}$ parameter is allowed to vary.

The values of the shift and scaling parameters of the signal mass peak, which accommodate some of the remaining simulation-data differences, are fixed from separate mass-only fits of the control mode candidates. In this fit, the B^0_s decay is modelled using the same mass distribution as that of the control mode signal, but with a shift in the mean of $+87.2\,\mathrm{MeV/c^2}$, corresponding to the difference between the nominal masses of the B^0 and the B^0_s mesons [7]. The contribution from this highly suppressed mode, estimated via

$$f^S_{B^0_s} = \frac{f_s}{f_d} \cdot \frac{\mathcal{B}(B^0_s \to K^{*0} J/\psi)}{\mathcal{B}(B^0 \to K^{*0} J/\psi)} \,, \tag{8.18}$$

is fixed in the fit.[8] Contamination from misidentified Λ^0_b decays, suppressed by the small ratio of $f_{\Lambda^0_b}/f_d$ [177] (and the misidentification probability), is low. Contribution from this mode is similarly fixed in the fit, and is estimated via

$$f^S_{\Lambda^0_b\,\mathrm{misid}} = \frac{f_{\Lambda^0_b}}{f_d} \cdot \frac{\epsilon_{\Lambda^0_b\,\mathrm{misid}}}{\epsilon_{\mathrm{control}}} \cdot \frac{3}{2} \cdot \frac{\mathcal{B}(\Lambda^0_b \to pKJ/\psi)}{\mathcal{B}(B^0 \to K^{*0} J/\psi)}, \tag{8.19}$$

where the efficiencies of $\epsilon_{\Lambda^0_b\,\mathrm{misid}}$ and $\epsilon_{\mathrm{control}}$ are obtained from simulation, and the fraction of 3/2 comes from the Clebsch-Gordon factor for the decay of the vector K^{*0} meson to two pseudoscalar mesons. The combinatorial component of the control mode, which is expected to also contain residual partially reconstructed background and DSL decays, is modelled using a decreasing exponential function, and its slope is allowed to vary.

To address variations in, for example, the operating conditions, trigger thresholds and PID performances over the years, separate effective acceptance functions, signal mass models, background models and shift and scaling parameters are used for each Run period.[9] The final angular fit is performed simultaneously to all three data subsamples. The full log-likelihood is

[8] Note that here some simplifications are made, namely that the selection efficiencies for the two modes are assumed to be the same, and that the branching fractions refer to the P-wave components only, while contributions from S-wave, and S and P interference, are present in data, and cannot be separated from the P-wave 'signal'. Nevertheless, the impact of these assumptions is expected to be marginal.

[9] Note that while treating the two L0 (or even bremsstrahlung) categories separately has multiple advantages (e.g. improved background modelling, effective acceptance correction), this is not taken as the nominal strategy, as achieving good stability given the limited statistics is a priority. Needless to say, a more precise analysis of this mode using larger statistics, for example, would benefit significantly from this type of optimisation.

given as the sum of the log-likelihoods of the three subsamples,

$$
\begin{aligned}
-\ln\mathcal{L}_{\text{rare}} = &-\sum_{\text{e}_1=1}^{N_{\text{Run1}}} \omega_e \cdot \ln\mathcal{P}(S_i,\, \theta^{\text{Run 1}},\, f_{\text{PR}}^{\text{S}};\, m(K\pi ee)_{\text{PV}},\, \bar{\Omega}) \\
&-\sum_{\text{e}_2=1}^{\text{Run 2p1}} \omega_e \cdot \ln\mathcal{P}(S_i,\, \theta^{\text{Run 2p1}},\, f_{\text{PR}}^{\text{S}};\, m(K\pi ee)_{\text{PV}},\, \bar{\Omega}) \;, \\
&-\sum_{\text{e}_3=1}^{\text{Run 2p2}} \omega_e \cdot \ln\mathcal{P}(S_i,\, \theta^{\text{Run 2p2}},\, f_{\text{PR}}^{\text{S}};\, m(K\pi ee)_{\text{PV}},\, \bar{\Omega})
\end{aligned}
\tag{8.20}
$$

where $N_{\text{Run 1}}$, $N_{\text{Run 2p1}}$ and $N_{\text{Run 2p2}}$ refer to the events of the Run 1, Run 2p1 and Run 2p2 data samples, and $\theta_{\text{Run 1}}$, $\theta_{\text{Run 2p1}}$ and $\theta_{\text{Run 2p2}}$ refer to the nuisance parameters used, which differ for the Run periods with the exception of f_{PR}^{S}, which is shared. The shift and scaling parameters for the three subsamples are taken from three separate control mode mass fits.

Altogether, the parameters allowed to vary in the control mode mass fits are the shift and scaling parameters, and the combinatorial slopes. The parameters varied in the rare mode fit are the eight shared angular observables, three signal fractions, three DSL fractions, one partially reconstructed background fraction (with respect to the signal), and three slopes for the combinatorial backgrounds.

8.3.3 Weighted fit features

The angular fit is performed with per-event correction weights from effective acceptance functions. The use of weights gives rise to two main features, which have consequences for the analysis, the first of which is the potential difference between the component fractions determined from a weighted fit and the 'true' fractions of signal and background components, and the second is the random occurrence of pathological weights. The nature of these features and their impact will be discussed in the sections below.

Weighted fit fractions

If the number of (unweighted) events in a sample, N, can be described by a Poisson distribution with expectation value and variance of $E(N) = V(N) = \lambda$, where λ denotes the predicted number of events, after the addition of per-event weights, ω, the sum of the weights can generally be described by a compound Poisson distribution (CPD) [180]. In the limit of $\mu \to \infty$, by the central limit theorem, the CPD can be described by a Gaussian distribution with a mean, μ, and variance, σ^2, equalling

$$
\begin{aligned}
\mu &= E(N_\omega) = \lambda E(\omega) \;, \\
\sigma^2 &= V(N_\omega) = \lambda E(\omega^2) \;,
\end{aligned}
\tag{8.21}
$$

where $N_\omega = \sum_e^N \omega_e$ is the sum of the weights. For a data sample composed of a signal and a background component, the expectation value of the sum of weighted events can be written as

$$
E(N_\omega) = E(N_\omega^S + N_\omega^B) = \lambda_S E(\omega_S) + \lambda_B E(\omega_B), \tag{8.22}
$$

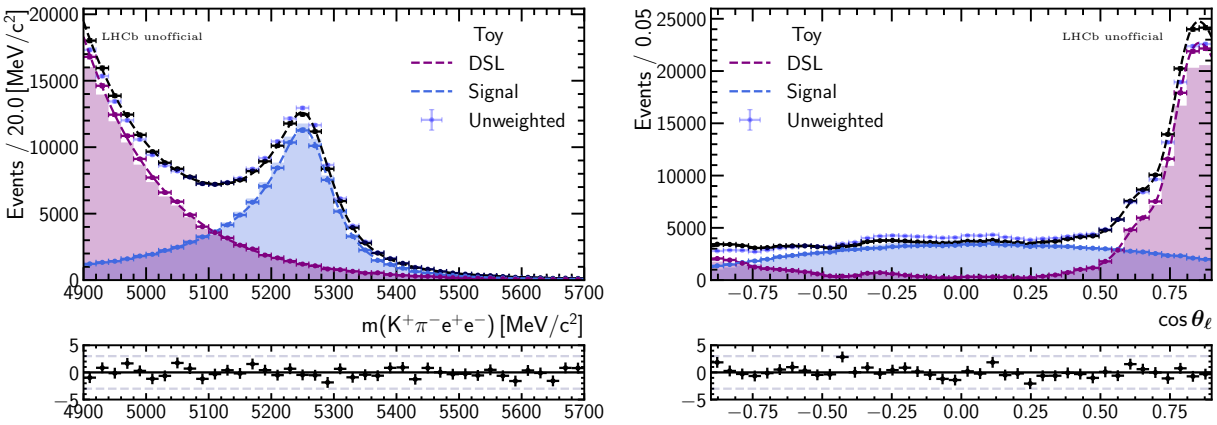


Figure 8.20: Result of the fit to a large statistics toy generated with the acceptance effect and fitted with correction weights (black). The unweighted total distribution is shown in light blue. The filled histograms show the unweighted distributions of the two components, and the coloured points show their weighted versions. Compared to the signal, more DSL events are distributed in a region where larger correction weights are assigned (edge of $\cos\theta_\ell$), therefore the weighted signal fraction is decreased from its unweighted value (the signal is 'weighed down', while the DSL is 'weighed up').

where S and B refer to the signal and background events, respectively. In general, $E(\omega_S) \neq E(\omega_B)$ for the correction weights, as the two samples populate the phase space differently. In this case, assuming perfect signal-background separation, the fractions of the two components determined from the fit are given by N^S_ω/N_ω and N^B_ω/N_ω. These are generally not equal to the unweighted fractions of signal and background events present in the sample of λ_S/λ and λ_B/λ. An example of this effect is illustrated using a high statistics toy featuring the signal and DSL components generated with the inclusion of the effective acceptance function (multiplied to the full pdf). The latter is chosen among the backgrounds as it tends to receive larger correction weights on average due to its $\cos\theta_\ell$ distribution. This toy has an unweighted signal fraction (N_S/N_{total}) of 0.509. The result of the weighted fit found 0.485 ± 0.001 instead. An adjusted value of 0.509 ± 0.002 can be obtained by dividing the weighted yields by the expectation values of the weights for both types of components, which can be estimated using high statistics pseudoexperiments. Note that an unweighted fit made for a toy generated without the acceptance effect produces a value of 0.508 ± 0.001 directly for a generation value of 0.508. The distorting effect of the acceptance weights can be seen in Figure 8.20.

Despite the existence of this feature, it is of no concern for the angular analysis. Component yields are only used for pseudoexperiment generation. In this case, corrections are made, which are, nevertheless, only approximate at low statistics. While conceptually interesting, these adjustments are found to be small in all cases, and their effect on the results is marginal.

Pathological weights

The effective acceptance functions may not be positive across the full phase space due to the limited statistics of the samples used for parametrisation. This causes some events to receive negative correction weights, which are not meaningful. The removal of these events leads to a systematic uncertainty, as the resulting 'holes' in the phase-space are not taken into account in the normalisation of any pdfs. Another type of pathological weights consist of very large

correction weights. They originate from low efficiency regions, which are often located close to where the function becomes negative. Large weights are less clearly defined and can also cause complications, especially since their distribution is not fully random. A few events with large weights may be sufficient to cause noticeable impact on the log-likelihood, and lead to enlarged uncertainties and bias. An example of an angular fit made to the Run 1 rare mode simulation samples without the phase space cut is shown in Figure 8.21, and the differences between the observable values found (with and without the removal of a small number of events) and those of the generator level fit are shown in Figure 8.22.

The 'origin' of large weights can be illustrated with a simple (binned) example, using ensembles of 'toys' that are generated to be uniformly distributed in a generic quantity that ranges between zero and ten in the absence of any 'acceptance effect'. This is akin to the generator level sample. After a particular 'selection' process, all 'events' with values ranging from zero to nine are unaffected, whereas those above nine are retained with a certain efficiency, ϵ. In practise, this effect is introduced to the sample by iterating over all events within this range, and keeping them if the value drawn at random between zero and one from a uniform distribution is lower than ϵ. This is done to mimic the effect of an acceptance function that is smooth (relative efficiency not low) in most regions of the phase-space, but has a small number of very low efficiency regions. The shape of the distorted 'post-selection' sample can be corrected using per-event weights corresponding to $1/\epsilon$ for events in the last bin (nine to ten). At high statistics (where the Gaussian approximation holds), the mean and standard deviation of the sum of the weights in the last bin are given by

$$\begin{aligned} \mu &= N_{\text{total}}/n_{\text{bin}} \cdot (1/\epsilon) \ , \\ \sigma &= \sqrt{N_{\text{total}}/n_{\text{bin}} \cdot (1/\epsilon)} \ , \end{aligned} \tag{8.23}$$

based on Equation 8.21, where N_{gen} is the (Poisson varied) yield used in generation, n_{bin} refers to the number of bins (in this case $n_{\text{bin}} = 10$).

Depending on the generation statistics and the value of the efficiency the correction strategy can be very successful, or it can almost always fail to retrieve the generator level shape (the correction strategy begins to break down as the number of post-selection events goes to zero). An example featuring a non-pathological case where N_{gen} is large and ϵ is not too small ($N_{\text{gen}} = 100000$ and $\epsilon = 0.4$) is shown in Figure 8.23. When a small value for N_{gen} is combined with a low ϵ (e.g. $N_{\text{gen}} = 100$ and $\epsilon = 0.01$), it is not possible to retrieve the generator level distribution, as the correction weights will always under or over-correct (Figure 8.24). In such cases, perhaps it is justifiable to veto events with large weights. However, it is difficult to determine when the corrections are no longer effective, as large weights can also appear in less problematic situations where the correct generator level values can still be retrieved (approximately) on average, as illustrated by a case featuring moderate statistics and limited efficiency ($N_{\text{gen}} = 10000$ and $\epsilon = 0.05$) (Figure 8.25).

Large weights can be ignored, although this is generally not desirable if their impact is significant. Associated events can also be vetoed. However, in contrast to negative weights, which are unambiguous, the threshold for a weight to be considered 'large' is less clear. Furthermore, the problem can be exacerbated when the signal effective acceptance function is applied to background events, which may populated regions in which little signal can be expected. Fortunately the region where both negative and large weights occur is relatively well confined to the edges of the phase space. After taking measures to reduce the impact of these effects by extending the q_c^2 parametrisation range and using a periodic function for the ϕ angle, the remaining instances are confined entirely to the edges of the $\cos\theta_\ell$ and near $\cos\theta_K = 1$. These

can be largely removed by the well-defined phase-space cut of $|\cos\theta_\ell| < 0.9$ and $\cos\theta_K < 0.9$ (Section 7.4).

The need to replicate the effects of correction weights and features of weighted fits in general, motivate the use of pseudoexperiments generated with effective acceptance functions in sensitivity as well as systematic studies. This configuration is more realistic as the distribution of signal events in data across the phase space follows $\sim \epsilon_{\text{eff}} \cdot P_S$ rather than P_S. For backgrounds, ideally the matching background acceptance functions should be used in generation. However, as they are not trivial to obtain, the functions for the signal are used in all cases. This may cause the pseudoexperiments to behaviour slightly better (e.g. fewer instances of background-related large weights) compared to data due to the lack of possible mismatches in the underlying acceptance effects, but it is not expected to alter the general picture.

Asymptotically correct uncertainties

For unweighted maximum likelihood fits, the inverse of the matrix of second derivatives of the log-likelihood (Hessian) provides asymptotically correct uncertainties. This is no longer true for weighted fits, as the presence of non-trivial weights ($\omega \neq 1$) prevents the simplification of the expression of the parameter variance to the inverse Hessian. The asymptotically correct expression in this case is given by [181]

$$\begin{aligned} V_{ij} = (\sum_e^N \omega_e \frac{\partial^2 \ln P(x_e|\vec{\lambda})}{\partial\lambda_i \partial\lambda_k}\Big|_{\hat{\vec{\lambda}}})^{-1} \times \Big(\sum_e^N \omega_e^2 \frac{\partial \ln P(x_e|\vec{\lambda})}{\partial\lambda_k}\Big|_{\hat{\vec{\lambda}}} \frac{\partial \ln P(x_e|\vec{\lambda})}{\partial\lambda_l}\Big|_{\hat{\vec{\lambda}}}\Big) \\ \times (\sum_e^N \omega_e \frac{\partial^2 \ln P(x_e|\vec{\lambda})}{\partial\lambda_l \partial\lambda_j}\Big|_{\hat{\vec{\lambda}}})^{-1} \ , \end{aligned} \tag{8.24}$$

where the first and last terms are the standard Hessian matrices. Note that the appearance of the sum of squared weights in the numerator, and the square of the sum of the weights in the denominator ensure that any overall normalisation factor will have no impact on the uncertainties.

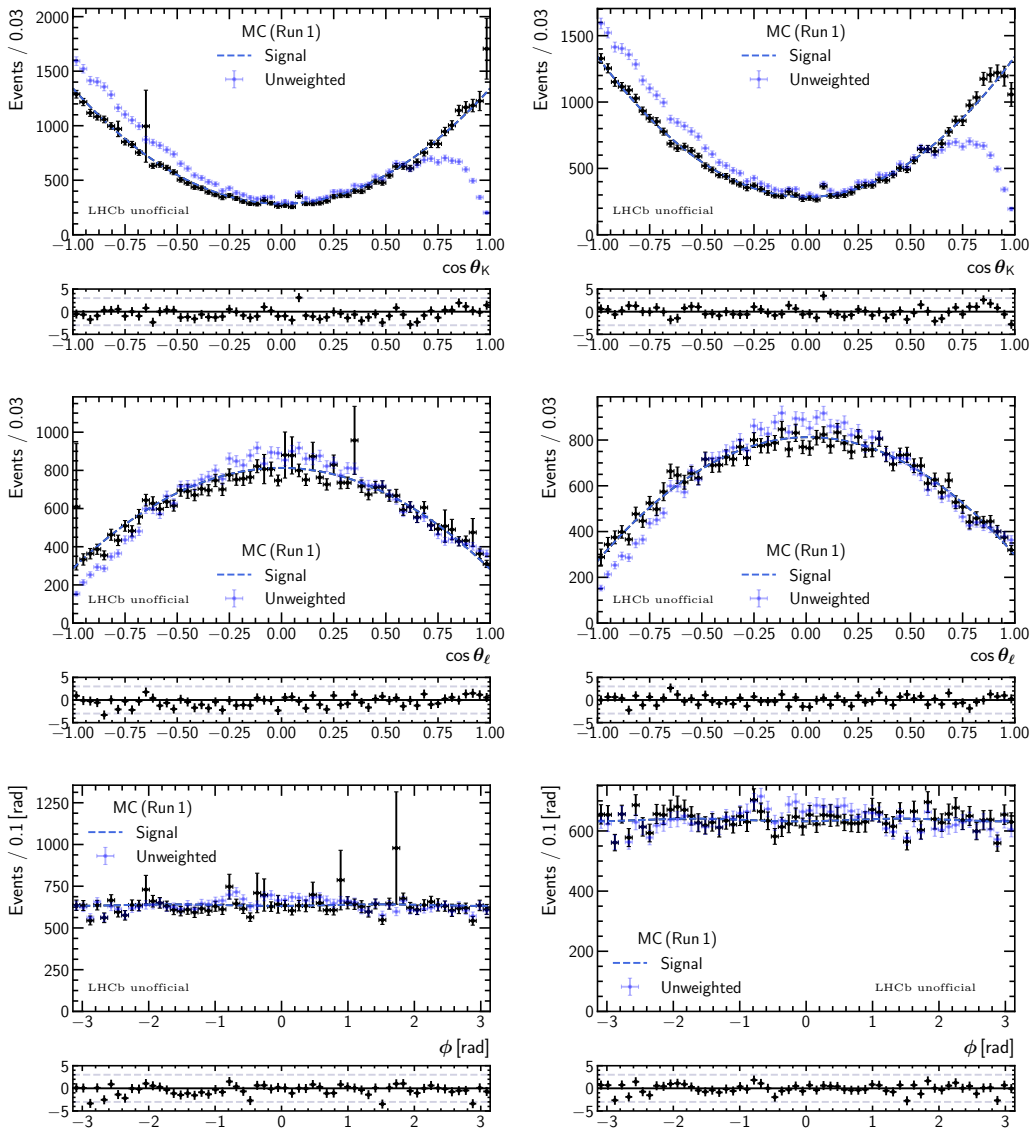


Figure 8.21: Fits to the simulated rare mode candidates of the Run 1 sample (the lowest statistics subsample), with acceptance weights from the k-folding approach (Section 8.1.4) without (left) and with (right) the removal of the 13 events with the largest weights (0.03% of the sample). The nominal phase-space cut to remove regions of low efficiency has been removed.

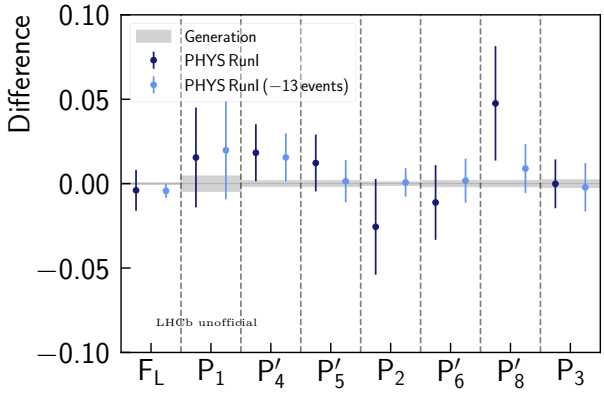


Figure 8.22: Differences with respect to generator level values for the fit to the Run 1 rare mode simulation with and without the removal of 13 events with the largest weights (0.03% of the sample). A relatively small fraction of (non-trivially) distributed events can have a noticeable impact on both the central value and the uncertainties of the fit.

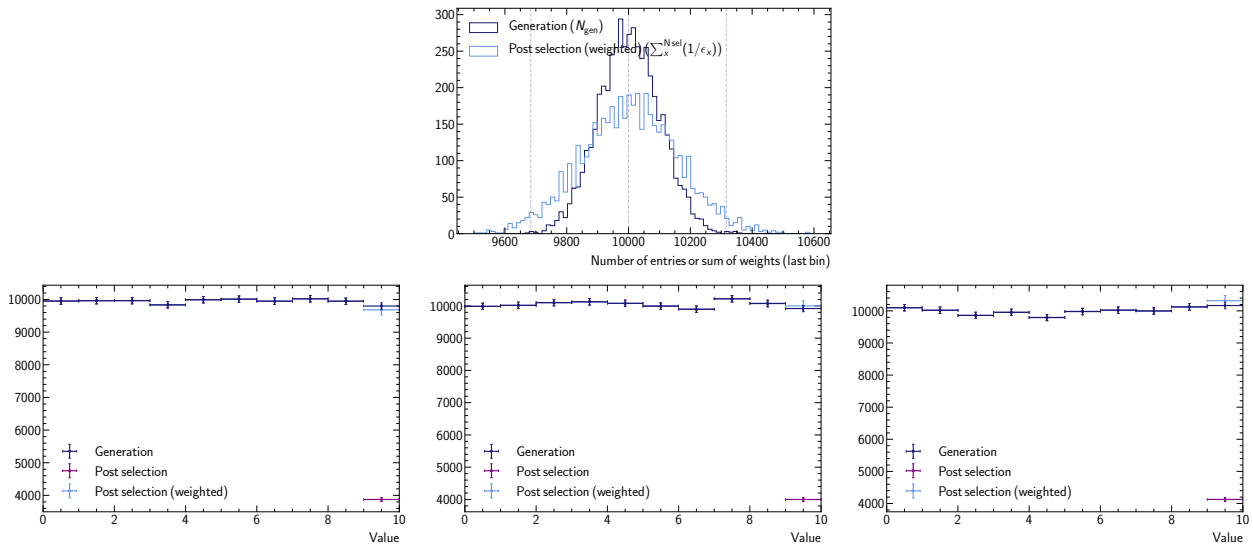


Figure 8.23: Simple example featuring 5000 'toys' generated with $N_{\text{gen}} = 100000$ where 'events' in the last bin are subjected to an 'acceptance' effect of $\epsilon = 0.4$. The distribution of the number of generator level events (sum of correction weights) in the last bin is shown in dark (light) blue (top). Examples of toys for which the sum of the weights in the last bin are $\sim 2\sigma$ below the mean (bottom left), close to the mean (bottom middle), and $\sim 2\sigma$ above the mean (bottom right), are also shown. No large weights can be seen.

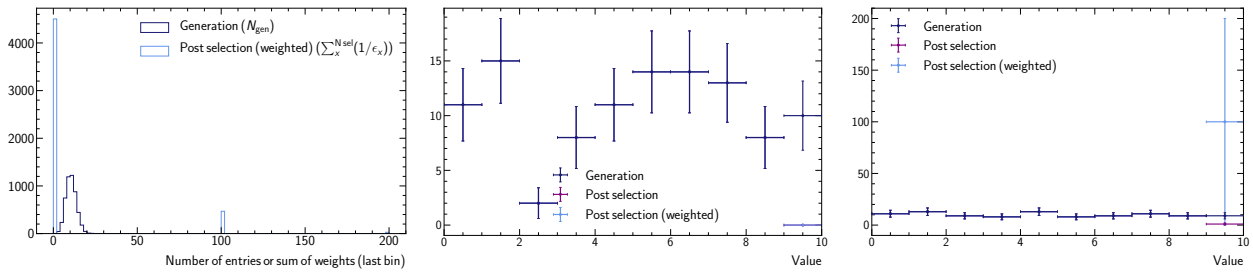


Figure 8.24: Toys generated with $N_{\text{gen}} = 100$ and subjected to $\epsilon = 0.01$. Due to the the combined effect of low statistics and low efficiency, the correct generator level shape can not be retrieved, as the corrections either undershoot (bottom left), or overshoot (bottom right). Large weights can be seen in the latter case.

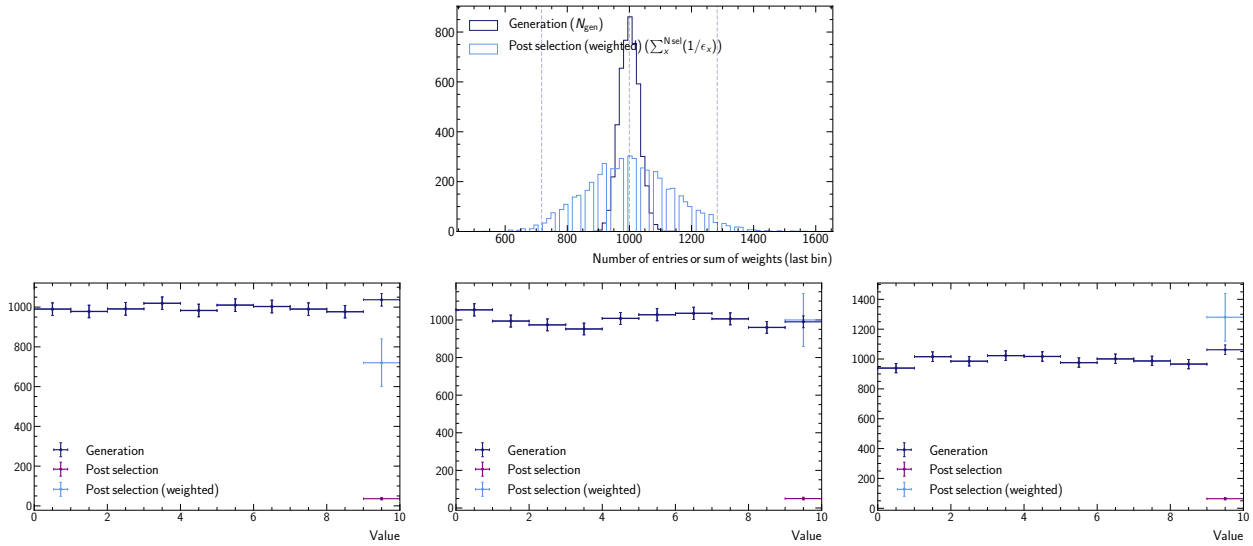


Figure 8.25: Toys generated with $N_{\text{gen}} = 10000$ and subjected to $\epsilon = 0.05$. Although relatively large correction weights occasionally occur, reasonably good retrieval of the generator level distribution is still possible.

8.4 Pseudoexperiment studies

A pseudoexperiment (also referred to as a 'toy') in this context is a sample of fake signal candidates, generated based on realistic component yields and (simplified) models to replicate the characteristics of the actual data sample, as much as it is possible to do so. A large number of toys can be used to study the behaviour associated with a given fit strategy. These include the expected statistical uncertainties with which the observables can be determined, biases, soundness of the uncertainty estimations, and fit pathologies. As the nominal fit is made with per-event weights, instead of generating toys directly from the pdfs used, each component is generated separately from the product of its pdf with the effective acceptance function of the signal, and the number of generated events is varied according to a Poisson distribution, with a mean value equal to the (adjusted) yields determined from the blinded data fit. Due to the presence of q^2 in the effective acceptance functions, it also needs to be included as an additional dimension. This requires the use of an amplitude model for the signal, and the modelling of the q^2 distribution for all backgrounds. Details on the set-up of the toy studies and the results obtained can be found in the following sections.

8.4.1 Component yields

The blinded data fit in the larger q^2 range is used to obtain yields for the signal and background components. The component ratios and yields used for toy generation should ideally approximate their unweighted values in data. The results of the fit are therefore corrected based on the expectation values of the weights from the effective acceptance functions, calculated using high statistics samples via $\sum_e^N \omega_e/N$, where N is the total number of events of a particular type, and ω_e are the per-event weights. The values used are shown in Table 8.3.

Table 8.3: Adjusted component yields used in toy generation. Note that the uncertainties do not take correlations between the fit fractions into account.

Component	Run1	Run2p1	Run2p2
Signal	116 ± 17	178 ± 23	374 ± 29
Combinatorial	104 ± 24	143 ± 27	110 ± 36
DSL	113 ± 15	95 ± 13	183 ± 18
PR	14 ± 1	22 ± 2	47 ± 4

8.4.2 Signal and background models

For toy generation, four dimensional models (angles and q^2) are needed for all components. For the signal, the amplitude model based on [172] is used. Toys are generated using SM values for the WCs without the inclusion of charm loop or S-wave effects. The corresponding observable values are given in Table 8.4. For the backgrounds, the fits described in Section 8.2 are re-performed with the inclusion of q^2 as an added dimension. In all cases, it is parametrised using Chebyshev polynomials up to second order and assumed to factorise. The results for the DSL, combinatorial and partially reconstructed backgrounds are shown in Figure 8.26 for the Run 2p2 subsample as an example.

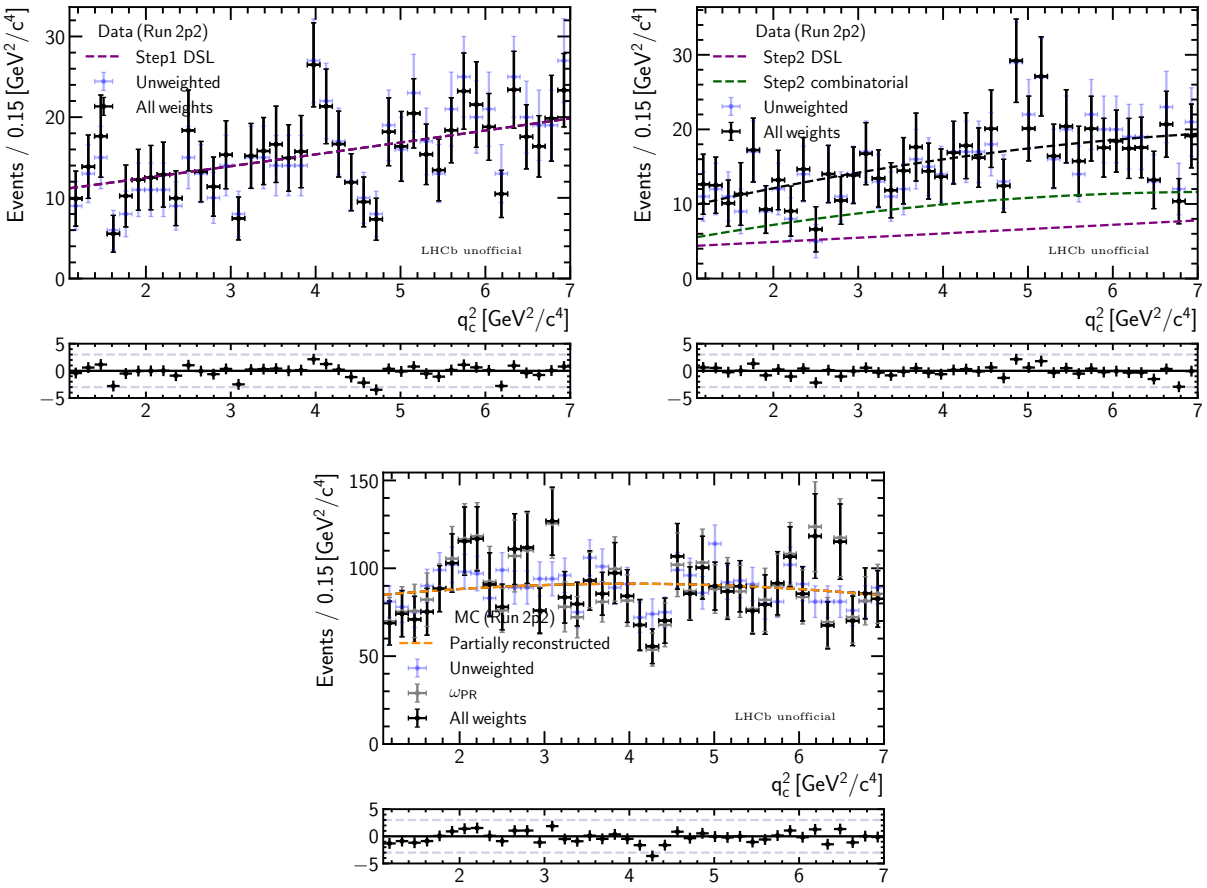


Figure 8.26: The fitted q^2 distributions of the DSL and combinatorial backgrounds (Step 1 of the two-step procedure is shown on the top left, and Step 2 on the top right), and the partially reconstructed background (bottom) of the Run 2p2 subsample. These results are only used for toy generation.

Table 8.4: Calculated values of the S and P-basis angular observables in the large and small q^2 ranges associated with the pseudoexperiments generated from the amplitude model with the nominal SM configuration.

	$1.1 < q^2 < 7.0$	$1.1 < q^2 < 6.0$
F_L	0.744	0.771
S_3	−0.019	−0.015
S_4	−0.177	−0.157
S_5	−0.251	−0.215
A_{FB}	0.066	0.028
S_7	0.000	0.000
S_8	0.000	0.000
S_9	0.000	0.000
P_1	−0.146	−0.128
P_4'	−0.405	−0.374
P_5'	−0.576	−0.512
P_2	0.171	0.081
P_6'	0.000	0.000
P_8'	0.000	0.000
P_3	0.000	0.000

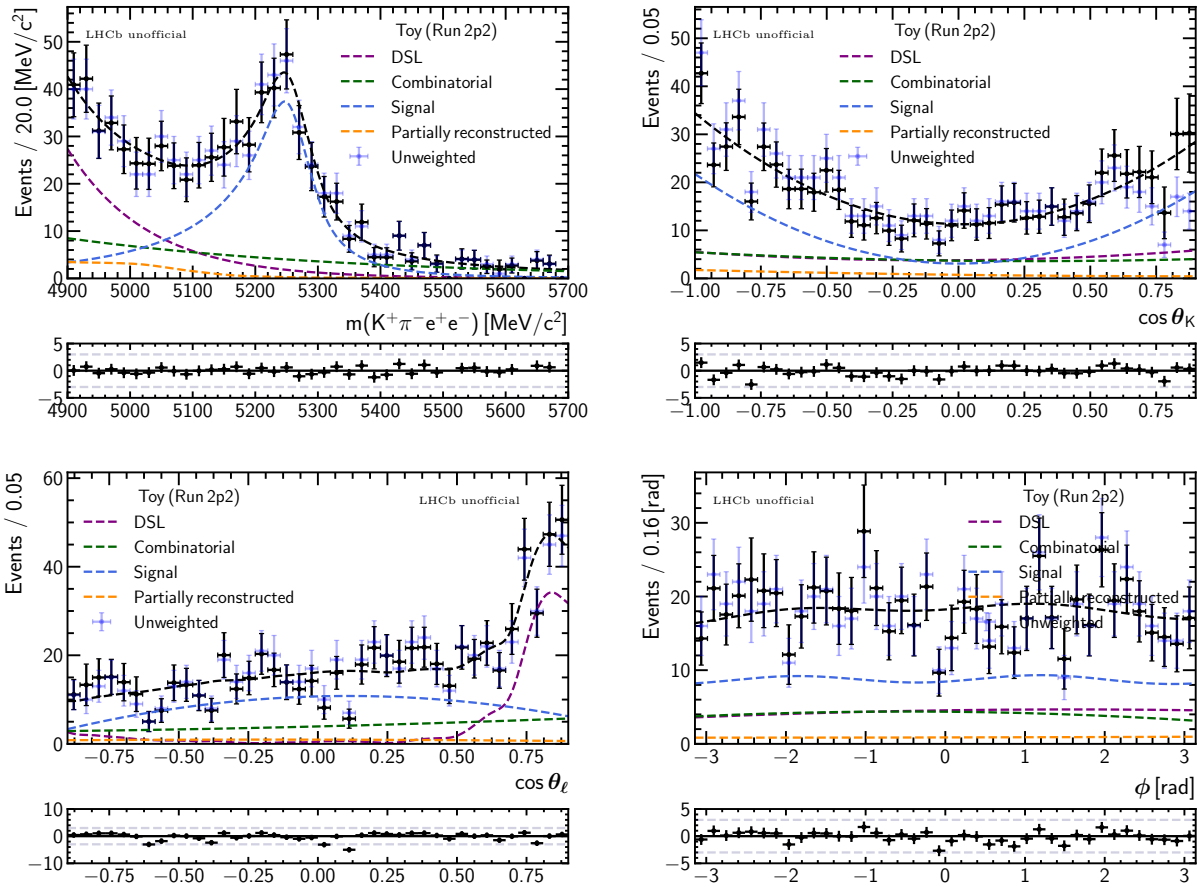


Figure 8.27: Result of a toy fit (Run 2p2 models). The unweighted distributions are shown in blue, and the distribution after effective acceptance correction is shown in black.

8.4.3 Sensitivity

To study the behaviour of the nominal fit, around one thousand toys are generated in the larger q^2 range of $1.1 < q^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$, with the phase-space cut and using the yields in Table 8.3. For simplicity, the control mode mass fit is not included. Any impact resulting from this is expected to be marginal. The toys are fitted using the angular pdf with correction weights obtained from the effective acceptance functions used in generation, an example of which is shown in Figure 8.27. The same set of toys is used to study fit behaviour in the smaller q^2 range, in which case the cut of $q^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ is applied. The asymptotically correct approach (Section 8.3.3) is used for uncertainty calculations.

The key information to be extracted from the results include the expected statistical uncertainties of the observables ('sensitivity'), the existence and size of any fit biases, and whether or not the uncertainties have been properly estimated. The convergence rate of the toys can also provide important information on the stability of the fit.

Due to the asymptotic properties of the maximum likelihood estimator, as the toy sample size becomes infinitely large, the distribution of the estimated quantity goes to that of a Gaussian with a mean equalling the generation value (plus possible bias), and a width given by the uncertainty of the fit. The distribution of the pulls, p, calculated via

$$p = \frac{S_{\mathrm{fit}} - S_{\mathrm{gen}}}{\sigma_{S_{\mathrm{fit}}}} \,, \tag{8.25}$$

where S_{fit} refers to the fitted value of an observable, S_{gen} its generation value, and $\sigma_{S_{\mathrm{fit}}}$ the fit uncertainty, is a Gaussian centred at zero (in the event of no bias), with unit width (if the calculated uncertainties agree with the fit uncertainties). The sensitivity values can be determined from the width of the distribution of observable values from the individual toy fits. Biases and uncertainties can be checked by comparing the mean and width of the pull distributions to zero and one, respectively. The results for the observables of interest are summarised in Table 8.5.

The sensitivity values suggest that the statistical uncertainties of the measurement are approximately twice that of the muon mode observables measured using full Run 1 and 2016 data within $1.1 < q^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ [57], while they are comparable to the older analysis using Run 1 data only [63]. Nevertheless, they are lower than that of the most recent electron mode measurement [62]. No large biases or serious coverage issues are found. The convergence rate of the toys is around 99%, which shows that the chosen configuration is stable, and that the data fit is unlikely to fail.[10]

[10] Nevertheless, the physics model (and hence observable values) used for toy generation is not expected to correspond exactly to that of the data. The behaviour of the fit may show sensitivity to observable values, in particular at relatively low statistics. For example, in the event that the preferred values lie close to the edge of the physical region of the angular pdf, biases and coverage problems can appear, as seen in e.g. [63]. Therefore, additional toy studies are envisaged following the unblinding of the data fit result.

Table 8.5: Results of around a thousand toys generated in the larger q^2 range and fitted in the larger (top), and the smaller (bottom) q^2 ranges.

$1.1 < q^2 < 7.0$, SM			
	Sensitivity	Pull mean	Pull width
F_L	0.0400 ± 0.0009	-0.030 ± 0.031	1.016 ± 0.022
S_3	0.0370 ± 0.0008	-0.013 ± 0.031	1.011 ± 0.022
S_4	0.0563 ± 0.0012	0.023 ± 0.031	1.001 ± 0.022
S_5	0.0497 ± 0.0011	0.015 ± 0.032	1.040 ± 0.023
A_{FB}	0.0372 ± 0.0008	-0.028 ± 0.032	1.027 ± 0.022
S_7	0.0496 ± 0.0011	0.059 ± 0.030	0.985 ± 0.022
S_8	0.0584 ± 0.0013	-0.041 ± 0.030	0.982 ± 0.021
S_9	0.0376 ± 0.0008	0.002 ± 0.032	1.027 ± 0.022
F_L	0.0401 ± 0.0009	-0.028 ± 0.032	1.024 ± 0.022
P_1	0.3010 ± 0.0070	-0.013 ± 0.031	0.990 ± 0.022
P_4'	0.1332 ± 0.0029	0.021 ± 0.031	0.993 ± 0.022
P_5'	0.1221 ± 0.0027	0.020 ± 0.031	1.011 ± 0.022
P_2	0.1003 ± 0.0022	-0.025 ± 0.031	1.004 ± 0.022
P_6'	0.1151 ± 0.0025	0.059 ± 0.030	0.979 ± 0.021
P_8'	0.1359 ± 0.0030	-0.040 ± 0.030	0.976 ± 0.021
P_3	0.1533 ± 0.0033	-0.004 ± 0.031	1.000 ± 0.021

$1.1 < q^2 < 6.0$, SM			
	Sensitivity	Pull mean	Pull width
F_L	0.0449 ± 0.0010	-0.047 ± 0.032	1.017 ± 0.022
S_3	0.0421 ± 0.0009	-0.028 ± 0.032	1.042 ± 0.023
S_4	0.0634 ± 0.0014	0.041 ± 0.025	0.995 ± 0.021
S_5	0.0551 ± 0.0012	0.030 ± 0.032	1.024 ± 0.023
A_{FB}	0.0418 ± 0.0009	-0.008 ± 0.034	1.031 ± 0.023
S_7	0.0544 ± 0.0012	0.061 ± 0.030	0.967 ± 0.021
S_8	0.0658 ± 0.0015	-0.015 ± 0.031	0.989 ± 0.022
S_9	0.0417 ± 0.0009	-0.030 ± 0.032	1.033 ± 0.023
F_L	0.0450 ± 0.0010	-0.053 ± 0.032	1.028 ± 0.023
P_1	0.3910 ± 0.0090	-0.024 ± 0.031	0.998 ± 0.022
P_4'	0.1577 ± 0.0035	0.044 ± 0.031	0.989 ± 0.021
P_5'	0.1398 ± 0.0031	0.038 ± 0.030	0.978 ± 0.021
P_2	0.1327 ± 0.0029	0.009 ± 0.031	0.997 ± 0.022
P_6'	0.1323 ± 0.0029	0.057 ± 0.030	0.954 ± 0.021
P_8'	0.1601 ± 0.0034	-0.019 ± 0.030	0.976 ± 0.021
P_3	0.1960 ± 0.0040	0.025 ± 0.030	0.981 ± 0.022

9 Data fits

To validate important aspects of the analysis strategy, a data fit is made to the control mode, which is known to obey LFU to high precision [182, 183]. The results of the fit are compared against the angular observables of the muonic charmonium mode [57], taking statistical and systematic uncertainties into account. An angular fit is made to the rare mode candidates in data, and important sources of systematic uncertainties are quantified. While the results of the measurement cannot be revealed at this stage, the expected observable uncertainties are discussed, and the projections of the fit are shown.

9.1 Control mode validation

The fit strategy for the rare mode can be validated to some extent using the decay of $B^0 \to K^{*0}(\to K^+\pi^-)J/\psi(\to e^+e^-)$. This control mode has the same final state particles as the signal, almost identical topology,[1] and comparable kinematics. As it is a tree-level $b \to sc\bar{c}$ process, it has a much larger branching fraction, and is known to be consistent with the SM. The angular distributions of the muon mode of $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ has been studied by multiple experiments, namely LHCb [184], BaBar [185] and Belle [186, 183], and LFU is also established to high precision in the branching fraction measurements of the $B^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)$ decays, where $\ell = e, \mu$ [182, 7]. The muonic charmonium mode has been used internally for the validation of the $B^0 \to K^{*0}\mu^+\mu^-$ angular analyses [63, 57]. The availability of these results means that fit validation can be made to some extent by applying the rare mode strategy to the control mode, as much as it is possible to do so, and comparing the resulting observable values with those of the muon mode.

Note that the focus of this 'validation' is unavoidably on the treatment of the signal (e.g. the assumption of factorisation between mass and angles, simulation-data corrections). Due to significant differences in background composition and yields, it is not possible to validate the rare mode background strategy properly. In addition, differences in the q^2 distributions of the two means that the effective acceptance approach can only be validated to some extent. These limitations need to be considered in the interpretation of the results.

The strategy used for the control mode angular fit is discussed in Section 9.1.1. The main systematic uncertainties are quantified in Section 9.1.2, and the result is given in Section 9.1.3.

9.1.1 Angular fit

Control mode candidates are selected in the q_c^2 range of $7.0 < q_c^2 < 11.0\,\mathrm{GeV}^2/\mathrm{c}^4$, and are made to pass the same selection requirements as the rare mode, with the exception of an added cut of $m(K\pi ee)_{\mathrm{PV,J}/\psi} > 5150\,\mathrm{MeV}/\mathrm{c}^2$, which is used to strongly suppress partially reconstructed background.[2] While enforcing alignment is generally beneficial, in this case it is

[1] Due to the short lifetime of the J/ψ resonance the two electrons effectively originate from the decay point of the B^0 meson.

[2] As $m(K\pi ee)_{\mathrm{PV,J}/\psi}$ is highly correlated with the constrained q_c^2, this requirement is almost equivalent to reducing the q_c^2 range to a maximum of $10.0\,\mathrm{GeV}^2/\mathrm{c}^4$.

not done for two main reasons. Firstly, the difference between the signal-to-background ratios of the two is unavoidably large even in its absence (~ 0.85 compared to ~ 0.5). Secondly, removing the cut would lead to complications due to the lack of reliable angular models for some of these backgrounds, which could introduce significant discrepancies that may not be easy to understand or quantify. Furthermore, due to non-trivial differences in composition, background systematic uncertainties are not directly comparable between these two modes. As the large statistics of the charmonium mode (~ 400 times higher than that of the rare mode) provides sensitivity to contributions beyond the P-wave amplitude, the differential decay rate equation with additional terms (Equation 2.18) is used. Note that these terms are also included in the muon mode fit. In order to allow for a more stable determination of the combinatorial background fraction, a larger mass window of $4500 < m(K\pi ee)_{\text{PV}} < 6200\,\text{MeV/c}^2$ is used.

Three background components are considered (Section 8.2.2): $B_s^0 \to K^{*0}J/\psi(\to e^+e^-)$ decays, misidentified $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$ decays and the combinatorial background. The fraction of B_s^0 decays with respect to the signal is fixed to calculated values (Equation 8.18), i.e. $f_{B_s^0}$ is expressed as $f_{B_s^0}^S \cdot f_S$, where f_S is the signal fraction, and the value of $f_{B_s^0}^S$ is fixed. Separate values of 0.009 ± 0.001 and 0.008 ± 0.001 are used for Run 1 and Run 2 subsamples due to the dependency of the hadronisation fraction on the CoM energy [187, 177]. The misidentified Λ_b^0 component is modelled using the $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$ phase-space simulation reconstructed as $B^0 \to K^{*0}J/\psi(\to e^+e^-)$, and with the addition of data-driven correction weights. The modelling of its angular distributions is carried out using factorised Chebyshev polynomials up to second order for $\cos\theta_\ell$ and ϕ, and a KDE lineshape for $\cos\theta_K$. The fit result for the Run 2p2 subsample is shown in Figure 9.1 as an example. Like the B_s^0 component, its contribution is fixed as a fraction of the signal. The values of $f_{\Lambda_b^0}^S$ used for Run 1, Run 2p1 and Run 2p2 are 0.0057 ± 0.0006, 0.0018 ± 0.0002 and 0.0017 ± 0.0002, respectively. For the combinatorial background, exponential functions are used to describe the mass distributions, and Chebyshev polynomials up to second order are used for the angles. As it is expected to take in contributions from DSL and residual partially reconstructed decays, all its parameters are allowed to vary.

The full log-likelihood used for the control mode is formulated in the same way as that of the rare mode (Equation 8.20). The fit is made with per-event weights from the control mode effective acceptance functions. Eight main angular observables and six additional S-wave and S and P-wave interference parameters are allowed to vary, as are the slopes of the combinatorial backgrounds for the three subsamples, three signal fractions, the shift and scaling parameters, and a total of eighteen parameters of the Chebyshev polynomials (six for each Run subsample) that describe the angular distributions of the combinatorial background.

9.1.2 Control mode systematic uncertainties

The validation is made by checking the consistency of the electron and muon mode observables, for which both statistical and systematic uncertainties need to be taken into account. In this section, dominant sources of the latter are quantified.[3] These include aspects of the effective acceptance functions and the strategy of describing the partially reconstructed and DSL backgrounds using a single model.

The procedure used to quantify the uncertainties is conceptually identical to that of the rare mode (Section 9.2). Besides intrinsic biases introduced by the effective acceptance model (Section 8.1), the uncertainties of which are used directly as systematic uncertainties, and the background strategy, which is computationally expensive to assess (Section 9.1.2), for each

[3]This is only done for the S-basis observables, as no muon mode P-basis results are available.

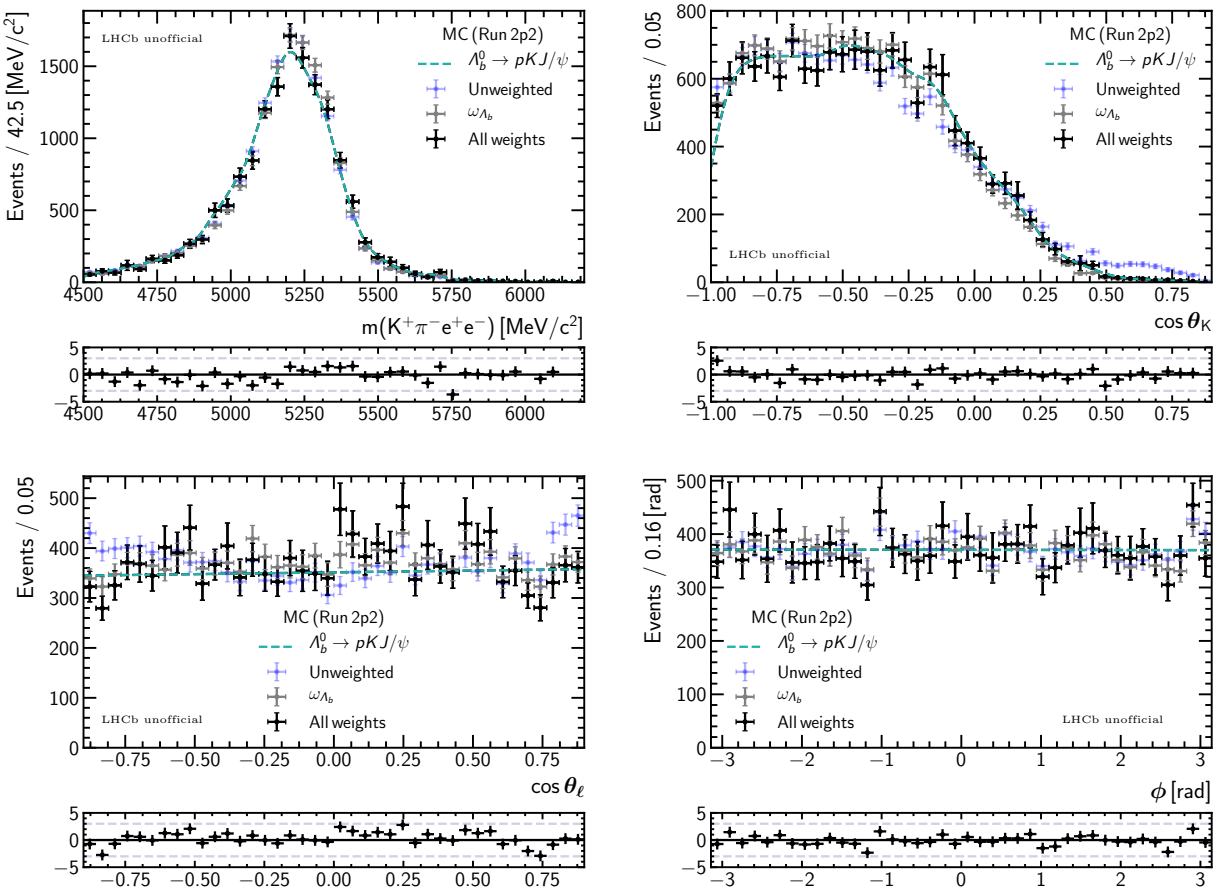


Figure 9.1: Results of fits to the mass and angular distributions of the Run 2p2 phase-space $\Lambda_b^0 \to pKJ/\psi(\to e^+e^-)$ simulation reconstructed as the signal. The unweighted sample is shown in blue, the distribution after including the data-driven correction weights ($\omega_{\Lambda_b^0}$) is shown in gray, and the final distribution used for modelling that includes both $\omega_{\Lambda_b^0}$ and acceptance weights is shown in black.

source, an alternative configuration is set up and a large number of signal-only toys are generated using realistic component yields. These are then fitted using the alternative as well as the nominal configuration. The differences between the observables obtained are used to quantify the associated systematic uncertainty. The results of these studies are summarised in Table 9.1, and additional information can be found in the following sections.

Table 9.1: Summary of the systematic uncertainties of the control mode angular fit.

	Coefficient uncertainties	Higher orders	Alternative correction	Additional backgrounds	Bias uncertainties	Total (σ_{syst})
F_L	0.0006	0.0006	0.0003	0.0016	0.0008	0.0020
S_3	0.0007	0.0001	0.0008	0.0002	0.0008	0.0013
S_4	0.0008	0.0003	0.0022	0.0027	0.0011	0.0038
S_5	0.0007	0.0007	0.0002	0.0003	0.0009	0.0014
A_{FB}	0.0005	0.0017	0.0006	0.0007	0.0006	0.0021
S_7	0.0007	0.0005	0.0003	0.0002	0.0010	0.0014
S_8	0.0008	0.0002	0.0010	0.0007	0.0011	0.0018
S_9	0.0008	0.0004	0.0015	0.0008	0.0008	0.0021
F_S	0.0015	0.0027	0.0056	0.0013	0.0017	0.0068
S_{S_1}	0.0014	0.0030	0.0150	0.0004	0.0018	0.0154
S_{S_2}	0.0009	0.0004	0.0005	0.0002	0.0011	0.0016
S_{S_3}	0.0009	0.0001	0.0002	0.0001	0.0009	0.0013
S_{S_4}	0.0008	0.0003	0.0004	0.0003	0.0010	0.0014
S_{S_5}	0.0009	0.0001	0.0012	0.0006	0.0012	0.0020

Parametrisation limitations

The results of the acceptance parametrisation validation provided in Section 8.1.4 show that the chosen models cannot fully capture all features present in the samples, leading to the imperfect retrieval of generator level observables. While part of the discrepancies is expected to be statistical in nature, it is possible to link shifts in certain observables to the imperfect parametrisation of the effective acceptance effect. Therefore, as a conservative measure, the observable shifts are treated as biases and used to correct the final result. To obtain values relevant for the simultaneous fit, the weighted average of the differences found for the three Run subsamples is calculated using control mode yields in data for each observable. The uncertainties of the biases are treated as systematic uncertainties.

Higher order description

The nominal choices for the orders of the Legendre and Fourier polynomials used in the parametrisation are biased towards lower orders in order to reduce the sizes of possible pathological regions, and to avoid problems at the edges of the parametrisation range. To investigate the systematic effect of this choice, the orders used for all angles are increased by three to terms of order eight and lower for $\cos\theta_K$, seven and lower for $\cos\theta_\ell$ and seven and lower for ϕ, and alternative effective acceptance functions are parametrised. As expected, the previously poorly described region of $\cos\theta_K$ near one has improved, as shown in Figure 9.2.

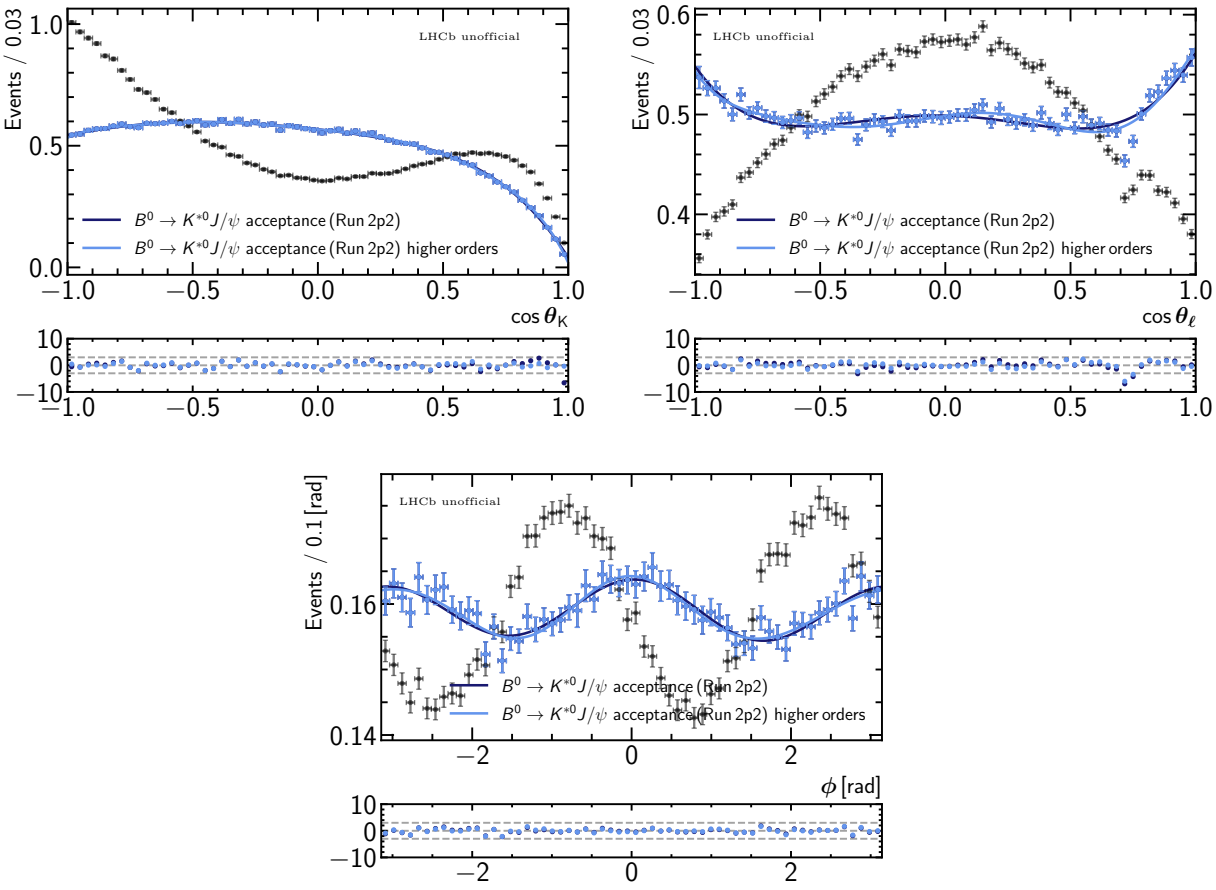


Figure 9.2: Comparison between the projections of the nominal Run 2p2 control mode acceptance function and that of the alternative made with all maximum polynomial orders increased by three.

Three hundred toys are generated with the alternative acceptance functions and fitted with corrections from either the matching functions or the nominal ones. All distributions of observable differences are found to be well described by Gaussian functions. The differences between the matching and the mismatched fits are reflected mainly in the shifts of the mean from zero. However, in a few cases the widths of the difference distributions are comparable to the mean, which may be related to slightly larger weights that can arise randomly in some regions due to the acceptance mismatch. Therefore, the systematic uncertainties are taken as the sum in quadrature of the mean and the width values.

Coefficient uncertainties

To quantify systematic uncertainties associated with the uncertainties of the coefficients of the acceptance functions, the simulation samples used are bootstrapped three hundred times to produce the same number of alternative acceptance functions. These are used to generate toys, which are fitted with correction weights derived from either its matching randomised functions, or the nominal ones. The mean values of the difference distributions are generally consistent with zero, which is expected, as the alternative acceptances can be seen as random variations about the nominal ones, while the widths tend to deviate from zero. Note that this behaviour is nevertheless not guaranteed, as in the absence of the phase-space cut to remove low efficiency

regions, larger weights at the edges of the phase-space can lead to biases, in particular for F_L. In this case, the widths values are taken to be the systematic uncertainties.

Simulation correction strategy

The systematic uncertainties associated with the simulation-data correction strategy is estimated in the same way as the rare mode (Section 9.2.2) using an alternative correction chain in the parametrisation of the effective acceptance functions [67]. This strategy differs from the nominal in two main ways. Firstly, the efficiencies of the main PID cuts are extracted directly from data in bins of the `PIDCalib` samples, and the correction is made by assigning corresponding per-event weights to simulated events.

Secondly, an attempt is made to correct the (pre-trigger) generator level distributions of kinematic and multiplicity variables using a BDT reweighting procedure based on PID and trigger corrected muon mode samples. Furthermore, additional corrections are made for the electron tracking efficiency and reconstructed quantities, include variables that are not addressed in the nominal chain.

Three hundred signal-only toys are generated using the alternative effective acceptance functions, and fitted with the alternative and nominal acceptance weights. The distributions of the differences in the observable values between the two fits are found to be well described by Gaussian distributions. In all cases, the shifts in the mean values are much larger than the widths, and are therefore taken directly as systematic uncertainties.

Additional backgrounds

In the control mode fit, the combinatorial component is allowed to vary in order to accommodate residual non-peaking DSL and partially reconstructed decays. This is expected to be a source of systematic uncertainty. For example, the use of Chebyshev polynomials up to second order may not describe the sharp features in the $\cos\theta_\ell$ distribution of the former sufficiently well, and the assumption that the mass distribution is exponential is known to not hold for the latter. To quantify the impact of this approximation, first, a mass-only fit is made without the cut of $m(K\pi ee)_{\mathrm{PV},J/\psi} > 5150\,\mathrm{MeV/c^2}$ to allow the yield of the partially reconstructed background to be determined. The mass model of this component is obtained from inclusive simulations of $B_{d,u,s} \to K\pi X J/\psi(\to e^+e^-)$ decays, where X refers to missing particles. These include cases where the X originate from the decays of higher $c\bar{c}$ resonances, and ones where they originate from the decays of various K_J mesons. The latter is the dominant contribution in data. Due to the lack of angular information, no separate component can be used for the DSL background. The yields of the partially reconstructed background from the fit are scaled to their expected values after requiring $m(K\pi ee)_{\mathrm{PV},J/\psi} > 5150\,\mathrm{MeV/c^2}$ using simulation. To obtain reasonable estimations for the DSL and combinatorial backgrounds, the total yield of the effective 'combinatorial background' (which contains DSL events) is scaled based on the $K\pi e\mu$ data sample, and the ratio between them in that sample is used to separate the two components.

For toy generation, models for the DSL and combinatorial backgrounds are obtained by applying the two-step procedure (Section 8.2.1) to the control mode q^2_c region. The angular distributions of the partially reconstructed backgrounds are modelled from simulation using Chebyshev polynomials. Note that this model is expected to be approximate to some extent due to difficulties in replicating the non-trivial angular distributions of a mixture of modes.[4]

[4]The simulations also do not include S-wave contributions.

The nominal models (and yields) are used for components that are present in the main fit. As these fits are computationally expensive due to the high statistics, the number of parameters varied (44), and the low convergence rate ($\sim 70\%$), the toys are fitted once using the nominal configuration. The differences between the generation values and the mean values of the parameter distribution from these fits are used directly as systematic uncertainties.

9.1.3 Results and discussions

The result of the control mode fit is shown in Figure 9.6 for all subsamples. Several interesting features are immediately apparent. Firstly, the mass fits show large pulls in the region of the signal peak. This is mainly due to the limited ability of the chosen model to accommodate remaining simulation-data differences that become prominent at high statistics, rather than the slight mismodelling of the simulation samples or unaccounted for peaking backgrounds, both of which have very limited impact. This type of discrepancy has been seen and studied by similar analyses (e.g. [67, 142]). One way by which it can be reduced is through the addition of extra degree(s) of freedom in the data fit. This is not used for the angular fit due to its limited effectiveness, and the low impact of this type of mismodelling on the angular observables.

Secondly, the two main issues seen during the validation of the acceptance functions on simulation (Section 8.1.4), namely the (slight) tendency for the corrections to overshoot near $\cos\theta_K \sim 0.85 - 0.90$, and the kink near $\cos\theta_\ell = 0.75$ caused by the vetoes against semileptonic decays, are clearly present in data. The impact of these effects are expected to be taken into account by the bias correction discussed in Section 9.1.2.

A third noteworthy feature is the shape of the angular distribution of the combinatorial background, especially that of $\cos\theta_\ell$, for which an asymmetric shape that peaks at both ends of its range, but more strongly near $\cos\theta_\ell = 0.9$, is preferred for all three subsamples. This suggests that by allowing the shape to vary, the 'combinatorial' background has indeed been able accommodate some DSL events.

The fourth feature is the poor description of the data near $\cos\theta_K = -0.9$. This effect is also present in the muon mode fit, and is related to the presence of contributions from exotic charmonium states that decay to $J/\psi\pi^+$ pairs (e.g. $Z_c^+(4200)$ and $Z_c^+(4430)$) [183, 188]. While this component can be expected to be lepton flavour universal, experimental effects can introduce electron-muon differences. Nevertheless, in the absence of a greater level of understanding, it is difficult to estimate the size of this systematic uncertainty.

The values of the angular observables obtained directly from the fit (without bias correction) and their biases are given in Table 9.2 alongside their statistical and systematic uncertainties, the statistical uncertainties of the muon mode fit [57], the differences with respect to the muon mode values, and the differences divided by the total uncertainties (in the number of σs). The differences are also displayed in Figure 9.3. Most observables are found to lie within $\sim 2\sigma$ of the muon mode results. Slightly larger differences are seen for F_L and S_3. This is expected to be mainly related to the treatment of backgrounds. While an attempt has been made to quantify the leading source of background-related systematic uncertainty, it is complicated by the approximate nature of the models used. A more thorough and comprehensive assessment, which makes use of improved simulations and considers other factors (e.g. neglecting residual peaking backgrounds, possible experimental electron-muon differences in the angular shape of the exotics background) is expected to explain most of the remaining differences. Given that conservative approaches are used to assess the impact of backgrounds to the rare mode fit, potential analogous effects are expected to be covered by the systematic uncertainties.

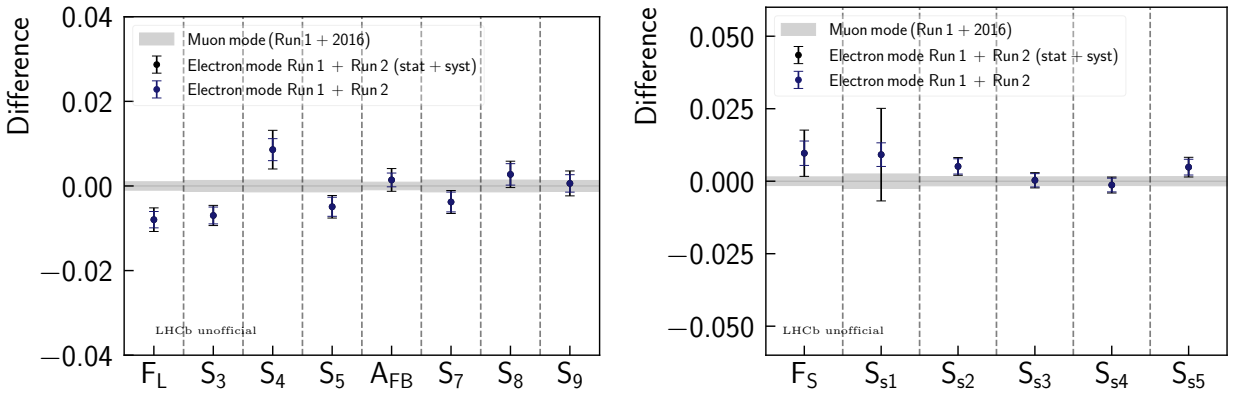


Figure 9.3: Differences between the observable values found from the muon mode angular fit made using Run 1 and 2016 data [57] and the (bias corrected) electron mode values. The error bars show the statistical (blue) and total (black) uncertainties of the electron mode fit. The statistical uncertainties of the muon mode results are shown by the gray band (and are not included in the error bars).

Table 9.2: Values of the S-basis observables found directly from the fit to the control mode candidates and their associated biases. The uncertainties shown are the statistical uncertainties of the electron mode fit (σ_{stat}) and that of the muon mode fit [57] ($\sigma^{\mu}_{\text{stat}}$), as well as the (electron mode) systematic uncertainties (σ_{syst}). The total uncertainties (σ_{tot}) are the sum in quadrature of these sources. The absolute values of the differences (ΔS_i) and the differences divided by the total uncertainties ($\Delta S_i/\sigma_{\text{tot}}$) are also shown.

	Result	Bias	σ_{stat}	$\sigma^{\mu}_{\text{stat}}$	σ_{syst}	σ_{tot}	ΔS_i	$\Delta S_i/\sigma_{\text{tot}}$
F_L	0.5539	−0.0010	0.0019	0.0010	0.0020	0.0030	0.0080	2.7
S_3	−0.0074	0.0002	0.0020	0.0013	0.0013	0.0027	0.0070	2.6
S_4	−0.2393	−0.0010	0.0026	0.0014	0.0038	0.0048	−0.0086	−1.8
S_5	−0.0036	−0.0007	0.0023	0.0014	0.0014	0.0030	0.0049	1.6
A_{FB}	0.0008	0.0011	0.0016	0.0009	0.0021	0.0028	−0.0014	−0.5
S_7	−0.0022	−0.0005	0.0023	0.0014	0.0014	0.0031	0.0038	1.2
S_8	−0.0517	0.0000	0.0025	0.0014	0.0018	0.0034	−0.0027	−0.8
S_9	−0.0839	−0.0007	0.0021	0.0013	0.0021	0.0032	−0.0006	−0.2
F_S	0.0693	0.0020	0.0042	0.0015	0.0068	0.0081	−0.0096	−1.2
S_{S_1}	−0.2154	−0.0036	0.0041	0.0025	0.0154	0.0162	−0.0092	−0.6
S_{S_2}	0.0278	−0.0006	0.0026	0.0016	0.0016	0.0035	−0.0051	−1.5
S_{S_3}	0.0014	−0.0003	0.0023	0.0015	0.0013	0.0030	−0.0003	−0.1
S_{S_4}	−0.0012	−0.0001	0.0024	0.0015	0.0014	0.0031	0.0013	0.4
S_{S_5}	−0.0619	−0.0003	0.0027	0.0016	0.0020	0.0038	−0.0049	−1.3

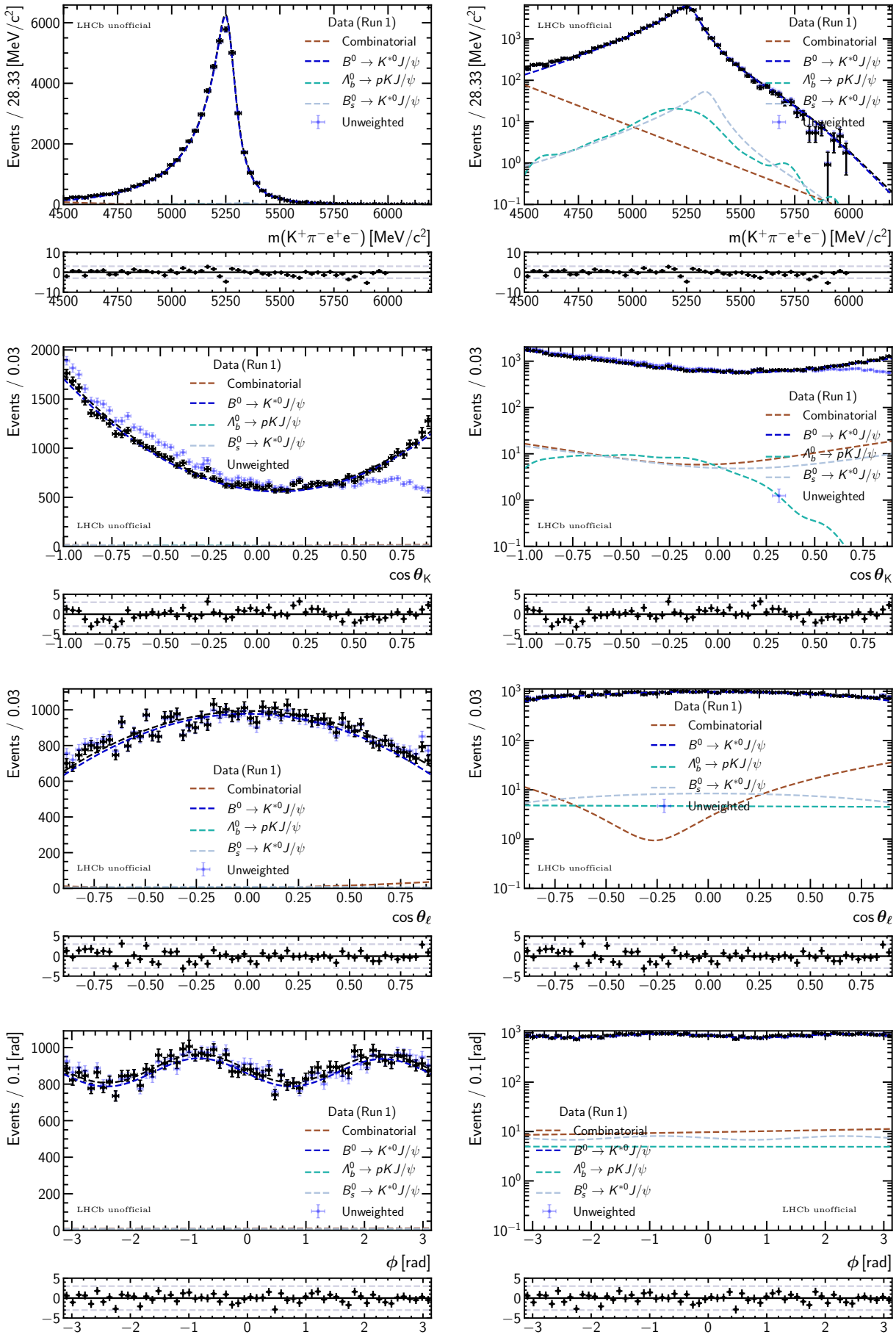


Figure 9.4: Result of the simultaneous control mode angular fit (Run 1).

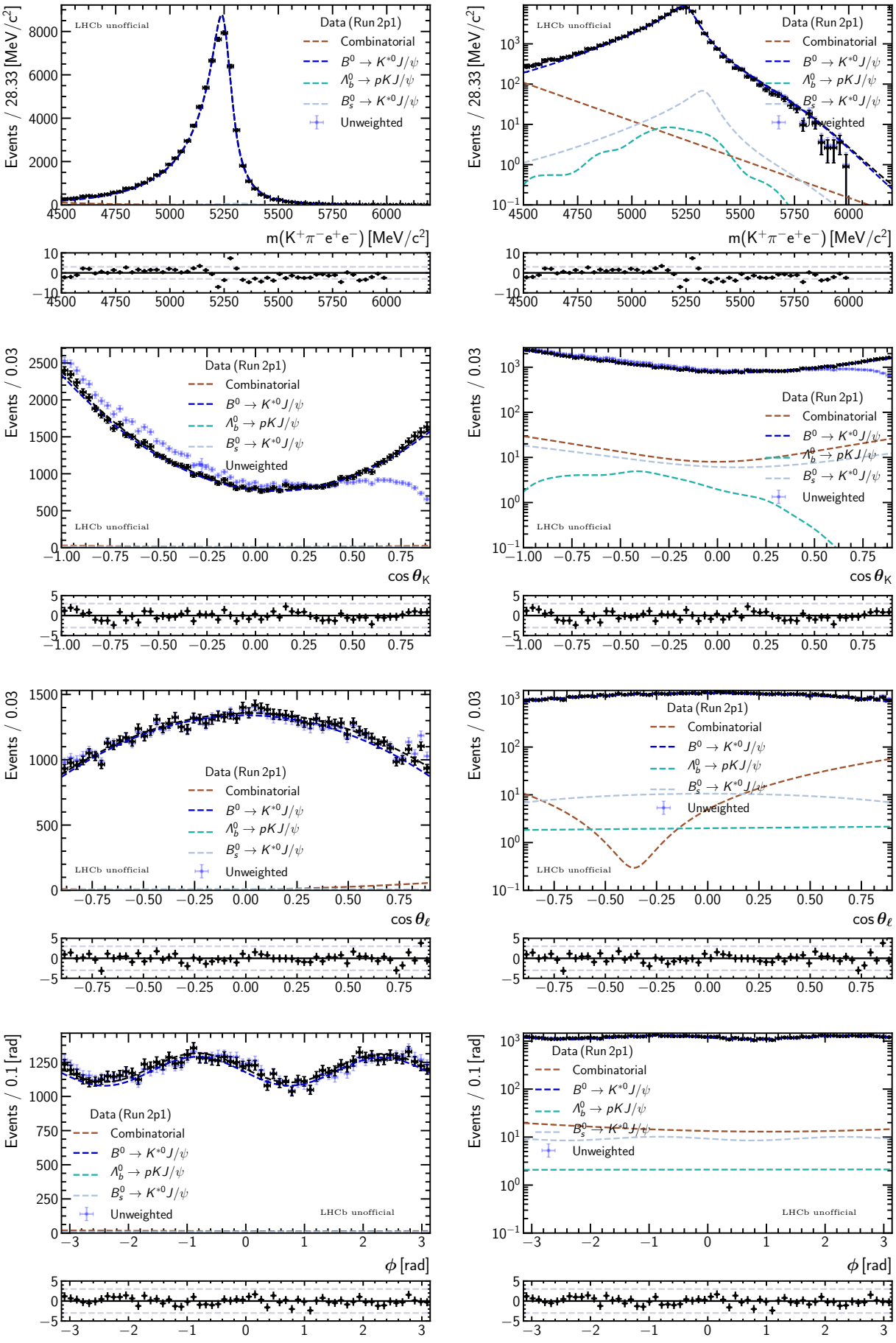


Figure 9.5: Result of the simultaneous control mode angular fit (Run 2p1).

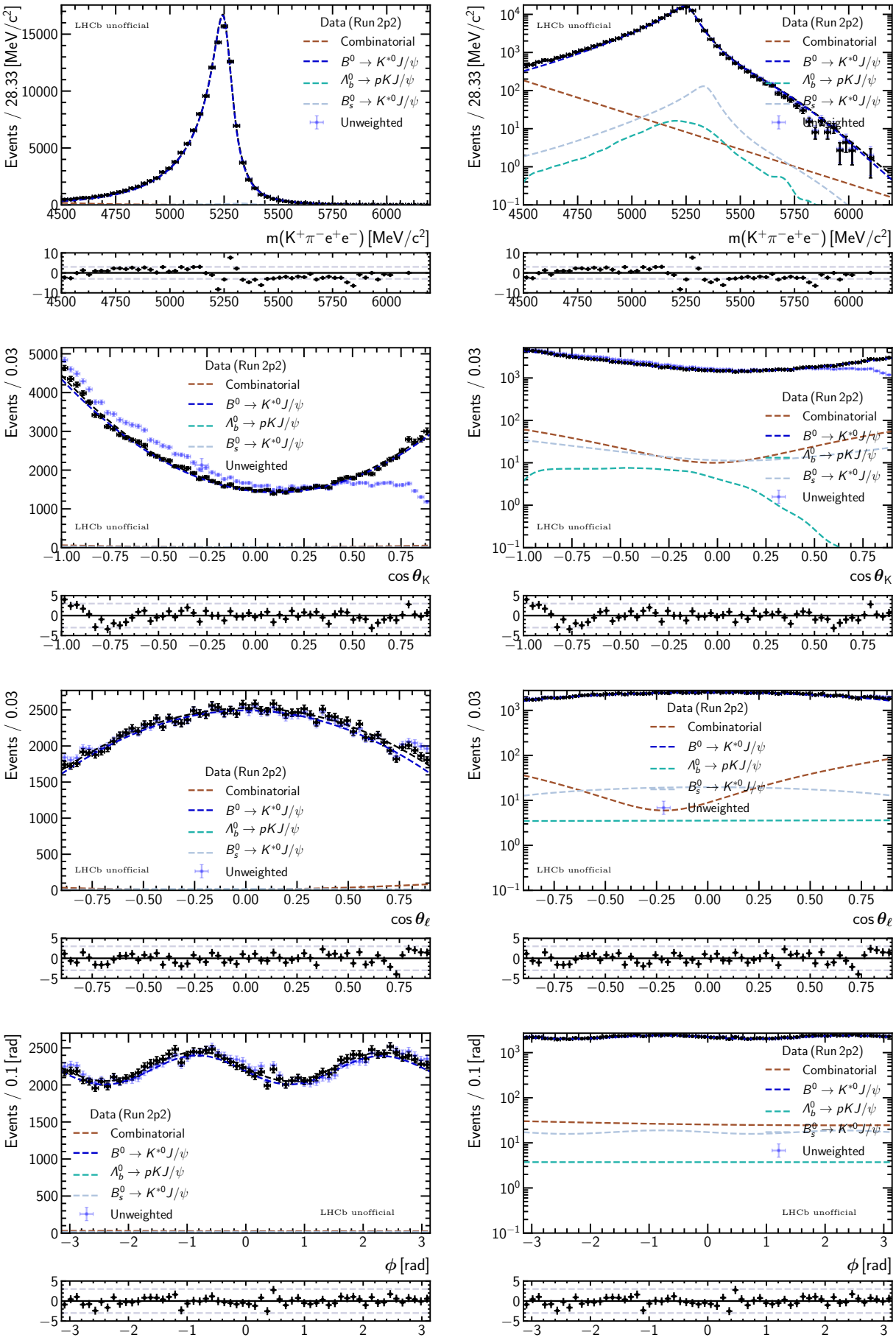


Figure 9.6: Result of the simultaneous control mode angular fit (Run 2p2).

9.2 Systematic uncertainties

The uncertainties associated with the measured angular observables include the statistical uncertainties of the fits, as well as a number of systematic uncertainties, which originate from the strategies used in the design of the analysis and data selection. The choice of which sources to evaluate out of a large range of possible options for which comprehensive assessment is not feasible, is based on its expected size compared to the statistical uncertainties.[5] Significant sources include aspects of background modelling, the effective acceptance parametrisation strategy, the impact of neglecting residual backgrounds, and the veto against B^+ decays, which leads to the breaking of the factorisation assumption between mass and $\cos\theta_K$.

The evaluation of most systematic uncertainties involves the generation of toys. First, an alternative configuration, or configuration A, that is representative of the range of variation of the source in question is defined. Then, it is used to generate a large number of toys, which are fitted twice – once with the matching alternative and once with the nominal configuration (configuration N), which is aligned to the main analysis strategy. This gives rise to two sets of results – one from toys generated with A and fitted with A ('AA' results), and the other from toys generated with A and fitted with N ('AN' results). The latter show the impact of using the nominal configuration when, for example, the data actually follows the alternative. This may lead to biases in the central values of the observables or changes to the widths of the observable distributions. However, in addition to the systematic effect, these results are also affected by statistical fluctuations. To take these into account, the result of the matching AA fit is subtracted from that of the AN fit. The deviation of the mean of the observable differences (the bias when N is used instead of A) can be taken as the systematic uncertainty. However, at times the width of this distribution may be comparable in size to the shift in the mean. This indicates a difference in the statistical uncertainties between the AA and the AN fits, and can occur when, for example, the mismatch in the AN fit leads to greater difficulties in determining the values of certain observables compared to the matching AA fit. When this occurs, the widths of the difference distributions are added in quadrature to the mean values. Not taking the width into account when it is comparable to the mean can lead to the underestimation of the possible range of fluctuations in the result if the data follows A.

Note that in the application of this strategy, ideally the AA toys are unbiased, such that any biases appearing the AN results can be attributed purely to the effect of the mismatched configurations. However, AA toys may be biased, especially at the low statistics of the rare mode. This is not a significant problem if the underlying mechanism causing the bias is unchanged to a good approximation in the AN toys, as they can approximately cancel together with the statistical effect when the differences are calculated. However, if the mechanism responsible disappears in the AN fit, then the differences would include its effects. This type of situation rarely occurs, but for cases where significant biases are found and are not expected to be shared, to be slightly conservative, their absolute values are added to the systematic uncertainties found.

To ensure good reproduction of the features of the nominal fit, including potential statistics-dependent ones, all toys are generated using realistic component yields found from the fit to the larger q^2 range. Full toys including both signal and background components are used to quantify systematic uncertainties related to background modelling, while signal-only toys are used to study the ones related to the effective acceptance functions, as they mainly affect the signal. The assessments for the smaller q^2 range are carried out after removing events

[5] As the total uncertainties are given by the sum of the statistical and systematic uncertainties in quadrature, effects that are much smaller than the former are mostly irrelevant for the interpretation of the results.

with $q^2 > 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ from larger q^2 toys. The nature of the main sources of systematic uncertainties, and the methods used to quantity them, are discussed in detail below. For clarity, the numerical results of these studies are summarised in Tables 9.3 and 9.4 for the large and small q_c^2 ranges, respectively, and shown graphically as fraction of the expected statistical uncertainties in Figure 9.7. The total systematic uncertainties are calculated from the sum in quadrature of the listed sources.

Table 9.3: Summary of systematic uncertainty of the rare mode angular fit for the large q^2 range.

	DSL	Comb. DSL-like	Part. reco.	Bias uncert.	Higher orders	Coef. uncert.	Sim. corr.	Control leak	J/ψ comb.	B^+ veto	Peaking bkg	Total (σ_{syst})
F_L	0.021	0.025	0.006	0.002	0.008	0.003	0.009	0.006	0.015	0.018	0.017	0.046
S_3	0.005	0.005	0.001	0.002	0.006	0.002	0.004	0.001	0.002	0.005	0.012	0.017
S_4	0.008	0.008	0.003	0.003	0.014	0.005	0.012	0.001	0.004	0.009	0.038	0.045
S_5	0.011	0.007	0.004	0.003	0.009	0.004	0.008	0.002	0.004	0.009	0.045	0.050
A_{FB}	0.022	0.059	0.003	0.002	0.008	0.002	0.009	0.006	0.013	0.011	0.011	0.067
S_7	0.005	0.005	0.001	0.003	0.011	0.004	0.010	0.001	0.003	0.007	0.018	0.026
S_8	0.007	0.007	0.002	0.003	0.015	0.006	0.016	0.001	0.003	0.009	0.038	0.047
S_9	0.005	0.006	0.001	0.002	0.006	0.002	0.005	0.001	0.003	0.004	0.041	0.043
P_1	0.036	0.059	0.013	0.014	0.046	0.017	0.035	0.007	0.025	0.051	0.054	0.122
P_4'	0.021	0.031	0.010	0.007	0.033	0.012	0.026	0.003	0.010	0.027	0.077	0.101
P_5'	0.019	0.031	0.013	0.006	0.022	0.009	0.018	0.005	0.014	0.028	0.092	0.109
P_2	0.050	0.164	0.005	0.004	0.022	0.006	0.018	0.016	0.043	0.023	0.024	0.183
P_6'	0.012	0.011	0.003	0.007	0.025	0.010	0.024	0.002	0.006	0.018	0.036	0.057
P_8'	0.017	0.016	0.004	0.007	0.033	0.013	0.037	0.003	0.006	0.023	0.077	0.099
P_3	0.017	0.027	0.007	0.007	0.022	0.009	0.019	0.004	0.015	0.024	0.098	0.112

9.2.1 Background models

The systematic uncertainty related to component modelling is dominated by that of the backgrounds rather than the signal, which has comparatively little effect on the analysis. Among these, the largest impact originates from the strategy used to model DSL decays due both to its abundance in the final data sample, and its asymmetric $\cos\theta_\ell$ distribution. The strategies used to estimate systematic uncertainties related to the three background components are discussed in the following sections.

Combinatorial and double semileptonic

The nominal models of the 'DSL' and 'combinatorial' backgrounds are designed to be effective in nature and complementary. The former is dominated by genuine DSL decays with additional contributions from misidentified and partly combinatorial events (potentially also semileptonic decays), and the latter by standard combinatorial background, but includes residual DSL-like events that are not taken into account by the 'DSL' component. The main limitations of this

Table 9.4: Summary of systematic uncertainty of the rare mode angular fit for the small q^2 range.

	DSL	Comb. DSL-like	Part. reco.	Bias uncert.	Higher orders	Coef. uncert.	Sim. corr.	Control leak	J/ψ comb.	B^+ veto	Peaking bkg	Total (σ_{syst})
F_L	0.021	0.027	0.006	0.002	0.009	0.003	0.010	0.008	0.006	0.018	0.022	0.048
S_3	0.006	0.005	0.001	0.002	0.006	0.002	0.005	0.001	0.001	0.005	0.021	0.025
S_4	0.009	0.009	0.003	0.004	0.016	0.006	0.013	0.001	0.002	0.010	0.049	0.056
S_5	0.010	0.007	0.004	0.003	0.010	0.004	0.009	0.002	0.002	0.009	0.048	0.053
A_{FB}	0.022	0.058	0.003	0.002	0.009	0.003	0.010	0.003	0.006	0.011	0.013	0.066
S_7	0.006	0.005	0.002	0.003	0.012	0.005	0.011	0.001	0.001	0.008	0.026	0.034
S_8	0.008	0.008	0.002	0.004	0.016	0.006	0.017	0.001	0.001	0.010	0.035	0.046
S_9	0.006	0.006	0.001	0.002	0.006	0.002	0.005	0.001	0.001	0.005	0.052	0.054
P_1	0.050	0.077	0.022	0.017	0.055	0.022	0.046	0.021	0.020	0.074	0.096	0.174
P_4'	0.027	0.041	0.013	0.008	0.039	0.014	0.031	0.003	0.005	0.032	0.098	0.126
P_5'	0.022	0.042	0.016	0.007	0.026	0.010	0.021	0.007	0.007	0.032	0.098	0.120
P_2	0.069	0.176	0.012	0.005	0.026	0.008	0.024	0.010	0.018	0.036	0.019	0.198
P_6'	0.014	0.016	0.006	0.007	0.029	0.011	0.026	0.002	0.003	0.021	0.052	0.073
P_8'	0.021	0.021	0.007	0.008	0.039	0.015	0.042	0.003	0.003	0.026	0.070	0.100
P_3	0.025	0.042	0.010	0.009	0.027	0.011	0.021	0.004	0.008	0.033	0.124	0.143

strategy originate from the assumptions that the distributions and ratios of DSL, DSL-like and combinatorial events are the same in the $K\pi e\mu$ and $K\pi ee$ data. In addition, one further assumption is that any differences between the (integrated) angular shape of the DSL model taken from the larger mass range of $4500 - 5200\,\text{MeV/c}^2$ and the one in the nominal range of $4900 - 5700\,\text{MeV/c}^2$ due to changes in the ratios of DSL to DSL-like contributions will be taken into account by the effective combinatorial model in Step 2 of the two-step procedure. Precise estimations of the systematic uncertainties associated with this strategy is challenging due to low statistics, which obstructs efforts to quantify the differences between the ee and $e\mu$ final states, and limited understanding of the composition of the effective DSL model. In light of this, a conservative approach is chosen, where the dominant DSL component is considered separately from that of the combinatorial, partly combinatorial and misidentified components. Simulation is used for the former, while an alternative data-based approach is used for the latter. The total systematic uncertainties of the nominal two-step approach is taken to be the sum in quadrature of both sources.

Double semileptonic

The distribution of the DSL background in the $K\pi e\mu$ sample may differ from that in the $K\pi ee$ due to experimental differences between electrons and muons. For example, the (slightly) improved resolution due to the presence of a muon in the final state may lead to better separation between DSL and signal in mass, such that the composition of events leaking into the signal region may be different. The presence of a muon in the final state may also lead to differences in the rates of hadron-lepton misidentification. Furthermore, for events belonging to the L0En

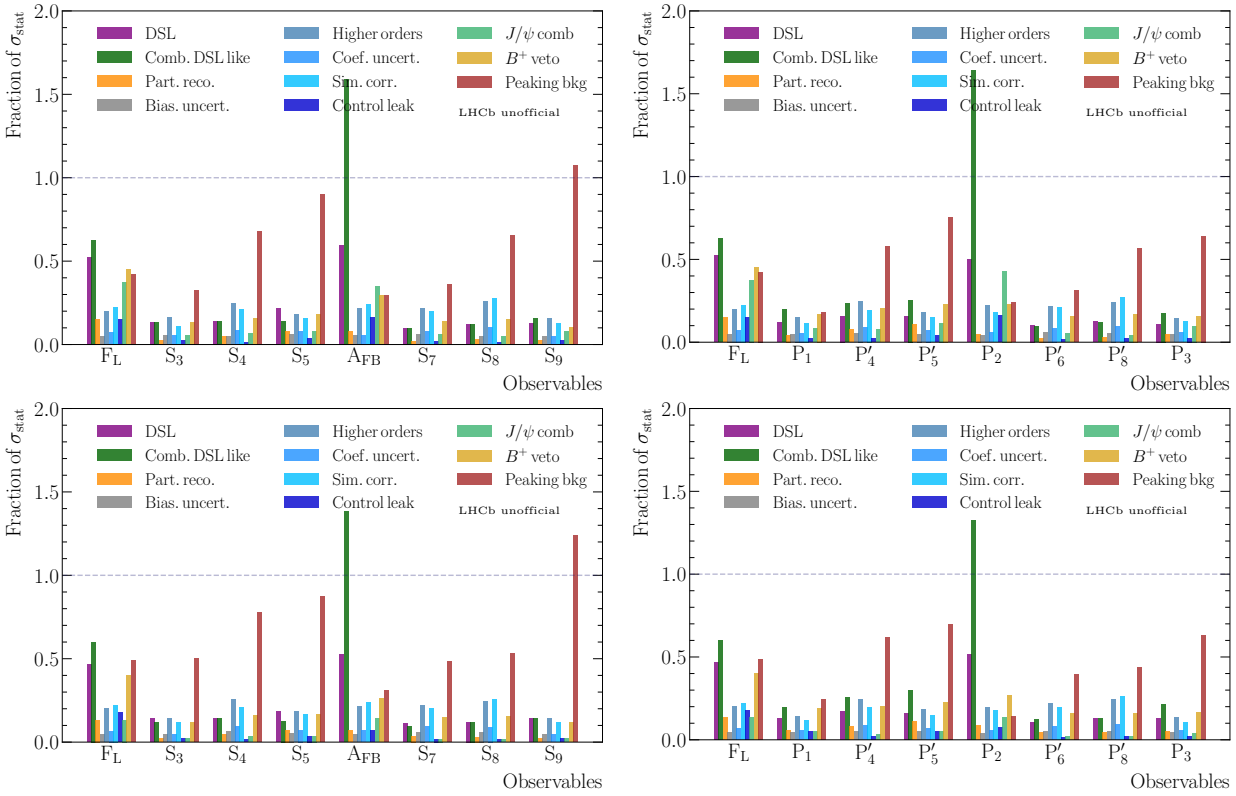


Figure 9.7: Different sources of systematic uncertainties as fractions of the expected statistical uncertainties for the larger and smaller q^2 ranges of $1.1 < q^2 < 7.0\,\text{GeV}^2/\text{c}^4$ (top) and $1.1 < q^2 < 6.0\,\text{GeV}^2/\text{c}^4$ (bottom).

trigger category (approximately half the data samples), candidates that are only retained due to positive L0 muon trigger decision are also included. These types of features introduce differences between the shape of the $\cos\theta_\ell$ distribution of the two samples, in particular its width (the sharpness of the peak). To assess the possible impact conservatively, the model obtained from simulated $B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)e^-\bar{\nu}_e)e^+\nu_e$ decays is used as an alternative. The difference, in particular in the width of the $\cos\theta_\ell$ peak, between truth-matched simulated DSL decays and the model obtained from the $K\pi e\mu$ sample is expected to be larger than possible differences between the effective DSL distributions in the $K\pi e\mu$ and $K\pi ee$ data. A comparison between the nominal and alternative model is shown in Figure 9.8.

The alternative model is used to generate three hundred toys, which are otherwise identical to those used for sensitivity studies. They are then fitted using both the alternative and the nominal configurations. The toys fitted with the matching alternative models show behaviour that is largely consistent with those of the sensitivity toys in both q^2 ranges. The distributions of the differences are generally well described by Gaussian functions, and show significant shifts in the mean values from zero. The widths are found to be generally comparable in size to the mean, therefore both values are added in quadrature.

This source of systematic uncertainty strongly affects the observables of A_{FB}/P_2 and F_L, amounting to around half of their statistical uncertainties. This can be expected as the alternative DSL model shows greater curvature than the nominal in the $\cos\theta_K$ distribution, which, in the absence of varied background parameters that can adjust to take this difference into account, leads to the modification of the signal shape (background has an approximately $\cos^2\theta_K$ distribution). As F_L is multiplied to the $\cos^2\theta_K$ term, it is particularly prone to be modified. Similarly, the increased asymmetry of the $\cos\theta_\ell$ distribution of the alternative model due to its

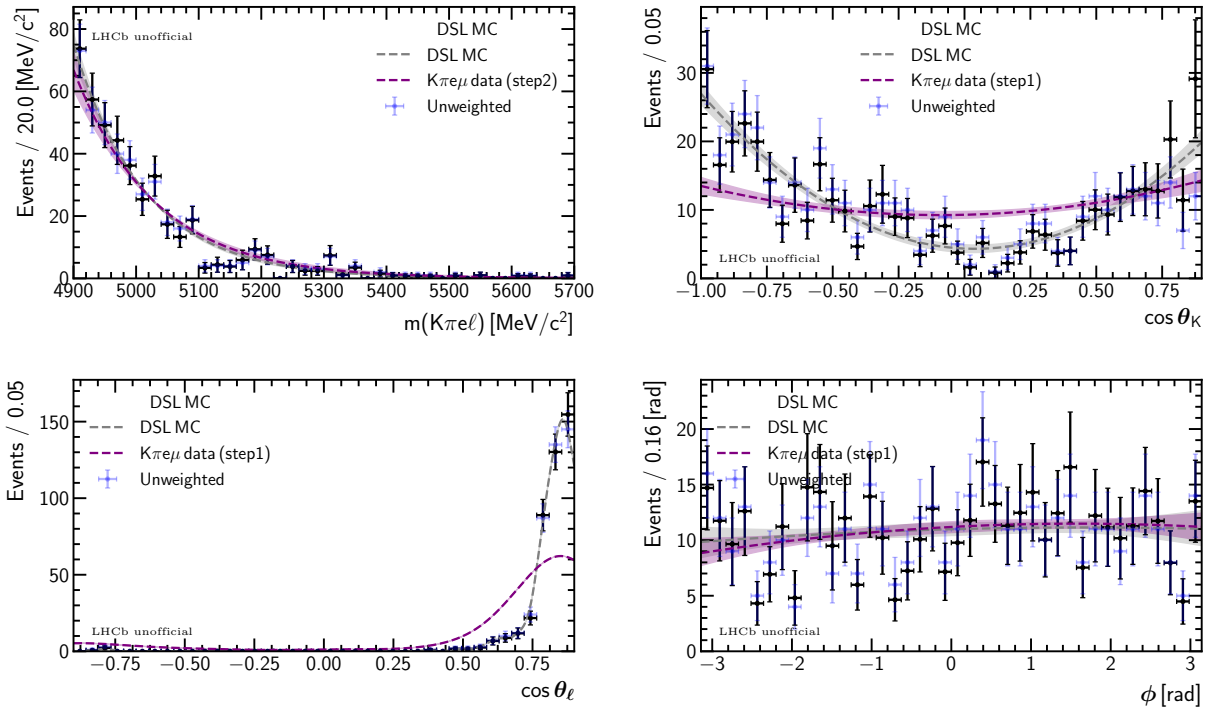


Figure 9.8: Comparison between the alternative (gray) and nominal (purple) models of the DSL component obtained from DSL simulation and $K\pi e\mu$ data, respectively. The lighter coloured bands represent 1σ uncertainty of the fit.

significantly narrower peak can be accommodated for by adjusting the value of A_{FB}, which is multiplied to a term that contains $\cos\theta_\ell$.

Combinatorial

In the nominal strategy, the description of DSL-like events that contains combinatorial and/or misidentified tracks is effectively split between the 'DSL' and 'combinatorial' models, which also includes standard combinatorial background. As discussed in Section 8.2.1, the main DSL component becomes largely negligible after $\sim 5200\,\mathrm{MeV/c^2}$, whereas the partly combinatorial DSL-like events populate the full signal mass window. The two-step procedure is designed to aid the separation of the main component (which unavoidably also includes DSL-like events that survive the tight MVA cut[6]) from the more combinatorial-like contributions by first extracting its angular shape from a region where it is known to dominate $(4500 - 5200\,\mathrm{MeV/c^2})$, before fitting the nominal mass range of $4900 - 5700\,\mathrm{MeV/c^2}$. As the ratio of the main DSL to DSL-like events is expected to differ between these mass ranges due to their slope differences (Figure 8.13), the shape obtained from Step 1 is expected to leave a part of the latter unaccounted for. By allowing the combinatorial component to very in the Step 2 fit, it is expected to take combinatorial and remaining partly combinatorial DSL events into account. Sources of systematic uncertainties include possible shape and ratio differences between the $K\pi e\mu$ and $K\pi ee$ samples for both DSL-like, and combinatorial backgrounds, and the assumption that the (relatively smooth) second order Chebyshev polynomials used for the combinatorial model can

[6]The combinational MVA appears to be less effective in removing DSL-like events compared to standard combinatorial. This may be due to part of this contribution consisting of misidentified candidates, which are more signal-like (e.g. have B^0 p_T distributions that are more similar to the signal/properly reconstructed DSL).

describe remaining DSL-like events, which may peak more sharply in $\cos\theta_\ell$, sufficiently well.

To address potential differences between the backgrounds in the $K\pi ee$ and $K\pi e\mu$ samples, as an alternative, models for the combinatorial and DSL-like events can be obtained from the same-sign sample where the main DSL component is absent using the two-step procedure. While this sample is not expected to describe all features present in this type of background in the (opposite-sign) data (Appendix C.1.1), it can provide an alternative effective model for the estimation of systematic uncertainties.

Besides shape differences due to electron and muon reconstruction effects, the two-step procedure assumes that the ratio between the standard combinatorial, DSL and the approximately symmetric DSL-like component is the same between $K\pi e\mu$ and $K\pi ee$. Furthermore, it assumes that the second order polynomial used for the combinatorial component is sufficient to describe a mixed sample of smooth combinatorial, and peaking DSL-like events. These assumptions are very important. If the $K\pi e\mu$ data after the nominal MVA cut (Step 2 sample) is dominated by standard combinatorial events, for example, then the sensitivity to peaking contributions from DSL-like background is lost. Consequently the resulting combinatorial angular model will be smooth. If the ratio of DSL-like events turns out to be different (e.g. much higher) in the $K\pi ee$ data, then there would be limited flexibility in the existing configuration to accommodate an excess near $\cos\theta_\ell = -1$ and thus the signal can be affected. The same type of problem may also occur if the polynomial used does not describe peaking structures sufficiently well.

To assess the impact of these assumptions, a strategy is devised to introduce a third symmetric 'quasi-DSL' component. In order to reduce overlaps in definitions and to allow for the possibility of shape differences between backgrounds in $K\pi e\mu$ and $K\pi ee$ samples to be taken into account at the same time, an attempt is made to remove the (assumed) symmetric part from the nominal $K\pi e\mu$ DSL model, leaving only the main peak near $\cos\theta_\ell = 1$. The symmetric part is then replaced by its same-sign version, and the combinatorial model from the same-sign is likewise used. An alternative data fit is made to obtain reasonable yields for all three effective components for toy generation (more details can be found in Appendix C.1).

The results of the fits made using the alternative configuration show behaviour that is consistent with that of the nominal sensitivity toys in both q_c^2 ranges. The mean values of the difference distributions are generally displaced from zero, and the widths are often of comparable sizes. Therefore, the systematic uncertainties are taken to be the sum in quadrature of both values.

Among the observables, A_{FB}/P_2 is found to be most strongly affected, with a systematic uncertainty that is around one and a half times the value of its statistical uncertainty. A_{FB}, which describes the forward-backward asymmetry of the dielectron system, is as expected, particularly sensitive to asymmetric differences in the $\cos\theta_\ell$ distribution. An excess near $\cos\theta_\ell = -1$ that is present in the alternative, but not in the nominal model, affects it directly in the absence of adjustable background components. F_L is the second most strongly affected observable with systematic uncertainties at the level of 60% of its statistical uncertainty. This is expected to be mainly due to its sensitivity to unaccounted for linear/curved discrepancies in the $\cos\theta_K$ description, which is in this case caused by the distribution of the DSL-like component.

These results suggest that modifications to the nominal fit strategy is necessary. At the same time, they indicate that this systematic uncertainty can be reduced by introducing relatively small changes (e.g. addition of a symmetric component), which largely preserves the current sensitivity and fit behaviour.

Partially reconstructed

In the nominal configuration, the partially reconstructed background is modelled using a data-driven approach with correction weights derived from the charmonium mode of $B^+ \to K^+\pi^+\pi^- J/\psi(\to \mu^+\mu^-)$. One potential shortcoming is that its hadronic structure may be different from that of the rare mode. Another complication is that the reweighting does not consider the full degrees of freedom of the system. In particular, it does not explicitly correct the angular distributions. Furthermore, while the charmonium mode is known to obey LFU, if the B anomalies are due to NP, then this may not hold for the rare mode. However, performing a detailed assessment is challenging given the complexities of the system and the current level of understanding. Nevertheless, due to the limited contribution of this component, the impact of mismodelling is expected to be small. Therefore, a simplified strategy is used, where an alternative model is parametrised instead from simulations of the two dominant modes of $B^+ \to K_1^+(1270)e^+e^-$ and $B^+ \to K_2^{*+}(1430)e^+e^-$, which are combined with a 2:1 ratio based on the results of [189]. A comparison between the distributions of these samples and that of the nominal are shown in Figure 9.9. The new model is parametrised in the same way as the nominal (using a KDE lineshape for the mass distribution, and Chebyshev polynomials up to second order for the angles). Toys are generated with standard signal and background components, with the exception that this new model is used for the partially reconstructed background. They are then fitted with the matching and the nominal models.

The toys fitted with the matching alternative models are generally consistent in behaviour with the nominal sensitivity toys for both q_c^2 ranges. The distributions of the differences are somewhat non-Gaussian. Fits to $\sim 3\%$ of the toys are found to result in differences for F_L, S_3/P_1, S_4/P_4' and S_5/P_5' that are much larger than the average. A shared characteristic present in most of these cases is that the difference in $f_{\rm PR}^{\rm sig}$ is large ($f_{\rm PR}^{\rm sig} > 0.1$), which is not seen in the alternative data fit. After removing these entries the distributions are found to be approximately Gaussian, therefore the systematic uncertainties are estimated using the sum in quadrature of the mean and the widths of the difference distributions. For all observables, these values are found to be less than 20% of their statistical uncertainties.

9.2.2 Effective acceptance parametrisation

The effective acceptance functions aim to parametrise angular and q^2 distortions caused by experimental (and FSR) effects in order to correct the measured distributions back to a form that can be described by the differential decay rate equation. As the correction weights modify the distributions of the signal directly, their associated uncertainties should be checked. The main sources include simulation-data corrections, the statistics of the samples, and the choice of using the particular polynomial based model.

Parametrisation limitations

The model used to parametrise the effective acceptance effect is not able to describe all features present in the samples. Examples include the shape of the $\cos\theta_K$ distribution close to $\cos\theta_K = 1$, the sharp dents in the $\cos\theta_\ell$ distribution created by the vetoes against semileptonic decays and the kink in the generator level q^2 distribution, which is related to the parametrisation of effective WCs in the `EvtGen` model. These imperfections lead to the imperfect retrieval of the generator level observables. Nevertheless, the differences are found to be relatively small (Section 8.1). Based on these results, for each observable, the average of the shift seen in the three Run subsamples, weighted using the control mode yields, is taken to be the bias due

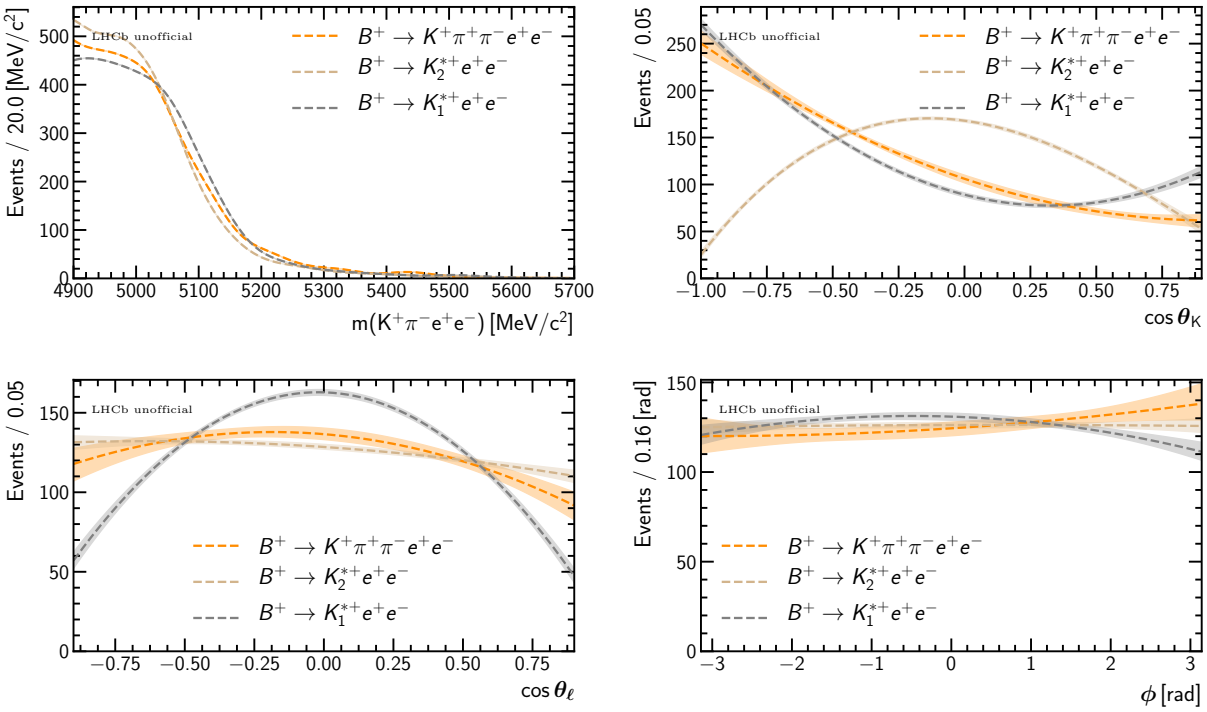


Figure 9.9: Comparison between the models of the partially reconstructed background obtained from the phase-space simulation with data-driven correction weights, and from simulated $B^+ \to K_1^+(1270)e^+e^-$ and $B^+ \to K_2^{*+}(1430)e^+e^-$ decays. The acceptance weights are included in all cases. The lighter coloured bands represent 1σ uncertainties of the fits.

to the limitations of the chosen model, and used to correct the final results. Their associated uncertainties are considered as systematic uncertainties.

Higher order description

For the parametrisation of the effective acceptance functions, the lowest orders of the Legendre and Fourier polynomials that can provide satisfactory modelling of the angular and q_c^2 distributions of the simulated samples are chosen. The maximum orders used are three for q_c^2, five for $\cos\theta_K$, four for $\cos\theta_\ell$ and four for ϕ. Lower orders are favoured to reduce the size of the low efficiency/pathological region (Section 7.4) that tends to enlarge with the use of higher orders, and to avoid fluctuations at the edges of the parametrisation ranges (Runge phenomenon [190]). However, using higher order polynomials may provide improved description of the effective acceptance in certain regions of the phase space, such as near $\cos\theta_K = 1$. To check the impact of the order choice, alternative acceptances are parametrised with all polynomial orders increased by three. A comparison between the alternative and nominal acceptance functions for the Run 2p2 sample is shown in Figure 9.10, and the enlargement of the pathological region is illustrated in Figure 9.11 (compare with Figure 8.3). Five hundred signal only toys are generated using these alternative functions, and fitted with acceptance corrections derived from the alternative, and the nominal functions.

The distribution of the observable differences are found to be Gaussian to a good approximation, and to have widths that are comparable to (or larger than) the mean values. The systematic uncertainties are therefore taken to be the sum in quadrature of both the mean and the widths. This source of systematic uncertainty is at a level of $15 - 25\%$ of the statistical

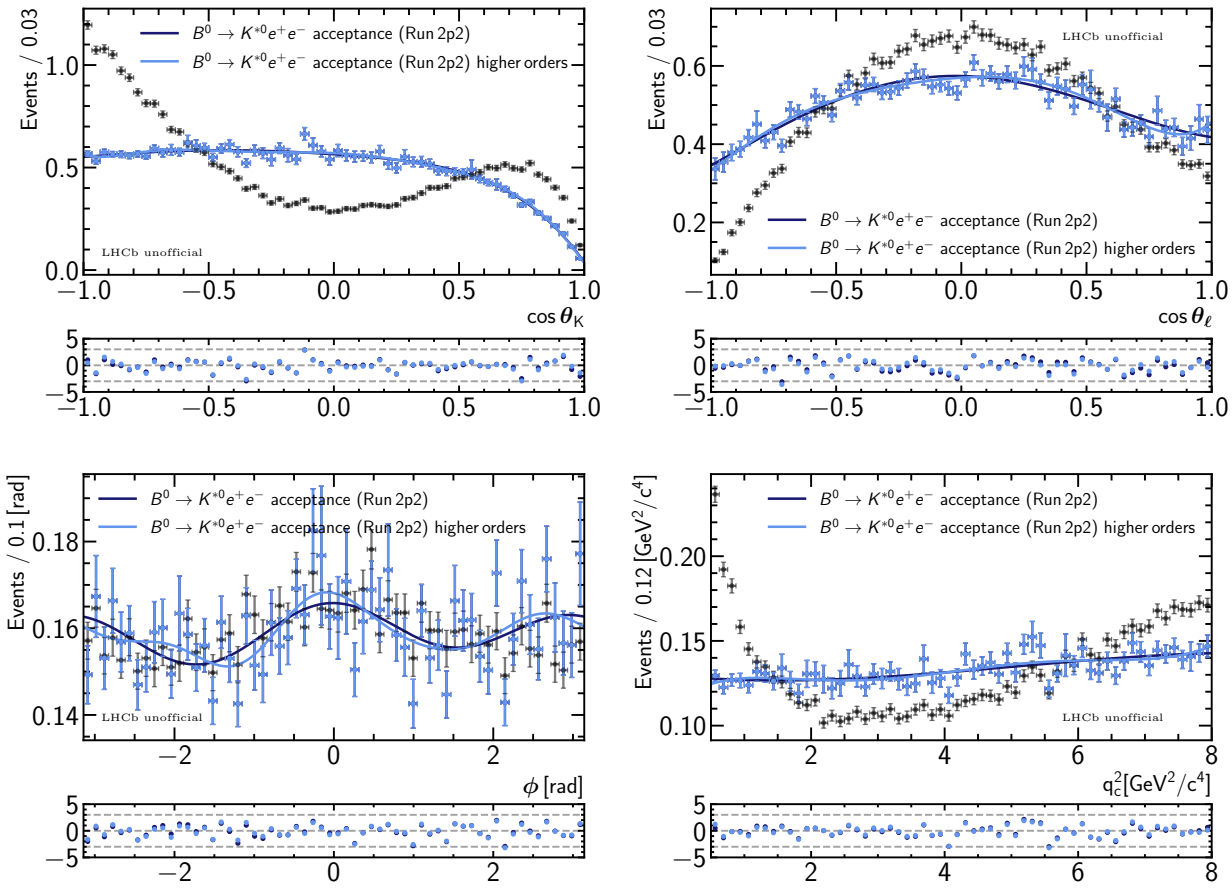


Figure 9.10: Comparison between the projections of the nominal Run 2p2 rare mode acceptance function and that of the alternative made with all maximum polynomial orders increased by three.

uncertainty for all observables.

Coefficient uncertainties

The uncertainties of the coefficients of the effective acceptance functions constitute a minor source of systematic uncertainty. To quantify this contribution, the nominal samples used to parametrise the three functions (one for each Run subsample), are bootstrapped and reparametrised five hundred times to produce five hundred sets of alternative acceptance functions. One signal-only toy is generated per set, and fitted with the matching corrections as well as the nominal ones.

The mean of the distributions of the observable differences are well described by Gaussian functions, and generally consistent with zero, which can be expected due to the random nature of the alternative functions. In this case the widths are taken directly as systematic uncertainties. For all observables, the values found are at the level of 10% of their statistical uncertainties or less.

Simulation correction strategy

The simulation-data correction strategy mainly affects the analysis through the effective acceptance functions. The nominal chain of corrections involves first resampling the distribution of

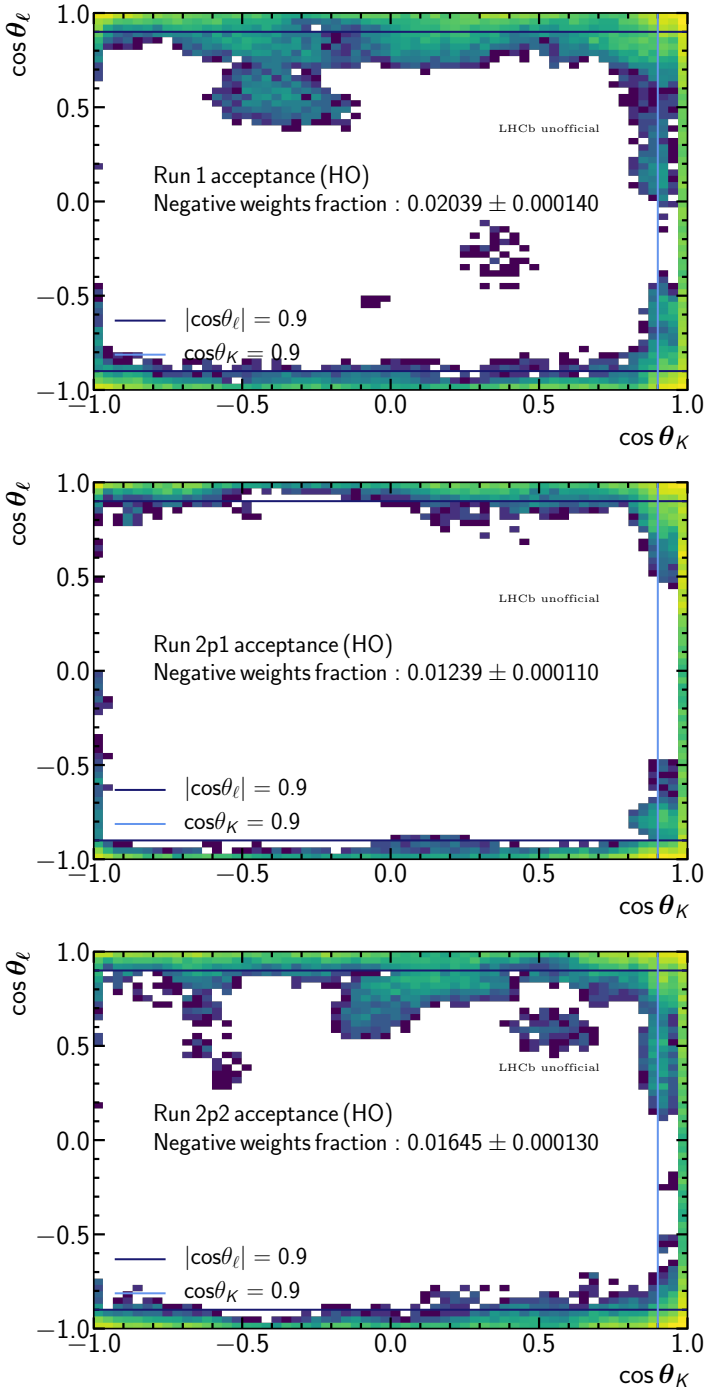


Figure 9.11: The distribution of 10^6 test events from a uniform sample in the angles and q_c^2 that received negative acceptance correction weights from the Run 1 (top), Run 2p1 (middle) and Run 2p2 (bottom) effective acceptance functions parametrised with all polynomial orders increased by three.

PID variables, then correcting the L0 and HLT trigger efficiencies, before reweighting the distributions of a number of kinematic and multiplicity variables to match those of the background subtracted control mode candidates in data.

An alternative correction chain [67] can be used to estimate the systematic uncertainties related to the chosen methodology. The key differences between this and the nominal strategy include the method used to obtain the kinematic and multiplicity weights, the PID correction approach, and the use of additional corrections for electron tracking efficiency and reconstructed variables. The main differences between the nominal and the alternative strategies are discussed below.

Multiplicity and kinematics

Candidates of the muon mode of $B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ can be selected with high efficiency by triggering on the final state muons, such that their reconstructed quantities provide good approximations of the generator level ones. Therefore, they can be used to correct, to a good approximation, the generator level distributions of important kinematic and multiplicity variables ($B^0\,p$, $B^0\,p_T$, $B^0\,\eta$ and n_{Tracks}). This is achieved by first correcting the PID distributions of the muon mode simulation, followed by L0 and HLT trigger efficiencies, and finally extracting the kinematic and multiplicity weights using background subtracted $B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ candidates. These weights are expected to be largely independent of the lepton flavour, and they are therefore also applied to the electron mode to correct its generator level distributions This approach largely avoids the 'double-counting' that can occur when distributions that are more strongly affected by the trigger are corrected directly (Appendix A), and allows for trigger corrections to be made based on a more correct set of kinematics variables.

PID

In the nominal approach, PID corrections are made by creating resampled variables, which are then used in cuts instead of the original versions. Alternatively, the weights that encode the efficiencies of a given PID cut can be extracted directly from the `PIDCalib` samples. These samples are binned in p_T, η and n_{Tracks}, and the efficiency in each bin is obtained by fitting the data with and without applying the cut in question and calculating the ratio. This 'fit-and-count' approach has some advantages over the nominal approach for the correction of electron PID variables. In the case of the latter, the $s\mathcal{P}$lot technique is used to statistically subtract background in the calibration samples, and the resulting (weighted) distribution in the PID variable of interest, p_T, η and n_{Tracks}, is described by a pdf. The (comparatively) statistically limited calibration sample together with the potential presence of correlations with the discriminating variable can lead to biases [191, 161]. However, disadvantages include the loss of correlations and difficulties in treating PID cuts used in complex vetoes.

Additional corrections

Additional stages present in the alternative correction chain include the correction of electron tracking efficiencies, and the reweighting of extra reconstructed variables, namely the significance of the impact parameters of the B^0 and the J/ψ meson (B^0, $J/\psi\,\chi^2_{\text{IP}}\,(\text{PV})$), in addition to (reconstructed) $B^0\,p$, $B^0\,p_T$ and $B^0\,\eta$. Unlike the previous stage, which aims to correct generator level multiplicity and kinematic quantities, the correction of these reconstructed variables is carried out specifically for the electron mode (as reconstruction effects are different between electron and muon modes). In this case, one important difference with respect to the nominal approach is that the correction is made by attempting to isolate a relatively pure signal sample through the application of the full selection requirements, and making a narrow cut around the

signal peak in the constrained invariant mass of $m(K\pi ee)_{\mathrm{PV},\,J/\psi}$, rather than via $s\mathcal{P}$lot background subtraction, which can be biased due to correlations between the B^0 invariant mass and quantities to be corrected.

To estimate the systematic uncertainties associated with the nominal correction chain, effective acceptance functions are parametrised including all corrections weights obtained using the alternative strategy. Toys are generated with the alternative functions and fitted with correction weights from the alternative as well as the nominal ones. The systematic uncertainties are taken to be the sum in quadrature of both the mean and the widths of the difference distributions. The level of uncertainty is around $10 - 25\%$ of the statistical uncertainty for all observables.

9.2.3 Control mode leakage

Leakage of control mode events into the rare mode q^2 region is reduced when the q_c^2 with improved resolution is used. In addition, it is further suppressed by the choice of the lower limit of the mass window ($4900\,\mathrm{MeV/c^2}$) as well as the cut on the absolute value of $\cos\theta_\ell$ ($|\cos\theta_\ell| < 0.9$). However, residual events are expected to remain and can be estimated by scaling the yields of the control mode data fits in the q_c^2 region of $7.0 - 11.0\,\mathrm{GeV^2/c^4}$ based on efficiency ratios from simulation. Around 12 events are expected to be present in the full data sample when estimated in this way.

Full toys with the same signal and background models as the sensitivity toys are generated with the addition of this component using estimated yields from the an alternative data fit, which is made by Gaussian-constraining its fraction based on the aforementioned estimation (Appendix C.2). The same constraint is also used when fitting toys with the matching configuration. The results of these fits show biases in F_L of around 15% of the statistical uncertainty, while all other observables are unaffected. The difference distributions show deviations from Gaussian behaviour in a few cases. Nevertheless, the magnitudes of the uncertainties are small, and the Gaussian approximation is still expected to provide reasonable estimations. The systematic uncertainties are therefore taken to be the sum in quadrature of their mean and width values. In the case of F_L, as the mechanism driving the bias is not fully understood, its absolute value is added directly to the calculated uncertainty.

This source of systematic uncertainty is generally negligible for most observables, as it amounts to less than 10% of their statistical uncertainties. However it is slightly more prominent for A_{FB}/P_2 in the larger q_c^2 region, where it constitutes around 15% of its statistical uncertainty.

9.2.4 Charmonium combinatorial

The use of the B^0 mass and PV constraints in the calculation of the q_c^2 variable cause the distribution of combinatorial events that contain real J/ψ decays to be shifted into the rare mode q_c^2 window (Section 7.2.3). This leads to the formation of a broad peaking structure in a region that otherwise contains limited numbers of combinatorial events from the tail of its exponential distribution, which can bias the determination of the combinatorial and signal yields when an exponential function is used.

The sample of background-like events in the control mode simulation is expected to provide a good approximation of this type of background, as its distinctive mass and q_c^2 distributions are shaped largely by kinematic effects. To obtain realistic yields, models are parametrised from

these events and used to perform an alternative data fit (Appendix C.3). As it is difficult to separate from the standard combinatorial background, its fraction is Gaussian constrained to the expected value (Section 7.2.3). The alternative fit found around 60 events of this type. Its component yields are used to generate toys, which are fitted once using the matching configuration, where the Gaussian constraint is applied, and once using the nominal configuration.

The behaviour of the fits made using the matching configurations is generally consistent with that of the nominal sensitivity toys, although only the mean of the pull distributions and the sensitivity values can be compared, as the (weighted) uncertainties are not calculated with the inclusion of the constraint term. The distribution of observable differences show shifts from zero, with widths that are comparable in size to the mean, therefore both are summed in quadrature. The resulting values are found to be reasonably large for F_L and A_{FB}/P_2, at the level of $30-40\%$ of their statistical uncertainties, while they are small (less than 10% of the statistical uncertainty) for all other observables.

9.2.5 $B^+ \to K^+e^+e^-$ veto

The veto against $B^+ \to K^+e^+e^-$ decays removes a region of the phase space at high B^0 reconstructed mass. The most important consequences of this include the creation of visible correlation between mass and the $\cos\theta_K$ angle leading to the breaking of the factorisation assumption, and the distortion of the combinatorial mass distribution from the assumed exponential shape. Note that the distortion caused by the veto cannot be corrected for fully by the effective acceptance function due to the involvement of the B^0 mass as well as the the full removal of events. Its effect can be reproduced using efficiency weights obtained from the $K\pi e\mu$ data sample in bins of the $m(K\pi ee)_{\rm PV}$, $\cos\theta_K$, and q_c^2 [142]. Due to the kinematic nature of this cut, the selections applied to these samples are relatively unimportant. To increase statistics, samples passing the stripping requirements are used directly, and all available data are combined. The parametrisation ranges are chosen to be between -1 to 1 for $\cos\theta_K$, $1-14\,{\rm GeV}^2/{\rm c}^4$ for q_c^2 and $4500-6800\,{\rm MeV}/{\rm c}^2$ for $m(K\pi ee)_{\rm PV}$, which are divided into 100, 40, and 50 bins, respectively.

To assess the impact on the observables, first an alternative data fit is made without the B^+ veto. In this case all effective acceptance functions are parametrised without it, and all signal and background models are updated with these weights. The result of this fit is used to generate around three hundred toys, which are then fitted with the matching as well as the nominal acceptance corrections and models. For the nominal fit, the efficiency values from the three dimensional map are used to randomly reject events to mimic the effect of the veto. Note that this introduces a difference in the samples used for the alternative and the nominal fits. However, as the efficiencies of the veto for background components present in the toy are comparable with that of the signal, potential changes in the sensitivity are expected to be small, and to be reflected in the widths of the observable differences. Therefore, for simplicity, the procedure of using the observable differences is kept unchanged.

The toys fitted using the alternative configuration are unbiased. The difference distributions are well described by Gaussian functions, and show shifts in the mean from zero and widths that are comparable to the mean values for most observables. The systematic uncertainties are taken to be the sum in quadrature of the mean and widths of these distributions. While this uncertainty amounts to around 20% of the statistical uncertainties of most observables, it is found to be much larger for F_L, at around 40%. The second most strongly affected observable is A_{FB}/P_2, for which it can contribute up to 30% of its statistical uncertainty.

9.2.6 Peaking backgrounds

Studies performed as part of the $R_{K^{(*0)}}$ analysis [67] show that even after full selection, residual backgrounds from various hadronic decays with single or double pion-electron and/or kaon-electron misidentification (e.g. decays of the type $B^0 \to K^{*0}\pi^-(\pi^0,\gamma)X$, where X denotes any number of other final state particles) are present for the $B^0 \to K^{*0}e^+e^-$ mode. While the contribution of each individual source is expected to be small, taken together, the total is found to be non-negligible. Furthermore, while many such decays populate the lower mass region due to incomplete reconstruction, some are known to peak (or be distributed non-trivially) in the proximity of the signal. The peaking contributions can come from fully reconstructed but misidentified hadronic decays, e.g. $B^0 \to K^{*0}\pi^+\pi^-$, where both pions are misidentified as electrons. These distributions cannot be fully accommodated by the exponential combinatorial distribution. The large number of possible modes, some of which are poorly known experimentally, complicates the brute force approach of generating simulation to assess the associated systematic uncertainties (or to estimate the type and amount present). For this reason, their impact is checked using a conservative data-driven approach. As these backgrounds mainly involve electron misidentification, the stability of the (blinded) observables as the electron PID requirements are progressively tightened (and the backgrounds become increasingly more suppressed) can allow the size of their effect on the observables. Note that this 'PID scan' method provided some of the first indications of the existence of those backgrounds [67].

To carry out this scan, the electron PID requirements, $\text{PID}_{e\pi}$ and `ProbNNe`, are tightened from the nominal values of $\text{PID}_{e\pi} > 2$ and `ProbNNe` > 0.2. First, the $\text{PID}_{e\pi}$ requirement is left unchanged, but the `ProbNNe` cut is increased from 0.2 to 0.6 in steps of 0.1. Then, similar scans of the `ProbNNe` values are ran at two tighter $\text{PID}_{e\pi}$ thresholds of 3 and 5. For each variation, the full angular fit, including the parametrisation of acceptance functions and the modelling of all components, is ran to determine the observable values. The stages of the nominal measurement that are not repeated are the trigger and kinematic corrections, which are expected to hold to a reasonable approximation.

The differences of the observables values found at tighter cuts compared to the nominal are shown in Figure 9.12. The shifts in all observables are found to be smaller than, or comparable to, their expected statistical uncertainties. In the large q^2 region, the largest shift is seen for S_9 at around 1.1 times its statistical uncertainty, followed by S_5/P_5' at 90%/75% and S_4 at 67%. In the smaller q_c^2 region the largest shifts are found for S_9 at around 1.3 times its statistical uncertainty, followed by S_5/P_5' at 87%/70% and S_4 at 77%. While part of these differences are expected to be statistical in nature, they can provide a conservative estimate of the extent of the impact. Therefore, the largest shift found for each observable (in absolute value) is considered as the systematic uncertainty related to neglecting misidentified hadronic backgrounds.[7]

[7] Nevertheless, while this approach can provide indications of its size, a more reliable estimation may produce different (and potentially smaller) values. One such possibility is to use a data-driven approach [67]. Background rich samples can be obtained by inverting the electron PID requirements. These events can be weighted with per-event weights from 'transfer functions', which are formulated using information from PIDCalib samples, in order to extrapolate their expected shape and yield after the nominal PID selection. A more rigorous approach is planned for the publication of this analysis. Note that while the inclusion of the misidentified backgrounds as an additional component in the angular fit is a possibility, it appears to be challenging at the present level of statistics. As the size of the samples available for its parametrisation will increase with increased statistics, this strategy is likely applicable (and potentially necessary) for future analyses of this mode. Another option that is more applicable to the current analysis, and is under consideration, is to tighten the nominal PID requirements for greater suppression of this type of background.

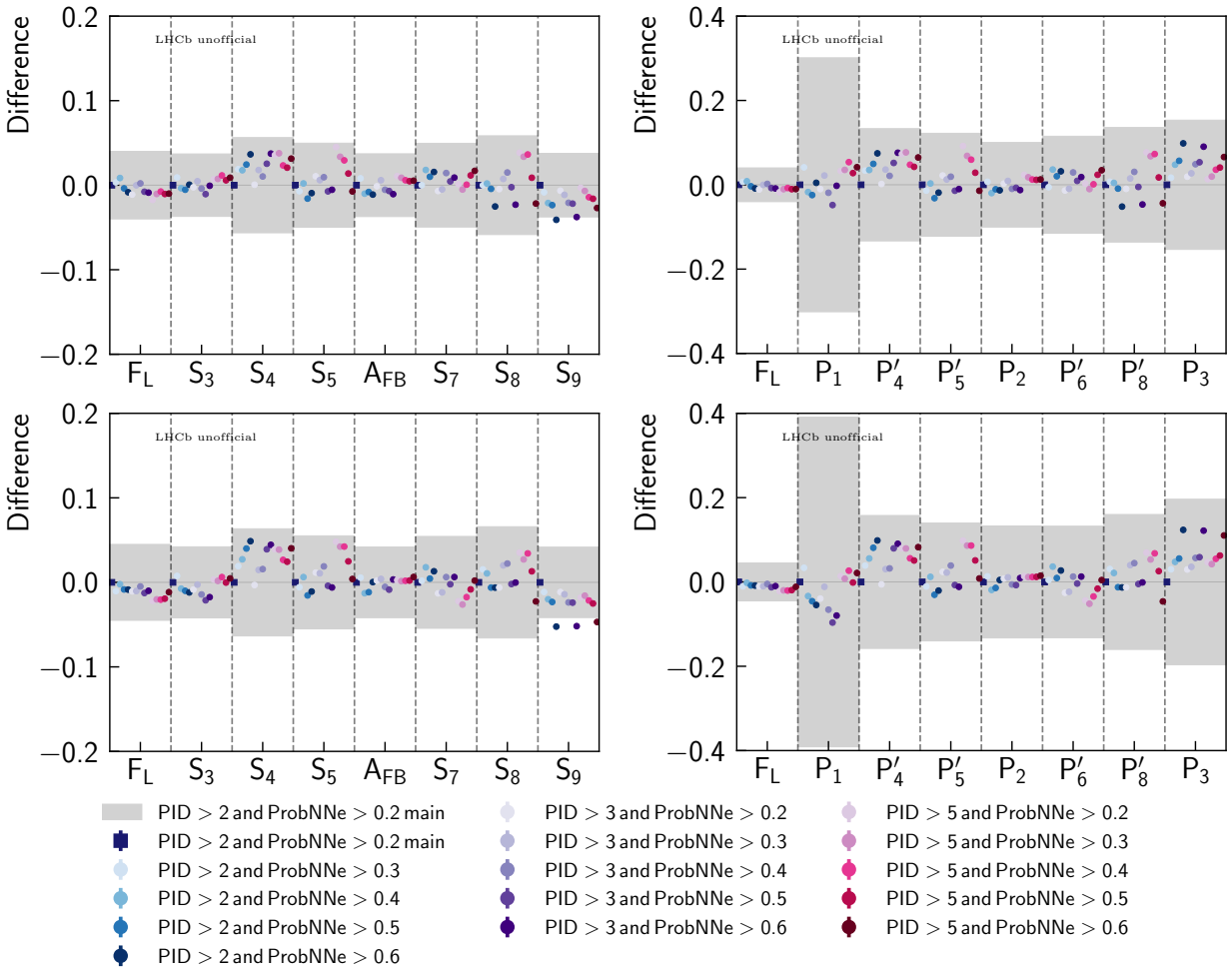


Figure 9.12: Differences in the observable values between the nominal fit and alternative fits made with different requirements on electron PID variables in the q^2 ranges of $1.1 < q^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (top) and $1.1 < q^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (bottom). In all cases, the gray band corresponds to the sensitivity estimates from toy studies.

9.3 Results and discussions

The angular analysis of $B^0 \to K^{*0}e^+e^-$ decays is currently under LHCb collaboration review, therefore its results need to remain blinded. To provide a picture of the present state of the analysis, the projections of the rare mode data fits in the q_c^2 ranges of $1.1 < q_c^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ and $1.1 < q_c^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ are shown in Figures 9.13 9.14 9.15 and 9.16 9.17 9.18, respectively, where the signal angular distributions are blinded (made invisible). The control mode fits used to determine the shift and scaling parameters of the mass distributions are shown in Figure 9.19. Table 9.5 provides a summary of the expected statistical uncertainties of all S and P-basis observables as well as their associated biases and total systematic uncertainties.

Notice that for most observables the systematic uncertainties are at the same level as the statistical ones. For all affected observables except F_L and A_{FB}/P_2, it is due to the conservative approach used to estimate the impact of remaining peaking backgrounds. Improving the estimation of such contributions can play an important role in the precision of the measurement. F_L and A_{FB}/P_2 are strongly affected by the limited ability of the nominal background models to accommodate for potential discrepancies related, for example, to the mismodelling of DSL-like contributions. In this case, relatively small changes to the fit strategy, such as the addition of an extra degree of freedom to the DSL model, are expected to reduce their impact. The current levels of total observable uncertainties are comparable to those of the first muon mode measurement of the optimised P-basis observables made using data collected in 2011 [50]. They are lower than those of the combined measurement of the $B^0 \to K^{*0}e^+e^-$ and $B^+ \to K^{*+}e^+e^-$ decays by the Belle collaboration [62], which reported uncertainties for P_4' and P_5' of ~ 0.4 in the central q^2 region of $1.0 < q^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$. Therefore, the upcoming measurement is expected to be the most precise of its kind to date.

Table 9.5: Summary of the biases, statistical and systematic uncertainties of the S and P-basis observables in the q_c^2 ranges of $1.1 < q_c^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ and $1.1 < q_c^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$.

	$1.1 < q_c^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$				$1.1 < q_c^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$			
	Bias	σ_{stat}	σ_{syst}	σ_{tot}	Bias	σ_{stat}	σ_{syst}	σ_{tot}
F_L	−0.002	0.040	0.046	0.061	−0.001	0.045	0.048	0.066
S_3	−0.001	0.037	0.017	0.041	−0.001	0.042	0.025	0.049
S_4	−0.001	0.056	0.045	0.072	−0.003	0.063	0.056	0.085
S_5	0.002	0.050	0.050	0.070	0.000	0.055	0.053	0.076
A_{FB}	0.000	0.037	0.067	0.077	−0.001	0.042	0.066	0.078
S_7	0.000	0.050	0.026	0.056	0.002	0.054	0.034	0.064
S_8	−0.002	0.058	0.047	0.075	−0.003	0.066	0.046	0.080
S_9	−0.001	0.038	0.043	0.057	0.000	0.042	0.054	0.068
P_1	−0.006	0.301	0.122	0.325	−0.005	0.391	0.174	0.428
P_4'	−0.001	0.133	0.101	0.167	−0.006	0.158	0.126	0.202
P_5'	0.005	0.122	0.109	0.164	0.001	0.140	0.120	0.184
P_2	0.000	0.100	0.183	0.208	−0.002	0.133	0.198	0.238
P_7'	0.001	0.115	0.057	0.129	0.004	0.132	0.073	0.151
P_8'	−0.005	0.136	0.099	0.168	−0.006	0.160	0.100	0.189
P_3	0.004	0.153	0.112	0.190	0.000	0.196	0.143	0.243

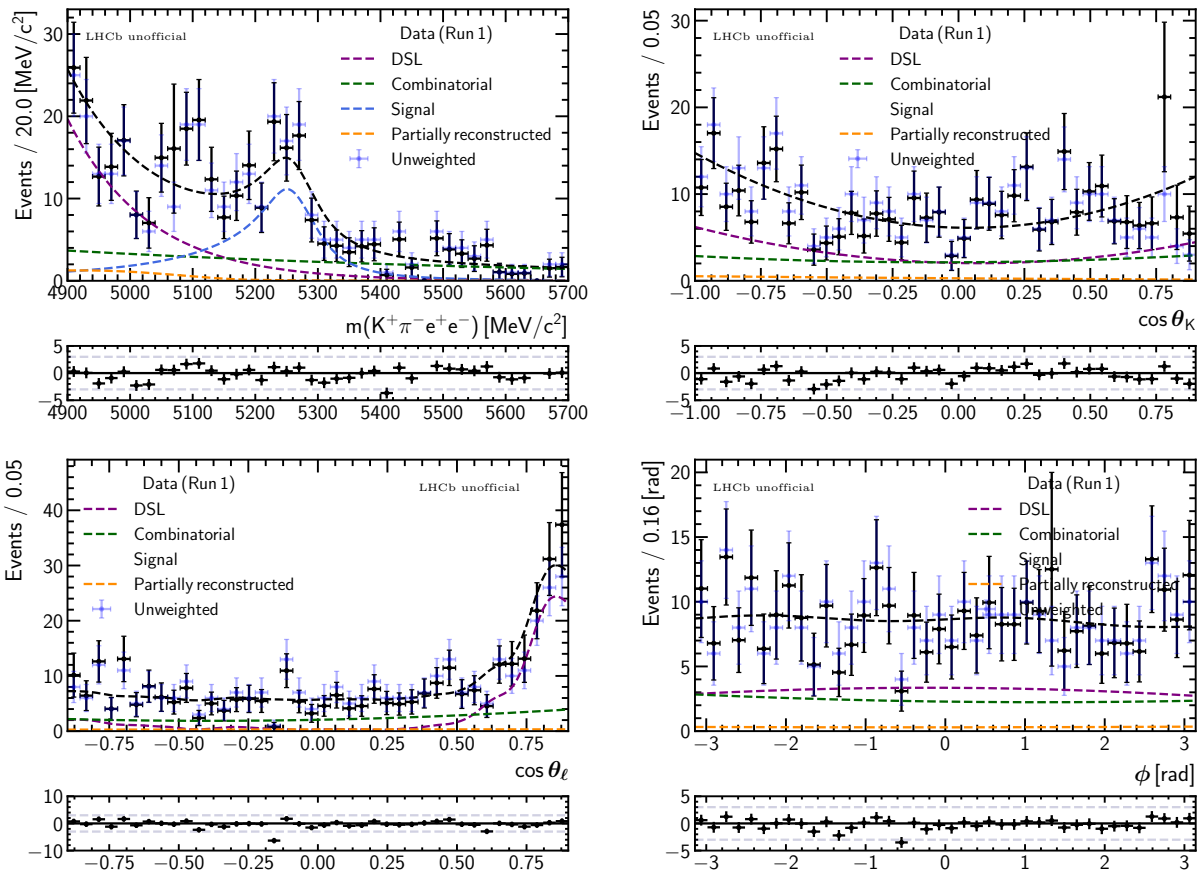


Figure 9.13: Result of the simultaneous rare mode angular fit in the q_c^2 range of $1.1 < q_c^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (Run 1).

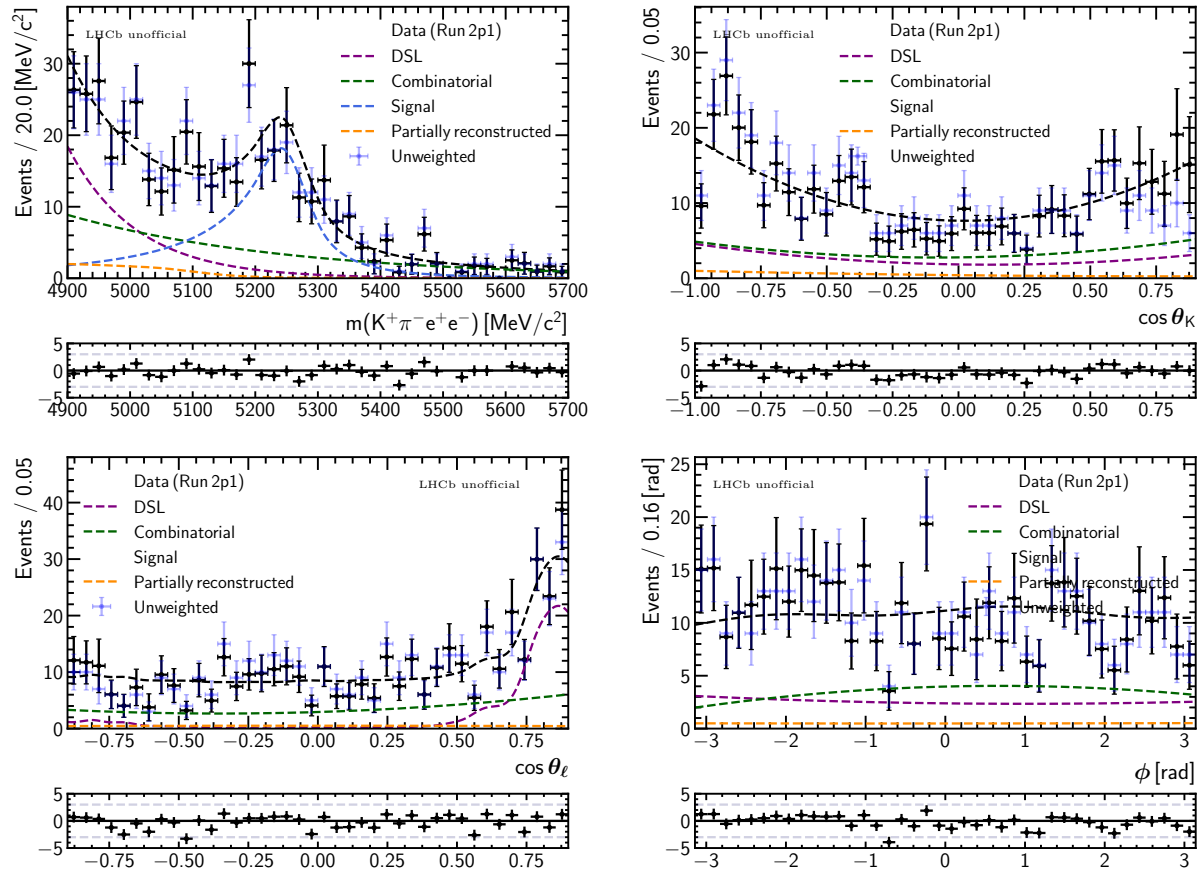


Figure 9.14: Result of the simultaneous rare mode angular fit in the q_c^2 range of $1.1 < q_c^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (Run 2p1).

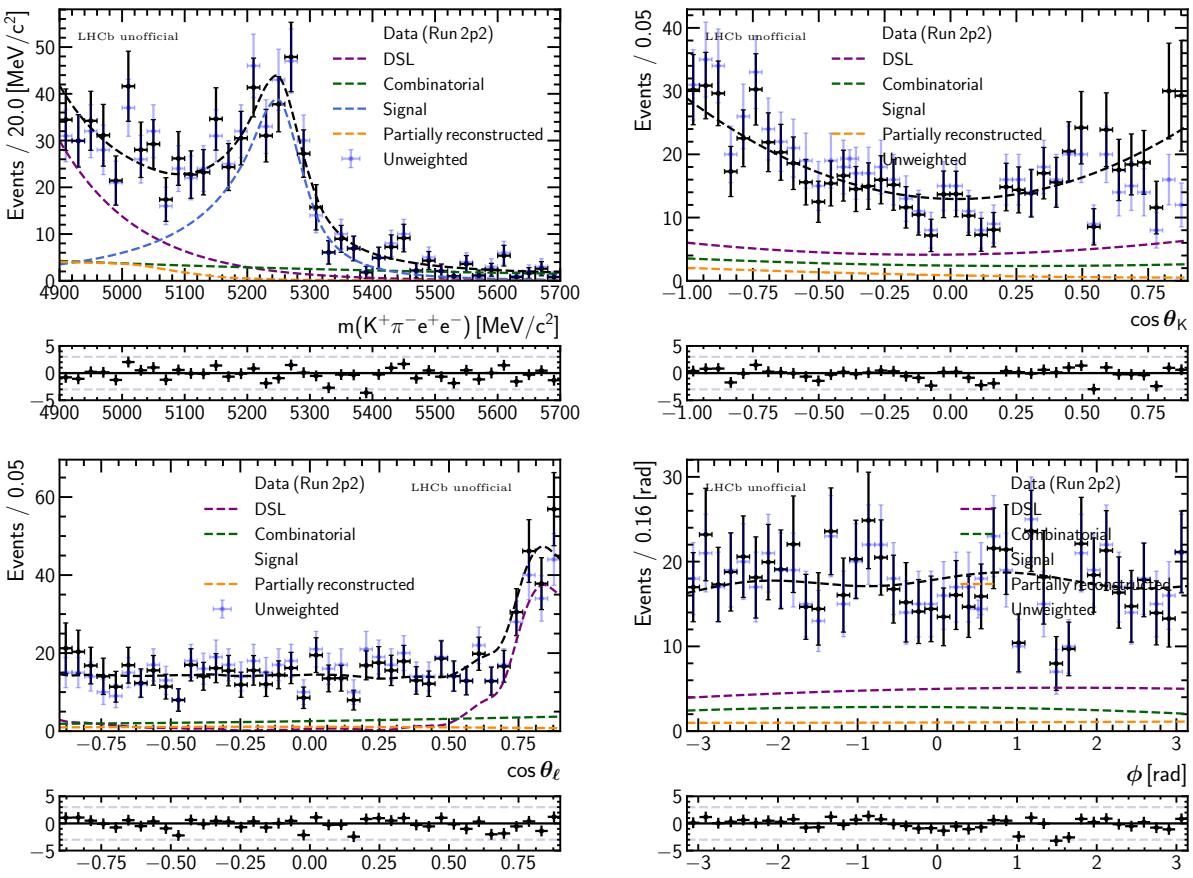


Figure 9.15: Result of the simultaneous rare mode angular fit in the q_c^2 range of $1.1 < q_c^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (Run 2p2).

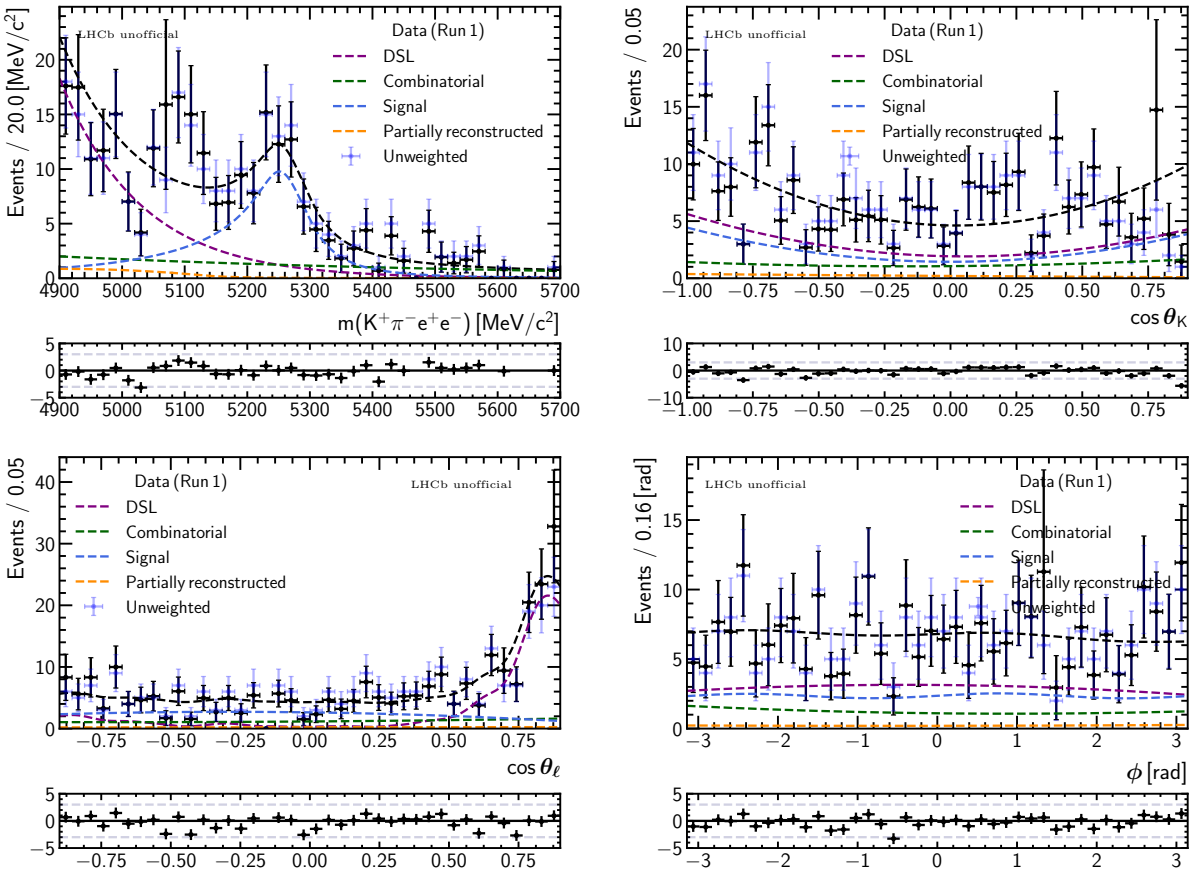


Figure 9.16: Result of the simultaneous rare mode angular fit in the q_c^2 range of $1.1 < q_c^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (Run 1).

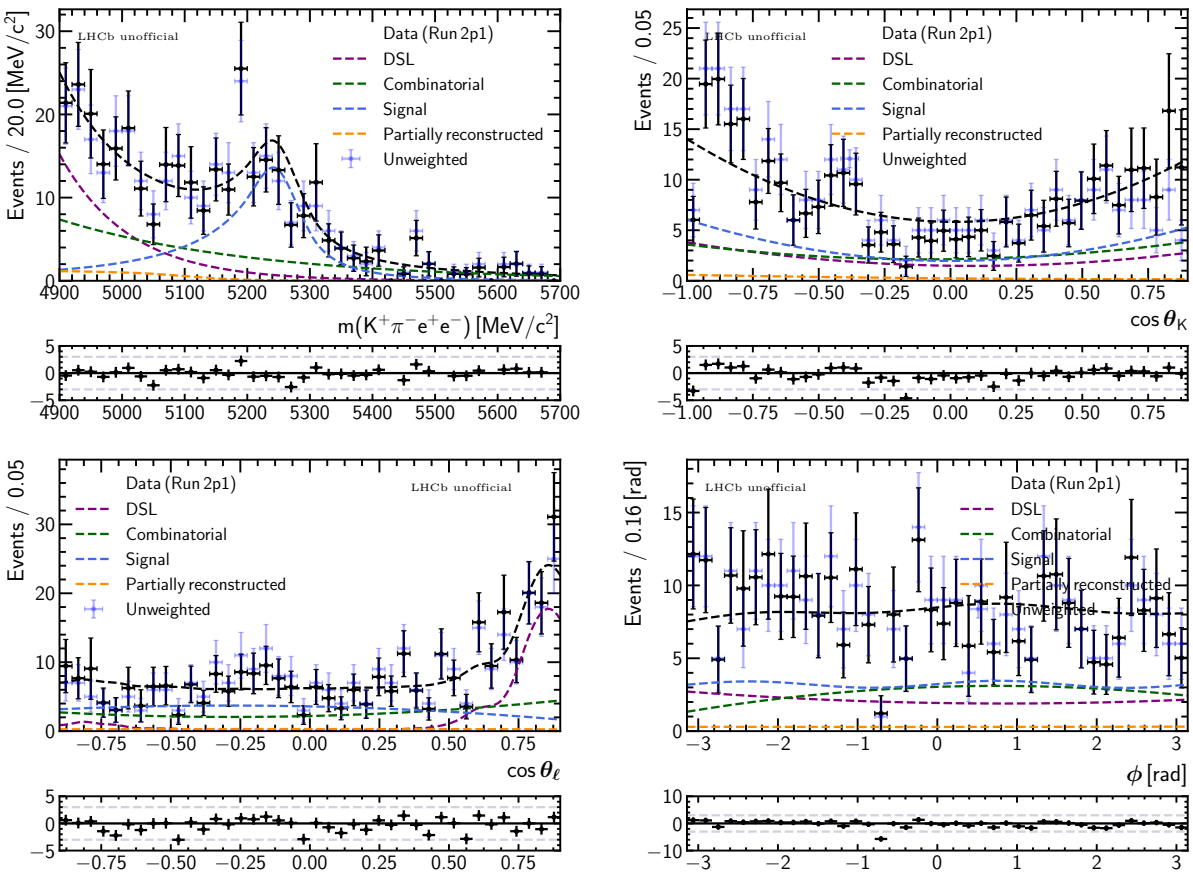


Figure 9.17: Result of the simultaneous rare mode angular fit in the q_c^2 range of $1.1 < q_c^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (Run 2p1).

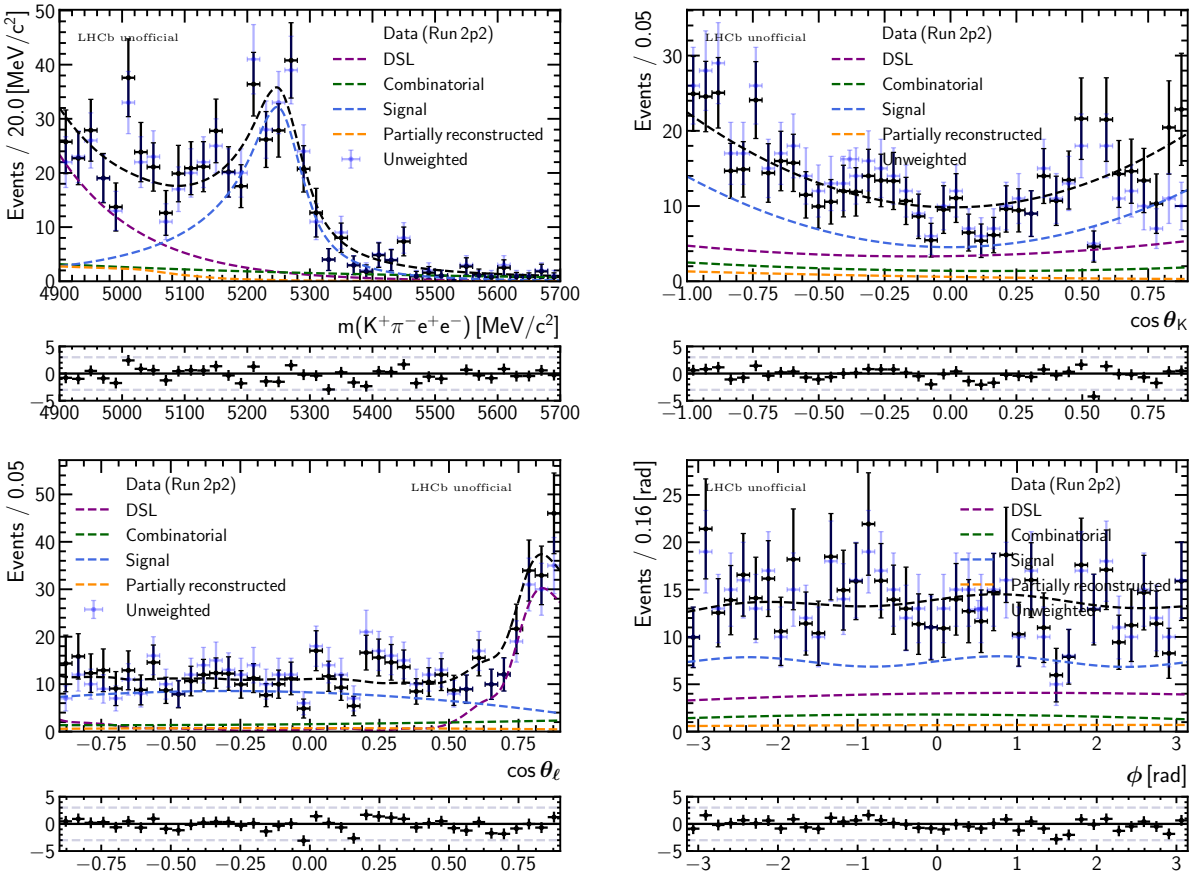


Figure 9.18: Result of the simultaneous rare mode angular fit in the q_c^2 range of $1.1 < q_c^2 < 6.0\,\mathrm{GeV}^2/\mathrm{c}^4$ (Run 2p2).

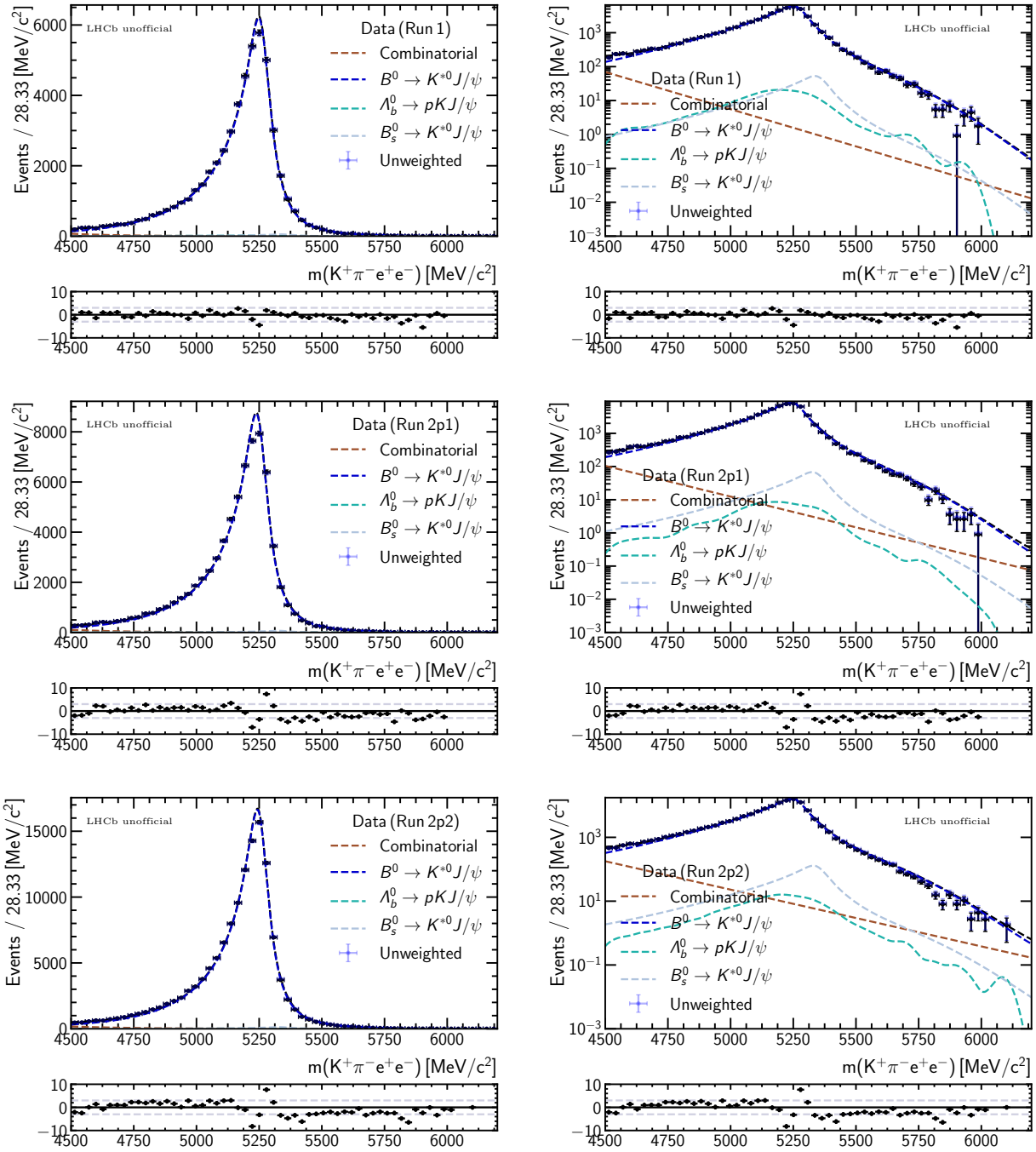


Figure 9.19: Results of the control mode mass fits used to determine shift and scaling parameters of the rare mode mass distribution.

10 Conclusion and outlook

The expectation that the SM, while incredibly successful, is an effective theory, motivates significant ongoing efforts to search for NP both directly and indirectly in the field of high energy physics. The rare decays of b hadrons, in particular decays that involve the $b \to s\ell^+\ell^-$ transition, provide good sensitivity to NP contributions, and therefore constitute a popular avenue for indirect searches. In this context, the emergence of a series of results that show tensions with SM predictions has engendered considerable interest. One of the first B anomalies is the angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ mode in 2013, which revealed a moderate local discrepancy with respect to the SM for the angular observable P_5' within the central q^2 region. While subsequent updates with increased statistics continue to confirm the initial discrepancy, its interpretation is complicated by theory uncertainties related to non-local SM hadronic effects. However, not long after the appearance of the first angular anomaly, measurements of the theoretically clean R_K ratio, followed by $R_{K^{*0}}$, started to provide intriguing hints for the violation of LFU.

For almost a decade, the B anomalies consisting of differential branching fraction measurements and angular observables of $b \to s\mu^+\mu^-$ modes, together with deviations from LFU in the R_X ratios (mainly $X = K^{\pm}$, K^{*0}), pointed towards a consistent pattern of NP featuring LFU violating contributions to muons, in particular to the vector coupling $C_{9\mu}^{\mathrm{NP}}$, which gained significance over the years with the addition of more measurements. However, this picture shifted in late 2022, when the simultaneously measurement of the R_K and $R_{K^{*0}}$ ratios made using improved experimental strategies for the electron modes revealed values that are highly compatible with the SM and therefore LFU. This suggests that the persistent anomalies in $b \to s\mu^+\mu^-$ decays may be driven by LFU NP contributions, or perhaps by unaccounted for SM QCD effects, rather than LFU violating NP. At present, this lack of clarity calls not only for improvements in the precision of commonly measured observables (through both increased statistics and improved analysis strategies) but also the exploration of hitherto underutilised quantities that can add to the ability to discriminate between SM and NP, such as the electron mode P_5' observable, which would allow for the determination of Q_5.

However, in contrast to the $B^0 \to K^{*0}\mu^+\mu^-$ decay, which has been explored in detail, few measurements of its corresponding electron mode exist. In particular, only one measurement of the angular observables of the $B^0 \to K^{*0}e^+e^-$ mode has been made in the central q^2 region (together with the $B^+ \to K^{*+}e^+e^-$ mode) [62]. The relative scarcity of $b \to se^+e^-$ measurements is closely linked to the effect of bremsstrahlung, which affects the less massive electrons to a much greater extent than muons. Energy lost in this way can cause more tracks to be bent out of the acceptance of the detector, which reduces reconstruction efficiency. For reconstructed tracks, the momenta resolution is worsened significantly, which reduces trigger efficiency, and complicates the separation of signal from backgrounds. Consequentially, much of the work described in this thesis is dedicated to understanding, mitigating and quantifying the impact of bremsstrahlung.

An important strategy that is key to enabling the analysis in its current form is the use of the constrained q_c^2 variable with improved resolution. This reduces the migration of signal events in q^2 and strongly limits contributions from the leakage of charmonium decays. The cost of

this approach, most prominently in the introduction of partly combinatorial events containing real J/ψ resonances to the upper mass region, is also found to be limited

The low statistics available strongly encourages the use of a simplified fit configuration, where the angular distributions of all background components are fixed, including that of the combinatorial. This greatly simplifies the analysis and has the advantage of improving fit stability, leading to a high convergence rate of almost 100% in toy studies. The cost of this 'template' based approach is that the impact of background mismodelling is increased. This is particularly relevant for the modelling of the DSL and DSL-like components, which is found to lead to large systematic uncertainties for the observables of A_{FB}/P_2 and F_L.

In view of the findings of the $R_{K^{*0}}$ analysis, systematic uncertainties due to residual misidentified hadronic backgrounds need to be assessed. A first attempt is made to obtain a conservative estimate of their impact on the angular analysis by performing fits to data after imposing progressively more stringent electron PID requirements. The largest difference found for each observable with respect to its value from the nominal fit is taken as the systematic uncertainty. As this procedure inevitably includes statistical effects, which may modify (and inflate) these values, this strategy will be revised to provide more realistic estimates.

The future of the analysis of the $B^0 \to K^{*0}e^+e^-$ decays is promising. An ambitious follow-up to the current work is already in progress, which aims to not only study the central q^2 region, but also the very low and high q^2 regions using full Run 1 and Run 2 data. Looking further ahead, Run 3 of the LHC, which is currently under way, aims to produce unprecedented statistics at the centre-of-mass energy of 13.6 TeV. The data gathered up to 2023 (full Run 3), which is anticipated to correspond to an integrated luminosity of $\sim 23\,\text{fb}^{-1}$ [192], will allow for measurements to be made with good precision in finer bins of q^2, provide sensitivity to additional observables, and allow for complex analysis strategies, such as a full amplitude analysis, to be used, leading to the extraction of more physics information. Larger statistics can also enable greater control over the residual misidentified hadronic backgrounds, by increasing the sample size available for their parametrisation through data-driven methods.

In addition, the Belle II experiment at the SuperKEKB electron-positron collider, which is dedicated to the precision study of the decays of B mesons, started data taking in 2019. It is expected to provide independent cross-checks of the LHCb results, in particular because in this case the B mesons are produced from the decays of the $\Upsilon(4S)$ resonance, which leads to reduced background complexity.

The current experimental status suggests that strong evidence to support or refute NP as an explanation of the remaining B anomalies can be expected in the near future. With the increase in statistics in the next few years and the refinement of existing analysis strategies, precision measurements of $b \to se^+e^-$ modes are expected to play an increasingly important role in this endeavour.

Appendix

A Simulation correction order

The ordering of the trigger and kinematic corrections (made using the nominal methods discussed in Section 6.5) affects the final result, in particular the $B^0\,p_T$ agreement. This variable is one of the most important features of the multivariate classifier, and it is slightly correlated with $\cos\theta_K$ due to the stripping level cut on the pion and kaon transverse momenta ($K\,p_T > 250\,\text{MeV/c}$ and $\pi\,p_T > 250\,\text{MeV/c}$). Large discrepancies in $B^0\,p_T$ may affect observables that are particularly sensitive to its curvature (e.g. F_L and S_{S1}).

In the nominal correction chain, the trigger corrections are made prior to the reweighting of the kinematic and multiplicity variables. Figure A.1 (left) shows the $B^0\,p_T$ distribution of simulated control mode candidates after full preselection compared to background subtracted data. Without corrections, the simulated distribution contains a slightly larger number of candidates with higher $B^0\,p_T$ values. The L0 correction weights exacerbate this difference, as they tend to increase the weighting of high $B^0\,p_T$ events (e.g. Figure 6.2). However, these differences are largely removed by the final kinematic (and multiplicity) weights, which are obtained on the basis of the trigger corrections.

If the correction order is reversed in favour of ensuring good $B^0\,p_T$ agreement prior to trigger correction, which can be justified by the dependency of the trigger efficiency on this variable, the final trigger correction weights would reduce the agreement, as shown in Figure A.1 (right). This occurs because the corrections are made to post-trigger rather than pre-trigger samples, where the discrepancy in the $B^0\,p_T$ distribution is due to a combination of generator level kinematic differences and differences in the trigger efficiencies. The BDT reweighter not only corrects for the former as intended, but also for the latter. While kinematic weights are used in the calculation of signal efficiency in simulation for the L0 trigger correction, the presence of weights in both the numerator and the denominator of the ratio allows for their effects to approximately cancel, leading to over-correction when both weights are combined.

A better way to perform trigger and kinematic corrections would be to first correct the pre-trigger distributions, and then apply the trigger corrections. However, this is in practise difficult to implement as no 'pre-trigger' data are available. A possible alternative is to use a 'prior-chain' of corrections, where kinematic and multiplicity weights are first obtained from trigger corrected muon mode samples (where the impact of the trigger is in any case small) and applied to the electron mode before producing the main chain of electron mode corrections [67]. Due to the complexity of this approach and its limited impact on the angular analysis (currently dominated by statistical and backgrounds-related systematic uncertainties), it is used for the evaluation of systematic uncertainties rather than as the nominal method.

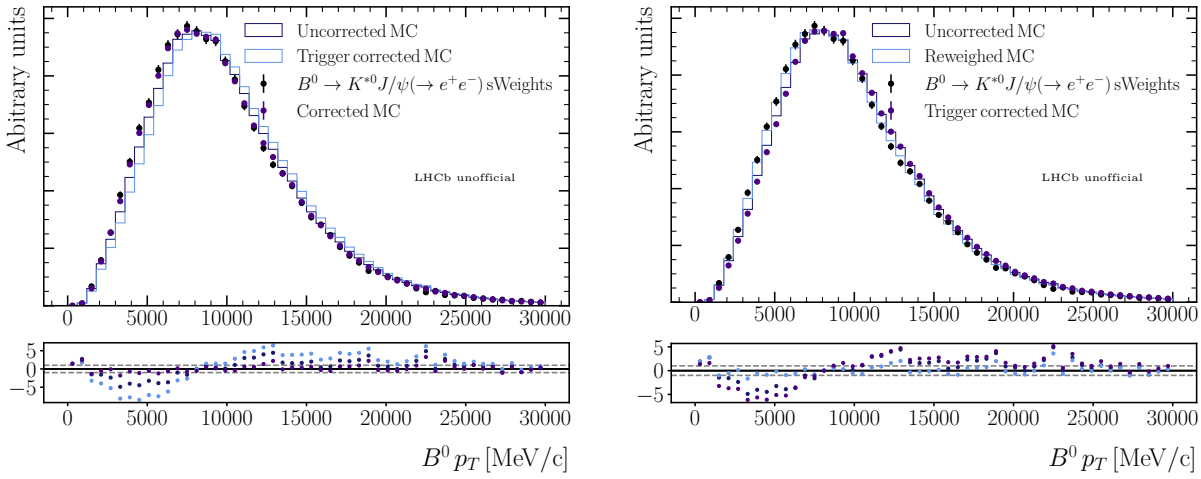


Figure A.1: $B^0\, p_T$ distribution of background subtracted control mode candidates (black) compared to simulation (violet) for which trigger corrections are made before kinematic corrections (left) and for which the reversed order is used (right). The results of the intermediate stages are shown in light blue for both cases.

B Main MVA threshold optimisation configuration

The main MVA threshold optimisation procedure discussed in Section 7.3.5, which involves using sensitivity to the angular observable P_5' as the FoM, is carried out at an early stage of the analysis. The set of preselection requirements differ from the ones discussed in Section 7.1, and a simplified background modelling strategy is used. Specifically, notable differences include the use of the HOP mass to reduce partially reconstructed background, the application of a veto against DSL decays ($|\cos\theta_\ell| < 0.8$), and the separation of the samples based on Run and L0 trigger categories. Furthermore, only Run 1 and Run 2p1 samples are used. Descriptions of the mass fit strategy and the pseudoexperiment configuration are provided below.

B.1 Mass fit set-up

The set-up of the mass fit is an earlier (and simplified) version of the nominal fit discussed in Section 8.3. The signal distribution is modelled via $B^0 \to K^{*0}e^+e^-$ simulation with kinematic and multiplicity correction weights only using the sum of two Crystal Ball functions (rather than the combined DCB model). The control mode mass fit used to provide values for the shift of the shared mean of the two CBs, and the scaling factor for its widths, is made without the $B_s^0 \to K^{*0}J/\psi$ and misidentified Λ_b^0 components. For the rare mode, only two types of background – partially reconstructed and combinatorial – are considered. The DSL background, which is suppressed via the veto of $|\cos\theta_\ell| < 0.8$ is not included as a separate component, in contrast to the nominal strategy. This choice is made as the separation of the DSL from combinatorial is generally not possible without angular information. The partially reconstructed background is also suppressed via the combined corrected B^0 mass (using the HOP approach) and B^0 flight distance significance ($B^0\,\chi^2_{\rm FD}\,({\rm PV})$) cut of $m_{\rm corr} > 4926 + 10\ln(\chi^2_{\rm FD}\,({\rm PV}))$. An additional difference is that separate fits are made for the L0 trigger categories of each Run period, resulting in two fits for Run 1 samples, and anther two for Run 2p1 samples. Later on the background modelling is no longer made separately for the L0 categories to increase fit stability. Examples of mass fits made with three very different classifier output cuts of 0.50, 0.99 and 0.9968 are shown in Figure B.1 for the L0En and L0TIS categories of Run 2p1.

B.2 Pseudoexperiment configuration

For each classifier cut, pseudoexperiments are generated based on the estimated signal and background yields. The mass models are identical to the ones used in the data mass fits. The angular model of the signal is given by Equation 2.14, after the cancellation of several terms via variable transformations that exploit the symmetries of the pdf.[1] Standard Model observable

[1] The 'folding' technique [50] can be used to reduce the number of observables in the signal pdf in order to improve the behaviour of the fit (increase stability and reduce bias) without the loss of statistical power. Different variable transformations can be used to cancel out different terms. The transformations applied

values from Flavio [193] for the q^2 range of $1.1 < q^2 < 7.0\,\mathrm{GeV}^2/\mathrm{c}^4$ are used. The model of the partially reconstructed background is obtained by fitting the distributions of the reweighted $B^+ \to K^+\pi^+\pi^- e^+ e^-$ phase-space simulation reconstructed as the signal (same as the current strategy) with factorised Chebyshev polynomials. Similarly, for the combinatorial background, Chebyshev polynomials are used to model the distributions of an earlier $K^+\pi^- e^+\mu^-$ data sample. Note that in contrast to the (current) nominal strategy, the acceptance functions are not included in the fits to obtain background models or in the pseudoexperiment generation. Higher order Chebyshev polynomials (higher than order two) often need to be used, and fits are made without weights using the extended maximum likelihood method in all cases. Separate background models are used for each L0 category of the Run 1 and Run 2p1 subsamples. The background angular shapes are obtained with a sanity cut on the classifier output of 0.2, and are not changed for the tighter thresholds, as the response of the classifier is relatively flat for the angles. Due to its shaping effect on the mass distribution, different signal mass models (and shift and scaling parameter values) are obtained for each classifier threshold. The slope of the combinatorial background and the shape of the partially reconstructed component are also modified based on the result of the data mass fit, and the shape of the simulation, respectively.

Around two to three hundred pseudoexperiments ('toys') are generated for each cut value, to ensure that about a hundred converged fit results are obtained. The fits are made simultaneously to the four subsamples sharing the signal observables. The yields of all components and the four slope parameters of the combinatorial background are allowed to vary, while the angular shape (and signal shift and scaling) parameters are fixed to generation values. The Gaussian width of the resulting spread of P_5' values from the fit is taken to be its expected sensitivity at the cut value. An example of a toy is shown in Figure B.2.

to the pdf (and data) in this case are: $\phi \to -\phi$ if $\phi < 0$, and $\theta_\ell \to \pi - \theta_\ell$ if $\theta_\ell > \pi/2$, which leaves only terms with F_L, P_1 and P_5'. This technique is no longer used after the unfolded pdf is found to be sufficiently stable for the nominal configuration, which greatly simplifies the analysis.

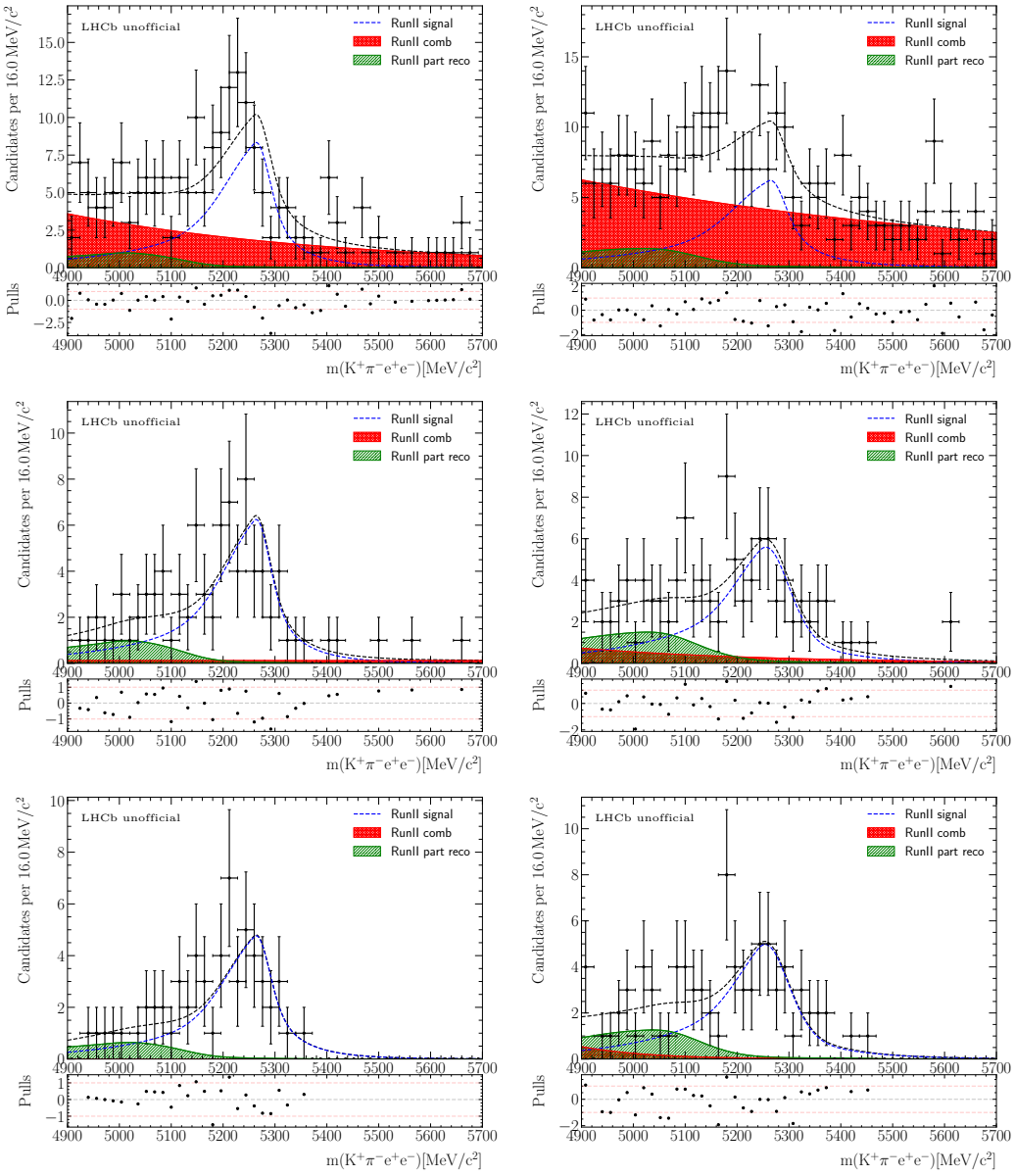


Figure B.1: Results of the fits to the Run 2p1 rare mode candidates belonging to the L0En (left column) and L0TIS (right column) categories. The classifier thresholds used are, in order from top to bottom, 0.50, 0.99 and 0.9968.

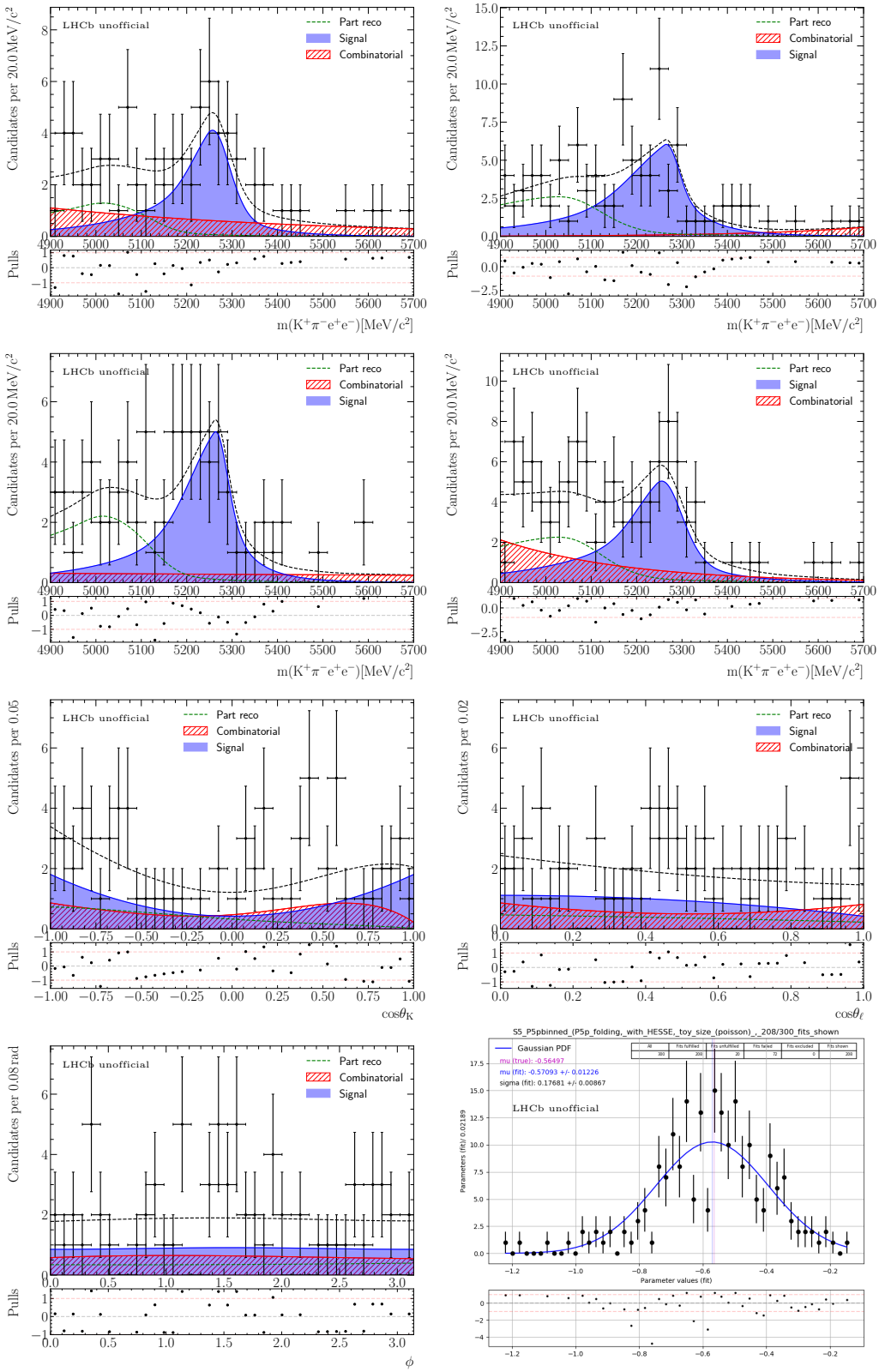


Figure B.2: Example of a toy generated with the yields calculated for the classifier cut of 0.99. The distribution of P_5' values from around two hundred converged fits are shown in the bottom right plot. In contrast to the data mass fits, the partially reconstructed component is not constrained, which leads to large fluctuations in the combinatorial and partially reconstructed background yields

C Additional details on systematic uncertainties

The following sections contain additional information that complements the discussions on the assessment of systematic uncertainties (Section 9.2).

C.1 Alternative combinatorial and DSL modelling strategy

The alternative model used for assessment of systematic uncertainties (Section 9.2.1) is made up of three parts, which have the following types of $\cos\theta_\ell$ distributions and are listed based on the slopes of their mass distributions (from most negative, i.e. the fastest decreasing exponential, to the least negative):

- **Asymmetric** – mainly DSL events;
- **Symmetric** – mainly partly combinatorial and misidentified DSL events;
- **Approximately uniform** – combinatorial events,

where the symmetric and approximately uniform components are obtained from same-sign data, and the asymmetric component is extracted from the $K\pi e\mu$ sample by subtracting away the symmetric contribution (in that sample). The procedures used to obtain the three models are discussed below. Note that while this alternative data-based strategy contains slightly different assumptions compared to the nominal two-step approach, it is similarly an effective strategy, where approximate models are used for a mixture of background contributions that are not always fully distinct. The key difference here is that an additional degree of freedom is introduced by the symmetric component, which adds the flexibility to account for possible differences in particular in the $\cos\theta_\ell = -1$ region.

C.1.1 Combinatorial and DSL-like backgrounds

The same-sign sample is used to extract models for the symmetric DSL-like events as well as the approximately uniform combinatorial background that does not have distinctive features. This is done by applying the nominal two-step procedure to the same-sign sample, the results of which are shown in Figure C.1 (Step 1) and C.2 (Step 2).

Minor differences between the distributions of the combinatorial background obtained from the same-sign compared to the $K\pi e\mu$ sample can be seen (Figure C.3). Comparing the $\cos\theta_\ell$ distribution in the $\cos\theta_\ell < 0$ region, the approximately symmetric DSL-like component appears to be more sharply peaking in the same-sign sample than the $K\pi e\mu$ (Figure C.4). However, it should be noted that even in the absence of experimental electron-muon differences, the same-sign sample cannot be expected to include all features of this type of background in the $K\pi e\mu$ or opposite-sign data due to the absence of candidates for which the kaon and/or pion is combinatorial in nature but the electrons come from a real DSL decay. In addition, even

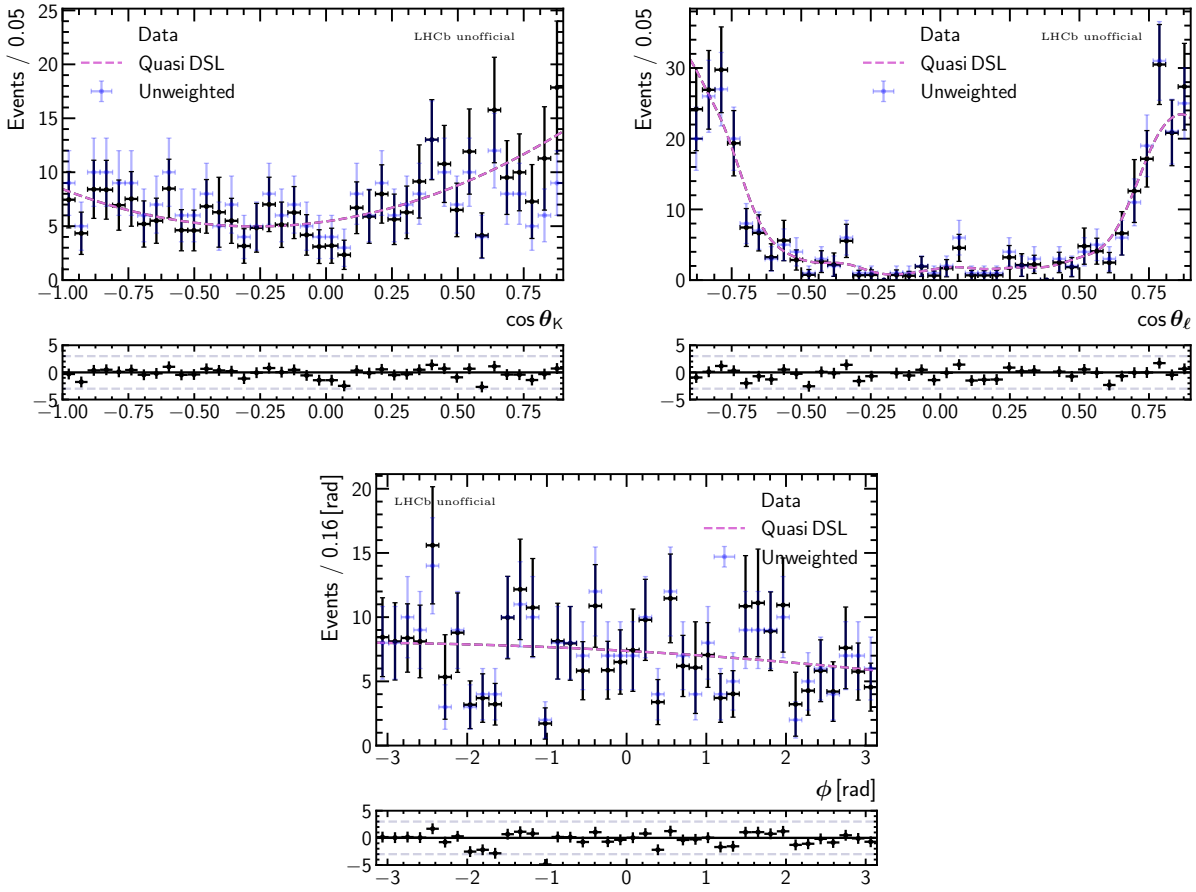


Figure C.1: Results of the Step 1 fit to the same-sign data sample to extract the angular shape of the DSL-like component. A tight MVA cut is applied to strongly suppress standard combinatorial background. All available samples are combined to increase the statistics available. Separate models are parametrised using the different acceptance functions of each Run subsample. The result for Run 2p2 is shown.

for cases where the hadronic part (and the identity of the $B^0/\bar{B}^0$) is properly identified, it is more symmetric than the analogous background in the opposite-sign data. For example, in the opposite-sign data (and $K\pi e\mu$), in the absence of electron charge misidentification, the $\cos\theta_\ell$ angle would never be calculated between the $B^0/\bar{B}^0$ and an electron from the D^-/D^+. However, this takes place in the same-sign sample if the $\cos\theta_\ell$ angles are calculated for $B^0 \to K^+\pi^-e^-e^-/\bar{B}^0 \to K^-\pi^+e^+e^+$ candidates, which would lead to an increase in events near $\cos\theta_\ell = -1$. These characteristics discourage the use of same-sign derived models in the nominal fit. However, including it as a component in the three-part model that is allowed to vary provides it with the freedom to account for possible sharply peaking structures on either sides of the $\cos\theta_\ell$ distribution, leaving the asymmetric component to take residual asymmetry into account.

C.1.2 Modified DSL model

To accommodate the introduction of the symmetric quasi-DSL component, the model for the main DSL background is modified by subtracting away the assumed symmetric component. The procedure used to separate the two from the $K\pi e\mu$ sample involves first obtaining KDE

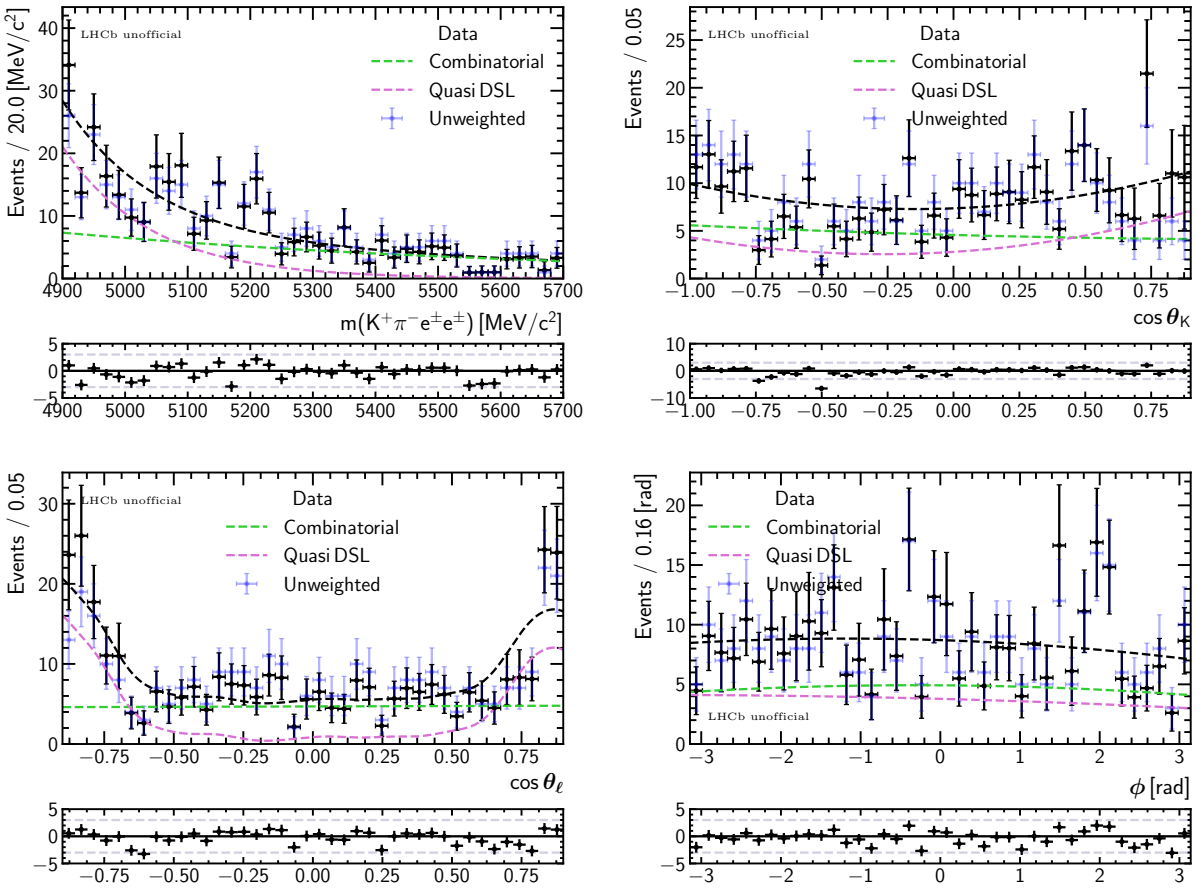


Figure C.2: Result of the Step 2 fit to the Run 2p2 same-sign data sample at the nominal MVA cut to determine the angular shape of the standard combinatorial component, and the slopes of both combinatorial and DSL-like backgrounds. The angular parameters of the DSL-like component are fixed from Step 1.

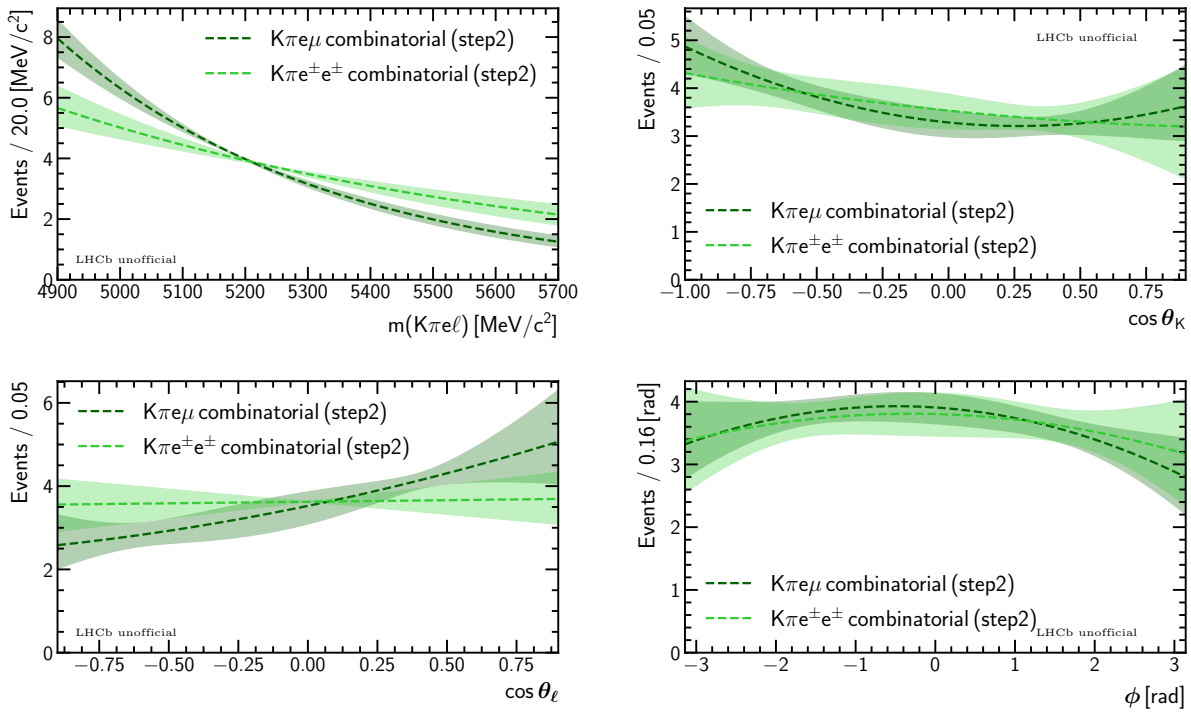


Figure C.3: Comparison between the standard combinatorial background obtained from the same-sign (light green) and $K\pi e\mu$ (dark green) data samples in Step 2 of the two-step procedure. The lighter coloured bands represent 1σ uncertainties of the fits.

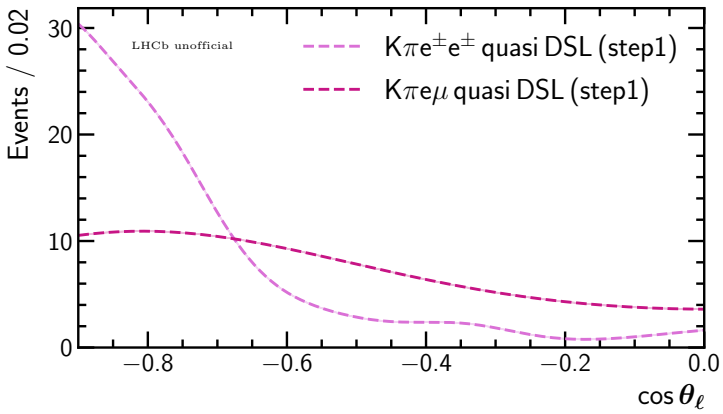


Figure C.4: Comparison between the KDE lineshapes used to describe the $\cos\theta_\ell$ distributions of the DSL-like component in the same-sign and $K\pi e\mu$ data in the $\cos\theta_\ell < 0$ region.

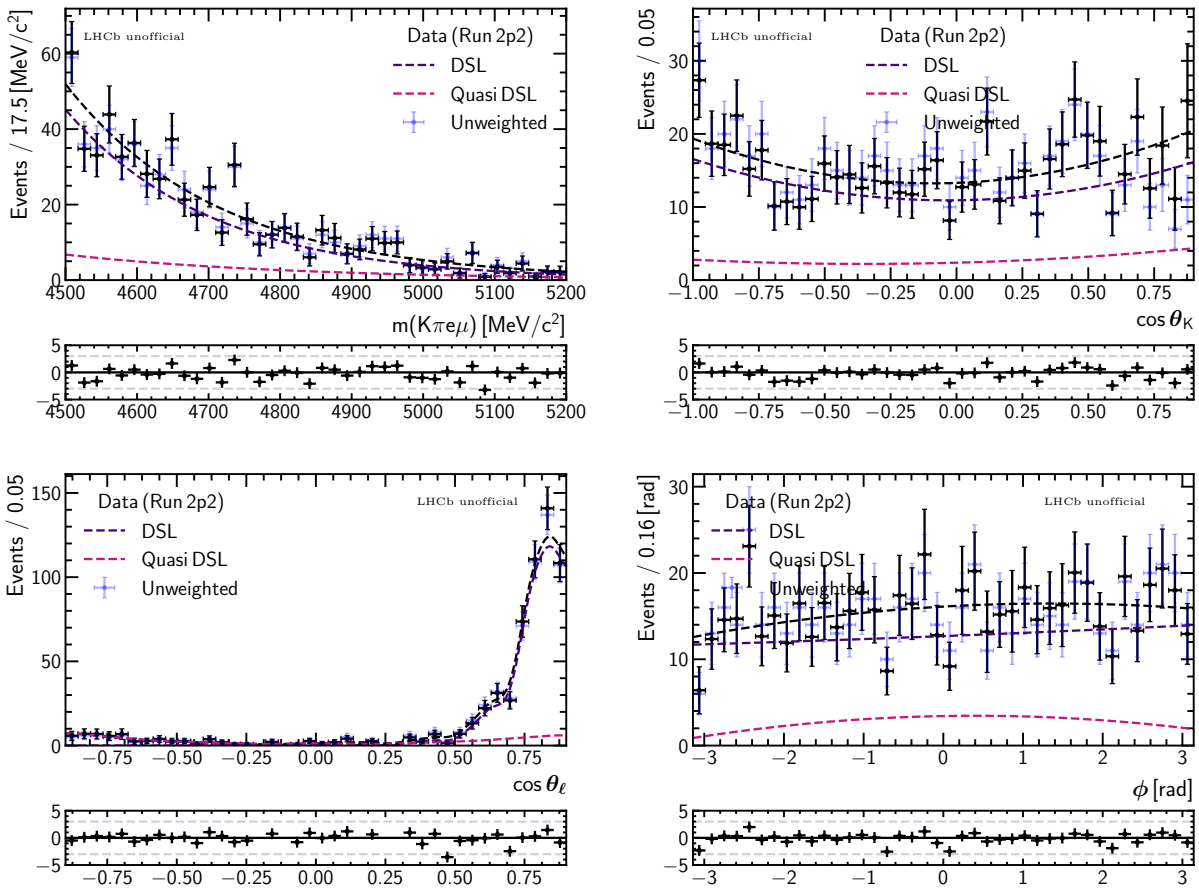


Figure C.5: The Run 2p2 $K\pi e\mu$ sample after the tight MVA cut, fitted using separate models that describe the symmetric and the asymmetric components.

descriptions of the $\cos\theta_\ell$ distribution of the latter by mirroring the data in the region of $\cos\theta_\ell < 0$ to > 0, and of the main DSL peak by subtracting away the mirrored distribution. Then, a fit is made with a tight MVA cut (MVA > 0.9985, Step 1 configuration) using two separate models, with angular and mass slope parameters allowed to very. The result of this fit to the Run 2p2 $K\pi e\mu$ sample is shown in Figure C.5.

C.1.3 Alternative data fit

In order to obtain reasonable yields for all three (effective) components, an alternative data fit is made with the three-part model. In this fit, the fraction of the symmetric quasi-DSL with respect to the main DSL component is allowed to vary. To improve stability, a single fraction is shared between all Run subsamples, and its value is Gaussian constrained to the ratio found in a separate $K\pi e\mu$ fit in the signal mass window with the nominal MVA cut (Step 2 sample), where the same strategy of sharing the quasi-DSL fraction is used. Note that this constraint is not necessary for convergence and it is only added to discourage the fit from incorporating other backgrounds, in particular control mode leakage events. The value of this fraction is found to be around 0.5. Interestingly, the ratio of quasi-DSL to combinatorial from the fit, which ranges from ~ 0.3 for Run 1 to ~ 0.5 for Run 2p2, with a weighted average of 0.4, is very comparable to the ratio found in the combined same-sign sample of 0.44 ± 0.05, and larger than the value of around 0.2 found from the $K\pi e\mu$ fit. This is indicative of potential ratio difference.

C.2 Control mode leakage modelling

To assess its impact of neglecting control mode candidates that remain in the signal q_c^2 region (Section 9.2.3), its distribution is first modelled from simulation, combining together all Run subsamples. Considering the low yield of this component in data, and the limited statistics of the simulation, a factorised model is expected to be sufficient. Chebyshev polynomials up to second order are used for the angles of $\cos\theta_K$ and ϕ, and KDE is used for $\cos\theta_\ell$, B^0 mass and q_c^2 (only used for the generation of toys with the acceptance effect). The result of the fit with Run 2p2 acceptance weights is shown in Figure C.6.

To obtain realistic estimations of background yields, an alternative data fit is made with the inclusion of this component. Due to difficulties in separating it from other sources of backgrounds, its fractions with respect to the signal are Gaussian constrained to their expected values for the three Run subsamples. The result of this fit favours significantly higher values of around 6 events for Run 1, 6 for Run 2p1 and 17 for Run 2p2. Given that the DSL-like background has similar mass and $\cos\theta_\ell$ distributions, it is possible that some of those events have been included. Nevertheless, this potential over-estimation is expected to lead to slightly more conservative uncertainty estimations.

C.3 Charmonium combinatorial modelling

In the absence of more reliable data-driven methods, simulated control mode candidates that contain correctly reconstructed J/ψ decays, but are otherwise combinatorial in nature, can provide a proxy for this type of background that is sufficiently realistic for the assessment of systematic uncertainties (Section 9.2.4). This is supported by the good agreement in the distribution of events in bins of $m(K\pi ee)_{\rm PV}$ and q_c^2 between the simulation and data (Figures 7.6 and 7.6). Due to limited statistics after the nominal selections, models of this background are obtained with the MVA requirement loosened from 0.96 to 0.9 in a slightly enlarged mass range of up to $5800\,{\rm MeV/c^2}$ (instead of $5700\,{\rm MeV/c^2}$), and with all samples combined. The $\cos\theta_K$ and B^0 mass distributions of these events are described using KDE lineshapes, Chebyshev polynomials up to second order are used for $\cos\theta_\ell$ and ϕ, and an exponential function is used for q_c^2. The result of the fit made using Run 2p2 acceptance weights is shown in Figure C.7.

To obtain reasonable estimations of its yield in data, an alternative data fit is made with its inclusion in the large q_c^2 range. Given the difficulties in separating it from the standard combinatorial in the angular fit (made without q_c^2, which is the most important discriminator), its fractions are Gaussian constrained based on the expected value obtained using q_c^2 information for the full dataset (Section 7.2.3). More specifically the value found is split up for the three Run subsamples using the ratio of signal yields from the control mode fit, leading to the central values of 11, 15 and 28 for Run 1, Run 2p1 and Run 2p2, respectively. Moderate widths corresponding approximately to ± 9 (Run 1), 11 (Run 2p1) and 18 (Run 2p2) events are used. Around 15, 15, and 31 events are found from the fit for the three subsamples.

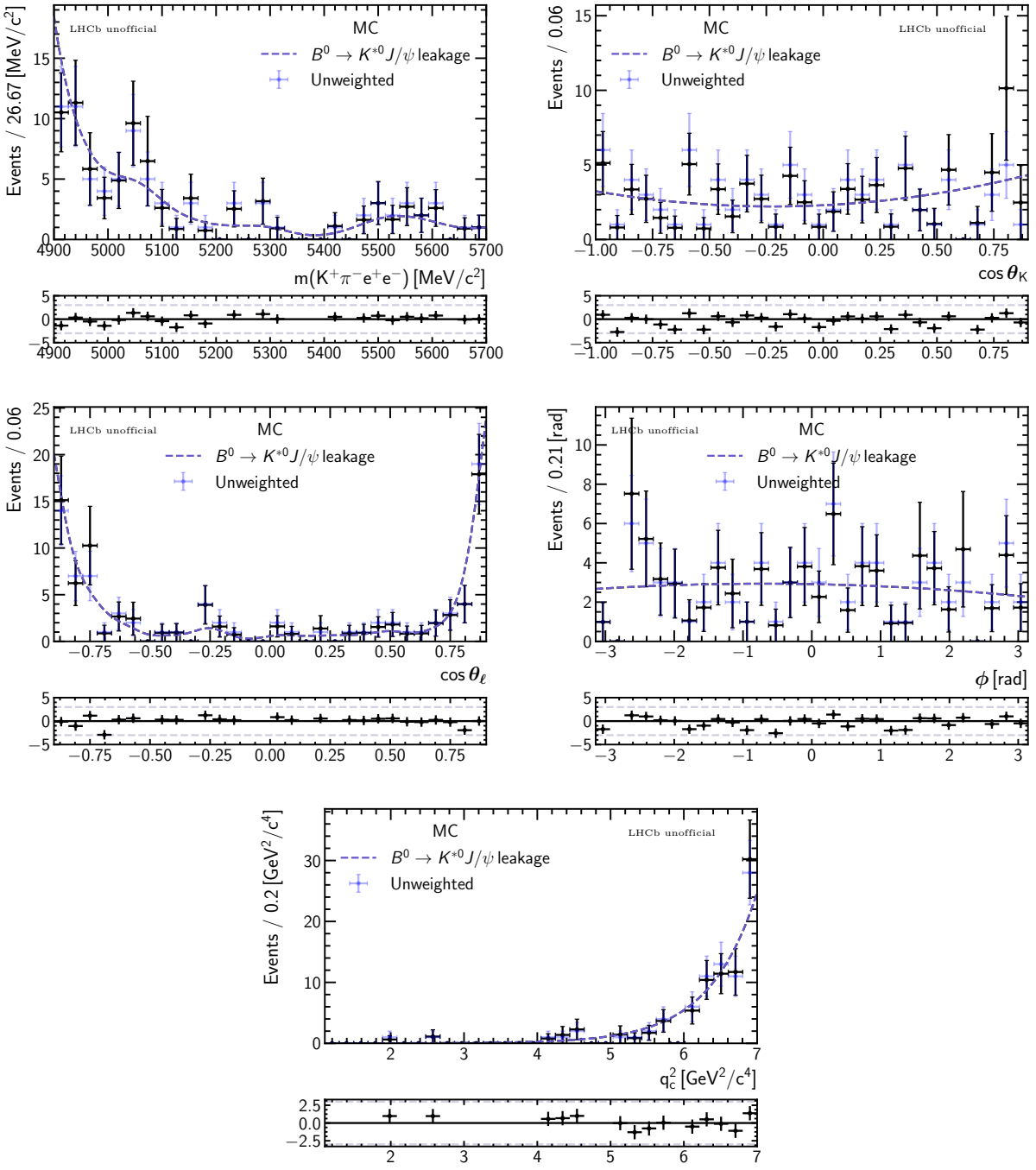


Figure C.6: Model used to describe control mode candidates that remain in the signal q^2_c window.

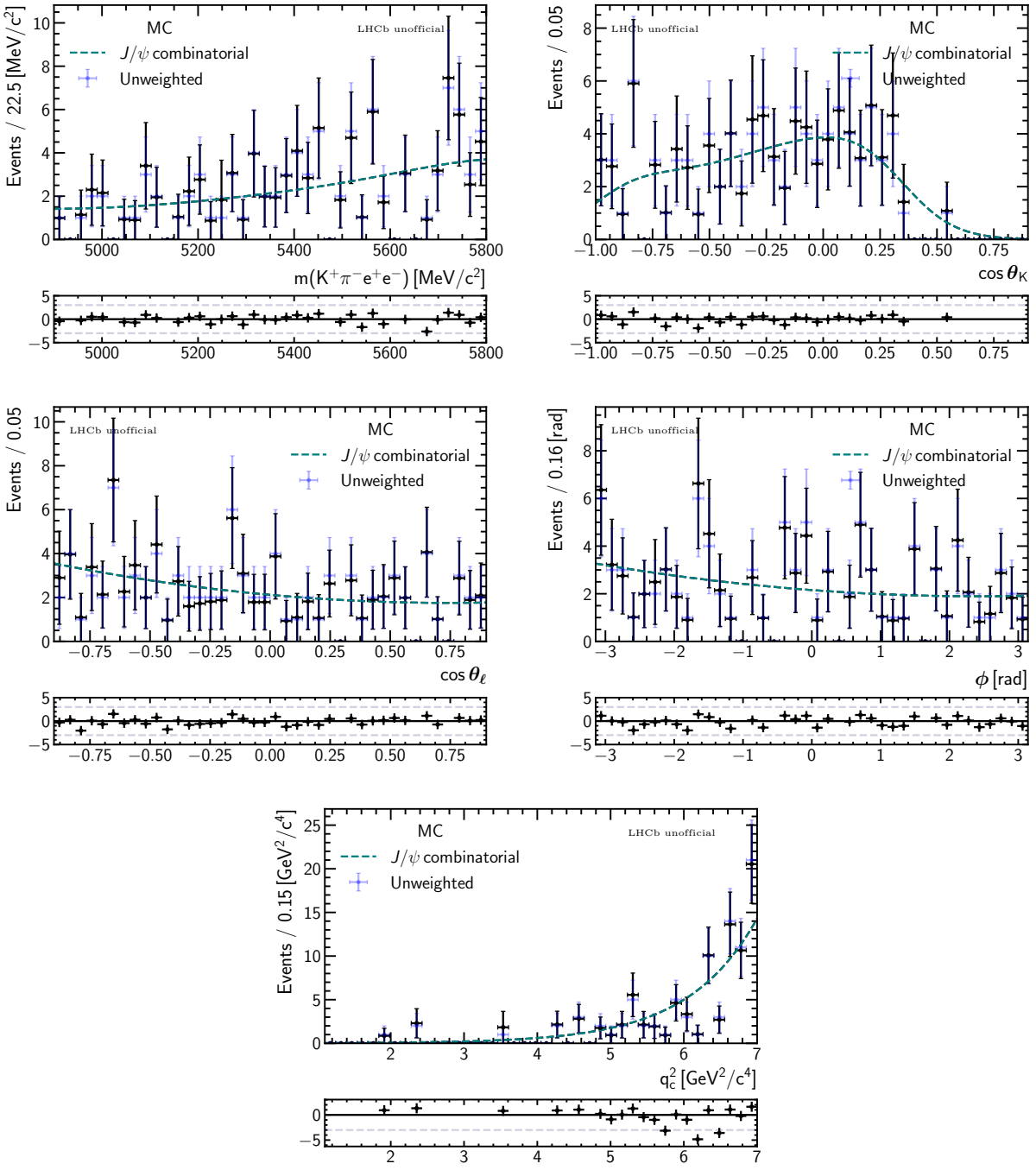


Figure C.7: Distribution of background-like events in the control mode simulation that enters the rare mode q^2 window due to the cut on the constrained q^2_c.

References

[1] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[2] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[3] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

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

[5] Antonio Pich. "The Standard Model of Electroweak Interactions". In: *2010 European School of High Energy Physics*. Jan. 2012, pp. 1–50. arXiv: 1201.0537 [hep-ph].

[6] Heather E. Logan. "TASI 2013 lectures on Higgs physics within and beyond the Standard Model". In: (June 2014). arXiv: 1406.1786 [hep-ph].

[7] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[8] F. L. Wilson. "Fermi's theory of beta decay". In: *American Journal of Physics* 36 (1968), p. 1150. ISSN: 12. DOI: https://doi.org/10.1119/1.1974382.

[9] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[10] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[11] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[12] A. Salam and J.C. Ward. "Electromagnetic and weak interactions". In: *Physics Letters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)90711-5. URL: https://www.sciencedirect.com/science/article/pii/0031916364907115.

[13] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[14] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. URL: https://doi.org/10.1143/PTP.49.652.

[15] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[16] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[17] D. Hanneke, S. Fogwell, and G. Gabrielse. "New Measurement of the Electron Magnetic Moment and the Fine Structure Constant". In: *Phys. Rev. Lett.* 100 (2008), p. 120801. DOI: 10.1103/PhysRevLett.100.120801. arXiv: 0801.1134 [physics.atom-ph].

[18] Tatsumi Aoyama et al. "Tenth-Order QED Contribution to the Electron g-2 and an Improved Value of the Fine Structure Constant". In: *Phys. Rev. Lett.* 109 (2012), p. 111807. DOI: 10.1103/PhysRevLett.109.111807. arXiv: 1205.5368 [hep-ph].

[19] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[20] N. Aghanim et al. "Planck 2018 results. VI. Cosmological parameters". In: *Astron. Astrophys.* 641 (2020). [Erratum: Astron.Astrophys. 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

[21] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". In: *Phys. Rev. Lett.* 81 (1998), pp. 1562–1567. DOI: 10.1103/PhysRevLett.81.1562. arXiv: hep-ex/9807003.

[22] R. D. Peccei. "The Strong CP problem and axions". In: *Lect. Notes Phys.* 741 (2008). Ed. by Markus Kuster, Georg Raffelt, and Berta Beltran, pp. 3–17. DOI: 10.1007/978-3-540-73518-2_1. arXiv: hep-ph/0607268.

[23] Eldad Gildener. "Gauge-symmetry hierarchies". In: *Phys. Rev. D* 14 (6 Sept. 1976), pp. 1667–1672. DOI: 10.1103/PhysRevD.14.1667. URL: https://link.aps.org/doi/10.1103/PhysRevD.14.1667.

[24] Steven Weinberg. "Implications of dynamical symmetry breaking". In: *Phys. Rev. D* 13 (4 Feb. 1976), pp. 974–996. DOI: 10.1103/PhysRevD.13.974. URL: https://link.aps.org/doi/10.1103/PhysRevD.13.974.

[25] S. Weinberg. "Implications of dynamical symmetry breaking: An addendum". In: *Phys. Rev. D* 19 (4 Feb. 1979), pp. 1277–1280. DOI: 10.1103/PhysRevD.19.1277. URL: https://link.aps.org/doi/10.1103/PhysRevD.19.1277.

[26] Gerhard Buchalla, Andrzej J. Buras, and Markus E. Lautenbacher. "Weak decays beyond leading logarithms". In: *Rev. Mod. Phys.* 68 (1996), pp. 1125–1144. DOI: 10.1103/RevModPhys.68.1125. arXiv: hep-ph/9512380.

[27] T. Blake, T. Gershon, and G. Hiller. "Rare b hadron decays at the LHC". In: *Ann. Rev. Nucl. Part. Sci.* 65 (2015), pp. 113–143. DOI: 10.1146/annurev-nucl-102014-022231. arXiv: 1501.03309 [hep-ex].

[28] Thomas Blake, Gaia Lanfranchi, and David M. Straub. "Rare B Decays as Tests of the Standard Model". In: *Prog. Part. Nucl. Phys.* 92 (2017), pp. 50–91. DOI: 10.1016/j.ppnp.2016.10.001. arXiv: 1606.00916 [hep-ph].

[29] A. Petermann. "La normalisation des constantes dan la théorie des quanta (normalisation of constants in the quanta theory)". In: *Helv. Phys. Acta.* 26 (499 1953). URL: https://www.e-periodica.ch/digbib/view?pid=hpa-001:1953:26::894#511.

[30] M. Gell-Mann and F. E. Low. "Quantum Electrodynamics at Small Distances". In: *Phys. Rev.* 95 (5 Sept. 1954), pp. 1300–1312. DOI: 10.1103/PhysRev.95.1300. URL: https://link.aps.org/doi/10.1103/PhysRev.95.1300.

[31] Christoph Bobeth, Mikolaj Misiak, and Jorg Urban. "Photonic penguins at two loops and m_t dependence of $BR[B \to X_s l^+ l^-]$". In: *Nucl. Phys. B* 574 (2000), pp. 291–330. DOI: 10.1016/S0550-3213(00)00007-9. arXiv: hep-ph/9910220.

[32] M. S. Alam et al. "First Measurement of the Rate for the Inclusive Radiative Penguin Decay $b \to s\gamma$". In: *Phys. Rev. Lett.* 74 (15 Apr. 1995), pp. 2885–2889. DOI: 10.1103/PhysRevLett.74.2885. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2885.

[33] J. P. Lees et al. "Precision Measurement of the $B \to X_s\gamma$ Photon Energy Spectrum, Branching Fraction, and Direct CP Asymmetry $A_{CP}(B \to X_{s+d}\gamma)$". In: *Phys. Rev. Lett.* 109 (2012), p. 191801. DOI: 10.1103/PhysRevLett.109.191801. arXiv: 1207.2690 [hep-ex].

[34] J. P. Lees et al. "Measurement of B($B \to X_s\gamma$), the $B \to X_s\gamma$ photon energy spectrum, and the direct CP asymmetry in $B \to X_{s+d}\gamma$ decays". In: *Phys. Rev. D* 86 (2012), p. 112008. DOI: 10.1103/PhysRevD.86.112008. arXiv: 1207.5772 [hep-ex].

[35] A. Limosani et al. "Measurement of Inclusive Radiative B-meson Decays with a Photon Energy Threshold of 1.7-GeV". In: *Phys. Rev. Lett.* 103 (2009), p. 241801. DOI: 10.1103/PhysRevLett.103.241801. arXiv: 0907.1384 [hep-ex].

[36] Roel Aaij et al. "Angular analysis of the $B^0 \to K^{*0}e^+e^-$ decay in the low-q^2 region". In: *JHEP* 04 (2015), p. 064. DOI: 10.1007/JHEP04(2015)064. arXiv: 1501.03038 [hep-ex].

[37] M. Misiak et al. "Estimate of $\mathcal{B}(\bar{B} \to X_s\gamma)$ at $O(\alpha_s^2)$". In: *Phys. Rev. Lett.* 98 (2007), p. 022002. DOI: 10.1103/PhysRevLett.98.022002. arXiv: hep-ph/0609232.

[38] Thomas Blake. *Cartoon of the q^2 spectrum in $b \to s\ell^+\ell^-$ decays*. 2022. URL: https://twiki.cern.ch/twiki/pub/LHCbPhysics/RareDecayConferenceMaterial/qsquare.pdf (visited on 11/15/2022).

[39] Gerhard Buchalla. "Heavy quark theory". In: *55th Scottish Universities Summer School in Physics: Heavy Flavor Physics (SUSSP 2001)*. Feb. 2002, pp. 57–104. arXiv: hep-ph/0202092.

[40] A. Khodjamirian et al. "Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$". In: *JHEP* 09 (2010), p. 089. DOI: 10.1007/JHEP09(2010)089. arXiv: 1006.4945 [hep-ph].

[41] Sébastien Descotes-Genon et al. "On the impact of power corrections in the prediction of $B \to K^*\mu^+\mu^-$ observables". In: *JHEP* 12 (2014), p. 125. DOI: 10.1007/JHEP12(2014)125. arXiv: 1407.8526 [hep-ph].

[42] David London and Joaquim Matias. "B Flavour Anomalies: 2021 Theoretical Status Report". In: *Ann. Rev. Nucl. Part. Sci.* 72 (2022), pp. 37–68. DOI: 10.1146/annurev-nucl-102020-090209. arXiv: 2110.13270 [hep-ph].

[43] Wolfgang Altmannshofer et al. "Symmetries and Asymmetries of $B \to K^* \mu^+ \mu^-$ Decays in the Standard Model and Beyond". In: *JHEP* 01 (2009), p. 019. DOI: 10.1088/1126-6708/2009/01/019. arXiv: 0811.1214 [hep-ph].

[44] R. Aaij et al. "Differential branching fraction and angular analysis of the decay $B^0 \to K^{*0} \mu^+ \mu^-$". In: *JHEP* 08 (2013), p. 131. DOI: 10.1007/JHEP08(2013)131. arXiv: 1304.6325 [hep-ex].

[45] Christoph Bobeth, Gudrun Hiller, and Giorgi Piranishvili. "CP Asymmetries in bar $B \to \bar{K}^*(\to \bar{K}\pi)\bar{\ell}\ell$ and Untagged $\bar{B}_s$, $B_s \to \phi(\to K^+K^-)\bar{\ell}\ell$ Decays at NLO". In: *JHEP* 07 (2008), p. 106. DOI: 10.1088/1126-6708/2008/07/106. arXiv: 0805.2525 [hep-ph].

[46] U. Egede et al. "New observables in the decay mode $\bar{B}_d \to \bar{K}^{*0} l^+ l^-$". In: *JHEP* 11 (2008), p. 032. DOI: 10.1088/1126-6708/2008/11/032. arXiv: 0807.2589 [hep-ph].

[47] Frank Kruger and Joaquim Matias. "Probing new physics via the transverse amplitudes of $B^0 \to K^{*0}(\to K^-\pi^+) l^+ l^-$ at large recoil". In: *Phys. Rev. D* 71 (2005), p. 094009. DOI: 10.1103/PhysRevD.71.094009. arXiv: hep-ph/0502060.

[48] Sebastien Descotes-Genon et al. "Optimizing the basis of $B \to K^* ll$ observables in the full kinematic range". In: *JHEP* 05 (2013), p. 137. DOI: 10.1007/JHEP05(2013)137. arXiv: 1303.5794 [hep-ph].

[49] Roel 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". In: *JHEP* 11 (2016). [Erratum: JHEP 04, 142 (2017)], p. 047. DOI: 10.1007/JHEP11(2016)047. arXiv: 1606.04731 [hep-ex].

[50] R Aaij et al. "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0} \mu^+ \mu^-$". In: *Phys. Rev. Lett.* 111 (2013), p. 191801. DOI: 10.1103/PhysRevLett.111.191801. arXiv: 1308.1707 [hep-ex].

[51] Heavy Flavor Averaging Group (HFLAV). *Semileptonic B Decays*. 2022. URL: https://hflav.web.cern.ch/content/semileptonic-b-decays (visited on 11/16/2022).

[52] Roel 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". In: *Phys. Rev. Lett.* 127.15 (2021), p. 151801. DOI: 10.1103/PhysRevLett.127.151801. arXiv: 2105.14007 [hep-ex].

[53] R. Aaij et al. "Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays". In: *JHEP* 06 (2014), p. 133. DOI: 10.1007/JHEP06(2014)133. arXiv: 1403.8044 [hep-ex].

[54] J. P. Lees et al. "Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays $B \to K^{(*)} l^+ l^-$". In: *Phys. Rev. D* 86 (2012), p. 032012. DOI: 10.1103/PhysRevD.86.032012. arXiv: 1204.3933 [hep-ex].

[55] J. -T. Wei et al. "Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for $B \to K^{(*)}\ell^+\ell^-$". In: *Phys. Rev. Lett.* 103 (2009), p. 171801. DOI: 10.1103/PhysRevLett.103.171801. arXiv: 0904.0770 [hep-ex].

[56] Vardan Khachatryan et al. "Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Lett. B* 753 (2016), pp. 424–448. DOI: 10.1016/j.physletb.2015.12.020. arXiv: 1507.08126 [hep-ex].

[57] Roel Aaij et al. "Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay". In: *Phys. Rev. Lett.* 125.1 (2020), p. 011802. DOI: 10.1103/PhysRevLett.125.011802. arXiv: 2003.04831 [hep-ex].

[58] Roel Aaij et al. "Angular analysis of the rare decay $B_s^0 \to \phi\mu^+\mu^-$". In: *JHEP* 11 (2021), p. 043. DOI: 10.1007/JHEP11(2021)043. arXiv: 2107.13428 [hep-ex].

[59] Roel Aaij et al. "Angular Analysis of the $B^+ \to K^{*+}\mu^+\mu^-$ Decay". In: *Phys. Rev. Lett.* 126.16 (2021), p. 161802. DOI: 10.1103/PhysRevLett.126.161802. arXiv: 2012.13241 [hep-ex].

[60] Morad 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". In: *JHEP* 10 (2018), p. 047. DOI: 10.1007/JHEP10(2018)047. arXiv: 1805.04000 [hep-ex].

[61] *Measurement of the P_1 and P_5' angular parameters of the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256738.

[62] S. Wehle et al. "Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$". In: *Phys. Rev. Lett.* 118.11 (2017), p. 111801. DOI: 10.1103/PhysRevLett.118.111801. arXiv: 1612.05014 [hep-ex].

[63] Roel Aaij et al. "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity". In: *JHEP* 02 (2016), p. 104. DOI: 10.1007/JHEP02(2016)104. arXiv: 1512.04442 [hep-ex].

[64] Marzia Bordone, Gino Isidori, and Andrea Pattori. "On the Standard Model predictions for R_K and R_{K^*}". In: *Eur. Phys. J. C* 76.8 (2016), p. 440. DOI: 10.1140/epjc/s10052-016-4274-7. arXiv: 1605.07633 [hep-ph].

[65] Roel Aaij et al. "Test of lepton universality in beauty-quark decays". In: *Nature Phys.* 18.3 (2022), pp. 277–282. DOI: 10.1038/s41567-021-01478-8. arXiv: 2103.11769 [hep-ex].

[66] R. Aaij et al. "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *JHEP* 08 (2017), p. 055. DOI: 10.1007/JHEP08(2017)055. arXiv: 1705.05802 [hep-ex].

[67] *Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays*. Tech. rep. All figures and tables, along with any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2022-045.html (LHCb public pages). Geneva: CERN, 2022. arXiv: 2212.09153. URL: https://cds.cern.ch/record/2845040.

[68] S. Choudhury et al. "Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays". In: *JHEP* 03 (2021), p. 105. DOI: 10.1007/JHEP03(2021)105. arXiv: 1908.01848 [hep-ex].

[69] A. Abdesselam et al. "Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle". In: *Phys. Rev. Lett.* 126.16 (2021), p. 161801. DOI: 10.1103/PhysRevLett.126.161801. arXiv: 1904.02440 [hep-ex].

[70] Bernat Capdevila et al. "Assessing lepton-flavour non-universality from $B \to K^*\ell\ell$ angular analyses". In: *JHEP* 10 (2016), p. 075. DOI: 10.1007/JHEP10(2016)075. arXiv: 1605.03156 [hep-ph].

[71] Marcel Algueró et al. "$b \to s\ell^+\ell^-$ global fits after R_{K_S} and $R_{K^{*+}}$". In: *Eur. Phys. J. C* 82.4 (2022), p. 326. DOI: 10.1140/epjc/s10052-022-10231-1. arXiv: 2104.08921 [hep-ph].

[72] Sébastien Descotes-Genon et al. "Global analysis of $b \to s\ell\ell$ anomalies". In: *JHEP* 06 (2016), p. 092. DOI: 10.1007/JHEP06(2016)092. arXiv: 1510.04239 [hep-ph].

[73] Marcel Algueró et al. "Emerging patterns of New Physics with and without Lepton Flavour Universal contributions". In: *Eur. Phys. J. C* 79.8 (2019). [Addendum: Eur. Phys. J. C 80, 511 (2020)], p. 714. DOI: 10.1140/epjc/s10052-019-7216-3. arXiv: 1903.09578 [hep-ph].

[74] Wolfgang Altmannshofer and Peter Stangl. "New physics in rare B decays after Moriond 2021". In: *Eur. Phys. J. C* 81.10 (2021), p. 952. DOI: 10.1140/epjc/s10052-021-09725-1. arXiv: 2103.13370 [hep-ph].

[75] T. Hurth et al. "More Indications for Lepton Nonuniversality in $b \to s\ell^+\ell^-$". In: *Phys. Lett. B* 824 (2022), p. 136838. DOI: 10.1016/j.physletb.2021.136838. arXiv: 2104.10058 [hep-ph].

[76] Marcel Algueró et al. "Disentangling lepton flavor universal and lepton flavor universality violating effects in $b \to s\ell^+\ell^-$ transitions". In: *Phys. Rev. D* 105.11 (2022), p. 113007. DOI: 10.1103/PhysRevD.105.113007. arXiv: 2205.15212 [hep-ph].

[77] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[78] "LHC Machine". In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[79] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*. LHC technical proposal. Geneva: CERN, 1994. DOI: 10.17181/CERN.NR4P.BG9K. URL: https://cds.cern.ch/record/290968.

[80] *Technical proposal*. LHC technical proposal. Cover title : CMS, the Compact Muon Solenoid : technical proposal. Geneva: CERN, 1994. URL: https://cds.cern.ch/record/290969.

[81] *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC technical proposal. Geneva: CERN, 1995. URL: https://cds.cern.ch/record/293391.

[82] *LHCb : Technical Proposal*. Geneva: CERN, 1998. URL: https://cds.cern.ch/record/622031.

[83] R Alemany-Fernandez et al. "Operation and Configuration of the LHC in Run 1". In: (2013). URL: https://cds.cern.ch/record/1631030.

[84] Jorg Wenninger. "Operation and Configuration of the LHC in Run 2". In: (2019). URL: https://cds.cern.ch/record/2668326.

[85] Monica Pepe Altarelli and Frederic Teubert. "B Physics at LHCb". In: *Int. J. Mod. Phys. A* 23 (2008). Ed. by Gordon Kane and Aaron Pierce, pp. 5117–5136. DOI: 10.1142/S0217751X08042791. arXiv: 0802.1901 [hep-ph].

[86] Roel Aaij et al. "LHCb Detector Performance". In: *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[87] Christian Elsässer. *$\bar{b}b$ production angle plots*. URL: %7Bhttps://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb%5C_ProductionAngles.html%7D.

[88] R Aaij et al. "The LHCb Trigger and its Performance in 2011". In: *JINST* 8 (2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. arXiv: 1211.3055 [hep-ex].

[89] R Alemany-Fernandez, F Follin, and R Jacobsson. "The LHCB Online Luminosity Control and Monitoring". In: (2013). URL: https://cds.cern.ch/record/1567250.

[90] Federico Alessio, Richard Jacobsson, and Renaud Le Gac Legac. *LHCb integrated luminosity*. 2022. URL: https://lbgroups.cern.ch/online/OperationsPlots/index.htm (visited on 11/17/2022).

[91] A Augusto Alves et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://cds.cern.ch/record/1129809.

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

[93] P R Barbosa-Marinho et al. *LHCb VELO (VErtex LOcator): Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/504321.

[94] Mark Tobin. "The LHCb Silicon Tracker". In: *Nucl. Instrum. Methods Phys. Res., A* 831 (2016), pp. 174–180. DOI: 10.1016/j.nima.2016.05.115. URL: https://cds.cern.ch/record/2287114.

[95] P R Barbosa-Marinho et al. *LHCb inner tracker: Technical Design Report*. Technical design report. LHCb. revised version number 1 submitted on 2002-11-13 14:14:34. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/582793.

[96] P R Barbosa-Marinho et al. *LHCb outer tracker: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/519146.

[97] Serena Oggero. "Beauty in the Crowd: Commissioning of the LHCb Pile-Up detector and First evidence of $B_s^0 \to \mu^+\mu^-$". Presented 03 Oct 2013. 2013. URL: https://cds.cern.ch/record/1635658.

[98] R Arink et al. "Performance of the LHCb Outer Tracker". In: *JINST* 9.01 (2014), P01002. DOI: 10.1088/1748-0221/9/01/P01002. arXiv: 1311.3893 [physics.ins-det].

[99] Ph. d'Argent et al. "Improved performance of the LHCb Outer Tracker in LHC Run 2". In: *JINST* 12.11 (2017), P11016. DOI: 10.1088/1748-0221/12/11/P11016. arXiv: 1708.00819 [physics.ins-det].

[100] Johannes Albrecht et al. *Upgrade trigger amp; reconstruction strategy: 2017 milestone*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2310579.

[101] Barbara Storaci. "Optimization of the LHCb track reconstruction". In: *J. Phys.: Conf. Ser.* 664.7 (2015), p. 072047. DOI: 10.1088/1742-6596/664/7/072047. URL: https://cds.cern.ch/record/2017850.

[102] Roel Aaij et al. "Measurement of the track reconstruction efficiency at LHCb". In: *JINST* 10.02 (2015), P02007. DOI: 10.1088/1748-0221/10/02/P02007. arXiv: 1408.1251 [hep-ex].

[103] O Callot. *FastVelo, a fast and efficient pattern recognition package for the Velo*. Tech. rep. LHCb. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1322644.

[104] O Callot and M Schiller. *PatSeeding: A Standalone Track Reconstruction Algorithm*. Tech. rep. Geneva: CERN, 2008. URL: https://cds.cern.ch/record/1119095.

[105] O Callot and S Hansmann-Menzemer. *The Forward Tracking: Algorithm and Performance Studies*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1033584.

[106] M Needham and J Van Tilburg. *Performance of the track matching*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1020304.

[107] M Needham. *Performance of the Track Matching*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1060807.

[108] R Kalman. "A new approach to linear filtering and prediction problems". In: *Journal of Basic Engineering* 35 (1960).

[109] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: https://www.sciencedirect.com/science/article/pii/0168900287908874.

[110] E Rodrigues. *Dealing with clones in the tracking*. Tech. rep. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/1000723.

[111] A Perieanu. *A Fast Algorithm to Identify and Remove Clone Tracks*. Tech. rep. Geneva: CERN, 2008. URL: https://cds.cern.ch/record/1102942.

[112] Michel De Cian et al. *Fast neural-net based fake track rejection in the LHCb reconstruction*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2255039.

[113] Roel Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *JINST* 14.04 (2019), P04013. DOI: 10.1088/1748-0221/14/04/P04013. arXiv: 1812.10790 [hep-ex].

[114] Marcin Kucharczyk, Piotr Morawski, and Mariusz Witek. *Primary Vertex Reconstruction at LHCb*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1756296.

[115] R. Aaij et al. "Performance of the LHCb Vertex Locator. Performance of the LHCb Vertex Locator". In: *JINST* 9 (2014). Comments: 61 pages, 33 figures, P09007. DOI: 10.1088/1748-0221/9/09/P09007. arXiv: 1405.7808. URL: https://cds.cern.ch/record/1707015.

[116] S Amato et al. *LHCb RICH: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/494263.

[117] S Amato et al. *LHCb calorimeters: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/494264.

[118] P R Barbosa-Marinho et al. *LHCb muon system: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/504326.

[119] *LHCb muon system: addendum to the Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/600536.

[120] *LHCb muon system: second addendum to the Technical Design Report*. Technical design report. LHCb. Submitted on 9 Apr 2005. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/831955.

[121] M. Adinolfi et al. "Performance of the LHCb RICH detector at the LHC". In: *Eur. Phys. J. C* 73 (2013), p. 2431. DOI: 10.1140/epjc/s10052-013-2431-9. arXiv: 1211.6759. URL: https://cds.cern.ch/record/1495721.

[122] R. Calabrese et al. "Performance of the LHCb RICH detectors during LHC Run 2". In: *JINST* 17.07 (2022), P07013. DOI: 10.1088/1748-0221/17/07/P07013. arXiv: 2205.13400. URL: https://cds.cern.ch/record/2811056.

[123] Valentin Niess. *Commissioning of the LHCb Preshower With Cosmic Rays and First LHC Collisions*. 2010. URL: http://clrwww.in2p3.fr/lhcb/prs/calor2010/CALOR-niess.pdf (visited on 11/20/2022).

[124] Carlos Abellán Beteta et al. *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*. Tech. rep. 2020. arXiv: 2008.11556. URL: https://cds.cern.ch/record/2729028.

[125] A. A. Alves Jr. et al. "Performance of the LHCb muon system". In: *JINST* 8 (2013), P02022. DOI: 10.1088/1748-0221/8/02/P02022. arXiv: 1211.1346 [physics.ins-det].

[126] F. Archilli et al. "Performance of the Muon Identification at LHCb". In: *JINST* 8 (2013), P10020. DOI: 10.1088/1748-0221/8/10/P10020. arXiv: 1306.0249 [physics.ins-det].

[127] R W Forty and O Schneider. *RICH pattern recognition*. Tech. rep. Geneva: CERN, 1998. URL: https://cds.cern.ch/record/684714.

[128] O Deschamps et al. *Photon and neutral pion reconstruction*. Tech. rep. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/691634.

[129] H Terrier and I Belyaev. *Particle identification with LHCb calorimeters*. Tech. rep. Geneva: CERN, 2003. URL: https://cds.cern.ch/record/691743.

[130] Daniel Alexander Berninghoff, Johannes Albrecht, and Vladimir Gligorov. *Bremsstrahlung Recovery of Electrons using Multivariate Methods*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2146447.

[131] Dmitry Golubkov and Victor Egorychev. *Electron particle identification with LHCb calorimeter system for 2011 data taking period*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1403098.

[132] Andreas Hoecker et al. "TMVA-toolkit for multivariate data analysis". In: *arXiv preprint physics/0703039* (2007).

[133] *Trigger Schemes*. URL: %7Bhttp://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html%7D%7Bhttp://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html%7D (visited on 03/04/2022).

[134] Alexandra Martin Sanchez, Patrick Robbe, and Marie-Hélène Schune. *Performances of the LHCb L0 Calorimeter Trigger*. Tech. rep. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1407893.

[135] R. Aaij et al. "Tesla : an application for real-time data analysis in High Energy Physics". In: *Comput. Phys. Commun.* 208 (2016), pp. 35–42. DOI: 10.1016/j.cpc.2016.07.022. arXiv: 1604.05596 [physics.ins-det].

[136] F Stagni, P Charpentier, and (On behalf ofthe LHCb Collaboration). "The LHCb DIRAC-based production and data management operations systems". In: *Journal of Physics: Conference Series* 368.1 (June 2012), p. 012010. DOI: 10.1088/1742-6596/368/1/012010. URL: https://dx.doi.org/10.1088/1742-6596/368/1/012010.

[137] Vladimir V Gligorov. *A single track HLT1 trigger*. Tech. rep. Geneva: CERN, 2010. URL: https://cds.cern.ch/record/1323812.

[138] Tatiana Likhomanenko et al. "LHCb Topological Trigger Reoptimization". In: *J. Phys.: Conf. Ser.* 664.8 (2015). 21st International Conference on Computing in High Energy Physics (CHEP2015), p. 082025. DOI: 10.1088/1742-6596/664/8/082025. arXiv: 1510.00572. URL: https://cds.cern.ch/record/2019813.

[139] M Williams et al. *The HLT2 Topological Lines*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1323557.

[140] Tatiana Likhomanenko et al. "LHCb Topological Trigger Reoptimization". In: *J. Phys. Conf. Ser.* 664.8 (2015), p. 082025. DOI: 10.1088/1742-6596/664/8/082025. arXiv: 1510.00572 [physics.ins-det].

[141] Michele Atzeni et al. "Angular analysis of $B^0 \to K^{*0}e^+e^-$ decays". In: *In preparation* (2021). URL: https://cds.cern.ch/record/2771986.

[142] Michele Atzeni et al. "Direct determination of the Lepton Flavour Universality observables $\Delta\mathcal{C}_{9,10}^{(\prime)}$ in $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *In preparation* (2022). URL: https://cds.cern.ch/record/2803661.

[143] Federica Lionetto. "Measurements of Angular Observables of $B^0 \to K^{*0}\mu^+\mu^-$ and $B^0 \to K^{*0}e^+e^-$ Decays and the Upgrade of LHCb". Presented 22 Mar 2018. 2018. URL: https://cds.cern.ch/record/2624938.

[144] Gloria Corti et al. "Software for the LHCb experiment". In: *IEEE Trans. Nucl. Sci.* 53 (2006), pp. 1323–1328. URL: https://cds.cern.ch/record/913834.

[145] G Barrand et al. "GAUDI: The software architecture and framework for building LHCb data processing applications". In: (2000). DOI: 10.1016/S0010-4655(01)00254-5. URL: https://cds.cern.ch/record/467678.

[146] Marco Clemencic et al. "Recent developments in the LHCb software framework gaudi". In: *Journal of Physics: Conference Series* 219.4 (Apr. 2010), p. 042006. DOI: 10.1088/1742-6596/219/4/042006. URL: https://dx.doi.org/10.1088/1742-6596/219/4/042006.

[147] M Clemencic et al. "The LHCb Simulation Application, Gauss: Design, Evolution and Experience". In: *Journal of Physics: Conference Series* 331.3 (Dec. 2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023. URL: https://dx.doi.org/10.1088/1742-6596/331/3/032023.

[148] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (May 2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. URL: https://dx.doi.org/10.1088/1126-6708/2006/05/026.

[149] David J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4. URL: https://www.sciencedirect.com/science/article/pii/S0168900201000894.

[150] N. Davidson, T. Przedzinski, and Z. Was. "PHOTOS interface in C++: Technical and Physics Documentation". In: *Comput. Phys. Commun.* 199 (2016), pp. 86–101. DOI: 10.1016/j.cpc.2015.09.013. arXiv: 1011.0937 [hep-ph].

[151] S. Agostinelli et al. "Geant4 - a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[152] *The BOOLE Project*. http://lhcbdoc.web.cern.ch/lhcbdoc/boole/. Accessed: 2022-11-05.

[153] *The MOORE Project*. https://lhcbdoc.web.cern.ch/lhcbdoc/moore/. Accessed: 2022-11-19.

[154] *The BRUNEL Project*. https://lhcbdoc.web.cern.ch/lhcbdoc/brunel/. Accessed: 2022-11-19.

[155] *The DAVINCI Project*. https://lhcbdoc.web.cern.ch/lhcbdoc/davinci/. Accessed: 2022-11-19.

[156] Jose I. Illana, M. Jack, and T. Riemann. "Predictions for $Z \to \mu\tau$ and related reactions". In: *2nd Workshop of the 2nd Joint ECFA / DESY Study on Physics and Detectors for a Linear Electron Positron Collider*. Dec. 1999, pp. 490–524. arXiv: hep-ph/0001273.

[157] Andrzej J. Buras and Manfred Munz. "Effective Hamiltonian for $B \to X(s)e^+e^-$ beyond leading logarithms in the NDR and HV schemes". In: *Phys. Rev. D* 52 (1995), pp. 186–195. DOI: 10.1103/PhysRevD.52.186. arXiv: hep-ph/9501281.

[158] H. H. Asatrian et al. "Two loop virtual corrections to $B \to X_s\ell^+\ell^-$ in the standard model". In: *Phys. Lett. B* 507 (2001), pp. 162–172. DOI: 10.1016/S0370-2693(01)00441-5. arXiv: hep-ph/0103087.

[159] Patricia Ball and Roman Zwicky. "$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited". In: *Phys. Rev. D* 71 (2005), p. 014029. DOI: 10.1103/PhysRevD.71.014029. arXiv: hep-ph/0412079.

[160] Vladimir Gligorov. *TupleToolMCBackgroundInfo*. 2009. URL: https://twiki.cern.ch/twiki/bin/view/LHCb/TupleToolMCBackgroundInfo (visited on 11/30/2022).

[161] Lucio Anderlini et al. *The PIDCalib package*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2202412.

[162] Roel Aaij et al. "Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2". In: *EPJ Tech. Instrum.* 6.1 (2019), p. 1. DOI: 10.1140/epjti/s40485-019-0050-z. arXiv: 1803.00824 [hep-ex].

[163] Anton Poluektov. "Kernel density estimation of a multidimensional efficiency profile". In: *JINST* 10.02 (2015), P02011. DOI: 10.1088/1748-0221/10/02/P02011. arXiv: 1411.5528 [physics.data-an].

[164] S Tolk et al. *Data driven trigger efficiency determination at LHCb*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1701134.

[165] Jonas Eschle. "raredecay: MVA and reweighting with Machine Learning". In: (Nov. 2018). DOI: 10.5281/zenodo.1491924.

[166] Muriel Pivk and Francois R. Le Diberder. "SPlot: A Statistical tool to unfold data distributions". In: *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083.

[167] Paul Avery. *Applied Fitting Theory VI Formulas for Kinematic Fitting*. https://www.phys.ufl.edu/~avery/fitting/kinematic.pdf.

[168] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, 2016, pp. 785–794. ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. URL: http://doi.acm.org/10.1145/2939672.2939785.

[169] Tatiana Likhomanenko et al. "Reproducible Experiment Platform". In: *J. Phys. Conf. Ser.* 664.5 (2015), p. 052022. DOI: 10.1088/1742-6596/664/5/052022. arXiv: 1510.00624 [physics.data-an].

[170] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[171] Patrick Koppenburg. "Statistical biases in measurements with multiple candidates". In: (Mar. 2017). arXiv: 1703.01128 [hep-ex].

[172] Christoph Bobeth et al. "Long-distance effects in $B \to K^* \ell\ell$ from analyticity". In: *Eur. Phys. J. C* 78.6 (2018), p. 451. DOI: 10.1140/epjc/s10052-018-5918-6. arXiv: 1707.07305 [hep-ph].

[173] M. Oreglia. "A Study of the Reactions $\psi' \to \gamma\gamma\psi$". Stanford University, 1980. URL: https://www.slac.stanford.edu/cgi-bin/getdoc/slac-r-236.pdf.

[174] Tomasz Skwarnicki. "A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances". PhD thesis. Cracow, INP, 1986.

[175] R. Aaij et al. "Observation of J/ψ pair production in pp collisions at $\sqrt{s} = 7TeV$". In: *Phys. Lett. B* 707 (2012), pp. 52–59. DOI: 10.1016/j.physletb.2011.12.015. arXiv: 1109.0963 [hep-ex].

[176] "Search for the lepton-flavour violating decays $B^0 \to K^{*0}\mu^{\pm}e^{\mp}$ and $B_s^0 \to \phi\mu^{\pm}e^{\mp}$". In: (July 2022). arXiv: 2207.04005 [hep-ex].

[177] Roel Aaij et al. "Measurement of b hadron fractions in 13 TeV pp collisions". In: *Phys. Rev. D* 100.3 (2019), p. 031102. DOI: 10.1103/PhysRevD.100.031102. arXiv: 1902.06794 [hep-ex].

[178] Elena Graverini et al. "Test of lepton flavour universality with $B^+ \to K^+\pi^+\pi^-\ell^+\ell^-$ decays". In: *In preparation* (2022). URL: https://cds.cern.ch/record/2804504.

[179] Roel Aaij et al. "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays". In: *Phys. Rev. Lett.* 115 (2015), p. 072001. DOI: 10.1103/PhysRevLett.115.072001. arXiv: 1507.03414 [hep-ex].

[180] G. Bohm and G. Zech. "Statistics of weighted Poisson events and its applications". In: *Nucl. Instrum. Meth. A* 748 (2014), pp. 1–6. DOI: 10.1016/j.nima.2014.02.021. arXiv: 1309.1287 [physics.data-an].

[181] Christoph Langenbruch. "Parameter uncertainties in weighted unbinned maximum likelihood fits". In: *Eur. Phys. J. C* 82.5 (2022), p. 393. DOI: 10.1140/epjc/s10052-022-10254-8. arXiv: 1911.01303 [physics.data-an].

[182] M. Ablikim et al. "Precision measurements of $\mathcal{B}[\psi(3686) \to \pi^+\pi^- J/\psi]$ and $\mathcal{B}[J/\psi \to l^+l^-]$". In: *Phys. Rev. D* 88.3 (2013), p. 032007. DOI: 10.1103/PhysRevD.88.032007. arXiv: 1307.1189 [hep-ex].

[183] K. Chilikin et al. "Observation of a new charged charmoniumlike state in $\bar{B}^0 \to J/\psi K^- \pi^+$ decays". In: *Phys. Rev. D* 90.11 (2014), p. 112009. DOI: 10.1103/PhysRevD.90.112009. arXiv: 1408.6457 [hep-ex].

[184] R Aaij et al. "Measurement of the polarization amplitudes in $B^0 \to J/\psi K^*(892)^0$ decays". In: *Phys. Rev. D* 88 (2013), p. 052002. DOI: 10.1103/PhysRevD.88.052002. arXiv: 1307.2782 [hep-ex].

[185] Bernard Aubert et al. "Measurement of decay amplitudes of $B \to J/\psi K^*, \psi(2S)K^*$, and $\chi_{c1}K^*$ with an angular analysis". In: *Phys. Rev. D* 76 (2007), p. 031102. DOI: 10.1103/PhysRevD.76.031102. arXiv: 0704.0522 [hep-ex].

[186] K. Chilikin et al. "Experimental constraints on the spin and parity of the $Z(4430)^+$". In: *Phys. Rev. D* 88.7 (2013), p. 074026. DOI: 10.1103/PhysRevD.88.074026. arXiv: 1306.4894 [hep-ex].

[187] R. Aaij et al. "Measurement of b-hadron production fractions in 7 TeV pp collisions". In: *Phys. Rev. D* 85 (2012), p. 032008. DOI: 10.1103/PhysRevD.85.032008. arXiv: 1111.2357 [hep-ex].

[188] Roel Aaij et al. "Observation of the resonant character of the $Z(4430)^-$ state". In: *Phys. Rev. Lett.* 112.22 (2014), p. 222002. DOI: 10.1103/PhysRevLett.112.222002. arXiv: 1404.1903 [hep-ex].

[189] H. Guler et al. "Study of the $K^+\pi^+\pi^-$ Final State in $B^+ \to J/\psi K^+\pi^+\pi^-$ and $B^+ \to \psi - primeK^+\pi^+\pi^-$". In: *Phys. Rev. D* 83 (2011), p. 032005. DOI: 10.1103/PhysRevD.83.032005. arXiv: 1009.5256 [hep-ex].

[190] C. Runge. "Ueber empirische Funktionen und die Interpolation zwischen qequidistanten Ordinaten". In: *Zeitschrift fuer Mathematik und Physik* 46 (1901), pp. 224–243.

[191] Eluned Anne Smith, Sebastian Schmitt, and Christoph Michael Langenbruch. *On the PID-response dependency for dielectron pairs*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2772374.

[192] Roel Aaij et al. "Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era". In: (Aug. 2018). arXiv: 1808.08865 [hep-ex].

[193] David M. Straub. "flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond". In: (Oct. 2018). arXiv: 1810.08132 [hep-ph].

Acknowledgements

Firstly, I would like to thank Nico and Rafael for giving me this amazing opportunity to experience the story of the B anomalies first hand, and of particle physics in general – you (probably) have no idea just how much this means to me. Nico, I admire your deep insight and passion (even though I was never good at showing it). And Rafael, your knowledge and tenacity and endless optimism was just what I needed to tackle a tricky analysis. There has been some crazy times during the last four years. The understanding and support that you both showed was truly incredible. Needless to say, it has been a joy and an honour to work with you.

A big thank you to my fellow '$b \to see$ teammates' – Michele, Andrea, Federica and Jonas. Michele, I am so grateful for the many, many times you listened patiently to my ideas – the good and the bad and the fuzzy – and for giving back more (and often better) ones in return. Seeing you in the office simply brightens my day, and I have to admit I enjoyed bothering you a little (sorry!). I learned quite a few things from you Andrea, in particular on the art of problem solving. You did join the muon camp, otherwise I would have bothered you a lot more! This analysis would not have progressed so quickly to its current state if it were not for the work of Federica. It was good to build upon the foundations you laid, and I hope you will see the culmination of your work soon! Jonas, your computing knowledge is incredible, and you taught and helped me so many times. I also have fond memories of the day a small hackathon turned out to be the beginning of `zfit`. Best of luck with it!

Thank you to Annarita, Patrick, Iaroslava and Vadym. Annarita, I truly enjoyed all the pleasant and relaxing conversations we had. Your wise words were so helpful during stressful times. Patrick, I learned many things from you, especially the value of clarity and simplicity in the communication of complicated topics. I enjoyed teaching with you, Iaros, it made the crazy situation that we faced much more bearable. And Vadym, I especially appreciate your amazing calmness and sense of humour, I had a good time working with you. I must also thank Davide, Martina, Dario and Martin, for the pleasant coffee breaks and fun conversations.

I would like to thank other (and former) members of the Zurich LHCb group – Abhijit, Eluned, Julian, Elena, Andreas, Oliver and Albert, and my theorist neighbours, especially Claudia and Marzia, for the great time we had together, be they long or short. And thank you to Roland, for teaching me how not to break the farm, and Olaf, for giving me the opportunity to try out a little detector-related project.

A very special thank you to all the quasi-Fluffy physicists for showing me glimpses of the Hallendar in real life. I wish I could have met all of you, but *a'la'song-ko'ter'ra* (this is how it is). Finally, a big thank you to my father Shouchun for all the amazing stories, and my mother Fengying for not only being a great mother but also a good friend. None of this would have been possible without your tireless work and your unconditional love and support. Also... sorry for the way too many hoodies.