Test of lepton flavour universality in b->sll decays

da Yu Tou

To cite this version:

da Yu Tou. Test of lepton flavour universality in b->sll decays. High Energy Physics - Theory [hep-th]. Sorbonne Université, 2020. English. NNT : 2020SORUS409 . tel-03580540

HAL Id: tel-03580540
https://theses.hal.science/tel-03580540v1

Submitted on 18 Feb 2022

Sorbonne Université

École doctorale des Sciences de la Terre et de l'environnement et Physique de l'Univers, Paris - ED 560

Laboratoire de Physique Nucléaire et de Hautes Énergies - UMR 7585

Test of Lepton Flavour Universality in $b \to s\ell\ell$ Decays

Test de l'universalité de la Saveur Lepton dans la Désintégration $b \to s\ell\ell$

Présenté par

Da Yu Tou

Thèse de Doctorat de Physique

Soutenue le 14/12/2020 devant le jury composé de :

Dr.	Delphine Hardin	Président du jury	LPNHE
Prof.	Gerhard Raven	Rapporteur	NIKHEF
Prof.	Maria Cavli	Rapporteur	LPNHE
Prof.	Tim Gershon	Examinateur	University of Warwick
Dr.	Francois Le Diberder	Examinateur	IJClab
Dr.	Aiofe Bharucha	Examinateur	CPT
Dr.	Vladimir Gligorov	Directeur de thèse	LPNHE
Dr.	Renato Quagliani	Co-directeur de thèse	LPNHE

Acknowledgement

First and foremost, my deepest gratitude goes to my supervisor, Vladimir Gligorov. Throughout my PhD journey, he has provided continuous guidance and endless support, and I would not have arrived at the end without his patience, especially during the final months spent writing my thesis. I am indebted to my generous co-supervisor, Renato Quagliani, for the regular private conversations which clarified my understanding of the many intricate parts involved in the measurement of lepton flavour universality documented in this thesis.

The simultaneous measurement of R_K and R_{K^*} is a culmination of the ideas and work by a brilliant team of physicist who I had the pleasure to work with. In addition to my supervisor and co-supervisor, many of the studies presented in this thesis were undertaken by Simone Bifani, Stephan Escher, Alex Seuthe and Ryan Calladine. They contributed much of the original code that was used to produce the plots in Chapter 4. The weekly meetings of this analysis were also regularly attended by Christoph Langenburch, Eluned Smith, Marie-Helene Schune, Francesco Polci and Nigel Watson who have regularly advised and supervised the progress of this analysis.

My past three years at LPNHE has been a pleasure and much of this is owed to the LHCb group at this lab. It's a little sad many of us are leaving the lab this year without a proper celebration due to the current pandemic.

My time in Paris would not have been complete without the friends I have made at this lab and in the past three years. Thank you for the beers, karaoke, hotpot and bouldering sessions throughout my time at Paris. I hope we can share another round of afterwork beer on Friday evening, gather round a steamy hotpot during a cold winter night, and sing *Boundless Oceans, Vast Skies* or a bitchy Britney Spears song again in the future.

I would like to extend gratitude to my loving parents who have tolerated 27 years of my stubbornness yet supported every decision I made in my life, including pursuing a PhD. Last but not least, words cannot describe how grateful I am to my girlfriend, Yajun He, who helped me maintained my sanity throughout the craziness that is 2020. Thank you for being the light at the end of my PhD tunnel.

Now, onto formalities. This thesis is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, under grant agreement No 724777 "RECEPT".

Contents

Chapter 1

Introduction

Decades of scientific endeavour by particle physics community have culminated in the Standard Model of particle physics. This theory encodes the mathematical structures and physical laws that govern the elementary particles and their interactions at the subatomic scales. Precise as it is, the Standard Model is far from the Theory of Everything. Theoretical and experimental efforts to search for physics beyond the Standard Model began before the final SM particle, the Higgs Boson, was discovered. As direct searches for new physics in the Large Hadron Collider (LHC) era have yielded null results, precision tests of the Standard Model offer an indirect method to probe theories beyond the reach of humanity's most advanced collider.

Among the indirect channels to search for new physics are the $b \to s\,\ell^+\ell^-$ decays, in particular the $b \to s\,\ell^+\ell^-$ lepton flavour universality measurements which will be the main body of this thesis. The first observations of $b \to s\,\ell^+\ell^-$ decays was made by the B-factory experiments, but their statistical limits prevented any precise interpretations of their results. Since the beginning of LHC operation, it has produced an unprecedented amount of $b\bar{b}$ events which herald a new age in precision measurements of flavour physics. Even with a partial LHCb dataset, published LHCb measurements of R_K and R_{K^*} are in tension with the SM at the $2-2.5\sigma$ level with a statistical sensitivity of about $7-17$ %.

The analysis will provide an updated simultaneous measurement of R_K and R_{K^*} with the full LHCb dataset (Chap. 4), in order to verify or invalidate the anomalies measured by previous analysis of these ratios. Nevertheless, the updated analysis will feature strategies that optimise the trigger selections and an efficiency calibration technique that decorrelates the signal and control modes. The mass fits in this analysis will fit R_K and R_{K^*} decay modes simultaneously, a significant departure from previous analysis that treat each channel independently. The author of this thesis is the primary contributor to the mass fits in this analysis and the pseudoexperiment generators as well, the latter which will be critical for numerous systematic studies. Given that this analysis is a collaboration between multiple investigators, the analysis chapter will contain figures not produced by the author of this thesis. Such figures will be captioned with ‘figure courtesy of a collaborator in this analysis’ and listed in Appendix G.

Whilst implementating the mass fits for the measurement of R_K and R_{K^*}, the author encountered a persistent bias when binned fits are used. Although this is a (not widely) known problem in RooFit, the author undertook an independent study to ascertain the nature of this bias. Furthermore, a solution is proposed to unbias the binned fits in RooFit, drawing inspiration from the per-bin numerical integration option offered by ROOT’s `TF1` fitter.

Chap. 4.5, Chap. 4.6 and Chap. 5 contain original work by the author of this thesis.

Chapter 2

The World as We Know It : Theory

The Standard Model (SM) is the most successful theory of particle physics that describes the fundamental interactions between particles at the subatomic scale (Sec. 2.1). This model catalogs indivisible fundamental particles and three of the four fundamental interaction (electromagnetic, weak, strong) between the fundamental particles. Its success owes to the symbiosis between experimental observations and theory calculations, the most precise of which, the electron magnetic dipole moment, agrees between theory (1) and experiment (2) at one part per trillion.

Despite its success, observations on cosmological and sub-atomic scales suggest the SM is incomplete (Sec. 2.2). The fourth force, gravity, has also eluded incorporation into the SM. These hints at physics beyond the SM and motivates the search for experimental evidence of new physics (NP).

One possible probe for NP are lepton flavour universality measurements in $b\to s\,\ell^+\ell^-$ decays (Sec. 2.3). Forbidden at the tree level, these decays proceed via electroweak penguin loops. The penguin loops can couple to virtual NP particles and probe beyond SM effects. As a matter of fact, state of the art SM predictions of lepton universality ratios hints at a possible tension with the corresponding experimental measurements (Sec. 2.4).

This tension has attracted theoretical attention (Sec. 2.5). Theoretical efforts to analyse $b \to s\,\ell^+\ell^-$ decays via model-independent Effective Field Theory approaches are compatible with some non-SM deviations in the effective Hamiltonian. These model-independent signatures can then be used to constrain or build NP models, such as the popular Z' or leptoquark models that attempt to explain our observations in $b \to s\,\ell^+\ell^-$ decays.

2.1 Standard Model of Particle Physics

The SM is built upon the theoretical framework of quantum field theory (QFT), born from the unification of classical field theory, special relativity and non-relativistic quantum mechanics. QFT describes fundamental particles (Fig. 2.1) as excitations of quantum fields and the dynamics of their interactions arise from the Lagrangian. The interaction forces are generated by gauge symmetries, that impose conservation laws of the theory [1].

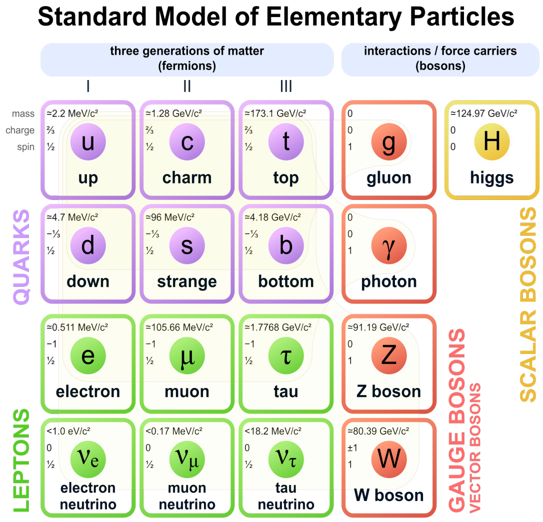


Figure 2.1 – The Standard Model fermions consists of 6 quarks and 6 leptons. The 4 vector bosons mediate the 3 interaction forces while the Higgs boson is responsible for the Higgs mechanism. Figure taken from (3).

The Standard Model is gauge invariant under the $U(1)_Y \times SU(2)_L \times SU(3)$ transformation. After spontaneous symmetry breaking, the electroweak $U(1)_Y \times SU(2)_L$ generates the electromagnetic photon and the weak $W^{\pm}$ and Z bosons. The strong force's $SU(3)$ color dynamics are quantised as 8 gluons, each with a unique color charge combination. Finally, the lone elementary scalar boson in the SM emerges from the Higgs mechanism, which generates particle masses without breaking the SM gauge symmetry.

---

[1] Via Noether's theorem.

These bosons dictate the interaction dynamics of the SM fermions. SM fermions are charged spin half particles that form the fundamental building block of all observable matter. The fundamental fermions are populated by two families: quarks and leptons with six members each.

The quarks are split into up-type and down-type quarks with three generations of increasing mass. The *up*, *charm* and *top* quarks make up the $+\frac{2}{3}e$ charged up-type quark while *down*, *strange* and *bottom* quarks are grouped into $-\frac{1}{3}e$ charged down-type quark. The quarks interact with all three fundamental forces via their electric charge (electromagnetic), color charge (strong) and quark flavour (weak).

The electrically charged lepton family members are, in order of increasing mass, *electron*, *muon* and *tau*. They are mirrored by their electrically neutral and massless neutrino counterparts [2]. In contrast to quarks, leptons do not interact with the strong force. All leptons interact with the weak force but only the electrically charged leptons interact with the electromagnetic force.

Quantum Electrodynamics

Historically, quantum electrodynamics (QED) was the first SM theory developed in the framework of QFT. Its Lagrangian (4) can be written as

$$\mathcal{L}_{QED} = \overline{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi + e\overline{\psi}\gamma^{\mu}A_{\mu}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{2.1}$$

where $F_{\mu\nu}$ is the electromagnetic field tensor, defined as

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}. \tag{2.2}$$

The QED Lagrangian is invariant under a local $U(1)$ gauge transformations of its fermion field ψ and photon field A_{μ}

$$\psi(x) \to e^{i\alpha(x)}\psi(x), \qquad A_{\mu} \to A_{\mu} - \frac{1}{e}\partial_{\mu}\alpha(x). \tag{2.3}$$

Quantum Chromodynamics

Subsequently, experimentalists began observing a plethora of strange and flavourless hadrons which theorists organised via the eightfold way (5). The patterns of

---

[2] Massless within the SM.

the eightfold way [3] then gave rise to an $SU(3)$ invariant theory known as quantum chromodynamics (QCD). QCD transforms the fermion fields as

$$\psi(x) \to \exp[ig_S\, \boldsymbol{\alpha}(x) \cdot \hat{\boldsymbol{T}}]\psi(x), \tag{2.4}$$

where $\hat{\boldsymbol{T}}$ are the eight generators of the $SU(3)$ symmetry group. This corresponds to the eight QCD gluons whose $SU(3)$ invariant fields transformation can be written as

$$G^k_\mu \to G^k_\mu - \partial_\mu \alpha_k - g_S f_{ijk} \alpha_i G^j_\mu. \tag{2.5}$$

These transformatoin define the $SU(3)$ gauge invariant Lagrangian

$$\mathcal{L}_{\rm QCD} = \bar{\psi}_i \left(i(\gamma^\mu D_\mu)_{ij} - m\, \delta_{ij} \right) \psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{2.6}$$

where the field strength tensor is

$$G^a_{\mu\nu} = \partial_\mu \mathcal{A}^a_\nu - \partial_\nu \mathcal{A}^a_\mu + g f^{abc} \mathcal{A}^b_\mu \mathcal{A}^c_\nu\,. \tag{2.7}$$

In contrast to the QED field strength tensor (Eq. 2.2), the QCD field strength tensor has an additional $gf^{abc}\mathcal{A}^b_\mu \mathcal{A}^c_\nu$. This term describes the self-interaction dynamics of the gluons which carry color charge unlike the charge neutral photons. Due to the gluon self-interaction, gluon field coupling between two color charges increases with charge separation. If the quarks are energetic enough to pull themselves apart, the energy in the gluon field instead hadronizes into colorless jets of quarks and gluons. This behaviour either traps quarks in colorless hadron states or the energetic quarks hadronize into colorless daugther hadrons. As a result, nature enforces hadronic systems into color singlet states, a phenomena known as color confinement. This qualitative argument explains why free quarks are not observed experimentally.

Contrary to its long distance behaviour, the non-Abelian $SU(3)$ structure of QCD causes asymtotic freedom (6; 7), a behaviour where particle interactions becomes asymptotically weaker as the energy scale increases (or length scale decreases). Although quarks are color confined to a hadronic bound state, the strong interaction between them becomes vanishingly small as their separation decreases. At this high energy regime, QCD calculations typically proceed pertubatively, such as the Heavy-Quark Effective Field Theory for b hadrons.

---

[3] The eightfold way organisation hinted that meson and baryons were composite particles, made up of spin-$\frac{1}{2}$ particles with fractional charge. Within this organisation, the spin-$\frac{3}{2}$ Δ^{++} particle seemingly had a symmetric quantum state despite being composed of elementary fermions. This is one of the early hints that quarks contain another quantum number, which was later shown to be the colour charge.

Discrete Symmetries

QCD and QED contain many similarities. They both respect gauges symmetries that generates the massless photon and gluon fields. QED and QCD also preserve the discrete class of symmetries, namely charge conjugation (C), parity conjugation (P) and time reversal (T).

Charge conjugation dictates that a quantum system behaves exactly the same when all charge quantum numbers are flipped. Time reversal symmetry preserves the quantum laws when the arrow of time is reversed. Parity inversion symmetry conserves physical laws when the spatial coordinates are flipped.

Observations of QCD and QED showed they respect all three symmetries in their interaction. This, however, is about to change in the weak interaction of the SM. Despite this, the multiplication of C, P and T quantum numbers, CPT, has been theoretically shown to be an exact symmetry of a Lorentz-invariant QFT model, known as the CPT theorem (8).

Weak Interaction

In contrast to QED and QCD, the weak interaction violates parity maximally and mediates flavour dynamics in the quark sector. Below the electroweak scale [4], weak decays are mediated by the $W^\pm$ bosons. The weak Lagrangian that couples $W^\pm$ and fermions is

$$\mathcal{L}_{weak} = -\frac{g_W}{\sqrt{2}}\left[\overline{u}_i\gamma^\mu\frac{1}{2}(1-\gamma^5)V_{ij}^{CKM}d_j + \overline{\nu_i}\gamma^\mu\frac{1}{2}(1-\gamma^5)e_i\right]W_\mu^+. \tag{2.8}$$

Both the quark (first RHS term of Eq. 2.8) and lepton (second RHS term of Eq. 2.8) contains a $\gamma^\mu(1-\gamma^5)$ term which encodes the vector minus axial current ($V-A$) structure of the weak interaction. Consequently, the $V-A$ term only couples to left-handed particles or right-handed antiparticles, violating parity conservation in weak decays.

The second term on the RHS contains a neutrino field $\overline{\nu_i}$ that can only couple to a charged lepton field e_i of the same generation i. This does not apply to quarks, where up-type quarks $\overline{u}_i$ and down-type quarks d_j of different generation couples via the CKM matrix V_{ij}^{CKM}, written as

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.9}$$

---

[4] Approximately 246 GeV.

The unitary V_{CKM} has four degrees of freedom, expressed under the Wolfenstein parameterisation (9) as

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1-\lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\lambda^2/2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} \tag{2.10}$$

where λ, A, ρ and η are real numbers. Due to the imaginary term $i\eta$, V_{CKM} gains a CP violating complex phase. This is experimentally observed as matter-antimatter asymmetry in hadronic decays.

Electroweak Unification

At the electroweak scale, the weak and electromagnetic interaction is unified into a $SU(2)_L \times U(1)_Y$ gauge group whose Lagrangian is

$$\mathcal{L}_{EW} = \overline{\psi} i\gamma^\mu D_\mu \psi \tag{2.11}$$

where ψ fermion fields denote a left-handed doublet ψ_L or a right-handed singlet ψ_R. The singlet and doublet fields are written as

$$\psi_L = \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \begin{pmatrix} \nu \\ e \end{pmatrix}_L \tag{2.12}$$

$$\psi_R = u_R, \quad d_R, \quad e_R \tag{2.13}$$

where u denotes the up-type quark, d the down-type quark, e the charged leptons and ν the neutrinos. The right-handed neutrino singlet is omitted because it does not couple to the electroweak sector. The $SU(2)_L \times U(1)_Y$ covariant derivative is expressed as

$$D_\mu = \partial_\mu - ig_W \frac{1}{2}\sigma_j W^j_\mu - i\frac{1}{2}g' B_\mu, \qquad \text{for } \psi_L \tag{2.14}$$

$$D_\mu = \partial_\mu - i\frac{1}{2}g' B_\mu, \qquad \text{for } \psi_R \tag{2.15}$$

where σ_j are the Pauli spin matrices. The three W^j boson fields are the generators of $SU(2)_L$ while B_μ generates the $U(1)_Y$ weak hypercharge field. These fields decompose into the physical $W^\pm$, Z and photon below the electroweak scale where W^1 and W^2 are related to the physical $W^\pm$ bosons by

$$W^+_\mu = \frac{1}{\sqrt{2}}(W^1_\mu - iW^2_\mu), \qquad W^-_\mu = \frac{1}{\sqrt{2}}(W^1_\mu + iW^2_\mu) \tag{2.16}$$

while W^3 and B_μ can be rotated into their physical Z boson and photon field by the weak mixing angle θ_W

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \frac{1}{\sqrt{g_W^2 + g'^2}} \begin{pmatrix} g_W W_\mu^{(3)} - g' B_\mu \\ g' W_\mu^{(3)} + g_W B_\mu \end{pmatrix} = \begin{pmatrix} \cos\,\theta_W & -\sin\,\theta_W \\ \sin\,\theta_W & \cos\,\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^{(3)} \\ B_\mu \end{pmatrix}. \tag{2.17}$$

The weak coupling constant g_W, hypercharge coupling constant g' and weak mixing angle θ_W in Eq. 2.17 are related to fundamental electric charge e by

$$e = g_W \sin\,\theta_W, \qquad e = g' \cos\,\theta_W. \tag{2.18}$$

Gauge bosons of a gauge invariant theory have zero mass, as is the case in QED and QCD. Yet at 80.4 GeV/c and 91.2 GeV/c the $W^\pm$ and Z of $SU(2)_L \times U(1)_Y$ are massive. In order to preserve gauge symmetry, these massive SM gauge bosons requires a spontaneous symmetry breaking mechanism in the form of a scalar Higgs field.

Higgs Mechanism

We can write the Lagrangian for a spontaneously broken electroweak theory (10)(11)(12) as

$$\mathcal{L}_{Higgs} = |D_\mu \phi|^2 - V(\phi) \tag{2.19}$$
$$= |D_\mu \phi|^2 - \mu^2 |\phi|^2 - \lambda |\phi|^4 \tag{2.20}$$

where the Higgs field ϕ is an $SU(2)$ doublet

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.21}$$

In the $\mu^2 < 0$ case, the minima of the potential has infinite degenerate solutions, which produces one massive scalar boson and three massless Goldstone bosons after spontaneous symmetry breaking. However, we can 'gauge away' the three massless Goldstone bosons with the unitary gauge. In this gauge choice, we expand the doublet scalar field (Eq. 2.21) about the minima potential v as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix}. \tag{2.22}$$

To obtain the masses of the EW gauge bosons, the above expression of the Higgs doublet and the EW covariant derivative (Eq. 2.14) are substitued into

$|D_\mu\phi|^2$ (Eq. 2.19). We will obtain terms that are quadratic in EW gauge boson fields

$$\frac{1}{8}v^2 g_W^2 (W_\mu^{(1)}W^{(1)\mu} + W_\mu^{(2)}W^{(2)\mu}) + \frac{1}{8}v^2 (g_W W_\mu^{(3)} - g' B_\mu)(g_W W_\mu^{(3)} - g' B_\mu). \quad (2.23)$$

where the quadratic $W^{(1)}$ and $W^{(2)}$ terms can be factorised as

$$\frac{1}{2}m_W^2 W_\mu^{(1)}W^{(1)\mu}, \qquad \frac{1}{2}m_W^2 W_\mu^{(2)}W^{(2)\mu} \quad (2.24)$$

and the mass of the physical $W^\pm$ is

$$m_W = \frac{1}{2}v g_W. \quad (2.25)$$

The term with $W^{(3)}$ and B_μ boson fields in Eq. 2.23 is orthogonal to the A_μ electromagnetic field (Eq. 2.17), hence a massless photon. Substituting Z_μ for $W^{(3)}$ and B_μ we arrive at the equation

$$\frac{1}{8}v^2(g_W^2 + g'^2)Z_\mu Z^\mu = \frac{1}{2}m_Z Z_\mu Z^\mu \quad (2.26)$$

which gives the mass of the Z boson

$$m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2}. \quad (2.27)$$

Another prediction of this scalar field is the existence of a massive scalar Higgs boson. Expanding the potential term in Eq. 2.19 using the Higgs doublet about the minima (Eq. 2.22) we can obtain the quadratic $h^2(x)$ term

$$v^2\lambda h^2(x) = \frac{1}{2}m_H^2 h^2(x), \qquad m_H = \sqrt{2\lambda}v \quad (2.28)$$

where the Higgs mass measurements average to $m_H = 125.10 \pm 0.14$GeV (13)(14)(15).

Fermion Masses

As we have seen, the Higgs field provides a spontaneous symmetry breaking mechanism to $SU(2)_L \times U(1)_Y$ invariant gauge theories and produce massive gauge bosons. Astonishingly, the Higgs field can also generate the fermion masses. Without the Higgs field, a vanilla fermion mass term of left- and right-handed fermions can take the form

$$-m\overline{\psi}\psi = -m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R) \quad (2.29)$$

in the Lagrangian. This formula presents a problem: it does not preserve $SU(2)_L \times U(1)_Y$ gauge invariance.

The $SU(2)$ invariant Higgs field can recover this. Under an infinitesimal local $SU(2)_L$ gauge transformation $\boldsymbol{\epsilon}(x)$, the Higgs field ϕ and Dirac conjugate of left-handed fermion doublet $\overline{\psi}_L$ transforms as

$$\phi \to (\boldsymbol{I} + ig_W \boldsymbol{\epsilon}(x) \cdot \boldsymbol{T})\phi, \tag{2.30}$$

$$\overline{\psi}_L \to \overline{\psi}_L(\boldsymbol{I} - ig_W \boldsymbol{\epsilon}(x) \cdot \boldsymbol{T}), \tag{2.31}$$

which leaves $\overline{\psi}_L \phi$ invariant under $SU(2)_L$ transformations. By adding the right-handed fields, we can write an $SU(2)_L \times U(1)_Y$ gauge invariant Yukawa interaction

$$\mathcal{L}_y = -g_f \left[\, \overline{\psi}_L \phi \psi_R + \overline{\psi}_R \phi \psi_L \, \right]. \tag{2.32}$$

In the unitary gauge (Eq. 2.22), the Lagrangian of the first generation leptons $\mathcal{L}_e$ is

$$\mathcal{L}_e = -\frac{1}{\sqrt{2}} g_e \left[\, (\overline{\nu}_e \; \overline{e})_L \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} e_R \; + \; \overline{e}_R \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \, \right] \tag{2.33}$$

$$= -\frac{1}{\sqrt{2}} g_e v (\overline{e}_L e_R + \overline{e}_R e_L) - \frac{1}{\sqrt{2}} g_e h (\overline{e}_L e_R + \overline{e}_R e_L) \tag{2.34}$$

where the first term gives the mass of the electron

$$m_e = \frac{1}{\sqrt{2}} g_e v \tag{2.35}$$

while the second term describes the electron coupling to Higgs field about the non-zero vacuum expectation value. Similarly, the down-type quarks can couple to the Higgs field and acquire a mass like the charged leptons. To generate the up-type quark masses, we construct the conjugate doublet ϕ_C which in the unitary gauge

$$\phi_C = -i\sigma_2 \phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} -v - h(x) \\ 0 \end{pmatrix} \tag{2.36}$$

couples to the up-type quarks via the Lagrangian

$$\mathcal{L}_u = \frac{1}{\sqrt{2}} g_u \left[\, (\overline{u} \; \overline{d})_L \begin{pmatrix} -v - h(x) \\ 0 \end{pmatrix} u_R \; + \; \overline{u}_R \begin{pmatrix} -v - h(x) \\ 0 \end{pmatrix} \begin{pmatrix} u \\ d \end{pmatrix}_L \, \right] \tag{2.37}$$

$$= -m_u (\overline{u}_L u_R + \overline{u}_R u_L) - \frac{1}{\sqrt{2}} g_u h (\overline{u}_L u_R + \overline{u}_R u_L) \tag{2.38}$$

and generate their masses. Likewise, this mechanism can be applied to the neutrino masses but this implies an extremely small Higgs-Yukawa coupling to the neutrinos ($g_\nu \leq 10^{-12}$), orders of magnitude smaller than the coupling to other fermions. This might hint at a mechanism beyond the SM that generates the neutrino masses.

2.2 Limitations of the Standard Model

How do we know that the Standard Model is an effective theory at our currently attainable energy scale with another theory looming beyond our direct reach? Ask an astronomer and he/she will say the SM does not provide a viable dark matter candidate to explain the rotational curves observed in galaxies (16). A cosmologist would point out a lack of dark energy mechanism needed for a universe in an accelerating expansion (17; 18). Consult a theorist and he would answer that the Standard Model is unreconcilable with general relativity.

As our telescopes scanned across the galaxy, we failed to find matter-antimatter annihilation, hints for local cluster of antimatter. Assuming equal amounts matter-antimatter in the early universe, some mechanism(s) must exists to tip the balance in matter's favour. Sakharov proposed whatever this mechanism(s) is, it has to satisfy the Sakharov conditions (19) to explain our matter dominated universe. These conditions are: baryon number violation, C and CP symmetry violation, and early universe interactions out of thermal equilibrium. Although CP violating processes have been observed in s, c and b decays, the size of their matter-antimatter asymmetry is insufficient to explain the matter-antimatter imbalance in the observable universe. Another challenge for high energy physics (HEP) community is to observe baryon number violation, a process beyond the explanation of the SM.

Within the Standard Model itself, neutrinos are assumed to be massless yet observations of neutrino oscillation has proved otherwise (20). Could neutrino masses be coupled to Beyond standard Model process instead? The Standard Model Higgs mass requires an unnatural fine tuning of its bare mass (21). Is there a new physics explanation for this? The electroweak unification happens at about $v = 246$ GeV. Could there be a strong-electroweak unification at a higher energy scale?

Our observations highlight gaps in our knowledge and our theories hint at hidden structure. The consensus among physicists is that the Standard Model would be superseeded by a more complete theory at a higher energy scale. Among the indirect searches for new physics, tests of lepton flavour universality (LFU) in $b \to s \ell^+ \ell^-$ decays have gained notable interests in the past few years.

2.3 Test of Lepton Universality with $b \to s\,\ell^+\ell^-$ Decays

Lepton Flavour Universality of the Standard Model

Tests of LFU are precision probes to beyond Standard Model physics because the SM has been experimentally observed and theoretically built to be lepton flavour universal. On the theoretical front, SM leptons and quarks interact weakly via Eq. 2.8 below the electroweak scale. Up-type and down-type quarks of different generations couple via the non-diagonal CKM matrix. This cross generation mixing does not conserve quark flavour and manifests as the flavourdynamics of hadron decays. In contrast, lepton couplings in Eq. 2.8 are diagonal and independent of lepton flavour. As a result, SM interactions conserve lepton flavour and respects LFU.

However, the neutrinos break the LFU (and lepton flavour conservation) via neutrino oscillation. The neutrino mass eigenstates, $\nu_{1,2,3}$ and flavour eigenstates, $\nu_{e,\mu,\tau}$ are related via the non-diagonal PMNS matrix (22)

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{2.39}$$

where the matrix U_{PMNS} is a unitary matrix like the CKM matrix. Naturally, this mechanism violates lepton flavour conservation and LFU, but the small mass differences between neutrino generations make such processes negligible in SM decays, as SM decays proceed over short distances. We thus assume neutrino oscillations play no role in $b \to s\,\ell^+\ell^-$ decays.

For an exhaustive review of experimental tests of LFU, we refer the reader to Ref. (23). Take note that this review does not contain the latest result of $\frac{\Gamma(W \to \tau\nu_\tau)}{\Gamma(W \to \mu\nu_\mu)}$ in Ref. (24) which was published subsequent to the review. For brevity, the measurements of $W^\pm$ and Z decays as test of LFU in the electroweak sector is given in Table 2.1, where they conserve LFU up to a precision of 0.3%.

$b \to s\,\ell^+\ell^-$ Decays

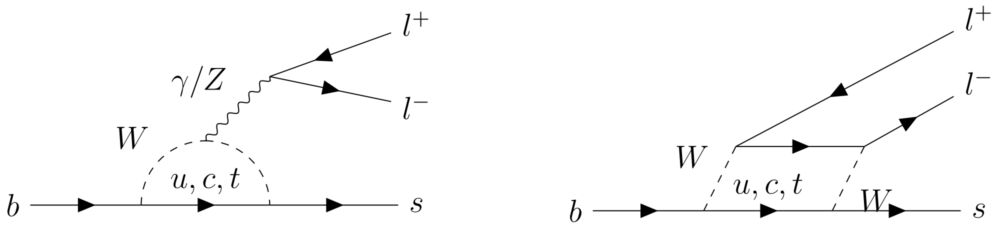
$b \to s\,\ell^+\ell^-$ decays are a class of flavour changing neutral current (FCNC) decay. FCNC decays do not proceed at the tree-level because the weak decay Lagrangian

$\frac{\Gamma(Z\to\mu\mu)}{\Gamma(Z\to ee)}$	Ref. (25)	1.0009 ± 0.0028
$\frac{\Gamma(Z\to\tau\tau)}{\Gamma(Z\to ee)}$	Ref. (25)	1.0019 ± 0.0032
$\frac{\Gamma(W\to\tau\nu_\tau)}{\Gamma(W\to e\nu_e)}$	Ref. (26)	1.063 ± 0.027
$\frac{\Gamma(W\to e\nu_e)}{\Gamma(W\to\mu\nu_\mu)}$	Ref. (27)	0.997 ± 0.010
$\frac{\Gamma(W\to\tau\nu_\tau)}{\Gamma(W\to\mu\nu_\mu)}$	Ref. (24)	0.992 ± 0.013

Table 2.1 – Experimental test of lepton flavour universality in the electroweak boson decays. Tests with dilepton final states conserve LFU at the 0.4% level while LFU in $W^\pm$ decays has been tested at least to the 1% level.

in Eq. 2.8 only couples up-type quarks to down-type quarks. The lowest order Feynman diagrams for FCNC are thus loop diagrams.

$b \to s\,\ell^+\ell^-$ transitions either proceed via an electroweak penguin loop that emits a neutral vector boson or a box diagram (Fig. 2.2). Any virtual up-type quark can contribute to the decay but the top quark contribution dominates due to its large mass. The branching fraction of these decays are small, $\mathcal{O}(10^{-6})$ in the SM (13). Although $b\to s\,\ell^+\ell^-$ decays are rare, they are sensitive probe to new physics coupling to the $b\to s\,\ell^+\ell^-$ virtual loop. For example, the heavy t, W and Z in $b\to s\,\ell^+\ell^-$ loops are very sensitive to non-SM extension with heavy charged Higgs or supersymmetric particles (28). These beyond SM coupling can lead to deviations in the SM predicted branching ratios or angular observables (29).



(a) Electroweak penguin.

(b) Box diagram.

Figure 2.2 – Lowest order Feynman diagrams of SM $b\to s\,\ell^+\ell^-$ decays.

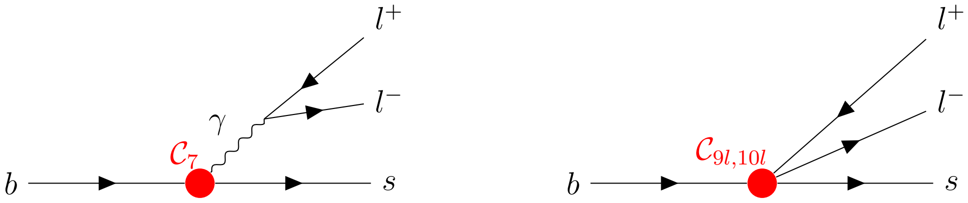
The Feynman diagrams of $b \to s\,\ell^+\ell^-$ decays incorporates a wide range of physics scales, from the $\mathcal{O}(100\,\text{GeV})$ electroweak scale down to the $\mathcal{O}(\,\text{MeV})$ s

quark mass. These contributions can be separated by taking an effective field theory (EFT) approach. This is similar to Fermi's theory of β decays where the $W^{\pm}$ propagator is integrated into a four-point weak interaction (30).

In our case, we are interested in a low energy EFT at the $\mu \approx m_b$ scale. Hence, short-distance propagation of massive degrees of freedoms ($W^{\pm}$, Z, t) are integrated into point-like interactions via operator product expansion (31). The light particles maintain their dynamical degrees of freedom and propagate over long distances. This EFT approach of integrating massive particle propagators is shown in Fig. 2.3, to be contrasted with the lowest order Feynman diagrams in Fig. 2.2. In this formalism, the leading terms of the effective Hamiltonian of $b \to s\,\ell^+\ell^-$ decays takes the form

$$\mathcal{H}_{eff} = -\frac{G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_i \mathcal{C}_i \mathcal{O}_i. \tag{2.40}$$

The contributions of short distance heavy fields are absorbed into the Wilson coefficients, C_i that act as effective coupling strength of the long distance operators $\mathcal{O}_i$. The four-fermion operators $\mathcal{O}_i$ are built from light fields propagating over long distances compared to the scale μ.



(a) $\mathcal{O}_7$ diagram. (b) $\mathcal{O}_{9l,10l}$ diagram.

Figure 2.3 – Feynman diagrams of the effective Hamiltoninan, where the t, $W^{\pm}$ and Z bosons are integrated into a point interaction.

Operators $i = 1, 2$ encode the current-current interactions, $i = 3$ .. 6 describes the gluonic penguin effects and $\mathcal{O}_8$ corresponds to the gluon pole. All these are irrelevant in $b \to s\,\ell^+\ell^-$ decays. We are only interested in the electroweak penguin operators $i = 7, 9, 10$

$$\mathcal{O}_7 = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_R b) F^{\mu\nu} \tag{2.41}$$

$$\mathcal{O}_{9l} = \frac{e}{16\pi^2} m_b (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu l) \tag{2.42}$$

$$\mathcal{O}_{10l} = \frac{e}{16\pi^2} m_b (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma^\mu \gamma_5 l) \tag{2.43}$$

where $\sigma_{\mu\nu} = \frac{i}{2}[\gamma_\mu, \gamma_\nu]$ and $F^{\mu\nu}$ is the electromagnetic field tensor. The $\mathcal{O}_7$ operator couples to the radiative contributions in $b \to s\,\ell^+\ell^-$ decays and can be tested with $b \to s\gamma$ probes. On the other hand, the $\mathcal{O}_{9l}$ and $\mathcal{O}_{10l}$ operators preserves the $V-A$ structure of the weak decays via the chirality operators on quark fields.

The Wilson coefficients and their operators are dependent on the Lorentz invariant lepton pair invariant mass, q^2. In phenomenological terms, probes at different q^2 regions are sensitive to different Wilson coefficients as illustrated in Fig. 2.4. At the J/ψ and $\psi(2S)$ resonances the branching fractions are dominated by SM tree-level process which makes them hard to disentangle from the electroweak process. Experimentalists therefore avoid these resonant q^2 regions in probes of $b \to s\,\ell^+\ell^-$ decays.

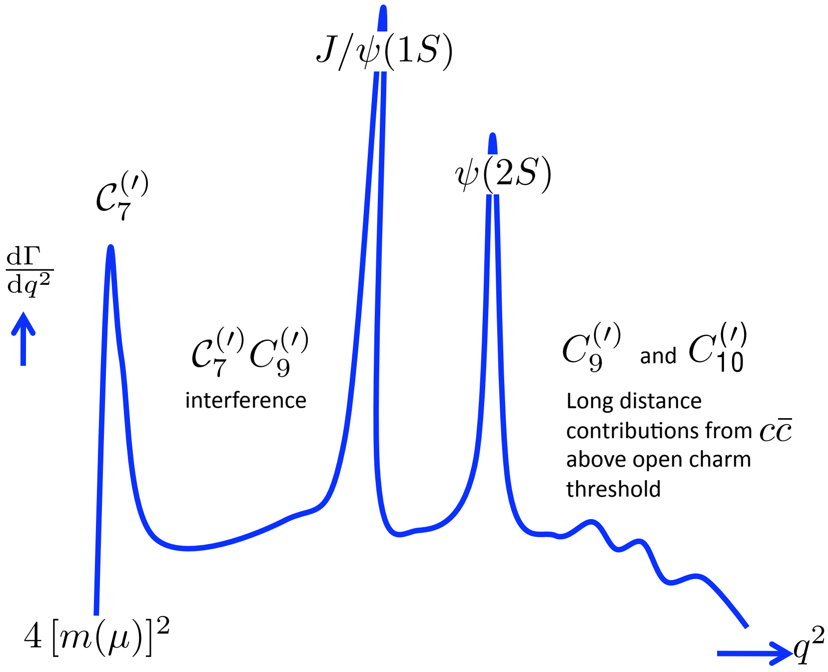


Figure 2.4 – Differential branching fraction of $b \to s\,\ell^+\ell^-$ decays as a function of q^2. q^2 region below the J/ψ resonance is sensitive to C_7 and C_9 while above the $\psi(2S)$ peak C_9 and C_{10} dominates. Image courtesy of Thomas Blake.

QCD contributions to these Feynman loops come in two forms: short distance hard gluons integrated into Wilson coefficients or non-pertubative long distance soft gluons interacting within the hadron system. Theoretical treatment of these form factors differs depending the q^2 region. Light-Cone Sum Rules (LCSR) are used in low q^2 regimes (32) while Lattice QCD simulations are employed at high q^2 (33; 34; 35). Furthermore, the calculations are expanded by EFTs (Soft-Collinear Effective Theory for low q^2 (36) and Heavy-Quark Effective Theory for high q^2 (37)) to reduce the number of independent hadronic inputs. Despite these theoretical treatments, the non-perturbative form factors remain as the dominant source of theoretical uncertainties in $b \to s\,\ell^+\ell^-$ predictions.

Given these large form factors uncertainties, one can try to define observables and the high precision LFU constraints in the SM, $b \to s\,\ell^+\ell^-$ lepton universality ratios offers an attractive route to probe new physics effect. These LFU ratios, R_X are integrated over a q^2 region

$$R_X = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma(H_b \to X\mu\mu)}{dq^2} dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma(H_b \to Xee)}{dq^2} dq^2} \tag{2.44}$$

where H_b denotes the initial beauty hadron while X the final state strange meson. The numerator and denominator in Eq. 2.44 shares common QCD form factors that cancels out, leaving R_X unaffected by form factor uncertainties. By canceling the dominant source of theory uncertainty, R_X predictions have very small residual theoretical uncertainties (Tab. 2.2).

Decay	q^2 range	Ref. (38)	Ref. (39)	Ref. (40)	Ref. (41)
$B^+ \to K^+\ell^+\ell^-$	$[1, 6]$	1.0 ± 0.01	$1.0004^{+0.0008}_{-0.0007}$	-	1.00 ± 0.01
$B^0 \to K^{*0}\ell^+\ell^-$	$[0.045, 1.1]$	0.92 ± 0.02	$0.920^{+0.007}_{-0.006}$	0.9259 ± 0.0041	0.906 ± 0.028
	$[0.1, 1.1]$	–	–	–	0.983 ± 0.014
	$[1.1, 6]$	1.00 ± 0.01	$0.996^{+0.002}_{-0.002}$	0.9965 ± 0.0006	1.00 ± 0.01
	$[15, 19]$	–	$0.998^{+0.001}_{-0.001}$	0.9981 ± 0.0001	–
$B^0_s \to \phi\ell^+\ell^-$	$[0.045, 1.1]$	–	–	0.9299 ± 0.0028	–
	$[1.1, 6.0]$	–	–	0.9970 ± 0.0002	–
	$[15, 19]$	–	–	0.9981 ± 0.0001	–

Table 2.2 – SM predictions of $B^+ \to K^+\ell^+\ell^-$, $B^0 \to K^{*0}\ell^+\ell^-$ and $B^0_s \to \phi\ell^+\ell^-$ in various q^2 bins. At most, the uncertainties are at the percent level. Notice that there are no predictions for R_K in the $[0.045, 1.1]$ q^2 bin. This is because $B^+ \to K^+\ell^+\ell^-$, for parity conservation reasons, do not couple to the photon pole Wilson coefficient C_7, which is the dominant contributor at low q^2.

Given their high precision theory predictions, measurements of R_X are sensitive probes for NP effects. Any deviations from SM predictions are strong hints for NP and lepton flavour universality violation (LFUV). Additionally, $b \to s\,\ell^+\ell^-$ observables can constrain the Wilson coefficients in Eq. 2.41-2.43 via global fits.

In turn, these constraints on Wilson coefficients can be used to rule out NP models or guide theorists in their model building quest.

2.4 Measurements of $b \to s\,\ell^+\ell^-$ Lepton Universality Ratios

2.4.1 B-factories

The B-factories Belle (42) and BaBar (43) were designed to study B meson decays, in particular the CP violating behaviour of B decays. The B-factories collide e^+e^- beams at a centre of mass energy near the $\Upsilon(4S)$resonance which decays into $B\overline{B}$ pairs more than 96% of the time (13). The main physics interest of B-factories were time-dependent measurements of CKM angles. To this end, the e^+e^- beam collides at asymmetric energies to produce boosted $\Upsilon(4S)$and subsequently, decays into boosted $B\overline{B}$ daughters. Since B factories were designed to study time dependent CP violation, both experiments commissioned high precision tracking detectors and designed flavour tagging algorithms to discern b flavour.

Belle was the world's first experiment to measure R_K and R_{K^*} using a dataset of 657 million $B\overline{B}$ pairs (44). Subsequently, BaBar published a complementary measurement using a dataset of 471 million $B\overline{B}$ pairs (45). Both experiments reconstructed charged and neutral kaons for their measurements while Belle also included final states with neutral pions to reconstruct $K^{*+} \to K^+\pi^0$ decays (see Tab. 2.3). Nine years after the end of Belle's data-taking, Belle collaboration updated their R_K and R_{K^*} measurement by analysing 772 million $B\overline{B}$ pairs (46; 47). Two significant differences in these updates are the finer q^2 binning as well as the split of R_X measurements into neutral and charged kaon final states (R_{K^+}-$R_{K^0_S}$ for R_K and $R_{K^{*0}}$-$R_{K^{*+}}$ for R_{K^*}).

Measurement	Final states used by BaBar and Belle	Final states used by Belle only
R_{K^*}	$B^0 \to K^{*0}(\to K^+\pi^-)l^+l^-$ $B^+ \to K^{*+}(\to K^0_S\pi^+)l^+l^-$	$B^+ \to K^{*+}(\to K^+\pi^0)l^+l^-$
R_K	$B^0 \to K^0_S l^+l^-$ $B^+ \to K^+l^+l^-$	-

Table 2.3 – Final states used in the measurements of R_K and R_{K^*} by the B-factories. The B-factories reconstruct both charged and neutral kaons. Belle included $K^{*+} \to K^+\pi^0$ decays to reconstruct K^{*+}.

2.4.2 *LHCb*

The LHCb experiment is the only experiment that has measured LFU ratios in $b \to s\,\ell^+\ell^-$ decays at the large hadron collider (LHC). The LHC collides symmetric proton-proton beams at various centre-of-mass energies. The LHC data-taking periods are subdivided two periods: Run 1 which comprises 2011 to 2012 and Run 2 which lasted from 2015 to 2018.

The main production mechanism of $b\bar{b}$ in the LHCb experiment is gluon-gluon fusion. As we shall see in Sec. 3.2, most of the $\bar{q}$uarks are produced collinearly and close to the beam direction. To take advantage of this production kinematic, LHCb detector is placed in the forward pseudorapidity region, $2 < \eta < 5$.

In comparison to e^+e^- colliders of B factories, proton-proton collisions are noisy events and the LHCb detectors have a higher occupancy. Furthermore, the high CoM energies produce energetic b hadrons. As a result, the energetic electron daughters emit bremsstrahlung radiation that dilutes momentum resolution. This is despite a dedicated recovery procedure to recover the momentum emitted by bremsstrahlung radiation. Both of these present a challenge in the offline resonstruction of b hadron events. Despite these disadvantages, the energetic b-quark can hadronize into all possible ground state hadron species namely B^0, B^+, B^0_s, B^+_c and $\mathit{\Lambda}^0_b$. In contrast, the $\mathit{\Upsilon}(4S)$resonance invariant mass only permits decays to B^0 or B^+ pairs. The hadronisation of b-quarks in LHCb therefore offers a wider range of initial and final hadronic states to probe $b \to s\,\ell^+\ell^-$ observables when compared to the B-factories.

LHCb has published measurements of R_K (48) and R_{K^*} (49) in the $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ final states using Run 1 dataset. Both of these measurements deviated from SM expectation at $> 2\sigma$. In 2019, LHCb updated the measurement of R_K using data collected over Run 1, 2015 and 2016 corresponding to an integrated luminosity $5.19\,\text{fb}^{-1}$ (50). LHCb then made the world's first test of LFU in the baryonic $\mathit{\Lambda}^0_b \to pK\ell^+\ell^-$ decays by measuring R_{pK} using Run 1 and 2016 data with an integrated luminosity of $4.7\,\text{fb}^{-1}$ (51).

2.4.3 Experimental Results

The experimental results for R_K, R_{K^*} and R_{pK} (LHCb only) are

$$R_{K^*} = \begin{cases} 0.66\,^{+\,0.11}_{-\,0.07} \pm 0.03 & \text{for} \quad 0.045 < q^2 < 1.1\,\mathrm{GeV}^2/c^4, \\ 0.69\,^{+\,0.11}_{-\,0.07} \pm 0.05 & \text{for} \quad 1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4, \end{cases}$$

$$R_K = 0.86\,^{+\,0.060\,+\,0.016}_{-\,0.054\,-\,0.014} \quad \text{for}\,1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4,$$

$$R_{pK} = 0.86\,^{+\,0.14}_{-\,0.11} \pm 0.05 \quad \text{for} \quad 0.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4,$$

for LHCb,

$$R_{K^*} = \begin{cases} 0.52\,^{+\,0.36}_{-\,0.26} \pm 0.05 & \text{for} \quad 0.045 < q^2 < 1.1\,\mathrm{GeV}^2/c^4, \\ 0.96\,^{+\,0.45}_{-\,0.29} \pm 0.11 & \text{for} \quad 1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4, \\ 0.90\,^{+\,0.27}_{-\,0.21} \pm 0.10 & \text{for} \quad 0.1 < q^2 < 8.0\,\mathrm{GeV}^2/c^4, \\ 1.18\,^{+\,0.52}_{-\,0.32} \pm 0.10 & \text{for} \quad 15 < q^2 < 19\,\mathrm{GeV}^2/c^4, \\ 0.94\,^{+\,0.17}_{-\,0.14} \pm 0.08 & \text{for} \quad 0.045 < q^2 < (m_B - m_{K^*})^2\,\mathrm{GeV}^2/c^4, \end{cases}$$

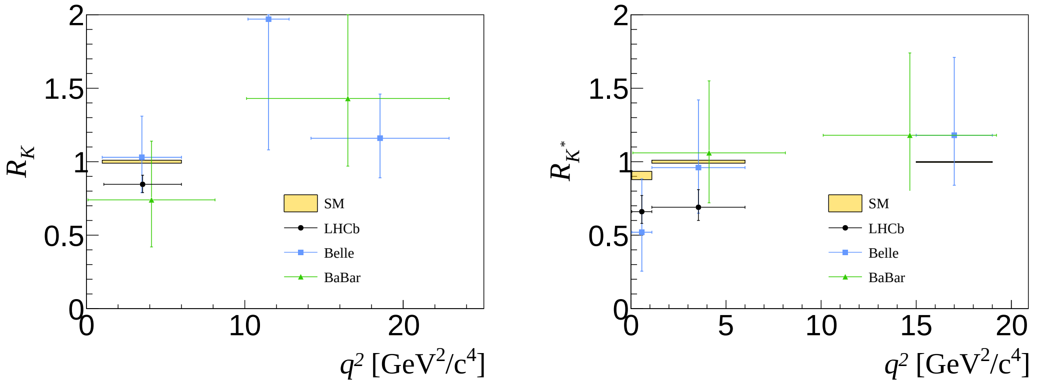
$$R_K = \begin{cases} 1.01\,^{+\,0.28}_{-\,0.25} \pm 0.02 & \text{for} \quad 0.1 < q^2 < 4.0\,\mathrm{GeV}^2/c^4, \\ 0.85\,^{+\,0.30}_{-\,0.24} \pm 0.01 & \text{for} \quad 4.0 < q^2 < 8.12\,\mathrm{GeV}^2/c^4, \\ 1.03\,^{+\,0.28}_{-\,0.24} \pm 0.01 & \text{for} \quad 1.0 < q^2 < 6.0\,\mathrm{GeV}^2/c^4, \\ 1.97\,^{+\,1.03}_{-\,0.89} \pm 0.02 & \text{for} \quad 10.2 < q^2 < 12.8\,\mathrm{GeV}^2/c^4, \\ 1.16\,^{+\,0.30}_{-\,0.27} \pm 0.01 & \text{for} \quad 14.18 < q^2 < (m_B - m_K)^2\,\mathrm{GeV}^2/c^4, \\ 1.10\,^{+\,0.16}_{-\,0.15} \pm 0.02 & \text{for} \quad 0.1 < q^2 < (m_B - m_K)^2\,\mathrm{GeV}^2/c^4, \end{cases}$$

for Belle and

$$R_{K^*} = \begin{cases} 1.06\,^{+\,0.48}_{-\,0.33} \pm 0.08 & \text{for} \quad 0.1 < q^2 < 8.12\,\mathrm{GeV}^2/c^4, \\ 1.18\,^{+\,0.55}_{-\,0.37} \pm 0.11 & \text{for} \quad 10.11 < q^2 < (m_B - m_{K^*})^2\,\mathrm{GeV}^2/c^4, \\ 1.13\,^{+\,0.34}_{-\,0.26} \pm 0.1 & \text{combined} \end{cases}$$

$$R_K = \begin{cases} 0.74\,^{+\,0.40}_{-\,0.31} \pm 0.06 & \text{for} \quad 0.1 < q^2 < 8.12\,\mathrm{GeV}^2/c^4, \\ 1.43\,^{+\,0.65}_{-\,0.44} \pm 0.12 & \text{for} \quad 10.11 < q^2 < (m_B - m_K)^2\,\mathrm{GeV}^2/c^4, \\ 1.00\,^{+\,0.31}_{-\,0.25} \pm 0.07 & \text{combined} \end{cases}$$

for BaBar. As we can see, $b \to s\,\ell^+\ell^-$ decays are very rare and LFU ratios are statistically limited. Compared to Belle and BaBar, LHCb results have much smaller statistical uncertainties. The results for R_K and R_{K^*} between experiments and their SM predictions are compared in Fig. 2.5. Note that these q^2 ranges apply a cut at the tree-level $c\bar{c}$ resonances, namely the region around J/ψ and $\psi(2S)$ resonances.



(a) R_K (b) R_{K^*}

Figure 2.5 – Experimental measurements of R_K (left) and R_{K^*} (right) by LHCb (red), Belle (blue) and BaBar (black). For the purpose of clarity, when an experiment reports results overlapping in q^2, we remove the largest q^2 range until no overlapping results remain. The theoretical predictions are extracted from Ref. (41).

2.5 New Physics Interpretation

2.5.1 Model Independent Analysis of Wilson Coefficients

NP beyond the SM are expected to propagate at distances equal to or shorter than the electroweak scale. As previously discussed in Sec. 2.3, we can decompose the effective Hamiltonian of $b \to s\,\ell^+\ell^-$ decays into different operators and integrate short distance processes into the Wilson coefficients. By expanding the effective Hamiltonian in the SM basis (Eq. 2.40), we can predict the SM Wilson coefficients and compare them to experimental observations.

The NP contribution to Wilson coefficients are assumed to be real as evidence of beyond SM sources of CP violation has not been detected. Although no theoretical argument forbids CP violating NP, CP-asymmetry measurements in $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ (52) are compatible with zero while global

CKM fits (53) agree with SM expectations. Under these assumptions, we can separate SM and NP contributions to Wilson coefficients as

$$\mathcal{C}_i = \mathcal{C}_i^{SM} + \mathcal{C}_i^{NP}. \tag{2.45}$$

Additional NP contributions can proceed via chirally flipped operators of Eq. 2.40

$$\mathcal{O}_{7'} = \frac{e}{16\pi^2} m_b(\bar{s}\sigma_{\mu\nu}P_L b)F^{\mu\nu} \tag{2.46}$$

$$\mathcal{O}_{9l'} = \frac{e}{16\pi^2} m_b(\bar{s}\gamma_\mu P_R b)(\bar{l}\gamma^\mu l) \tag{2.47}$$

$$\mathcal{O}_{10l'} = \frac{e}{16\pi^2} m_b(\bar{s}\gamma_\mu P_R b)(\bar{l}\gamma^\mu \gamma_5 l) \tag{2.48}$$

that are paired to $\mathcal{C}_{7'}$, $\mathcal{C}_{9'}$ and $\mathcal{C}_{10'}$. Due to the $SU(2)_L \times U(1)_Y$ symmetry of the SM, SM contributions to primed Wilson coefficients are negligible. This constraint does not apply to NP contributions so $\mathcal{C}_{i'}$ contributions are assumed to be entirely NP. Under these assumptions, we can constrain $\mathcal{C}_i^{NP}$ and $\mathcal{C}_{i'}$ via global fits that use experimental measurements of $b \to s\,\ell^+\ell^-$ observables as inputs.

Experimental inputs to $b \to s\,\ell^+\ell^-$ global fits include LFU ratios R_K and R_{K^*}, angular observables in $B \to K^{(*)}\ell^+\ell^-$ decays (54; 55; 56; 57; 58; 59) and branching ratios of b decays. Branching ratio inputs are available from $B_s^0 \to \mu^+\mu^-$ (60; 61; 62), $B \to K^{(*)}\ell^+\ell^-$ (63; 13), $B_s^0 \to \phi\mu\mu$ (64) and $b \to s\gamma$ (13) decays. These observables allow global fits to define two types of fit, either "All" the inputs or only "LFUV" observables are used as input. The two different global fits can highlight tensions between experimental observables or provide clues to NP structure.

Typically in these fits, the $\mathcal{C}_{7,7'}$ photon pole are strongly constrained by $b \to s\gamma$ measurements and $b \to s\,ee$ measurements at low-q^2 (65). In the global fits, hypothesis that leave $\mathcal{C}_{7,7'}$ floating agree very well with SM predictions (See Fig. 2.6) (65; 66; 67). This leaves $\mathcal{C}_{9,9',10,10'}$ as possible sources of NP. The $\mathcal{C}_{10,10'}$ coefficients, however, are strongly constrained by $B_s^0 \to \mu^+\mu^-$ measurements (68) whose current world average (13) of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.0 \pm 0.4) \times 10^{-9}$ is compatible with the SM prediction of $(3.65 \pm 0.23) \times 10^{-9}$ (69) at 1.4σ.

For global fits to generate LFUV effects $\mathcal{C}_{9(')}$ and $\mathcal{C}_{10(')}$ need to have different couplings to muon and electrons. A common hypothesis among global fits is to assume a hierachical NP coupling that favours muons with a negligible effect on the electrons. For simplicity and due to limited observable data, global fits usually test scenarios with modifications to only two or three Wilson coefficients and they assume NP couplings are LFUV in nature. However, 6D fits to $\mathcal{C}_{7,7',9,9',10,10'}$ have

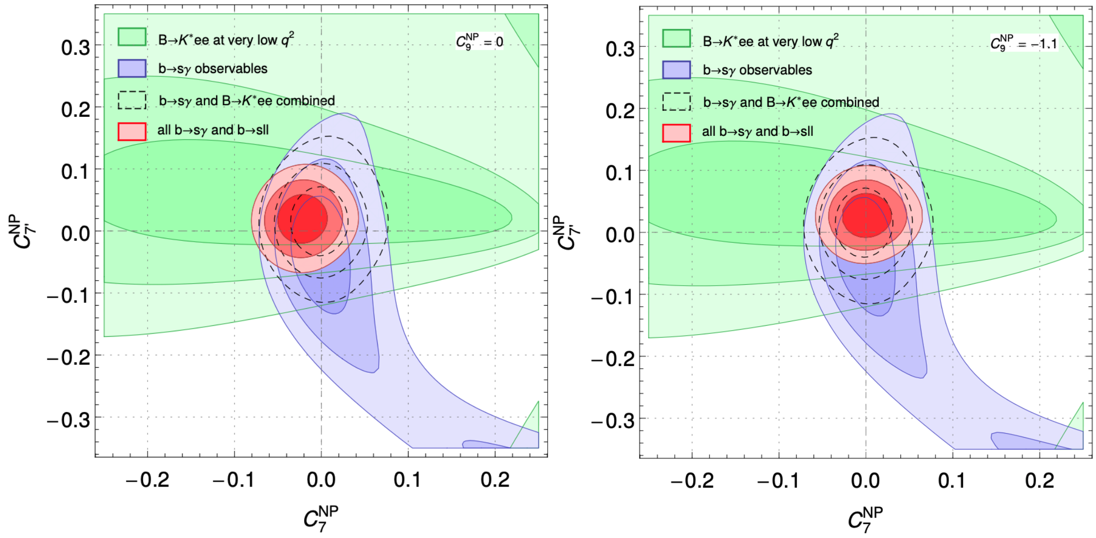


Figure 2.6 – Fits to deviation in Wilson coefficients from the SM, $\mathcal{C}_7^{NP}$ and $\mathcal{C}_{7'}^{NP}$, with $b \to s\gamma$ (blue) and $b \to s\,ee$ observables at very low q^2 (green) as experimental inputs. The combined fits to both set of inputs is shown with dashed contours. The global fits to all $b \to s\gamma$ and $b \to s\,\ell^+\ell^-$ data is shown by red contours. The contour boundaries mark the $1, 2, 3\sigma$ bounds. On the left, other Wilson coefficients take their SM values while $\mathcal{C}_9^{NP} = -1.1$ is assumed on the right. Figures extracted from Ref (65).

been performed (65; 66; 67) whereas Ref. (70; 67) has shown that fits to models with both LFU and LFUV components are possible with currently available observables.

Global fits to LFUV observables (65; 71; 72; 73; 66; 74; 75; 67) prefer solutions with negative $\mathcal{C}_{9\mu}^{NP}$. In 2D fits to $(\mathcal{C}_{9\mu}^{NP}, \mathcal{C}_{9e}^{NP})$, muon and electron couplings deviate in opposite direction: negative $\mathcal{C}_{9\mu}^{NP}$ and positive positive $\mathcal{C}_{9e}^{NP}$. The global fits also favor specific NP model motivated constraints, in particular $SU(2)_L$ invariant $\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{10\mu}^{NP}$ scenario and left-right chirality symmetric $\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{9\mu'}$ solution. Regardless, most of the best-fit scenarios favor a significant negative NP contributions to $\mathcal{C}_9$, usually at the 25% level. If the fits only consider LFUV observables, the most prominent NP scenarios have a significance of $3 - 4\sigma$ (66; 67).

This significance is increased to the $5 - 6\sigma$ level if the fits include all available $b \to s\,\ell^+\ell^-$ observables. Fig. 2.7 shows a common outcome of these global fits and compares results from "All" versus "LFUV" fits. In perspective, the pattern of $\mathcal{C}_{9\mu}^{NP}$ deviation from LFUV observables agrees with the deficiencies in $b \to s\,\mu\mu$

differential branching fractions measured by LHCb (76; 63; 64; 77). This non-SM behaviour is further supported by the P_5' anomaly in $B^0 \to K^{*0}\mu^+\mu^-$ angular analyses (59; 58; 57), where LHCb's latest $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis alone constrains $\mathcal{C}_{9\mu}^{NP}$ to deviate from the SM by 3.3σ. Concisely, a $\mathcal{C}_{9\mu}^{NP}$ deviation from SM is able to offer a common explanation for the deviations observed in $b \to s\,\ell^+\ell^-$ LFU ratios, $b \to s\,\mu\mu$ differential branching fractions and the $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis.

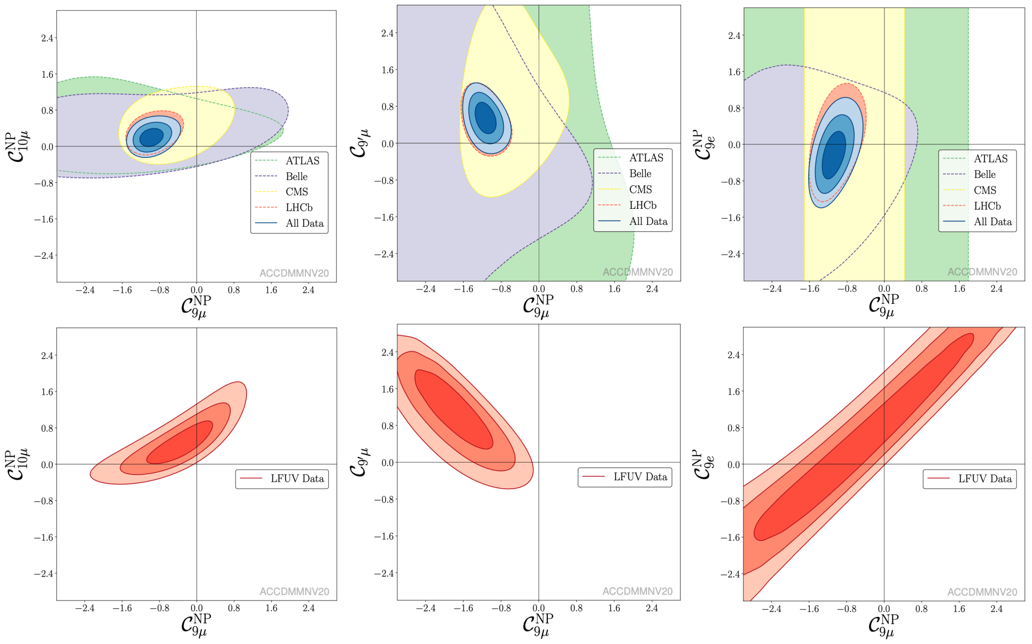


Figure 2.7 – From left to right: 2D fits to ($\mathcal{C}_{9\mu}^{NP}$,$\mathcal{C}_{10\mu}^{NP}$), ($\mathcal{C}_{9\mu}^{NP}$,$\mathcal{C}_{9\mu'}$) and ($\mathcal{C}_{9\mu}^{NP}$,$\mathcal{C}_{9e}^{NP}$) hypotheses using data from "All" $b \to s\,\ell^+\ell^-$ observables (top) or "LFUV" observables only (bottom). The allowed regions are bounded by $1, 2, 3\sigma$ contour lines. The "LFUV" fits deviate from SM at $2.9-3.7\sigma$ while "All" fits pulls, w.r.t. SM, are at $6.0-6.5\sigma$. Data from LHCb (dashed red) in the "All" fit currently dominates the global fit. Figures extracted from Ref (67).

Interestingly, there is a slight tension in the measurement of R_{K^*} in the lowest bin with other observables. Kinematically speaking, this bin is close to the $q^2 = 0$ photon pole and is mainly driven by $\mathcal{C}_{7,7'}$ coefficients which show no signs of NP contribution. The $\mathcal{C}_{9,9'}$ contributions, kinematically suppressed compared to the photon pole, can modify the experimental measurements of R_{K^*} in this bin (49; 47) but the central value is in tension with measurements at higher q^2. Furthermore, the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction measurement in this bin (76) lies within the range of SM prediction (see Fig. 2.8). As it stands current experimental uncertainties of R_{K^*} in this bin do not pose a problem for the global fits but an updated measurement will address this mild tension.

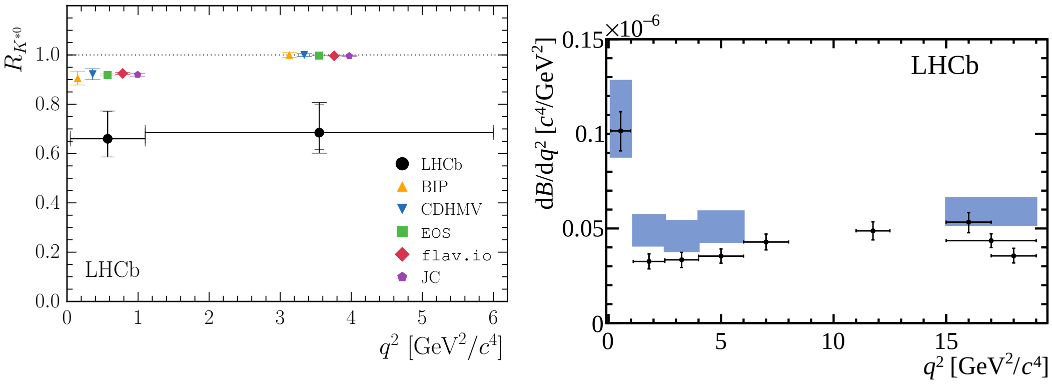


Figure 2.8 – LHCb measurements of R_{K^*} (left) and differential $\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)$. In this bin, the $B^0 \to K^{*0}\mu^+\mu^-$ branching fraction measurement agrees with SM prediction but central values of R_{K^*} measured by LHCb and Belle is less than unity. Figures extracted from Refs. (49) and (76) respectively.

2.5.2 New Physics Explanation

NP models that attempt to explain R_K and R_{K^*} anomalies also attempt to replicate other deviations in $b \to s\,\ell^+\ell^-$ decays, the angular observable P_5' for example. The popular models for the $b \to s\,\ell^+\ell^-$ anomalies revolve around leptoquarks, Z' and composite models. Theorists who build these NP models also evaluate their contributions to Wilson coefficients. Since model independent global fits favor $\mathcal{C}_{9\mu}^{NP}$ solutions and $\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{10\mu}^{NP}$ solutions preserve the $SU(2)_L \times U(1)_Y$ symmetry, we will focus our discussion on models that generate these scenarios. But first we touch upon models that are disfavored by current observables.

LFU Z' models (78) were developed to explain $b \to s\,\mu\mu$ anomalies (P_5' in particular) but these are strongly disfavored by R_K and R_{K^*}. These LFUV measurements also contradict four quark operator models that generates NP observables via charm loops (79). Models with right-handed currents (primed Wilson coefficients) are able to explain R_K (80; 81) but they predict $R_{K^*} > 1$, disfavoured by current measurements.

$\mathbf{\mathcal{C}_{9\mu}^{NP}}$ This pattern can be produced by Z' models that couple to leptons and avoid gauge anomalies via lepton number gauge symmetry. In this context, models that gauge $L_\mu - L_\tau$ numbers are popular because they do not affect electrons (82; 83; 84; 85). Even so, global fits (66; 67) also favor the $\mathcal{C}_{9\mu}^{NP} = -3\mathcal{C}_{9e}^{NP}$ scenario predicted by models that attempt to explain $b \to s\,\ell^+\ell^-$ anomalies and neutrino mixing (86). LQ models can generate a $\mathcal{C}_{9\mu}^{NP}$-like solution either with two scalar (an $SU(2)_L$ triplet and an $SU(2)_L$ doublet with $Y = 7/6$) or two vector representations (an $SU(2)_L$ singlet with $Y = 2/3$ and an $SU(2)_L$ doublet with $Y = 5/6$) (87).

$\mathbf{\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{10\mu}^{NP}}$ This scenario is popular as it preserves the $SU(2)_L \times U(1)_Y$ symmetry. This structure can arise from Z' models with loop-induced couplings (88) or heavy vector-like fermions (89; 90). LQs with a single representation (scalar $SU(2)_L$ triplet or vector $SU(2)_L$ singlet with $Y = 2/3$ (91; 92; 93; 94; 95; 96; 97)) can generate a $\mathcal{C}_{9\mu}^{NP} = -\mathcal{C}_{10\mu}^{NP}$ solution. Composite Higgs models (98) and models with loop contributions from three heavy new scalar and fermions (99; 100; 101) also predict this pattern.

2.5.3 Common Explanation with R_D and R_{D^*} Anomalies

The flavour changing charged current (FCCC) $b \to c\ell\bar{\nu}_\ell$ decays provide another suite of LFU tests, in the form of

$$R_D = \frac{\mathcal{B}(B \to D\tau\bar{\nu}_\tau)}{\mathcal{B}(B \to D\ell\bar{\nu}_\ell)} \tag{2.49}$$

$$R_{D^*} = \frac{\mathcal{B}(B \to D^*\tau\bar{\nu}_\tau)}{\mathcal{B}(B \to D^*\ell\bar{\nu}_\ell)} \tag{2.50}$$

measurements, where ℓ denotes both the electron and muon. The world average of this result by HFLAV (102) in Fig. 2.9 shows a combined $R_D - R_{D^*}$ deviation of 3.08σ from SM predictions.

With an analogous EFT approach, the effective hamiltonian can be written as

$$\mathcal{H}_{eff}(b \to c\ell\bar{\nu}_\ell) = \frac{4G_F}{\sqrt{2}} V_{cb} \sum_i \mathcal{C}_i \mathcal{O}_i, \tag{2.51}$$

and the dominant the long distance 4-fermion operator is

$$\mathcal{O}_{V\ell} = (\bar{c}\gamma_\mu P_L b)(\bar{\ell}\gamma^\mu P_L \nu_\ell). \tag{2.52}$$

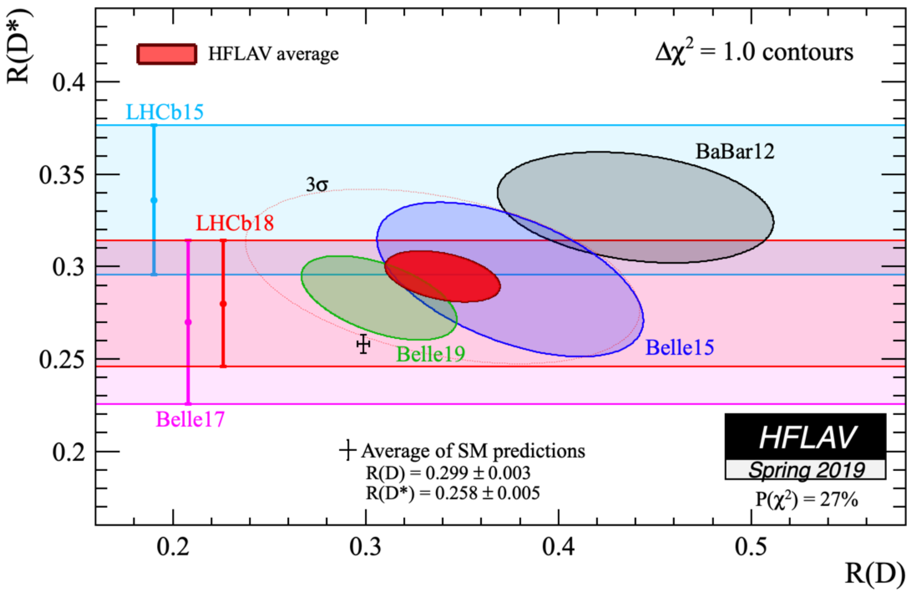


Figure 2.9 – The experimental results of R_D and R_{D^*} compared to their SM prediction. Figure courtesy of Ref. (102).

This operator can be unified with the $b\to s\,\ell^+\ell^-$ EFT operators (Eq. 2.41-2.43) via a general NP Lagrangian

$$\mathcal{L}_{NP} \propto \lambda^q_{il}\lambda^\ell_{\alpha\beta}[C_T(\bar{Q}^i_L\gamma_\mu\sigma^a Q^j_L)(\bar{L}^\alpha_L\gamma^\mu\sigma^a L^\beta_L) + C_S(\bar{Q}^i_L\gamma_\mu Q^j_L)(\bar{L}^\alpha_L\gamma^\mu L^\beta_L)] \qquad (2.53)$$

where $Q^i_L \sim (V^*_{ji}u^j_L, d_L)$ and $L^\alpha_L = (\nu^\alpha_L, \ell^\alpha_L)$ are the fermion doublets, C_S and C_T the singlet and triplet EFT couplings, and λ^q and λ^ℓ the flavour matrix couplings. The unified $(\bar{Q}^i_L\gamma_\mu Q^j_L)(\bar{L}^\alpha_L\gamma^\mu L^\beta_L)$ can now describe both FCNC and FCCC decays, and correlate the observations in $R_{D^{(*)}}$, $b\to s\,\ell^+\ell^-$ decays and $b\to s\,\nu\bar{\nu}$ decays. Since the $b \to c\ell\bar{\nu}_\ell$ SM decay is tree-level, a large contribution is needed to produce the experimental results of $R_{D^{(*)}}$. In turn, a large contribution is expected in $b \to s\,\nu\bar{\nu}$ decays, which will be probed by the upgraded Belle 2 experiment [5]. Given the semileptonic R measurements and by imposing $SU(2)_L$ invariance, a large enhancement is expected in $b\to s\,\tau\tau$ decays (103). Another interesting class of decays supported by this framework are lepton flavour violating decays, such as $B \to K\mu\tau$ decays which have been searched for by BaBar (104) and LHCb (105).

---

[5] Since there are 2 neutrinos in the final state, only the detectors in e^+e^- colliders can measured this decay as these experiment has a known initial state. LHCb, on the other hand, relies on hadron collisions whose initial state cannot be constrained.

The LFUV in semileptonic B decays require a NP $\mathcal{O}(10\ \%)$ enhancement of the tree-level SM contribution. On the other hand, $b \to s\,\ell^+\ell^-$ anomalies can be explained by a NP effect $\mathcal{O}(25\ \%)$ of the FCNC suppressed SM contributions. This suggests $\mathcal{O}(1)$ NP couplings to the τ, modest couplings to the μ and small or negligible couplings to the first generation.

2.6 Conclusion

The $SU(3) \times SU(2)_L \times U_Y$ invariant SM is the most comprehensible and precise theory available to particle physics. Despite that, evidence suggests that the SM is an effective theory at low energy, with NP physics theory waiting beyond the energy horizon. For example, popular new physics candidates attempts to provide a subatomic explanation for the dark matter and, for the more ambitious, dark energy observed at cosmological scales. Although such new physics models remain unobserved with direct methods, they can be probed with precision test of the SM.

In recent years, $b \to s\,\ell^+\ell^-$ decays have grown in popularity as one such probe. This is due to the deficiency in $b \to s\,\mu\mu$ differential branching fractions, the P_5' anomaly in $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis and deviations in LFU ratios with $b \to s\,\ell^+\ell^-$ decays. From an EFT perspective, the deviations in $b \to s\,\ell^+\ell^-$ decays agree with one another, and they unanimously support a NP contribution to $\mathcal{C}_{9\mu}^{NP}$. In fact, the global fits to $b \to s\,\ell^+\ell^-$ observables have reported a 6σ deviation from the SM. That being said, the tension of LFU measurements are at the $2 - 2.5\sigma$ level and they are statistically limited. This motivates an updated measurement of R_K and R_{K^*} with the full LHCb dataset, which will be presented in this thesis.

But first, the reasons measurements of such precision with $b \to s\,\ell^+\ell^-$ decays are possible in the last decade are presented in the following chapter. The precision of these measurements are supported by the copious amount of $b\bar{b}$ events produced by the LHC machine and the LHCb detector designed for precision tests of flavour physics.

2.7 Bibliography

[1] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, *Tenth-Order QED Contribution to the Electron g-2 and an Improved Value of the Fine Structure Constant*, Phys. Rev. Lett. **109** (2012) 111807, `arXiv:1205.5368`.

[2] B. C. Odom, D. Hanneke, B. D'Urso, and G. Gabrielse, *New Measurement of the Electron Magnetic Moment Using a One-Electron Quantum Cyclotron*, Phys. Rev. Lett. **97** (2006) 030801.

[3] C. MissMJ, *Pbs nova, fermilab, office of science, united states department of energy, particle data group*, 2019.

[4] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, Reading, USA, 1995.

[5] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, .

[6] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346.

[7] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343.

[8] M. Sozzi, *Discrete symmetries and CP violation: From experiment to theory*, Oxford University Press, 2008.

[9] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (1983) 1945.

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585.

[11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[12] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[13] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020), no. 8 083C01.

[14] ATLAS, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, `arXiv:1207.7214`.

[15] CMS, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, `arXiv:1207.7235`.

[16] V. C. Rubin, N. Thonnard, and J. Ford, W.$\tilde{K}$. *Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/*, Astrophys. J. **238** (1980) 471.

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

[18] Supernova Cosmology Project, S. Perlmutter *et al.*, *Discovery of a supernova explosion at half the age of the Universe and its cosmological implications*, Nature **391** (1998) 51, `arXiv:astro-ph/9712212`.

[19] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Sov. Phys. Usp. **34** (1991), no. 5 392.

[20] SNO, Q. R. Ahmad *et al.*, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301, `arXiv:nucl-ex/0204008`.

[21] P. Williams, *Naturalness, the autonomy of scales, and the 125 gev higgs*, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **51** (2015) 82.

[22] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, Prog. Theor. Phys. **28** (1962) 870.

[23] S. Bifani, S. Descotes-Genon, A. Romero Vidal, and M.-H. Schune, *Review of Lepton Universality tests in B decays*, J. Phys. G **46** (2019), no. 2 023001, `arXiv:1809.06229`.

[24] ATLAS Collaboration, *Test of the universality of τ and μ lepton couplings in W boson decays from $t\bar{t}$ events at 13 TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2020-014, CERN, Geneva, Jun, 2020.

[25] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group, S. Schael *et al.*, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, `arXiv:hep-ex/0509008`.

[26] ALEPH, DELPHI, L3, OPAL, LEP Electroweak, S. Schael *et al.*, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, Phys. Rept. **532** (2013) 119, `arXiv:1302.3415`.

[27] ATLAS, M. Aaboud *et al.*, *Precision measurement and interpretation of inclusive* W^+ *,* W^- *and* Z/γ^* *production cross sections with the ATLAS detector*, Eur. Phys. J. C **77** (2017), no. 6 367, `arXiv:1612.03016`.

[28] K. Lingel, T. Skwarnicki, and J. G. Smith, *Penguin decays of b-mesons*, Annual Review of Nuclear and Particle Science **48** (1998) 253–306.

[29] S. Descotes-Genon, T. Hurth, J. Matias, and J. Virto, *Optimizing the basis of* $B^0 \to K^{*0}ll$ *- observables in the full kinematic range*, Journal of High Energy Physics **2013** (2013) .

[30] E. Fermi, *Versuch einer Theorie der* β*-Strahlen. I*, Zeitschrift fur Physik **88** (1934) 161.

[31] K. G. Wilson and W. Zimmermann, *Operator product expansions and composite field operators in the general framework of quantum field theory*, Commun. Math. Phys. **24** (1972) 87.

[32] A. Bharucha, D. M. Straub, and R. Zwicky, $B \to Vl^+l^-$ *in the standard model from light-cone sum rules*, Journal of High Energy Physics **2016** (2016) .

[33] W. Detmold, C.-J. D. Lin, S. Meinel, and M. Wingate, $\Lambda_b \to \Lambda\ell^+\ell^-$ *form factors and differential branching fraction from lattice qcd*, Phys. Rev. D **87** (2013) 074502.

[34] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, *Lattice qcd calculation of form factors describing the rare decays* $b \to K^*\ell^+\ell^-$ *and* $B_s \to \phi\ell^+\ell^-$, Phys. Rev. D **89** (2014) 094501.

[35] C. Bouchard *et al.*, *Rare decay* $B \to Kl^+l^-$ *form factors from lattice qcd*, Physical Review D **88** (2013) .

[36] M. Beneke, T. Feldmann, and D. Seidel, *Systematic approach to exclusive* $B \to Vl^+l^-, V\gamma$ *decays*, Nuclear Physics B **612** (2001) 25–58.

[37] C. Bobeth, G. Hiller, and D. van Dyk, *The benefits of* $\bar{B} \to \bar{K}^* l^+ l^-$ *decays at low recoil*, Journal of High Energy Physics **2010** (2010) .

[38] B. Capdevila *et al.*, *Patterns of new physics in* $b \to sll$ *transitions in the light of recent data*, Journal of High Energy Physics **2018** (2018) .

[39] L.-S. Geng *et al.*, *Towards the discovery of new physics with lepton-universality ratios of* $b \to sll$ *decays*, Physical Review D **96** (2017) .

[40] W. Altmannshofer, C. Niehoff, P. Stangl, and D. M. Straub, *Status of the* $B \to K^* \mu^+ \mu^-$ *anomaly after moriond 2017*, The European Physical Journal C **77** (2017) .

[41] M. Bordone, G. Isidori, and A. Pattori, *On the standard model predictions for* R_K *and* R_{K^*}, The European Physical Journal C **76** (2016) .

[42] A. Abashian *et al.*, *The Belle Detector*, Nucl. Instrum. Meth. A **479** (2002) 117.

[43] BaBar, B. Aubert *et al.*, *The BABAR Detector: Upgrades, Operation and Performance*, Nucl. Instrum. Meth. A **729** (2013) 615, `arXiv:1305.3560`.

[44] Belle, J.-T. Wei *et al.*, *Measurement of the Differential Branching Fraction and Forward-Backword Asymmetry for* $B \to K^{(*)}\ell^+\ell^-$, Phys. Rev. Lett. **103** (2009) 171801, `arXiv:0904.0770`.

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

[46] Belle, A. Abdesselam *et al.*, *Test of lepton flavor universality in* $B \to K\ell^+\ell^-$ *decays*, `arXiv:1908.01848`.

[47] Belle, A. Abdesselam *et al.*, *Test of lepton flavor universality in* $B \to K^*\ell^+\ell^-$ *decays at Belle*, `arXiv:1904.02440`.

[48] LHCb, R. Aaij *et al.*, *Test of lepton universality using* $B^+ \to K^+\ell^+\ell^-$ *decays*, Phys. Rev. Lett. **113** (2014) 151601, `arXiv:1406.6482`.

[49] LHCb, R. Aaij *et al.*, *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, JHEP **08** (2017) 055, `arXiv:1705.05802`.

[50] LHCb, R. Aaij *et al.*, *Search for lepton-universality violation in* $B^+ \to K^+\ell^+\ell^-$ *decays*, Phys. Rev. Lett. **122** (2019), no. 19 191801, `arXiv:1903.09252`.

[51] LHCb, R. Aaij *et al.*, *Test of lepton universality with* $\Lambda_b^0 \to pK^-\ell^+\ell^-$ *decays*, JHEP **05** (2020) 040, `arXiv:1912.08139`.

[52] LHCb, R. Aaij *et al.*, *Measurement of CP asymmetries in the decays* $B^0 \to K^{*0}\mu^+\mu^-$ *and* $B^+ \to K^+\mu^+\mu^-$, JHEP **09** (2014) 177, `arXiv:1408.0978`.

[53] CKMfitter Group, J. Charles *et al.*, *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, Eur. Phys. J. C **41** (2005), no. 1 1, `arXiv:hep-ph/0406184`, Updated results and plots available at: http://ckmfitter.in2p3.fr.

[54] LHCb, R. Aaij *et al.*, *Angular analysis of the* $B^0 \to K^{*0}\mu^+\mu^-$ *decay using 3* fb^{-1} *of integrated luminosity*, JHEP **02** (2016) 104, `arXiv:1512.04442`.

[55] Belle, A. Abdesselam *et al.*, *Angular analysis of* $B^0 \to K^*(892)^0\ell^+\ell^-$, in *LHC Ski 2016: A First Discussion of 13 TeV Results*, 4, 2016. `arXiv:1604.04042`.

[56] Belle, S. Wehle *et al.*, *Lepton-Flavor-Dependent Angular Analysis of* $B \to K^*\ell^+\ell^-$, Phys. Rev. Lett. **118** (2017), no. 11 111801, `arXiv:1612.05014`.

[57] CMS, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay* $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ *in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Lett. B **781** (2018) 517, `arXiv:1710.02846`.

[58] ATLAS, M. Aaboud *et al.*, *Angular analysis of* $B_d^0 \to K^*\mu^+\mu^-$ *decays in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, JHEP **10** (2018) 047, `arXiv:1805.04000`.

[59] LHCb, R. Aaij *et al.*, *Measurement of CP-averaged observables in the* $B^0 \to K^{*0}\mu^+\mu^-$ *decay*, Phys. Rev. Lett. **125** (2020), no. 1 011802, `arXiv:2003.04831`.

[60] LHCb, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017), no. 19 191801, `arXiv:1703.05747`.

[61] ATLAS, M. Aaboud *et al.*, *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04** (2019) 098, `arXiv:1812.03017`.

[62] CMS, A. M. Sirunyan *et al.*, *Measurement of properties of $B_\mathrm{s}^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment*, JHEP **04** (2020) 188, `arXiv:1910.12127`.

[63] LHCb, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, JHEP **06** (2014) 133, `arXiv:1403.8044`.

[64] LHCb, R. Aaij *et al.*, *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, JHEP **09** (2015) 179, `arXiv:1506.08777`.

[65] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *Global analysis of $b \to s\ell\ell$ anomalies*, JHEP **06** (2016) 092, `arXiv:1510.04239`.

[66] B. Capdevila *et al.*, *Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data*, JHEP **01** (2018) 093, `arXiv:1704.05340`.

[67] M. Algueró *et al.*, *Emerging patterns of New Physics with and without Lepton Flavour Universal contributions*, Eur. Phys. J. C **79** (2019), no. 8 714, `arXiv:1903.09578`, [Addendum: Eur.Phys.J.C 80, 511 (2020)].

[68] W. Altmannshofer, C. Niehoff, and D. M. Straub, *$B_s \to \mu^+\mu^-$ as current and future probe of new physics*, JHEP **05** (2017) 076, `arXiv:1702.05498`.

[69] C. Bobeth *et al.*, *$B_{s,d} \to l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty*, Phys. Rev. Lett. **112** (2014) 101801, `arXiv:1311.0903`.

[70] M. Algueró *et al.*, *Are we overlooking lepton flavour universal new physics in $b \to s\ell\ell$?*, Phys. Rev. D **99** (2019), no. 7 075017, `arXiv:1809.08447`.

[71] W. Altmannshofer and D. M. Straub, *New physics in $b \to s$ transitions after LHC run 1*, Eur. Phys. J. C **75** (2015), no. 8 382, `arXiv:1411.3161`.

[72] M. Ciuchini *et al.*, *On Flavourful Easter eggs for New Physics hunger and Lepton Flavour Universality violation*, Eur. Phys. J. C **77** (2017), no. 10 688, `arXiv:1704.05447`.

[73] L.-S. Geng *et al.*, *Towards the discovery of new physics with lepton-universality ratios of $b \to s\ell\ell$ decays*, Phys. Rev. D **96** (2017), no. 9 093006, `arXiv:1704.05446`.

[74] T. Hurth, A. Arbey, F. Mahmoudi, and S. Neshatpour, *New global fits to $b \to s$ data with all relevant parameters*, Nucl. Part. Phys. Proc. **303-305** (2018) 2, `arXiv:1812.07602`.

[75] M. Algueró *et al.*, *What R_K and Q_5 can tell us about New Physics in $b \to s\ell\ell$ transitions?*, JHEP **07** (2019) 096, `arXiv:1902.04900`.

[76] LHCb, R. Aaij *et al.*, *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction*, JHEP **11** (2016) 047, `arXiv:1606.04731`, [Erratum: JHEP 04, 142 (2017)].

[77] LHCb, R. Aaij *et al.*, *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, JHEP **06** (2015) 115, `arXiv:1503.07138`, [Erratum: JHEP 09, 145 (2018)].

[78] R. Gauld, F. Goertz, and U. Haisch, *On minimal Z' explanations of the $B \to K^*\mu^+\mu^-$ anomaly*, Phys. Rev. D **89** (2014) 015005, `arXiv:1308.1959`.

[79] S. Jäger, M. Kirk, A. Lenz, and K. Leslie, *Charming new physics in rare B-decays and mixing?*, Phys. Rev. D **97** (2018), no. 1 015021, `arXiv:1701.09183`.

[80] D. Be ˇ cirević, S. Fajfer, N. Ko ˇ snik, and O. Sumensari, *Leptoquark model to explain the B-physics anomalies, R_K and R_D*, Phys. Rev. D **94** (2016), no. 11 115021, `arXiv:1608.08501`.

[81] P. Cox, A. Kusenko, O. Sumensari, and T. T. Yanagida, *SU(5) Unification with TeV-scale Leptoquarks*, JHEP **03** (2017) 035, `arXiv:1612.03923`.

[82] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, *Quark flavor transitions in $L_\mu - L_\tau$ models*, Phys. Rev. D **89** (2014) 095033, `arXiv:1403.1269`.

[83] A. Crivellin, G. D'Ambrosio, and J. Heeck, *Explaining $h \to \mu^{\pm}\tau^{\mp}$, $B \to K^*\mu^+\mu^-$ and $B \to K\mu^+\mu^-/B \to Ke^+e^-$ in a two-Higgs-doublet model with gauged $L_\mu - L_\tau$*, Phys. Rev. Lett. **114** (2015) 151801, `arXiv:1501.00993`.

[84] A. Crivellin, G. D'Ambrosio, and J. Heeck, *Addressing the LHC flavor anomalies with horizontal gauge symmetries*, Phys. Rev. D **91** (2015), no. 7 075006, `arXiv:1503.03477`.

[85] A. Crivellin, J. Fuentes-Martin, A. Greljo, and G. Isidori, *Lepton Flavor Non-Universality in B decays from Dynamical Yukawas*, Phys. Lett. B **766** (2017) 77, `arXiv:1611.02703`.

[86] D. Bhatia, S. Chakraborty, and A. Dighe, *Neutrino mixing and R_K anomaly in U(1)$_X$ models: a bottom-up approach*, JHEP **03** (2017) 117, `arXiv:1701.05825`.

[87] D. Buttazzo, A. Greljo, G. Isidori, and D. Marzocca, *B-physics anomalies: a guide to combined explanations*, JHEP **11** (2017) 044, `arXiv:1706.07808`.

[88] G. Bélanger, C. Delaunay, and S. Westhoff, *A Dark Matter Relic From Muon Anomalies*, Phys. Rev. D **92** (2015) 055021, `arXiv:1507.06660`.

[89] S. M. Boucenna *et al.*, *Non-abelian gauge extensions for B-decay anomalies*, Phys. Lett. B **760** (2016) 214, `arXiv:1604.03088`.

[90] S. M. Boucenna *et al.*, *Phenomenology of an $SU(2)\times SU(2)\times U(1)$ model with lepton-flavour non-universality*, JHEP **12** (2016) 059, `arXiv:1608.01349`.

[91] B. Gripaios, M. Nardecchia, and S. A. Renner, *Composite leptoquarks and anomalies in B-meson decays*, JHEP **05** (2015) 006, `arXiv:1412.1791`.

[92] S. Fajfer and N. Ko ˇ snik, *Vector leptoquark resolution of R_K and $R_{D^{(*)}}$ puzzles*, Phys. Lett. B **755** (2016) 270, `arXiv:1511.06024`.

[93] I. de Medeiros Varzielas and G. Hiller, *Clues for flavor from rare lepton and quark decays*, JHEP **06** (2015) 072, `arXiv:1503.01084`.

[94] R. Alonso, B. Grinstein, and J. Martin Camalich, *Lepton universality violation and lepton flavor conservation in B-meson decays*, JHEP **10** (2015) 184, `arXiv:1505.05164`.

[95] L. Calibbi, A. Crivellin, and T. Ota, *Effective Field Theory Approach to* $b \to s\ell\ell^{(\prime)}$*,* $B \to K^{(*)}\nu\overline{\nu}$ *and* $B \to D^{(*)}\tau\nu$ *with Third Generation Couplings*, Phys. Rev. Lett. **115** (2015) 181801, `arXiv:1506.02661`.

[96] R. Barbieri, G. Isidori, A. Pattori, and F. Senia, *Anomalies in* B*-decays and* $U(2)$ *flavour symmetry*, Eur. Phys. J. C **76** (2016), no. 2 67, `arXiv:1512.01560`.

[97] S. Sahoo, R. Mohanta, and A. K. Giri, *Explaining the* R_K *and* $R_{D^{(*)}}$ *anomalies with vector leptoquarks*, Phys. Rev. D **95** (2017), no. 3 035027, `arXiv:1609.04367`.

[98] C. Niehoff, P. Stangl, and D. M. Straub, *Violation of lepton flavour universality in composite Higgs models*, Phys. Lett. B **747** (2015) 182, `arXiv:1503.03865`.

[99] B. Gripaios, M. Nardecchia, and S. A. Renner, *Linear flavour violation and anomalies in B physics*, JHEP **06** (2016) 083, `arXiv:1509.05020`.

[100] P. Arnan, L. Hofer, F. Mescia, and A. Crivellin, *Loop effects of heavy new scalars and fermions in* $b \to s\mu^+\mu^-$, JHEP **04** (2017) 043, `arXiv:1608.07832`.

[101] F. Mahmoudi, S. Neshatpour, and J. Virto, $B \to K^*\mu^+\mu^-$ *optimised observables in the MSSM*, Eur. Phys. J. C **74** (2014), no. 6 2927, `arXiv:1401.2145`.

[102] HFLAV, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and* τ*-lepton properties as of 2018*, `arXiv:1909.12524`, updated results and plots available at `https://hflav.web.cern.ch/`.

[103] B. Capdevila *et al.*, *Searching for New Physics with* $b \to s\tau^+\tau^-$ *processes*, Phys. Rev. Lett. **120** (2018), no. 18 181802, `arXiv:1712.01919`.

[104] BaBar, J. P. Lees *et al.*, *A search for the decay modes* $B^{+-} \to h^{+-}\tau^{+-}l$, Phys. Rev. D **86** (2012) 012004, `arXiv:1204.2852`.

[105] LHCb, R. Aaij *et al.*, *Search for the lepton flavour violating decay* $B^+ \to K^+\mu^-\tau^+$ *using* B_{s2}^{*0} *decays*, JHEP **06** (2020) 129, `arXiv:2003.04352`.

Chapter 3

Beauty and the Machine: LHCb detector at LHC

The Large Hadron Collider (LHC) is hosted by *Organisation Européenne pour la Recherche Nucléaire* (CERN) [1] underneath the Swiss-French border near Geneva, and it currently holds the title for the longest and highest-energy particle collider. The LHC's initial primary goal was to produce and detect the Higgs boson, which was the final piece of the SM puzzle up till its discovery in 2012. Despite achieving its initial goal, the high luminosity delivered by the LHC is able to support precision measurements of the SM such as the exotic $t\bar{t}H^0$ production cross section (1) and Higgs boson branching fractions. The large amount of $b\bar{b}$ pairs produced by the proton-proton collision at the LHC underpin the sensitivity of $b \to s\,\ell^+\ell^-$ measurements performed by the LHC experimental collaborations and the world's first observation of the highly suppressed $B^0_s \to \mu^+\mu^-$ decay. The LHC accelerator also collides heavy ions, which have dedicated runs during the year to study the quark-gluon plasma from their collision.

The LHCb experiment is a single arm forward spectrometer dedicated to flavour physics, housed within the LHC accelerator complex to take advantage of LHC's large $b\bar{b}$ and $c\bar{c}$ production (Sec. 3.2.2). Being a flavour physics experiment, the LHCb experiment requires high performance vertexing, momentum reconstruction and particle identification capabilities. The vertex requirements is satisfied by a vertex locator around the proton-proton collision center while a magnet and tracking detectors after the vertex locator is able to measure the momentums

---

[1] CERN's original name was ***C**onseil **E**uropéen pour la **R**echerche **N**ucléaire* but after changing its name to ***O**rganisation **E**uropéenne pour la **R**echerche **N**ucléaire*, the abbreviation OERN does not sound as good.

of charge tracks (Sec. 3.3.1). The LHCb particle identification system relies on ring imaging Cherenkov detectors, calorimeters and muon stations to achieved to distinguish the charged particle species (Sec. 3.3.2). The readout of these detectors are triggered by a `L0` hardware trigger, and the events are further filtered by a software high level trigger for long term storage (Sec. 3.3.3). A novel, real time alignment and calibration was implemented in LHCb during Run 2, which is able to provide high quality detector calibration to the software trigger reconstruction that in Run 1, was only available offline.

3.1 The LHC collider

To understand LHC's primary physics goal we have to take a step back in time to the Large Electron-Positron Collider (LEP) era in the 1990s. Back then UA1 and UA2 at the Super Proton Synchrotron (SPS) proton-antiproton collider had detected the $W^{\pm}$ (2; 3) and Z (4; 5) bosons while the LEP detectors were running precision studies of electroweak bosons(6; 7). When the LHC was approved(8), two pieces of the Standard Model particle zoo were missing: the top quark and the Higgs boson. The CKM matrix predicted a third generation of quarks and the E288 experiment had detected the bottom quark (9). Discovery of the top quark was a matter of time (and collider energy). In fact about two months after the LHC's approval CDF (10) and D0 (11) announced the discovery of the heaviest SM quark. The final piece sought by LHC (12), the Higgs Boson, would conclusively prove the Higgs mechanism, so far a hypothetical particle to explain spontaneous symmetry breaking in the electroweak sector.

At the time of its proposal, the LHC was focused on the origin of mass (Higgs Boson discovery) but it was designed as an exploratory machine, with an objective to have sensitivity to mass scales up to 1 TeV. Due to the fact that protons are composite particles, this required an accelerator with significantly higher centre of mass energy and, equally important, much higher sustained instantaneous luminosity than the previous generations of colliders. The LHC design goal was set at 14 TeV centre of mass energy and $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ peak instantaneous luminosity. The former is achieved via an injection chain shown in Fig. 3.1. The latter limited LHC to a proton-proton collider since the highly inefficient antiproton production ruled out another proton-antiproton collider like the Tevatron, which was operating during LHC's proposal.

With the LHC approved, two general purpose detectors (GPD), ATLAS (14) and CMS (15), were built to detect and study the Higgs boson. As their name

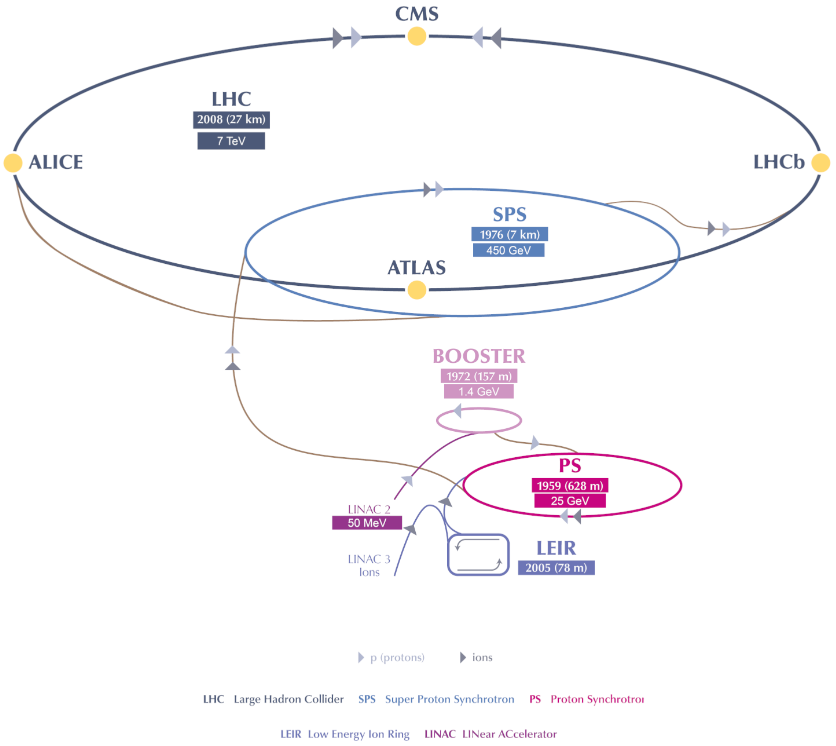


Figure 3.1 – The LHCb injection chain labeled with the maximum proton energy at each stage, edited from Ref. (13). The proton injection chain proceeds via Linac 2 — Booster — Proton Synchrotron (PS) — Super Proton Synchrotron (SPS) — LHC. Heavy ion injections are slightly different, using the Linac 3 — Low Energy Ion Ring (LEIR) as the first two stages before injection into the PS — SPS — LHC chain.

suggests, GPD detectors were also sensitive to a wide range of physics, notably SUSY searches, precision $W^{\pm}/Z$ studies and even B physics. A flavour experiment, the LHCb (16), was housed along the LHC to take advantage of the large $b\bar{b}$ production cross-section in LHC, a topic we will shortly discuss (Sec. 3.2.2). This detector is optimised in the forward region to fully exploit the B-physics potential of the LHC although its physics program diversified since its commissioning, adding areas such as charm physics and electroweak measurements in the forward region (complementary to those of ATLAS and CMS). The LHC was also designed to accelerate heavy lead ions up to 2.8 TeV per nucleon and achieve a peak luminosity of $\mathcal{L} = 10^{27}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. These heavy ion collisions are the main interest of a dedicated heavy ion experiment, A Large Ion Collider Experiment (ALICE) (17).

The LHC operation so far can be divied into two periods. Run 1 (18) covers 2010 to 2013 while Run 2 (19) lasts from 2015 to 2018. The Run 1 proton beam was commissioned at 3.5 TeV per beam until it was increased in 2012 to 4 TeV per beam, in order to increase the Higgs boson production (20). The peak instantaneous p-p collision achieved was $7.7 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ during this period. The beginning of Run 2 saw the beam energy bumped up to 6.5 TeV per proton beam. The LHC continuously tuned the accelerator in Run 2, reaching an instantaneous luminosity of $\mathcal{L} = 2.1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ in 2017, two times higher than its design luminosity. The annual integrated luminosity of LHC is shown in Fig. 3.2.

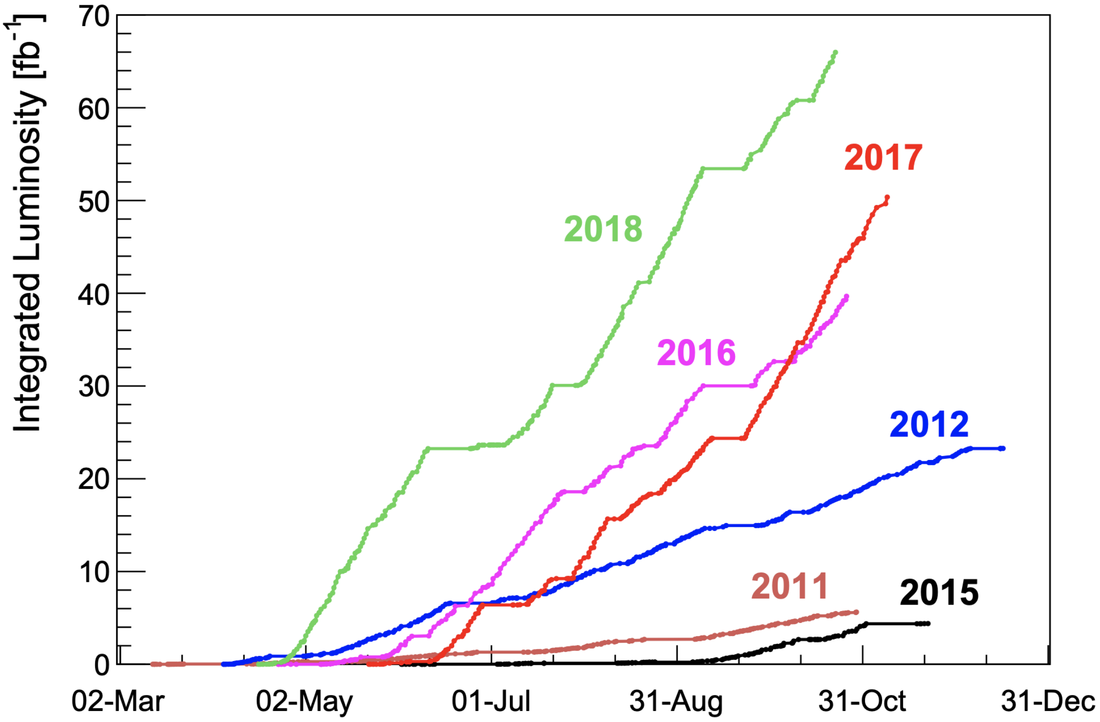


Figure 3.2 – Annual evolution of integrated luminosity during Run 1 and Run 2, extracted from Ref. (19). Notice how the first years in Run 1 (2011) and Run 2 (2015) had significantly smaller integrated luminosity than the rest of the run, due to the fact that first years of operation were dedicated to establishing machine performance.

During LHC operation, the accelerator complex repeats a cycle known as a "fill". A fill begins when a proton (or ion) beam is injected into the LHC injection chain. The protons are progressively accelerated in stages, reaching 450 GeV when they are injected in the LHC. In parallel to acceleration, the injection chain also congregates protons into bunches and bunch trains, and proton beams enter the LHC as bunch trains of 2808 proton bunches with 25 ns spacing, each bunch containing 1.15×10^{11} protons. Now injected with proton beams, LHC accelerates the protons to their nominal energies before adjusting the beam optics. Once LHC has achieved the desired beam parameters, the beams enter a “stable beam” state and proton-proton collisions begins at the various interaction points. The experiments now collect data until the beams have degraded significantly, which typically takes about 10 hours. At this point, the LHC dumps the beams and empties the LHC ring, thus completing a fill cycle.

3.2 b production at the LHCb interaction point

The precision of a particle physics measurement is limited by two general sources of uncertainty, systematic and statistical. Reducing the former depends on detector design and analysis technique but the latter is within the control of a collider, which can produce as many interesting physics event as possible. The production rate of a physics event in a collider is expressed as

$$N = \sigma \mathcal{L} \tag{3.1}$$

where $\mathcal{L}$ is the collider luminosity, σ the cross-section of the physics event and N the production rate of a physics event. Naturally, particle physics experiments want as much luminosity delivered as possible, albeit at a stable rate that does not saturate detector occupancy and overwhelm event reconstruction.

3.2.1 Instantaneous Luminosity at the LHCb

In the LHC, the instantaneous luminosity for head-on proton-proton collisions at an interaction region is

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta *} \cdot F \tag{3.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, $\beta*$ the beta function at the collision point and F the geometric luminosity reduction factor due to the crossing angle at the interaction point:

$$F = \big(1 + (\frac{\theta_c \sigma_z}{2\sigma *})^2\big)^{-1/2}. \tag{3.3}$$

θ_c is the full crossing angle at the OP, σ_z the RMS bunch length, and $\sigma*$ the transverse RMS beam size at the IP. This expression of F assumes round beams, $\sigma_z \ll \beta$, and equal beam parameters for both proton beams.

In the above equation, we can see that instantaneous luminosity decreases as beam intensity is lost to collisions, which is the main reason for luminosity decay in LHC. As a result, the interaction points at ATLAS and CMS have an instantaneous luminosity that decays over the length of a fill. In LHCb's case, it has a design luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, about two orders of magnitude lower than the peak luminosity at the LHC. This opens the possibility to keep the LHCb instantaneous luminosity constant by tuning the beam optics, with a technique known as luminosity leveling. The LHC levels the luminosity at LHCb via beam offset (21), where instead of colliding proton beams head-on, the proton beams are offset from their reference orbits, and the beams collide with a separation between the center of their transverse Gaussian profile. In the presence of offset luminosity levelling, the instantaneous luminosity equation in Eq. 3.2 is now modifed into

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta*} \cdot F \cdot e^{-\frac{(y_2 - y_1)^2}{4\sigma_y^2}} \tag{3.4}$$

where $y_{1,2}$ denotes the beam offset to their reference orbits and σ_y the Gaussian width of the beam along the offset, assuming equal beam parameters for both proton beams. As a fill progresses, the losses in beam density is mitigated by decreasing the offset term, $y_2 - y_1$, and a stable luminosity is achieved at the LHCb, which varies by less than 5% over the course of a fill in Run 2. For a comparison of luminosity evolution during a fill between ATLAS, CMS and LHCb, the reader is referred to Fig. 3.3.

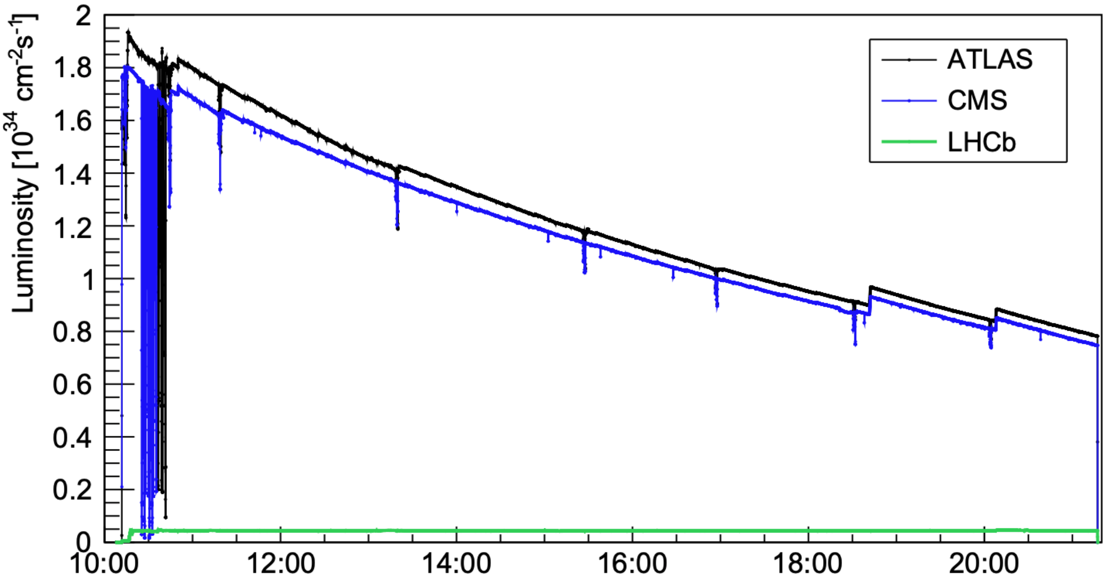


Figure 3.3 – Evolution of instantaneous luminosity at ATLAS, CMS and LHCb during a typical fill in 2018. The interaction point at LHCb maintains a consistent luminosity via offset luminosity leveling, as compared to the decaying luminosity at interaction points in ATLAS and CMS. Figure extracted from Ref. (19).

3.2.2 $b\bar{b}$ cross-section at the LHC

In addition to high luminosity, the LHC *pp* collisions have a high *b* production cross-section, which makes it a suitable collider to study *B*-physics. Fig. 3.4 plots the production cross-section for various processes in hadron colliders, showing that *b* production is orders of magnitude larger than corresponding electroweak or Higgs production in the LHC. Another fact to note is the *b* production cross-section increases with collider energy, and it approximately doubles between 7 TeV in Run 1 and 13 TeV in Run 2.

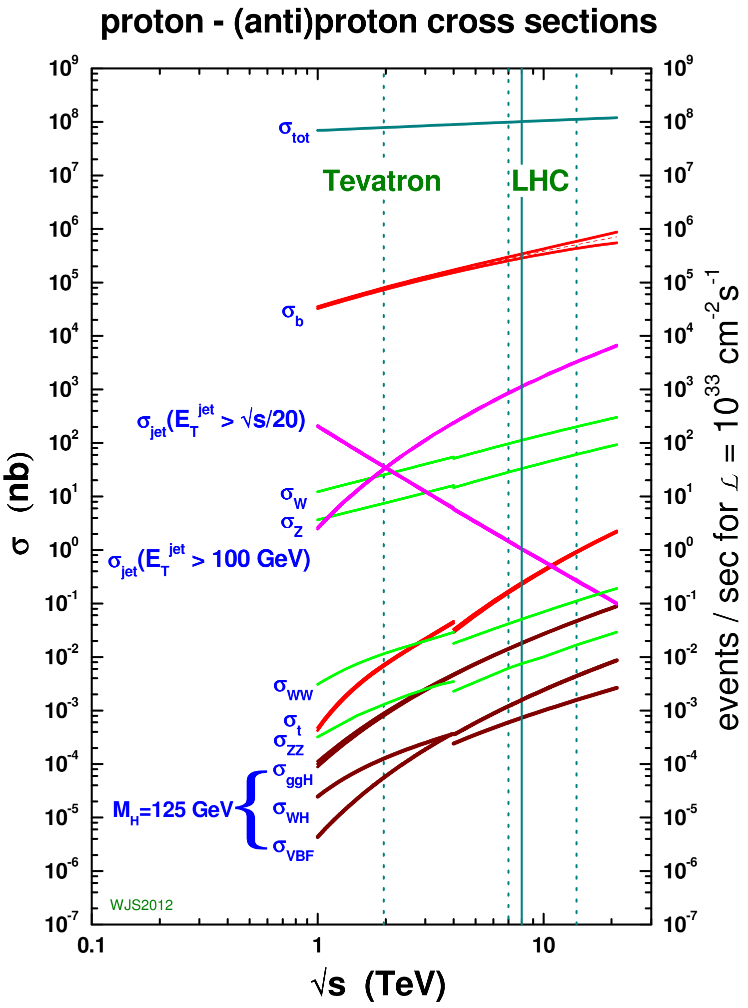
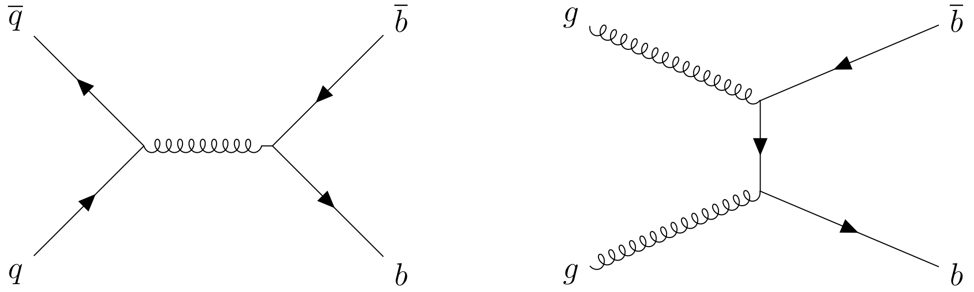
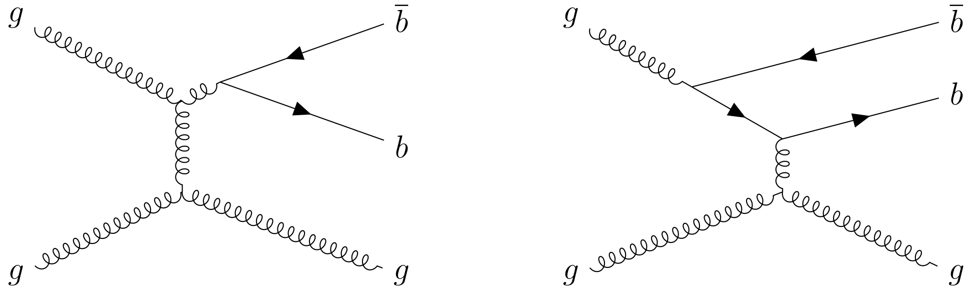


Figure 3.4 – The production cross-sections of heavy quarks, jets, electroweak bosons and Higgs boson at hadron colliders. At LHC energies, σ_b is much higher than σ_W, σ_Z or σ_H. The discontinuity at 4 TeV is due to the switch from proton-antiproton to proton-proton collisions at that energy. Figure extracted from Ref. (22)



(a) $q\overline{q}$ annihilation. (b) Gluon-gluon fusion.



(c) Gluon splitting. (d) Flavour excitation.

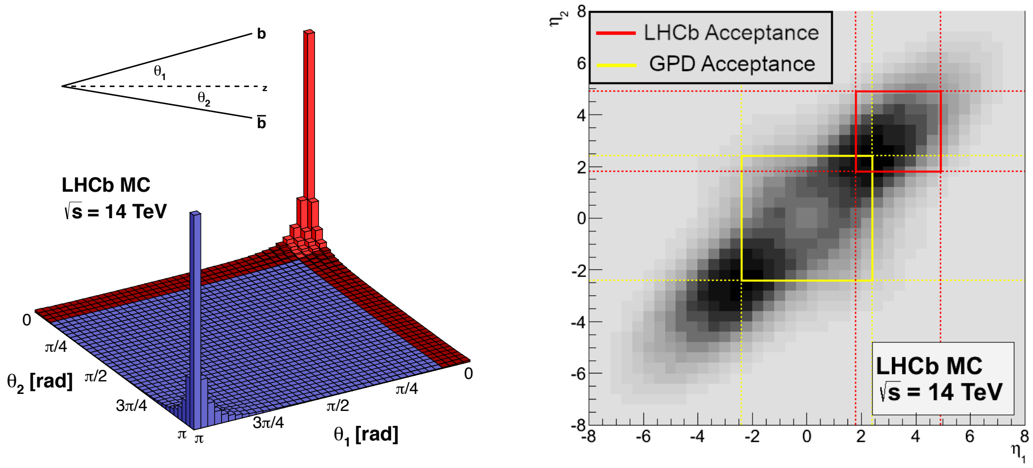
Figure 3.5 – The leading Feynman diagrams for $b\overline{b}$ production in hadron collisions.

Hadron colliders produce $b\overline{b}$ pairs via four main channels: $q\overline{q}$ annihilation, gluon fusion, gluon splitting and flavour excitation, shown in Fig. 3.5. Besides an energy dependence, these production mechanisms have an angular dependence which favours a direction along the p beam axis, shown in Fig. 3.6. Furthermore, the $b\overline{b}$ pairs are correlated and they are produced collinearly. This is why LHCb is instrumented in the forward region, where despite an angular coverage of approximately 0.1π, the LHCb acceptance covers about 27% all b produced in pp collisions (23; 24).

The total production cross-section in the LHCb acceptance has been measured experimentally (26),

$$\mathcal{L} \quad = \begin{cases} 72.0 \pm 0.02 \pm 0.26\,\mu\text{b}, & 7\,\text{TeV} \\ 144 \pm 1 \pm 21\,\mu\text{b}, & 13\,\text{TeV} \end{cases} \tag{3.5}$$

where the first uncertainty is statistical and the second systematic. Multiplying this with the Run 1 and Run 2 integrated luminosity at LHCb, approximately one trillion $b\overline{b}$ events have been produced in the LHCb acceptance. This has given LHCb a statistical advantage over measurements by Belle and BaBar especially



[b] [b]

Figure 3.6 – $b\bar{b}$ production angles at the LHC. The z-axis is defined along the LHC proton beams. On the left, the $b\bar{b}$ produced are highly correlated, and they favour directions parallel or antiparallel to the beam axis. On the right, the $b\bar{b}$ production are concentrated around the LHCb acceptance and the GPD acceptance is shown as a comparison. Figures extracted from Ref. (25).

in precision measurements of rare decays, such as the lepton flavour universality measurements in $b \to s\,\ell^+\ell^-$ decays we discussed in the Theory chapter (Chap. 2.4), where LHCb measurements dominate world averages due to its statistical reach.

3.3 The LHCb Detector

The LHCb detector is a single-arm spectrometer optimised to reconstruct the vast amount of $b\bar{b}$ produced in the forward region. The LHCb dipole magnet has an integrated field of roughly 4 Tm to deflect charged particles in the horizontal plane. Its acceptance in the bending plane is 15 mrad to 300 mrad while the non-bending plane detects particles within the 15 mrad to 250 mrad region, corresponding to a pseudorapidity acceptance of $2 < \eta < 5$. The LHCb detector cross-section is shown in Fig. 3.7.

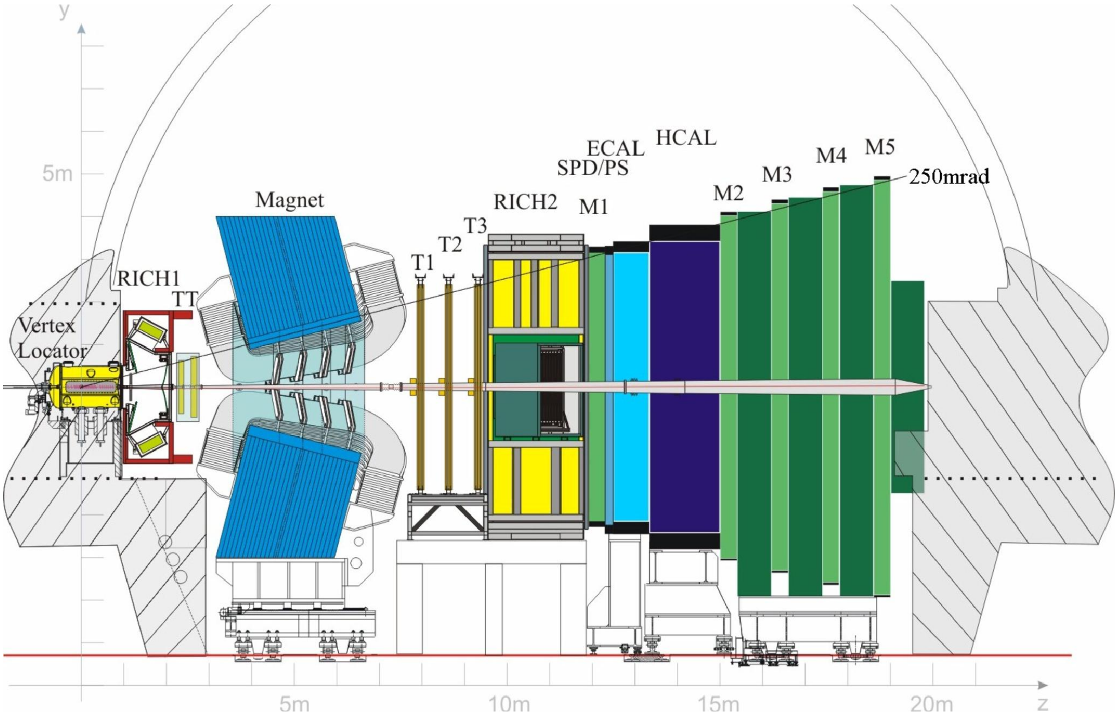


Figure 3.7 – The LHCb detector (27). The "forward" region points from the interaction region into the LHCb acceptance while "backward" region lies in the opposite direction. The "downstream"("upstream") direction refers to the positive(negative) z direction.

3.3.1 Tracking System

Tracking Subdetectors

The *Vertex Locator* (VELO) subsystem consists of interleaved $r-\phi$ silicon strip detectors located around the interaction region, outside the magnetic field such that particles in the VELO travel in straight lines. The VELO measures the position of proton-proton collisions, known as primary vertex (PV), and the distance of closest approach between track-vertex pairs, more commonly known as impact parameter (IP). Tracks with small IP are classified as direct products from pp collisions while tracks with large IP are likely decay products from particles of interest, such as strange, charmed or beauty hadrons, whose lifetimes and boosts are large enough to decay a measurable distance from their PV.

The VELO detector contains 42 silicon strip modules, and has an inter-strip pitch ranging from 40 µm to 100 µm from the innermost to outermost region, shown in Fig. 3.8. Each pair of modules make up a VELO station, for a total of 21 VELO station arranged perpendicularly to the beam axis. Two additional stations, known as the pile-up system, only contain R sensors and are placed in the most upstream location. The innermost strips are only 7mm away from the beam center, which is smaller than the aperture allowed during LHC beam injection. In order to avoid radiation damage during LHC beam injection, the 21 stations are arranged into two retractable VELO halves. The VELO stations are retracted during the injection stage and waits for LHC to enter “stable beam” state, such that the beam widths are small enough for the VELO to close around the interaction region. Fig. 3.9 shows the VELO station positions and a schematic of two retracable VELO halves.

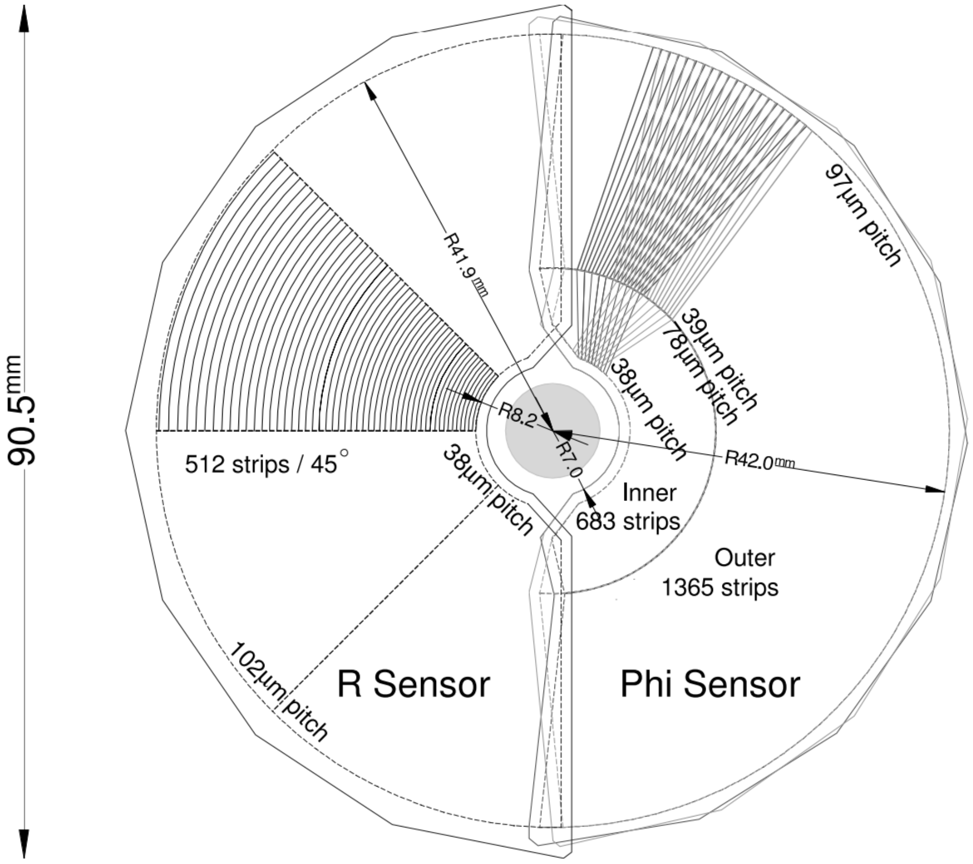


Figure 3.8 – The $r - \phi$ geometry of VELO sensors. For clarity, only a portion of the strips are produced. Figure extracted from Ref. (28).

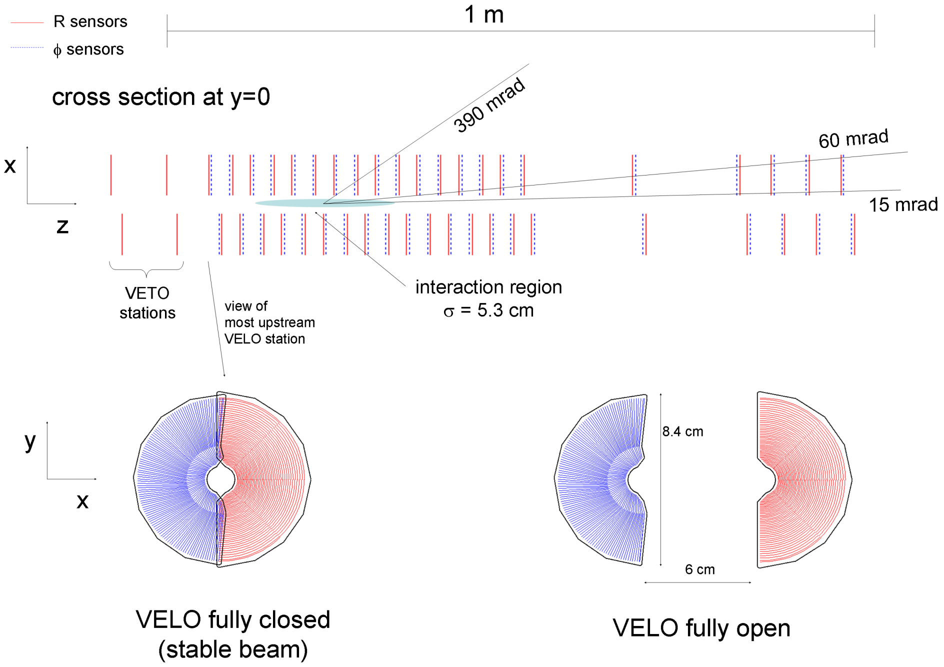


Figure 3.9 – Top: Arrangement of VELO modules along the beam axis. Bottom: VELO halves are closed once stable beam is achieved and opened outside of data-taking to reduce radiation damage. Figure extracted from Ref. (28).

Three tracking stations (T1-T3, collectively known as T-stations) are placed downstream of the LHCb magnet to measure particle momentum. The T-station regions close to the beam axis, known as the *Inner Tracker* (IT), are made up of silicon strips while straw-tubes occupy the regions further out, known as the *Outer Tracker*, OT. Fig. 3.10 shows the geometry of a T-station and the dimensions of the IT module. The IT only covers a small fraction of T-station surface area, about 1.3%. However, approximately 20% of all charged tracks from the interaction region pass through the IT, which is the reason silicon strips are used in this region to achieve higher spatial resolution. Another silicon strip tracking station, the *Tracker Turicensis* (TT), is placed before the magnet. The TT, IT and OT are made up of four detector layers in each station, where the first and the last vertical layers sandwich two ±5° stereo layers in between, giving the tracking detectors two-dimensional spatial information. This four-layer scheme is shown in Fig. 3.11.

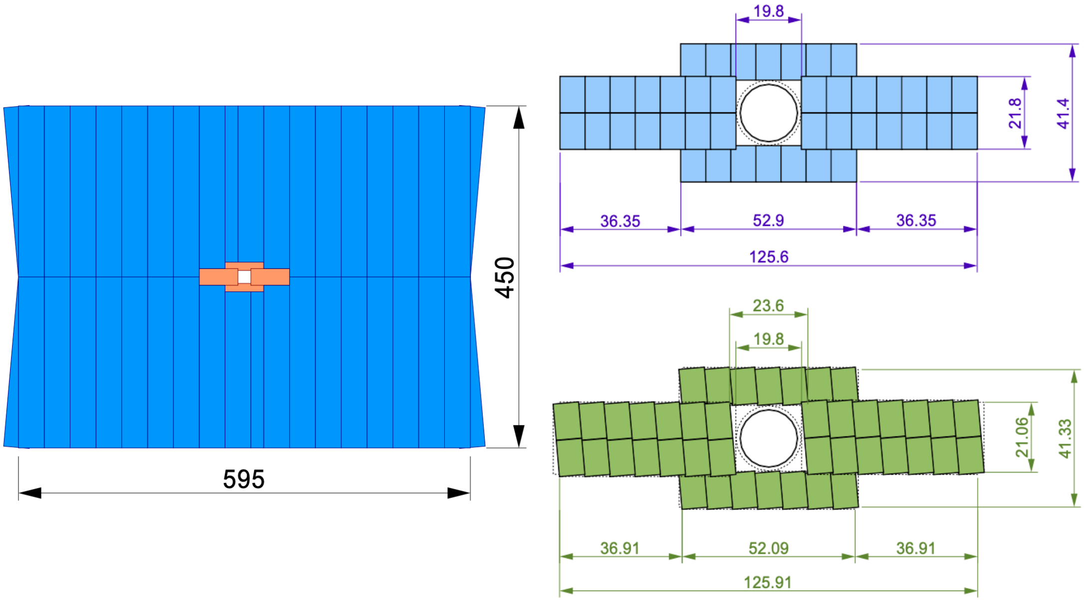


Figure 3.10 – Left: a LHCb T-station viewed from the front, drawn to scale. The four IT boxes surrounds the beam axis (orange) and the OT covers the rest of the T-station (blue). Right: dimensions (in cm) of the four IT boxes surrounding the LHC beampipe, for a vertical layer (top) and a stereo layer (bottom). Figure extracted from Ref. (29).

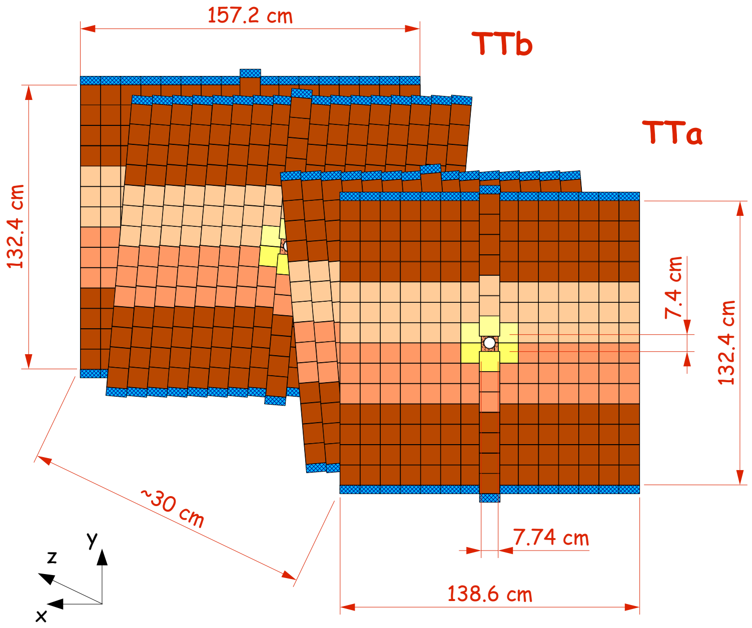


Figure 3.11 – Dimensions of the TT. The TT, IT and OT use the same four-layer arrangement shown here, with two vertical layers at the front and back while the two center layers are rotated by ±5° w.r.t. the vertical layers. Figure extracted from Ref. (30).

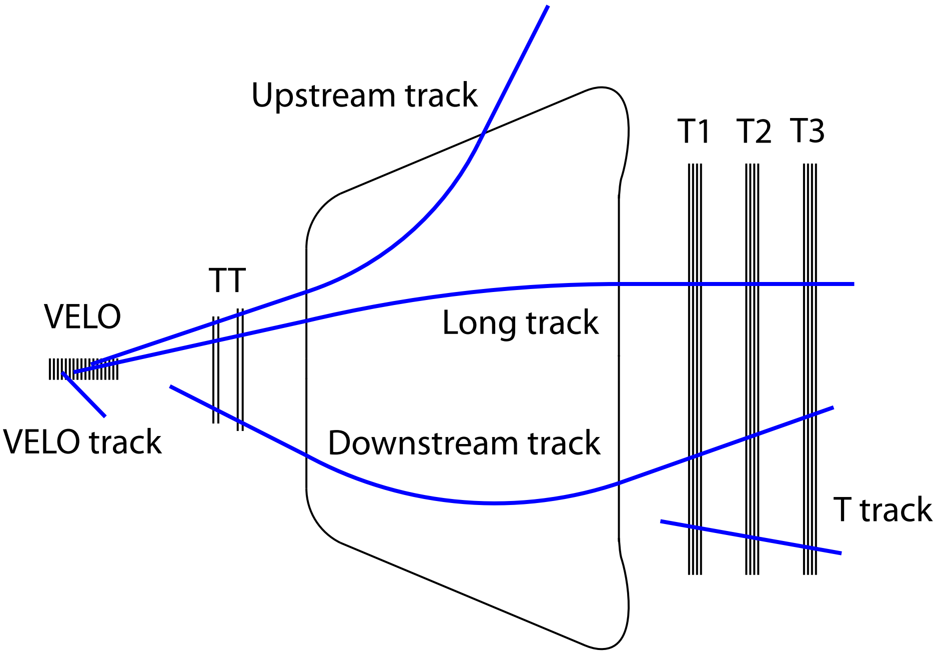


Figure 3.12 – The types of track that are reconsturcted in LHCb, extracted from Ref. (31).

Track Reconstruction

The LHCb tracking software consists of various algorithms broadly classified into two subclass, the pattern recognition algorithms and the track fitting algorithms. The pattern recognition algorithms build track candidates from raw subdetector signals. The track fit then applies a Kalman filter on these track candidates to estimate the track parameters.

The categories of charged tracks in LHCb, built by the pattern recognition algorithms, are illustrated in Fig. 3.12. The most important of these are the ‘long’ tracks. Long tracks contain a VELO track segment and associated T-station hits. This gives long tracks high resolution vertex and momentum information, critical for the majority of LHCb analysis which reconstructs b or c hadron decays. The second-most important are ‘downstream’ tracks. Downstream tracks are found by extrapolating T-tracks to hits in the TT. These tracks are particularly important for long-lived neutral particles, such as Λ and K^0_{S}, that decays without leaving VELO station signatures.

The upstream and T-tracks play minor roles in LHCb. Upstream tracks are made from VELO tracks matched to hits in the TT, but are swept out from the LHCb acceptance by the magnetic field. Because the TT sits in the fringe of the LHCb magnetic dipole, upstream tracks have a charge estimate and a momentum resolution that is slightly worse than 10%. However, this momentum resolution makes it diffucult to use upstream tracks for precision measurements. T-tracks have an even worse momentum resolution, about 25%. Despite this, T-tracks provide vital information for electron-photon separation in *Electromagnetic Calorimeter* (ECAL) clusters. Finally, the VELO tracks propagate outside the LHCb magnetic field, which simplifies their pattern recognition algorithms but makes their momentum impossible to determine. The VELO tracks are critical for PV reconstruction, especially since the VELO is the only subdetector that detects track in the backward region. Furthermore, VELO tracks are used in track isolation algorithms, whose purpose is to suppress partially-reconstructed backgrounds in offline physics analysis.

The track fits, besides providing a measurement of track parameters, also estimate an associated χ^2 of the Kalman fit. The track fit χ^2 is vital for rejecting fake track reconstructed by the pattern recognition algorithms. A dedicated neural network provides additional fake track supression, which uses the Kalman fit χ^2 and information from all tracking subdetectors to assign each track a 'ghost probability'. The typical performance of the LHCb fake track rejection algorithm is shown in Fig. 3.13.

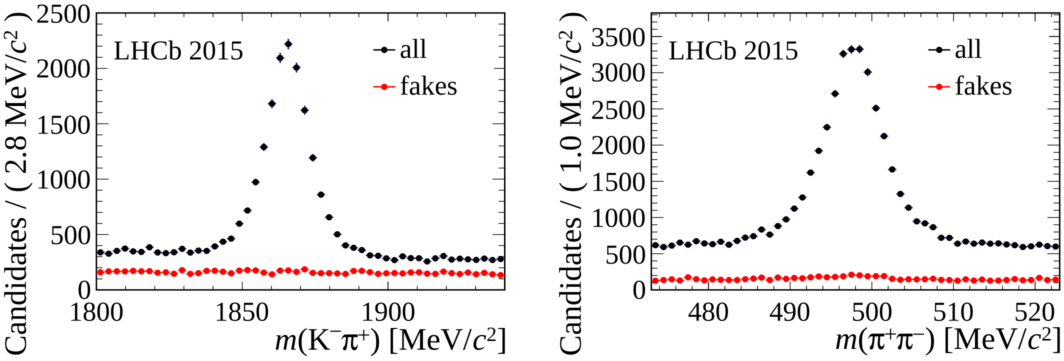


Figure 3.13 – LHCb fake track classifier performance, extracted from Ref. (32).

LHCb Tracking Performance

LHCb has published dedicated papers on tracking performance, such as Ref. (32). Nevertheless, we will present some important tracking performance metrics here. The LHCb tracking performance in $J/\psi \to \mu^+\mu^-$ decays from b hadrons were extensively studied by Ref. (33). This study looked at long track reconstruction efficiency and long track momentum resolution, shown in Fig. 3.14, and how it affects LHCb's mass resolution, shown in Fig. 3.15. In these plots, the LHCb long track reconstruction achieves a 0.5% momentum resolution when the momentum is below 20 GeV, and slowly degrades to about 1% for particles around 150 GeV. The long track reconstruction efficiency exceeds 95% except for tracks in the $p <$ 5 GeV momentum region. These result in a mass resolution of about 12.5 MeV in $J/\psi \to \mu^+\mu^-$ decays.

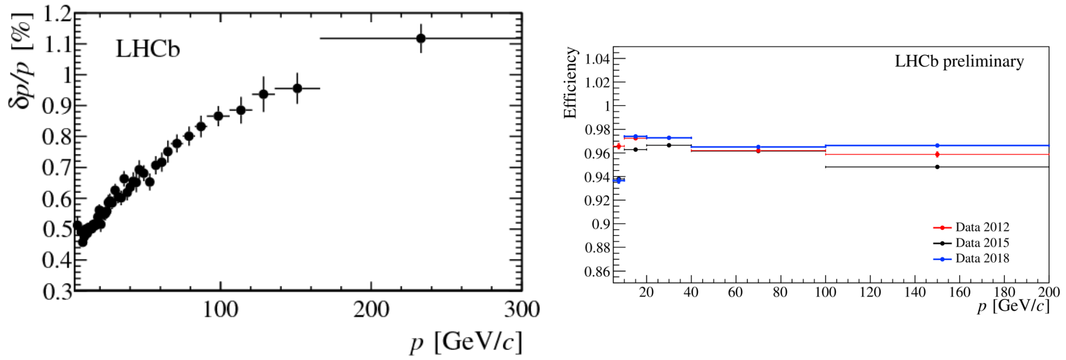


Figure 3.14 – LHCb momentum resolution (left) and the long track reconstruction effiency (right) as a function of the particle momentum, extracted from Ref. (34) and Ref. (35).

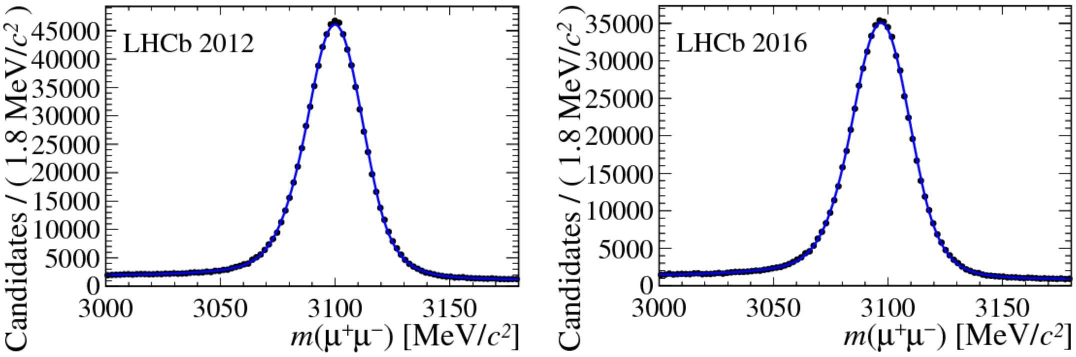


Figure 3.15 – LHCb mass resolution, extracted from Ref. (32). The LHCb mass resolution at the J/ψ mass is 12.4 MeV in 2012 (left) and 12.7 MeV in 2016.

LHCb's impact parameter resolution is shown in Fig. 3.16 and primary vertex resolution in Fig. 3.17, both of which strongly affect the decay time resolution shown in Fig. 3.18. As we can see, LHCb has a base IP resolution of 12 µm that increases by 24 µm per $\text{GeV}^{-1}c$ of $1/p_{\mathrm{T}}$. LHCb PV resolution strongly depends on how many tracks are used to reconstruct the PV, which is roughly 40(290) µm along the $x(z)$ direction with 5 tracks and decreases asymtotically to 9(50) µm once 50 tracks are associated to a PV. Together, the IP resolution and PV resolution translate into a 40 − 50 fs decay time resolution in $B_s^0 \to J/\psi\,\phi$ decays.

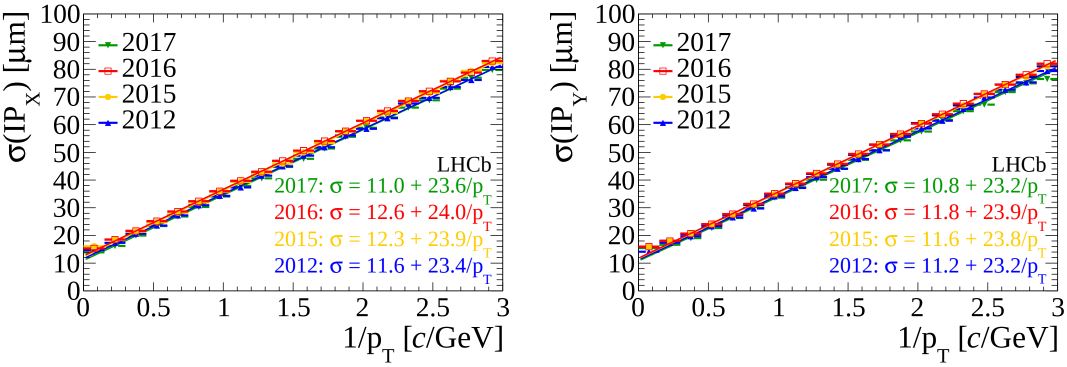


Figure 3.16 – LHCb IP resolution along the x(left) and y(right) direction as a function of $1/p_T$, extracted from Ref. (32).

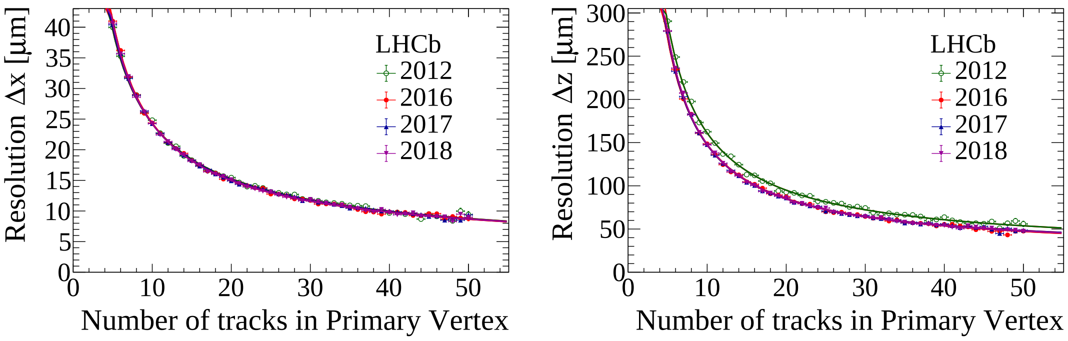


Figure 3.17 – LHCb PV resolution along the x(left) and z(right) direction as a function of the number of tracks in the PV, extracted from Ref. (32).

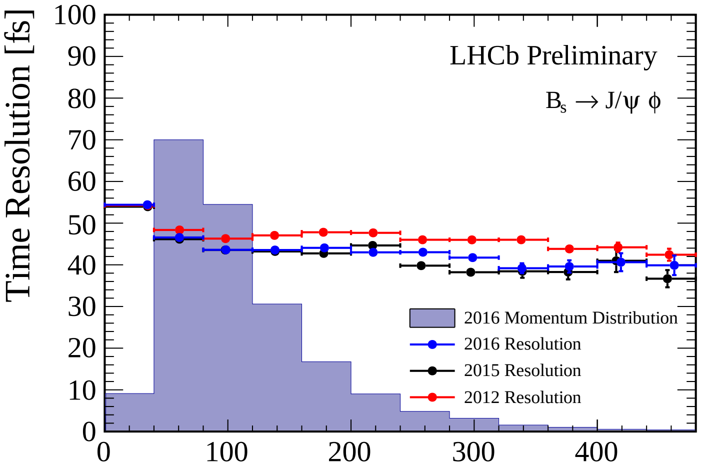


Figure 3.18 – LHCb decay time resolution in $B_s^0 \to J/\psi\,\phi$ decays as a function of momentum, extracted from Ref. (32).

3.3.2 Particle Identification System

Complementing LHCb's tracking system is the particle identification (PID) system that assigns a particle species to a charged track, and detects the presence of photons. The stable charged particles within LHCb acceptance are pions, kaons, protons, electrons and muons.

Particle Identification Subdetectors

Muon identification is provided by the muon system, which consists of five rectangularly shaped stations. Four of these sit downstream of the calorimeter (M2-M5) while one (M1) sits upstream of the calorimeter system in order to improve the p_{T} resolution of the muon system [2]. The muon stations are made up of multiwire proportional chambers except for the inner region of M1, which uses a more radiation hard triple gas electron multiplier design to cope with the increased particle flux. The M1-M3 stations have high spatial resolution along the bending plane to achieve a muon p_{T} resolution of 20% while the main purpose of M4-M5 is to identify highly penetrating particles. Each muon station contains four regions, R1 to R4 with increasing distance from the beampipe. The segmentation ensures each region experiences similar particle fluxes and to cope with the multiplicity increase, the spatial resolution of each region increases towards the beampipe. To illustrate this point, a front view of the M1 station is shown if Fig. 3.19.

Electron, photons and hadrons are distinguished by the LHCb calorimeter system, situated downstream of RICH2 and the first muon station. The calorimeter system consists of a *Scintillator Pad Detector* (SPD), a *PreShower* (PS), an *Electromagnetic Calorimeter* (ECAL) and a *Hadronic Calorimeter* (HCAL). Structurally, the SPD and PS are walls of scintillator pads. The ECAL cells are arranged as alternating layers of lead and scintillator arranged in a shashlik structure. The HCAL has alternating iron and scintillating tiles arranged in a sampling structure, where the tile plane is parallel to the beam axis. The ECAL and HCAL cell placements are drawn in Fig. 3.20.

---

[2] The muon system can approximate the p_{T} of a muon, independently of the tracking system measurements. We shall delve deeper into this in the trigger section.

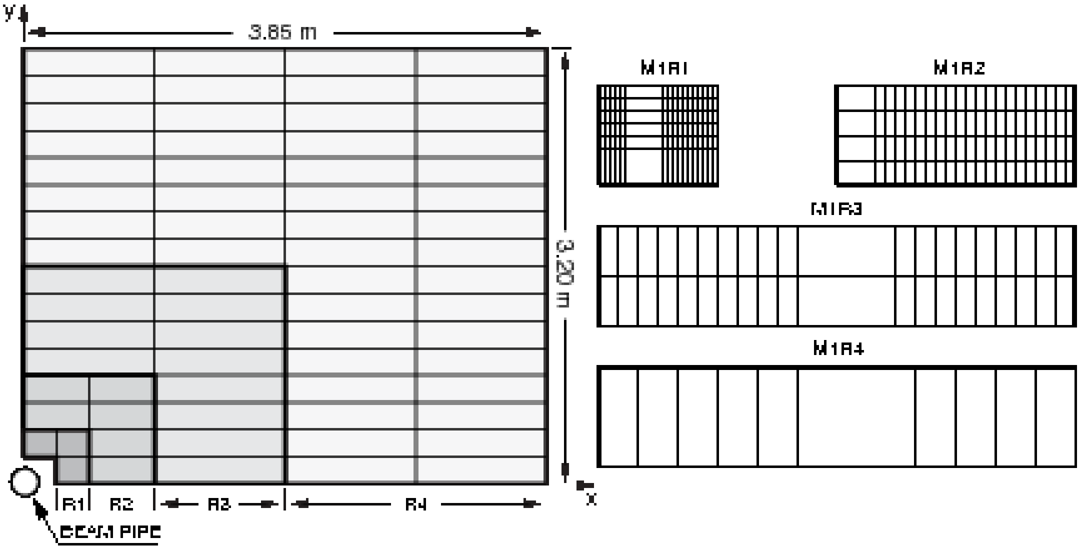


Figure 3.19 – Left: a quadrant of the M1 station. Each rectangle represents one chamber. Right: the division of chamber into logical pads in each muon station region. The number of logical pad columns in each chamber in muon stations M1:M2-M3:M4-M5 follows a ratio 2:4:1. The number of logical pad rows per chamber is the same between muon stations. Figure extracted from Ref. (36).

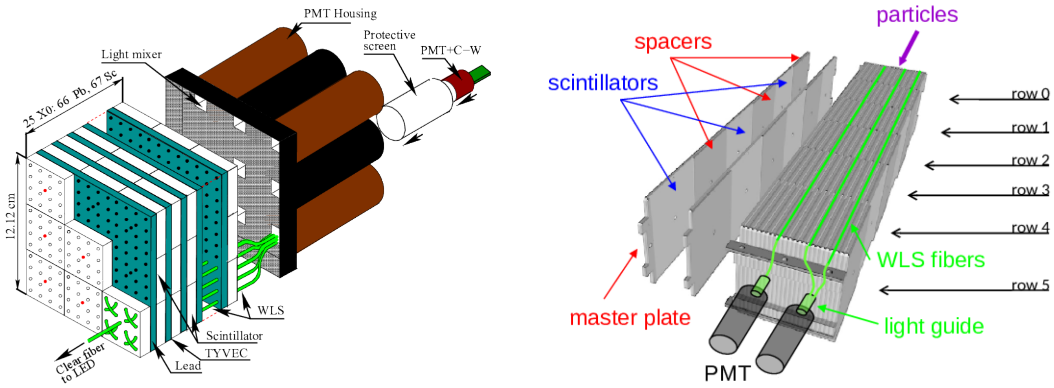


Figure 3.20 – The ECAL (left) and HCAL (right) cell arrangement, extracted from Ref. (37). The ECAL is arranged in a shashlik structure while HCAL lead and scintillating tile planes are arranged parallel to the beam axis.

Similar to the T-stations and muon stations, the calorimeter occupancy peaks in the region surrounding the beampipe and decreases as we move outwards. Thus, the ECAL (HCAL) calorimeter system has a three (two) different regions with the highest granularity in the middle, and the granularity decreases as we move further away from the beam axis. This region dependent granularity is illustrated in Fig. 3.21 for the ECAL and HCAL.

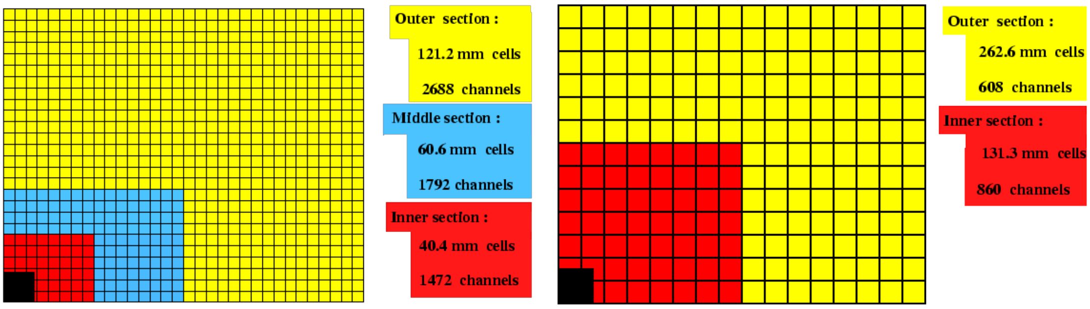


Figure 3.21 – The SPD, PS and ECAL cell dimensions (left) and HCAL cell dimension (right), extracted from Ref. (37).

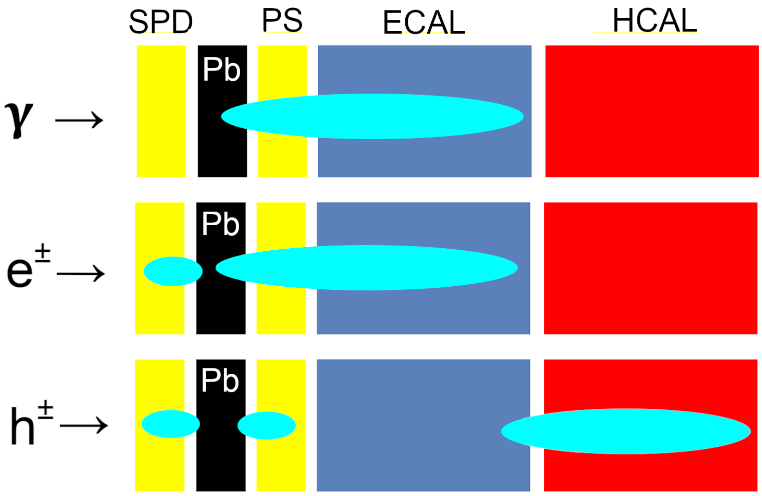


Figure 3.22 – Schematic showing the photon, electron and hadron shower patterns in the calorimeter system.

Electron, charged hadron and photon interactions in the LHCb calorimeter are illustrated in Fig. 3.22. Charged particles leave a track within the SPD, useful to discriminate neutral photons from electrons. The particles then arrive at the lead plate between the SPD and PS, designed to initiate particle showers in the ECAL. Electrons and photons are then detected as showers in the ECAL. Charged hadrons showers begins towards the end of the ECAL and are detected in the HCAL. The ECAL and HCAL energy resolution are $\frac{\sigma_E^{ECAL}}{E} = \frac{10\%}{\sqrt{E}} \oplus 1\%$ and $\frac{\sigma_E^{HCAL}}{E} = \frac{65\%}{\sqrt{E}} \oplus 9\%$ respectively, where the E is given in GeV. This resolution means that the HCAL is mainly used as a hardware trigger, whereas the ECAL resolution is vital for reconstructing photons and neutral pions used by offline analyses. In addition, the ECAL plays an important role for recovering Bremsstrahlung photons emitted by electrons due to material interaction in the VELO station.

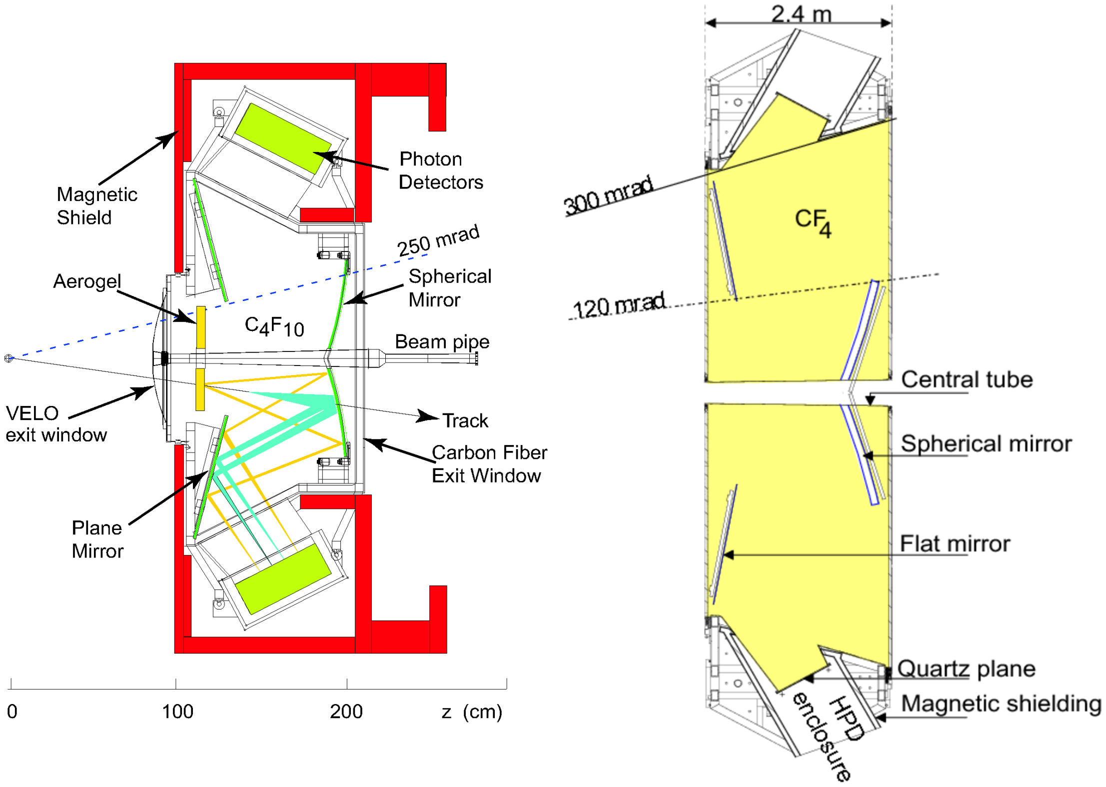


Figure 3.23 – Schematic of the RICH1 (left) and RICH2 (right) detectors, extracted from Ref. (16).

The LHCb's two Ring Imaging Cherenkov (RICH) detectors are shown in Fig. 3.23. The RICH1 and RICH2 detectors vital for kaon-pion-proton separation in LHCb. RICH1 sits immediately after the VELO detector and targets particles in the low-intermediate $2-40\,\mathrm{GeV}/c$ momentum range over the full $25-300\,\mathrm{mrad}$ LHCb acceptance. RICH2 sits downstream of the LHCb magnet and covers the

$15 - 100\,\mathrm{GeV}/c$ high momentum region over a reduced $15 - 120\,\mathrm{mrad}$ angular acceptance, a cost-cutting measure justified by the fact that high momentum particles tend to have larger boost and therefore larger η. RICH1 and RICH2 both use fluorocarbon gases as Cherenkov radiators, C_4F_{10} and CF_4 respectively. The Cherenkov photons are focused by a primary spherical mirror and a secondary flat mirror onto a matrix of Hybrid Photon Detectors (HPD) enclosed by magnetic shields, in order to attenuate the LHCb magnetic field into a < 3 mT regime where the HPD photocathodes can operate.

Software-based Particle Identification

The PID reconstruction strategy depends on the purpose and the available computing resources. The earliest, fast hardware triggers rely on requirements within single subdetectors, and they will be the main topic of discussion in the `L0` trigger (Sec. 3.3.3). The software triggers and offline physics analysis have access to two, more computationally demanding, algorithms: a log likelihood variable and a multivariate (MVA) classifier probability. The log likelihood approach sums the individual log likelihoods from each subdetector linearly into a global likelihood. A drawback of this approach is that correlations between subdetectors are not accounted for. In contrast, the MVA classifiers infer a single, global probability based on information from all particle identification subdetectors, which accounts for correlation among subdetectors. Interestingly, some physics analyses report an increased performance with neural network probabilities compared to global likelihoods. This is the case for example in $\Sigma^+ \to p\mu^+\mu^-$ decays, which is shown in Fig. 3.24.

Despite their edge, the MVA algorithms still rely on individual subdetector likelihoods as an input variable. Among them, the most important likelihoods are provided by the RICH reconstruction. The RICH reconstruction begins with the charged track found by the tracking algorithms, and assumes a mass hypothesis for each track. Given the mass hypothesis, the RICH reconstruction then computes the expected photon yield at each HPD cell, which takes photocathode efficiencies and electronics noise into account. The expected HPD photon yield is then compared to the detected signature to assign a likelihood, assuming Poisson statistics. For each track, the RICH reconstruction then iterates over five possible particle species, which are the electrons, muons, kaons, pions and protons [3]. The mass hypothesis with the highest likelihood is assigned to the track. The RICH

---

[3] The deuteron was added in Run 2.

reconstruction performance in kaon-pion and proton-kaon separation are shown in Fig. 3.25.

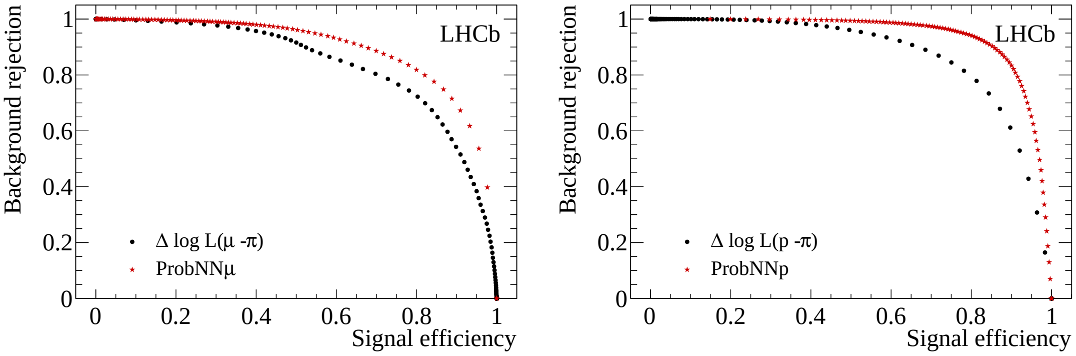


Figure 3.24 – Background misidentification rates versus muon (left) and proton (right) identification efficiency measured in $\Sigma^+ \to p\mu^+\mu^-$ decays. The log likelihood difference w.r.t. pion mass hypothesis (black) and neural network (red) probability is evaluated on $5-10\,\mathrm{GeV}/c$ muon candidates and $5-50\,\mathrm{GeV}/c$ proton candidates. The background sample uses data sidebands while the signal sample uses Monte Carlo simulation. Figure extracted from Ref. (38).

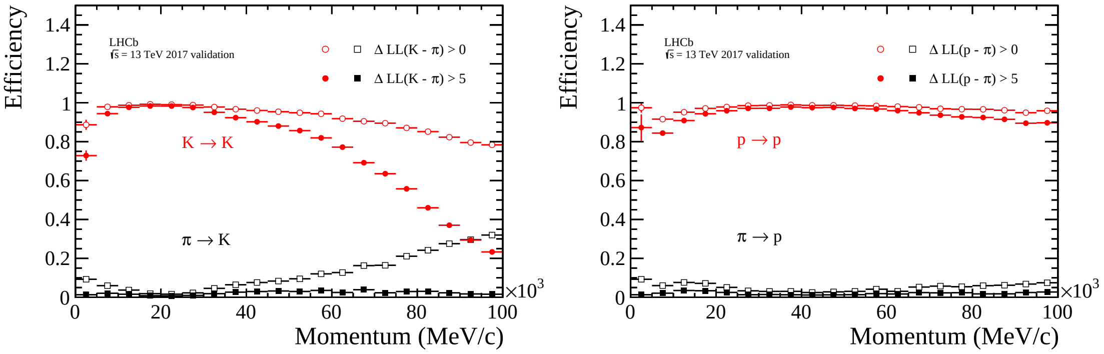


Figure 3.25 – LHCb RICH particle identification performance in kaon-pion separation (left) and proton-pion separation (right) up to a momentum of $100\,\mathrm{GeV}/c$ using 2017 data, extracted from Ref. (38). Kaon-pion separation peaks around the $20-30\,\mathrm{GeV}/c$ range and degrades significantly at higher momentum.

3.3.3 LHCb Trigger

The LHCb trigger reduces LHCb's detector event rate from the LHC's 40 MHz *pp* bunch crossing rate down to a manageable rate for long-term storage. This reduction proceeds via a dual-level trigger, where a Level-0 (`L0`) fast hardware trigger selects the bunch crossings to be read out to a software High Level Trigger (HLT) farm that decides which events are kept in long-term storage. LHCb employed slightly different trigger strategies in Run 1 and Run 2, illustrated in Fig. 3.26. In Run 2, LHCb commissioned an online detector calibration and alignment system which gave the online trigger access to high quality alignment and calibrations that in Run 1 were only available for offline analysis.

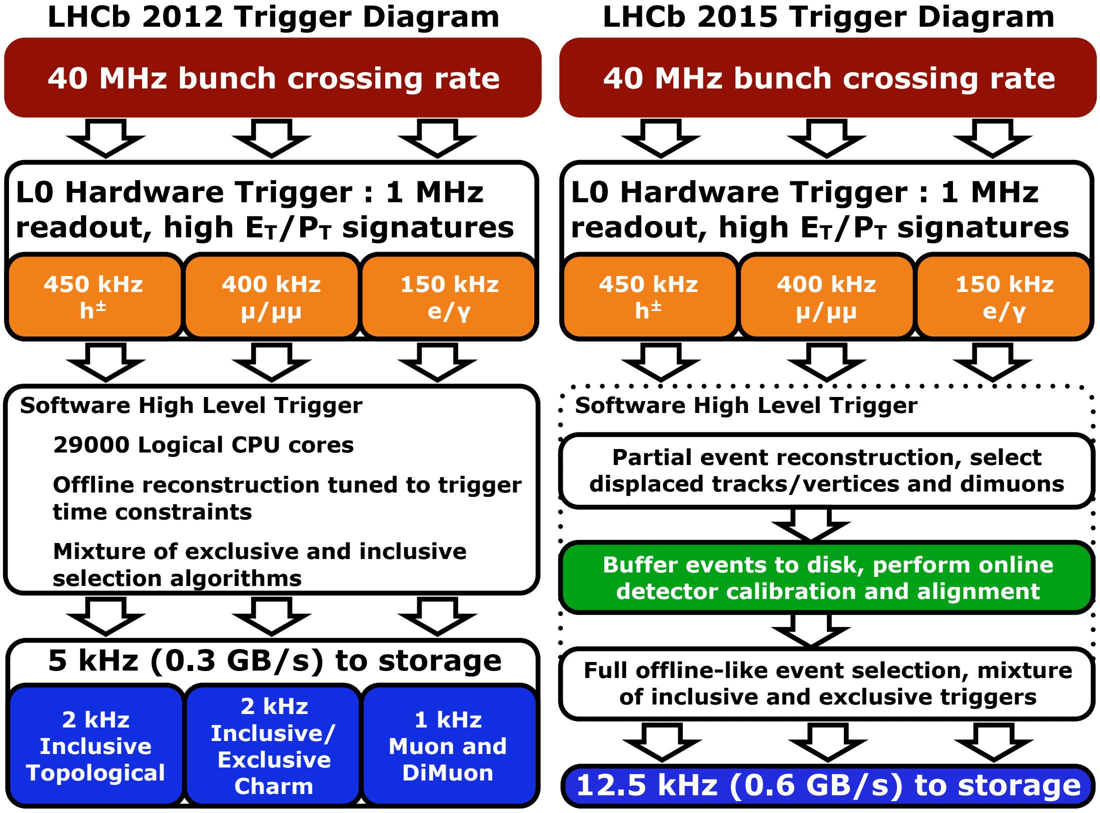


Figure 3.26 – The LHCb trigger strategies used during Run 1 (left) and Run 2(right), extracted from Ref. (39). Aside from the increased storage rate, the LHCb trigger added an online detector calibration and alignment system in Run 2.

Of the jargons used by LHCb members, "trigger line" and "Turbo" will be relevant for our discussion about the LHCb trigger, and we shall give their definitions here. A "trigger line" is single set of selection criteria which targets a specific characteristic expected from interesting physics events. The trigger lines in `L0` and HLT uses the reconstructed quantities at each level for their selection criteria, and the overall `L0` and HLT triggers are inclusive ORs of their respective trigger lines. The "Turbo" stream is a set of trigger lines performing real-time analysis. These streams provide a decay candidate with offline quality full reconstruction and if the candidate passes the Turbo line criteria, only a subset of the full event information is saved. The Turbo streams are introduced in Run 2 because Turbo reconstruction requires a real-time, online calibration and alignment system which, as we mentioned, was commissioned in Run 2.

L0 Trigger

The `L0` trigger reduces the LHC bunch crossing rate down to 1 MHz, the read out rate of the LHCb detector. The `L0` hardware is implemented in the LHCb muon system and calorimeters. The `L0` muon system has two trigger lines, the `L0Muon` and `L0DiMuon`, which detects a single and a pair of high p_{T} muons respectively. The calorimeters trigger on high energy deposits by electrons, photons, and hadrons, which trigger the `L0Electron`, `L0Photon` and `L0Hadron` lines respectively. This design to trigger on highly energetic stable particles is motivated by the fact that b hadrons have invariant masses above 5 GeV and are highly boosted in the LHCb acceptance. Thus, b hadron decays will produce energetic daughters that are likely to fire the `L0` triggers. The thresholds of the various `L0` triggers are shown in Tab. 3.1.

`L0` trigger	$E_{\mathrm{T}}/p_{\mathrm{T}}$ threshold 2012	2016	`nSPDHits` 2012	2016
Electron	> 3.0 GeV	> 2.4 GeV	< 600	< 450
Muon	> 1.76 GeV	> 1.8 GeV	< 600	None
Muon high p_{T}	N/A	> 6.0 GeV	N/A	< 450
Dimuon	> 2.56 GeV^2	> 2.25 GeV^2	< 900	< 900
Hadron	> 3.7 GeV	> 3.7 GeV	< 600	< 450
Photon	> 3.0 GeV	> 2.78 GeV	< 600	< 450

Table 3.1 – The trigger thresholds used in the majority of 2012 and 2016, which are representative thresholds of the Run 1 and Run 2 period. The `L0` high p_{T} muon trigger was not available in Run 1.

The `L0` trigger lines have an additional requirement to remove high multiplicity events. These busy events are difficult to reconstruct and they drain sigfinicant computing resources, only to end up with a lower quality reconstruction. These multiplicity requirements cut on the number of SPD hits, and are looser in the low trigger rate `L0DiMuon`.

The `L0` muon triggers first look for straight line tracks in the five muon stations. For each straight line candidate, the hardware trigger then estimates the p_{T} of the track. This estimation assumes the particle receives a single kick in the magnetic field and the track originates from the center of the interaction region, since PVs are not reconstructed at this stage. Finally, the `L0Muon` triggers on the muon with the highest p_{T} while the `L0DiMuon` triggers on the product of p_{T} from the two highest p_{T} muons.

The calorimeter-based `L0Electron`, `L0Photon` and `L0Hadron` trigger lines search for energetic clusters of 2×2 calorimeter cells in size. These triggers then calculate the transverse energy of the cluster, E_T, defined as

$$E_T = \sum_{i=1}^{4} E_i \sin(\theta_i), \tag{3.6}$$

where E_i is the energy deposited in cell i and θ_i is the angle between the beamline and a line from the cell center to the center of the interaction region. `L0Electron` and `L0Photon` triggers apply an E_T threshold on ECAL clusters and requires the presence of PS cell hits in front of the ECAL clusters. Electrons and photons are differentiated by the presence of SPD hits in front of the PS cells. The `L0Hadron` trigger searches for a HCAL cluster and measures its E_T. However, hadron showers can begin towards the end of the ECAL and leave energy deposits in the ECAL. Thus, the `L0Hadron` also looks for ECAL clusters in front of each HCAL cluster and adds their E_T measurements before making a decision.

High Level Triggers

Once the bunch crossings are selected by the `L0` trigger, the detector information is read out to the HLT that further reduces the event rate down to 5 (12.5) kHz in Run 1 (Run 2). The HLT software runs on an Event Filter Farm (EFF) which gradually increased in size over the course of Run 1 and Run 2. At the end of Run 2, the EFF has 1700 CPU nodes housing 27000 physical cores, and runs about 50000 single-threaded processes concurrently. The HLT is further split into HLT1, which selects events based on information from a partial reconstruction, and HLT2, which relies on fully reconstructed events. The HLT processing differed between Run 1 and Run 2, not just with an introduction of the online detector calibration and alignment but also in terms of disk buffer usage.

In Run 1, the `HLT` is a single process and the `HLT2` sub-process ran directly after an event had passed the `HLT1` . The Run 1 disk buffer received about 20% of `L0` triggered events and the HLT processing of these was deferred until the LHC fill was over, spreading the EFF computing load over in-fill and out-of-fill periods.

In Run 2, the `HLT1` and `HLT2` are split into independent processes. All events read out from `L0` were processed directly by `HLT1` but the events selected by `HLT1` were saved to a disk buffer. Once there is sufficient statistics for the full detector alignment and calibration, these alignment and calibration constants are passed to the HLT2. HLT2 then reads the HLT1-selected events from the disk buffer, reconstructs them with the latest alignement and calibration constants, and selects events for permanent storage. HLT2 typically runs out-of-fill but in-fill HLT2 processes are allowed when there are available computing resources.

HLT1

The `HLT1` sequence begins by reconstructing VELO tracks, then applies a simplified Kalman filter on these tracks before reconstructing PVs with them. The `HLT1` in Run 1 then selects VELO tracks with significant IP w.r.t. all reconstructed PVs, a requirement removed in Run 2. Next, the forward tracking algorithm propagates the selected VELO tracks to the TT to create upstream tracks before further extrapolation into the T-stations. The benefit of this is two-fold. First, long tracks matched to TT hits have a lower fake track rate. Secondly, the upstream tracks have a charge estimate to narrow the search window in the T-stations. This search window has a minimum threshold at $> 500(1200)$ MeV of p_{T} in Run 2 (Run 1). Once the long tracks are found, they are fitted with a Kalman filter and pruned by fake track removals. Finally, clone tracks, defined as a pair of tracks who share more than 70% of their detector hits, are suppressed by only keeping the track more hits.

The `HLT1` muon identification in Run 2 extrapolates fully fitted longs tracks to the muon stations. A muon track is required to leave a hit from M2 up to M3, M4 or M5 depending on the momentum estimate, due to the fact that high momentum muons are more penetrating. However, this strategy misses out on tracks below the p_{T} threshold of the T-station search window. To recover them, the Run 2 `HLT1` sequence uses a similar muon identification strategy in Run 1.

The upstream tracks (VELO tracks in Run 1) are extrapolated directly to the muon stations. If a match is found, the upstream track (VELO track in Run 1) is extrapolated using the forward tracking algorithm, regardless of its p_{T}. In Run 2, this extends the muon identification reach down to a p_{T} of 80 MeV, significantly improving low-momentum muon performance at little computing cost.

The `HLT1` set of trigger lines includes inclusive b and c lines, four muon lines, a radiative decay line (Run 1 only), calibration lines (for online calibration and alignment in Run 2) and a few exclusive lines. Since the physics analysis in this thesis uses the inclusive trigger lines, we shall briefly discuss them here. The inclusive lines target highly boosted c and b hadrons coming from the interaction region and decaying at a vertex significantly displaced from the beamline. Therefore, this line searches for a high p_{T} daughter track with significant displacement and its selection criteria also include track quality information. The Run 1 inclusive line operates as a set of simple selection cuts, known as the `HLT1TrackAllL0`. Run 2 uses a MVA classifier instead, but the classifier input uses many of the same selection variables. The Run 2 inclusive trigger is called `HLT1TrackMVA`. Although Run 2 has another `HLT1TwoTrackMVA` line, which triggers on two high p_{T} tracks from a single displaced vertex, this line is not used in our analysis.

HLT2

The `HLT1` filters the LHCb event rate down to 110 (80) kHz in Run 2 (Run 1). This significantly increases the computing budget allowance per event for HLT2 to fully reconstruct each event. HLT2 first repeats the `HLT1` VELO tracking and PV reconstruction algorithms. The forward tracking algorithm is similar to `HLT1` but the T-station search window now has a smaller p_{T} threshold at 80 MeV. Next, `HLT2` reconstructs T-tracks with a standalone algorithm and the T-track candidates are Kalman filtered. These T-tracks seed the track matching algorithm, which extrapolates backwards and matches T-tracks to VELO tracks leftover by the forward tracking algorithm. This two-fold long track redundancy recovers a few percent of efficiency, albeit at a cost of higher clone rate. The remaining T-tracks are used by the downstream track pattern recognition algorithm, extrapolating T-tracks to TT hits to form downstream tracks. `HLT2` then applies the Kalman fit on reconstructed long and downstream tracks, followed by the fake track suppression and clone killer algorithm.

The `HLT2` muon identification algorithm is the same as in `HLT1` . However, with the relaxed selections in the `HLT2` tracking algorithms, the `HLT2` tries to match a larger amount of tracks to the muon stations. Next, the RICH reconstruction, which is unavailable in `HLT1` due to its prohibitive computing time, assigns each reconstructed track a likelihood to be either a muon, electron, pion, kaon or proton. The `HLT2` attempts to further improve the photon, electron and hadron likelihood by matching charged tracks to calorimeter clusters. Finally, the PID classifiers integrate all associated subdetectors information to calculate the particle species probabilities for each track.

The electrons in LHCb have an undesirable side effect of producing bremstrahlung radiation due to material interactions. The HLT2 sequence has a dedicated bremsstrahlung algorithm to recover bremstrahlung photons emitted in the VELO and TT. The recovery procedure extrapolates a straight line from the VELO to an ECAL region, taking advantange of the fact bremsstrahlung photons travel collinearly to the electrons which produced them. If a photon cluster is found within the search window, its energy deposit is added to the electron momentum. The bremsstrahlung photons emitted after the magnet, mainly in RICH2 and T-stations, typically end up in the same ECAL cluster as the electrons, and require no recovery procedure.

`HLT2` contains an even larger set of trigger lines, about 1000, in order to include the wide range of LHCb physics analysis. The lines dedicated to charm, electroweak and jet physics are beyond the scope of this thesis, and we will focus on the topological b triggers. These topological triggers are MVA classifiers that selects a vertex significantly displaced from reconstructed PVs, with two, three or four high p_{T} tracks and a topology similar to a b hadron. The topological trigger variants in the analysis have an extra requirement that at least one or two of the tracks are identified as a muon or electron.

Real Time Tracking and Alignment

The real time alignment and calibration (40; 41) sample uses dedicated `HLT1` trigger lines. The alignment and calibration measures new constants at regular intervals. Should these new constants differ significantly from previous values, their values are updated in real-time. The intervals depend on the specific alignment or calibration constant computed, and the typical time it takes for each calibration or alignment to collect a sufficient sample size is shown in Fig. 3.27.

The VELO, tracking and muon station alignments are based on a χ^2 minimization of the Kalman track fit residuals. The RICH mirror alignment fits the Cherenkov angle of Cherenkov rings as a function of azimuthal angle of the ring,

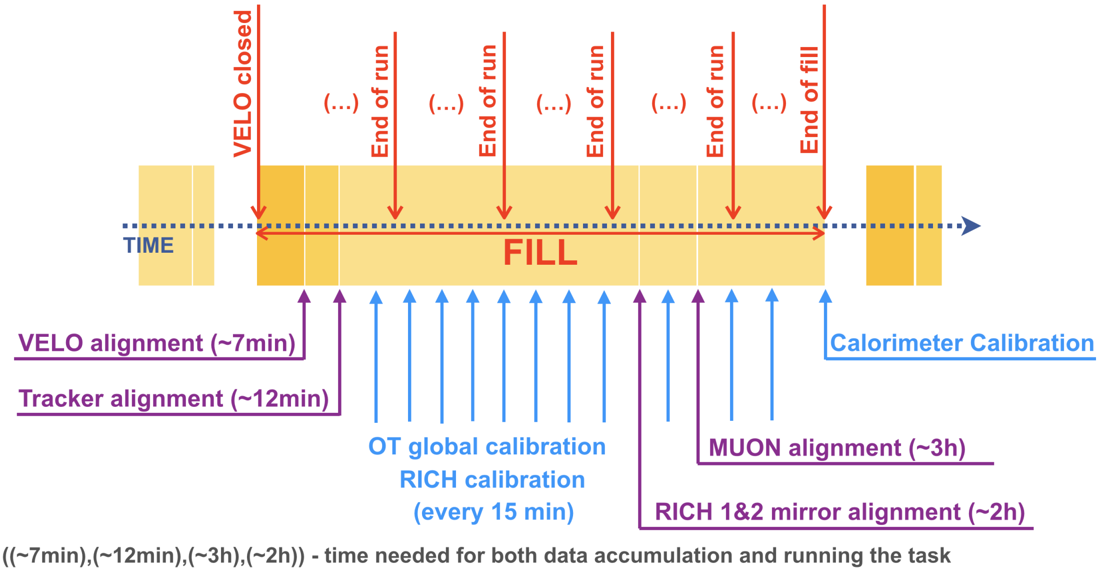


Figure 3.27 – The LHCb online alignment and calibration procedure that starts at the beginning of each fill, used in 2018 data-taking. Figure extracted from Ref. (32).

and iteratively corrects the alignment until the Cherenkov angle has no azimuthal angle dependence. The RICH calibration corrects the Cherenkov radiator refractive index by comparing the reconstructed Cherenkov angles with the angles from an expected refractive index. The expected refractive index is inferred from gas temperature, pressure and Cherenkov radiator gas composition monitors. The OT straw tubes drift time calibration fits the difference between measured and estimated drift time. The HCAL is calibrated via a Cs source scan. Finally, the ECAL calibration constrains the reconstructed π^0 mass to the PDG mass to obtain a calibration constant.

LHCb Trigger Performance

Typically, LHCb calculates trigger efficiencies using the TISTOS method, which is the case in the analysis of this thesis and in the LHCb trigger performance papers, Ref. (34) and (32). The TISTOS method uses two categories of trigger,

- **TOS** events, acronym for *Trigger On Signal*, refers to events triggered by the signal (decay of interest).
- **TIS** events, acronym for *Trigger Independent of Signal*, refers to events trig-

gered by the ‘underlying event’, in other words event objects that are unrelated to the signal decay.

The TISTOS method assumes the TIS and TOS categories are uncorrelated, and measures the TISTOS efficiency,

$$\epsilon_{\text{TISTOS}} = \frac{N(\text{TIS \texttt{\&\&} TOS})}{N(\text{TIS})}, \tag{3.7}$$

where $N(\text{TIS})$ is the number of events passing the TIS trigger and $N(\text{TIS \texttt{\&\&} TOS})$ the number of events passing both TOS and TIS triggers. Recall our previous discussions that b quarks in LHCb are produced as correlated pairs. This invalidates the assumption that the TOS and TIS categories are totally uncorrelated, since the opposite B is part of the underlying event. Nevertheless, the correlation can be accounted for by binning the efficiency in small regions of phase space, typically the B p_{T}. The residual bias of this method was studied by Ref. (42), reporting an efficiency bias of $0.25 \pm 0.1\%$ with the optimal binning scheme.

The Run 2 efficiencies of `L0Hadron`, `L0Electron`, `L0Muon`, and `L0DiMuon` are shown in Fig. 3.28, plotted as a function of B (or D^0) η and p_{T}. These efficiencies are evaluated using the TISTOS method with different signal decays,

`L0Hadron` : $D^0 \to K^-\pi^+$,

`L0Electron` : $B^+ \to K^+ J/\psi\,(\to e^+e^-)$,

`L0Muon` and `L0DiMuon` : $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$.

The muon `L0` efficiencies are relatively independent of B η, and they show a slight dependence with B p_{T}. The `L0Hadron` and `L0Electron` have a lower efficiency at the high η bins, the calorimeter regions with higher multiplicity. The calorimeter-based `L0` triggers also have much lower efficiencies than the muon triggers, especially at low B p_{T}. This is a consequence of the tighter calorimeter thresholds.

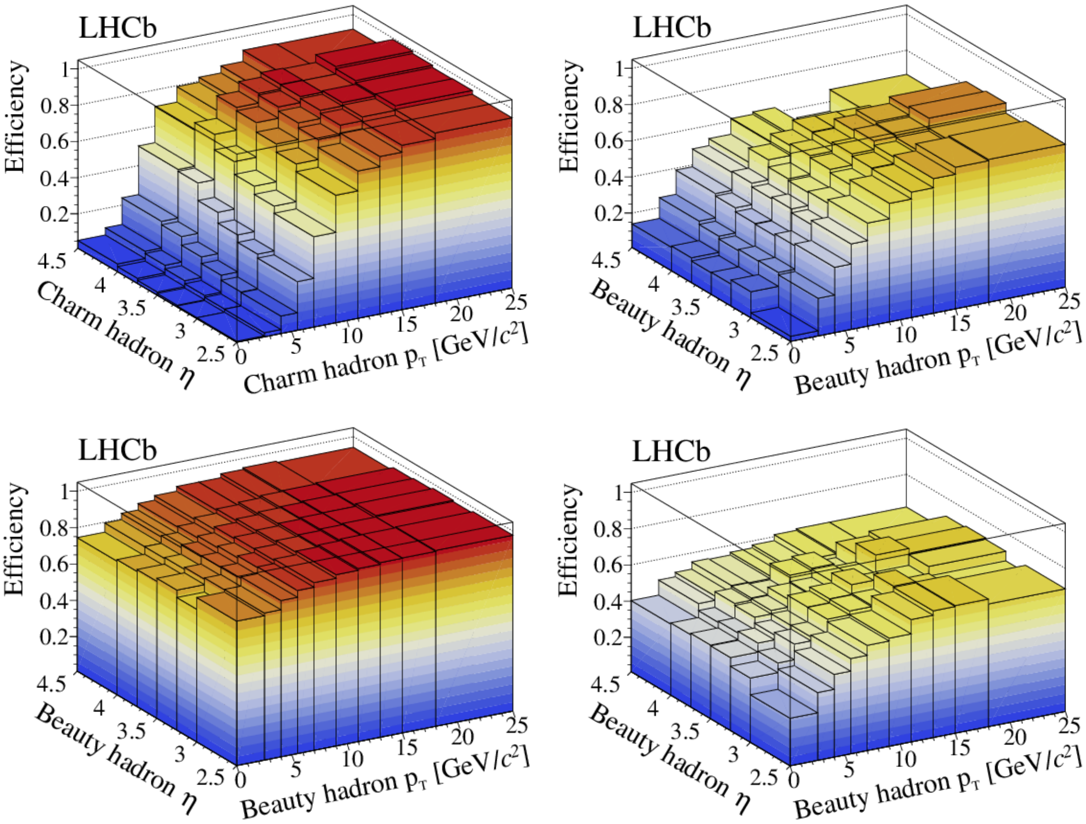


Figure 3.28 – The two-dimensional efficiencies of `L0Hadron` (top left), `L0Electron` (top right), `L0Muon` (bottom left), and `L0DiMuon` (bottom right) in Run 2 data. Figure extracted from Ref. (32).

Fig. 3.29 shows the inclusive `HLT1TrackMVA` line performance on B decays in Run 2. Of relevance to the analysis chapter, the `HLT1TrackMVA` line has a good efficiency across the p_{T} spectrum. The `HLT1TrackMVA` efficiency is more than 95% for $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi(\to e^+e^-)$ decays at high B p_{T} but they degrade significantly below $p_{\mathrm{T}} < 10$ GeV. Although the $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ efficiency decreases to 85% at the lowest p_{T} bin, the $B^+ \to K^+ J/\psi(\to e^+e^-)$ efficiency degradation is much worse at 70% in this bin. The performance of `HLT1TwoTrackMVA`, irrelevant to the analysis, can be found in Appendix A.

The `HLT2` topological lines efficiencies for the same decay modes are shown in Fig. 3.30. The $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi(\to e^+e^-)$ decays both have an efficiency above 90% at above 10 GeV of p_{T}. Similar to the `HLT1` performance, their efficiencies degrade at low p_{T}, reaching a low of about 70% and 58% respectively at the lowest p_{T} bin.

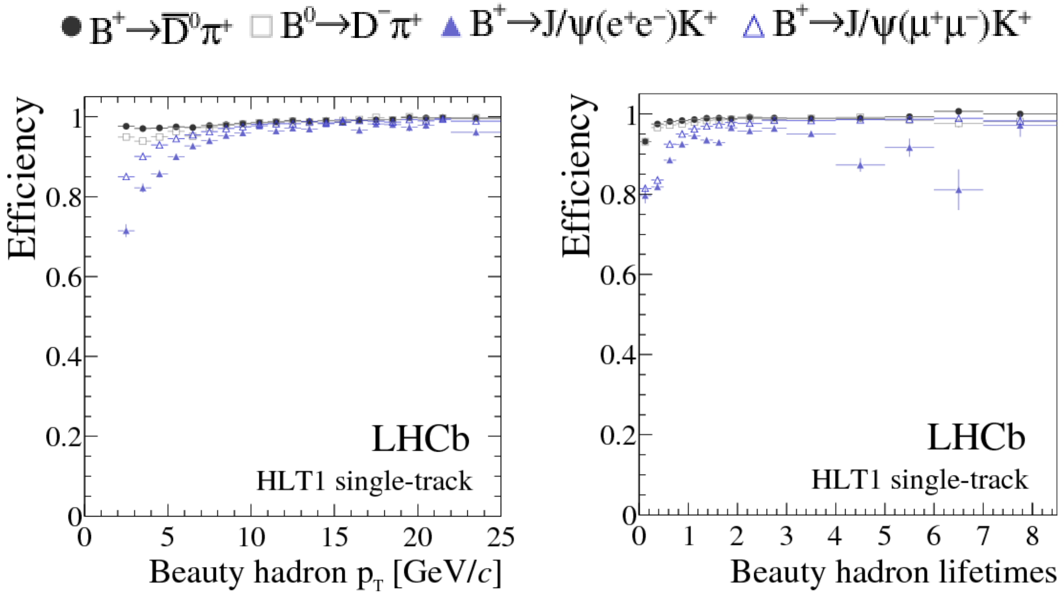


Figure 3.29 – The Run 2 efficiencies of `HLT1TrackMVA` as a function of B p_{T} (left) and units of B lifetime (right). Figure extracted from Ref. (32).

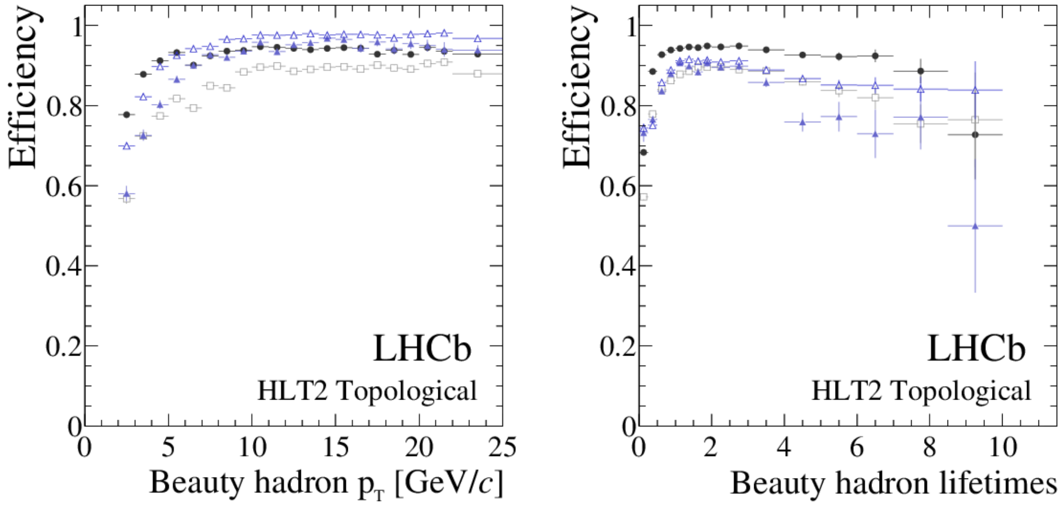


Figure 3.30 – The Run 2 efficiencies of `HLT2` topological lines as a function of B p_{T} (left) and units of B lifetime (right). Figure extracted from Ref. (32).

3.4 Conclusion

The LHC machine is able to collide proton-proton beams at an unprecedented center-of-mass energy of 13 TeV. At this energy, the $b\bar{b}$ production in the LHCb acceptance is about 144 μb (72.0 μb at 7 TeV). Coupled with the high luminosities sustained by the LHC machine, the LHC has produced approximately one trillion $b\bar{b}$ events for the LHCb detector.

The LHCb tracking system has a good momentum resolution which is able to reconstruct $J/\psi \to \mu^+\mu^-$ decays with a mass resolution of $12.4 - 12.7$ MeV. The LHCb data acquisition has efficient triggers for $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ decays at each stage. At worst it is 70% efficient at the low p_{T} regime but above $p_{\mathrm{T}} > 10$ GeV, $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ are selected with 90 % efficiency or more.

The same cannot be said for electrons. Electrons undergoes bremsstrahlung radiation in the LHCb detector and the dedicated bremsstrahlung recovery procedure is not 100 % efficient. This leads to a degradation in $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ `HLT1` and `HLT2` trigger efficiencies, which are worse than corresponding $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ trigger efficiencies especially in the low p_{T} region. Also, the bremsstrahlung radiation leads to significant degradation in the momentum and mass resolution of reconstructed B candidates. As for the `L0Electron` trigger, the electrons have significantly lower `L0` trigger efficiencies than muons. Even though electrons might have lost energies via bremsstrahlung radiation, it has to face a tighter ECAL threshold as well [4]. This means that the LFU tests with $b \to s\,\ell^+\ell^-$ decays, rare as it is, will be statistically limited by the electron modes.

---

[4] These are electrons that emit bremsstrahlung before the magnetic field. Bremsstrahlung photons emitted after the magnet ends up in the same ECAL cluster as the electron.

3.5 Bibliography

[1] CMS, A. M. Sirunyan *et al.*, *Observation of* $\mathrm{t\bar{t}}H$ *production*, Phys. Rev. Lett. **120** (2018), no. 23 231801, `arXiv:1804.02610`.

[2] UA1, G. Arnison *et al.*, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, Phys. Lett. B **122** (1983) 103.

[3] UA2, M. Banner *et al.*, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys. Lett. B **122** (1983) 476.

[4] UA1, G. Arnison *et al.*, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. B **126** (1983) 398.

[5] UA2, P. Bagnaia *et al.*, *Evidence for* $Z^0 \to e^+e^-$ *at the CERN* $\bar{p}p$ *Collider*, Phys. Lett. B **129** (1983) 130.

[6] ALEPH, DELPHI, L3, OPAL, LEP Electroweak, S. Schael *et al.*, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, Phys. Rept. **532** (2013) 119, `arXiv:1302.3415`.

[7] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group, S. Schael *et al.*, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) 257, `arXiv:hep-ex/0509008`.

[8] *Resolution Approval of the Large Hadron Collider (LHC) Project. Résolution Approbation du projet de grand collisionneur de hadrons (LHC). 100th Session of Council*, .

[9] S. W. Herb *et al.*, *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39** (1977) 252.

[10] CDF, F. Abe *et al.*, *Observation of top quark production in* $\bar{p}p$ *collisions*, Phys. Rev. Lett. **74** (1995) 2626, `arXiv:hep-ex/9503002`.

[11] D0, S. Abachi *et al.*, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632, `arXiv:hep-ex/9503003`.

[12] CERN, , (Geneva), CERN, 1992.

[13] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, , General Photo.

[14] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[15] CMS, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[16] LHCb, J. Alves, A.Ãugusto *et al.*, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[17] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[18] R. Alemany-Fernandez *et al.*, *Operation and Configuration of the LHC in Run 1*, .

[19] J. Wenninger, *Operation and Configuration of the LHC in Run 2*, .

[20] CERN, *Chamonix 2012 Workshop on LHC Performance: Chamonix, France 6 - 10 Feb 2012. Chamonix 2012 Workshop on LHC Performance*, (Geneva), CERN, 2012. doi: 10.5170/CERN-2012-006.

[21] R. Alemany-Fernandez, F. Follin, and R. Jacobsson, *The LHCb Online Luminosity Control and Monitoring*, .

[22] J. Stirling, *Parton luminosity and cross section plots*, .

[23] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, `arXiv:hep-ph/0603175`.

[24] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.

[25] LHCb collaboration, C. Elsässer, *$\bar{b}b$ production angle plots*, .

[26] LHCb, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017), no. 5 052002, `arXiv:1612.05140`, [Erratum: Phys.Rev.Lett. 119, 169901 (2017)].

[27] LHCb, R. Antunes-Nobrega *et al.*, *LHCb technical design report: Reoptimized detector design and performance*, .

[28] LHCb VELO Group, T. Latham, *Performance of the LHCb Vertex Locator*, Phys. Procedia **37** (2012) 830.

[29] LHCb Collaboration, P. R. Barbosa-Marinho *et al.*, *LHCb inner tracker: Technical Design Report*, Technical Design Report LHCb, CERN, Geneva, 2002. revised version number 1 submitted on 2002-11-13 14:14:34.

[30] LHCb Collaboration, *Lhcb silicon tracker - material for publications*, 3, 2012.

[31] LHCb Collaboration, *Tracking and alignment plots for conferences*, 10, 2019.

[32] LHCb, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019), no. 04 P04013, `arXiv:1812.10790`.

[33] LHCb, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015), no. 02 P02007, `arXiv:1408.1251`.

[34] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. A **30** (2015), no. 07 1530022, `arXiv:1412.6352`.

[35] LHCb Collaboration, *Muon track reconstruction efficiency at LHCb*, .

[36] J. Alves, A.Ã. *et al.*, *Performance of the LHCb muon system*, JINST **8** (2013) P02022, `arXiv:1211.1346`.

[37] C. A. Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, `arXiv:2008.11556`.

[38] LHCb Collaboration, *Pid plots for conference*, 06, 2020.

[39] LHCb Collaboration, *Trigger plots and diagrams for conferences*, 06, 2018.

[40] G. Dujany and B. Storaci, *Real-time alignment and calibration of the LHCb Detector in Run II*, J. Phys. Conf. Ser. **664** (2015), no. 8 082010.

[41] LHCb, S. Borghi, *Novel real-time alignment and calibration of the LHCb detector and its performance*, Nucl. Instrum. Meth. A **845** (2017) 560.

[42] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at lhcb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014.

Chapter 4

Sensing an Imbalance in the Force : Test of Lepton Flavour Universality with R_K and R_{K^*}

This analysis aims to measure the values of R_K and R_{K^*}, defined as

$$R_K = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma(B^+ \to K^+ \mu^+ \mu^-)}{dq^2} dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma(B^+ \to K^+ e^+ e^-)}{dq^2} dq^2}, \tag{4.1}$$

$$R_{K^*} = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma(B^0 \to K^{*0} \mu^+ \mu^-)}{dq^2} dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{d\Gamma(B^0 \to K^{*0} e^+ e^-)}{dq^2} dq^2}, \tag{4.2}$$

which are essentially measurements of branching fraction ratios in different regions of q^2 [1].

As a recap of Chap. 2.4 on the experimental status of R_K and R_{K^*}, the Belle and BaBar measurements are statistically limited around the $30-50\%$ level. Although the LHCb measurements achieve sensitivities of around 7% for R_K and 17% for R_{K^*}, statistical errors remain the dominant uncertainty. Despite that, LHCb's measurements are in tension with the SM at $2.1-2.5\sigma$.

Given that previous LHCb measurements were statistically limited, this updated analysis aims to measure R_K and R_{K^*} with the full $9\,\text{fb}^{-1}$ LHCb dataset collected over Run 1 and Run 2 [2]. Nevertheless, this update will redefine the

---

[1] Recap: The invariant mass of the di-lepton system

[2] Within LHCb, an R_K only analysis is running in parallel to the analysis of this thesis but they use different analysis strategies, especially in the hardware trigger. These measurements of R_K serve as complementary cross-check of each other.

analysis procedure and change some strategies, the first of which is the q^2 bin definitions.

4.1 Analysis Strategy

The previous LHCb measurement of R_{K^*} defines the q^2 bins as

$$\begin{aligned} low - q^2 &: 0.045 < q^2 < 1.1\,\mathrm{GeV}^2/c^4, \\ central - q^2 &: \quad 1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4. \end{aligned}$$

The deviation in central-q^2 corroborates the patterns of NP physics hinted at by other $b \to s\,\mu\mu$ decays, in particular the branching fraction measurements and the P_5' anomaly in the $B^0 \to K^{*0}\mu^+\mu^-$ angular analysis. These measurements favour a NP contribution to $\mathcal{C}_9$.

However, the R_{K^*} result at low-q^2 was a bit of a surprise. As we discussed in Sec. 2.5, the low-q^2 bin, being near the photon pole, is strongly influenced by $\mathcal{C}_7$ and $\mathcal{C}_{7'}$ values but these Wilson coefficients are strongly constrained to be consistent with the SM by $b \to s\gamma$ BF measurements and $b \to s\,ee$ angular analysis at very small q^2. Although $\mathcal{C}_9$ can contribute to this anomaly, constraints from other $b \to s\,\ell^+\ell^-$ measurements do not agree with LHCb's measurement of R_{K^*} at low-q^2 .

On the theoretical side, SM predictions of R_{K^*} at low-q^2 have uncertainties of about $2-3\%$ (see Tab. 2.2) due its proximity to the photon pole. As a result, low-q^2 predictions of R_{K^*} are dominated by QED corrections. A study of QED corrections to R_K and R_{K^*} by Ref. (1) predicts that

$$R_{K^*} = \begin{cases} 0.906 \pm 0.028, & \text{for } 0.045 < q^2 < 1.1\,\mathrm{GeV}^2/c^4, \\ 0.983 \pm 0.014, & \text{for } \quad 0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4, \end{cases} \tag{4.3}$$

where a small increase in low-q^2 bin lower limit halves the theoretical uncertainty of R_{K^*}. Based on this result, Ref. (1) strongly suggests that future R_{K^*} measurements should adopt the latter low-q^2 definition, a suggestion this analysis will follow.

This analysis will also measure the value of R_K in the low-q^2 region. Recall that $B^+ \to K^+\ell^+\ell^-$ decays do not couple to the photon pole due to parity conservation. Thus in the low-q^2 bin, a measurement of R_K is complementary to a measurement of R_{K^*}, especially since the $\mathcal{C}_7$ related observables have been observed to be SM-like. Therefore, this analysis will use the same q^2 definitions for R_K and R_{K^*}

measurements, which are

$$\begin{aligned} low - q^2 &: 0.1 < q^2 < 1.1\,\mathrm{GeV}^2/c^4, \\ central - q^2 &: 1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4. \end{aligned}$$

Experimentally, measurements of R_K and R_{K^*} can be written as

$$\begin{aligned} R_K^{SR} &= \frac{\mathcal{B}(B^+ \to K^+\mu^+\mu^-)}{\mathcal{B}(B^+ \to K^+e^+e^-)} \\ &= \frac{N(B^+ \to K^+\mu^+\mu^-)}{N(B^+ \to K^+e^+e^-)} \times \frac{\epsilon(B^+ \to K^+e^+e^-)}{\epsilon(B^+ \to K^+\mu^+\mu^-)}, \end{aligned} \tag{4.4}$$

$$\begin{aligned} R_{K^*}^{SR} &= \frac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} \\ &= \frac{N(B^0 \to K^{*0}\mu^+\mu^-)}{N(B^0 \to K^{*0}e^+e^-)} \times \frac{\epsilon(B^0 \to K^{*0}e^+e^-)}{\epsilon(B^0 \to K^{*0}\mu^+\mu^-)}, \end{aligned} \tag{4.5}$$

where the yields, N, are measured from mass fits to real data and the efficiencies, ϵ, are estimated from simulation. We shall label these as 'single ratios', SR, for reasons that shall be apparent soon. Because efficiencies rely on LHCb simulation that is known to be imperfect, a significant part of this analysis consists of correcting simulation using data driven methods. These corrections use collaboration-wide calibration samples and dedicated control modes [3] to weight the simulated events.

As per the norm of $b \to s\,\ell^+\ell^-$ analyses in LHCb, this analysis uses $B^+ \to K^+ J/\psi(\to \ell^+\ell^-)$ and $B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-)$ decays as control modes, where the J/ψ decays into an ee or $\mu\mu$ pair. These tree-level decays have the same $K^+\ell^+\ell^-$ and $K^{*0}\ell\ell$ final states but they are produced abundantly in the LHCb experiment [4], with branching fractions roughly $3-4$ orders of magnitude higher than $b \to s\,\ell^+\ell^-$ decays (2). Furthermore, the $b \to s\,\ell^+\ell^-$ (rare) and control modes [5] have similar kinematics in the LHCb detector. The J/ψ resonance has a short lifetime since it decays electromagnetically and the leptons are emitted at the B decay vertex. Coupled with the fact that B mesons, and their daughters, in LHCb are highly boosted, the kinematic distributions between rare and control modes are similar

---

[3] Rather than highly suppressed $b \to s\,\ell^+\ell^-$ modes.

[4] The abundance of these decays and the fact that the leptons are produced in pairs from the J/ψ decays makes them the decays of choice for LHCb tracking efficiency studies and PID calibrations, for both electrons and muons.

[5] This thesis avoids referring to $b \to s\,\ell^+\ell^-$ decays as 'signal' modes because this analysis will fit both control and $b \to s\,\ell^+\ell^-$ modes candidates, and it will be very confusing to differentiate the control or $b \to s\,\ell^+\ell^-$ signal in the mass fit from 'signal' mode.

in the LHCb lab frame [6]. Thus, this analysis uses the abundant yet kinematically similar control modes to build correction maps in a data-driven way before porting the corrections to simulated rare modes.

Additionally, the control modes provide a stringent closure test of the efficiency calculation by measuring $r_{J/\psi}$, defined as

$$\begin{aligned} r^{K}_{J/\psi} &= \frac{\mathcal{B}(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))}{\mathcal{B}(B^+ \to K^+ J/\psi(\to e^+e^-))} \\ &= \frac{N(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))}{N(B^+ \to K^+ J/\psi(\to e^+e^-))} \times \frac{\epsilon(B^+ \to K^+ J/\psi(\to e^+e^-))}{\epsilon(B^+ \to K^+ J/\psi(\to \mu^+\mu^-))}, \end{aligned} \tag{4.6}$$

$$\begin{aligned} r^{K^*}_{J/\psi} &= \frac{\mathcal{B}(B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-))}{\mathcal{B}(B^0 \to K^{*0} J/\psi(\to e^+e^-))} \\ &= \frac{N(B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-))}{N(B^0 \to K^{*0} J/\psi(\to e^+e^-))} \times \frac{\epsilon(B^0 \to K^{*0} J/\psi(\to e^+e^-))}{\epsilon(B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-))}. \end{aligned} \tag{4.7}$$

In the $r_{J/\psi}$ formulae, the $B \to K^{(*)} J/\psi$ branching fractions in the numerator and denominator cancel out, leaving the $r_{J/\psi} = \frac{\Gamma(J/\psi(\mu\mu))}{\Gamma(J/\psi(ee))}$ equality. This ratio's most precise measurement is provided by BES III (3)

$$r^{-1}_{J/\psi} = \frac{\Gamma(J/\psi(ee))}{\Gamma(J/\psi(\mu\mu))} = 1.0017 \pm 0.0017 \pm 0.0033$$

which is lepton flavour universal at the sub-percent level. Thus, if the efficiency ratios in Eq. 4.6 and 4.7 are correctly calibrated, the measured value of $r_{J/\psi}$ will be consistent with unity. We shall refer to this as the 'integrated measurement of $r_{J/\psi}$', which will be presented in Sec. 4.7.1. Another powerful closure test is to measure $r_{J/\psi}$ differentially by binning the control mode samples in variables of interest. The differential values of $r_{J/\psi}$ are expected to be flat across any variable of choice and any trends would hint at local efficiency miscalibration, which might cancel out in the integrated $r_{J/\psi}$ measurement. The differential test of $r_{J/\psi}$ is referred to as the '$r_{J/\psi}$ flatness test', which will be presented in Sec. 4.7.2.

Recall our discussion from the Chap. 3.3, that electrons and muons behave differently in the LHCb detector. At the hardware stage, the ECAL triggers electron events while the muon stations provide fast muon identification. The ECAL experiences a higher multiplicity and it therefore sets a higher energy threshold on the electrons. With their reduced multiplicity, muons stations can set looser

---

[6] The fraction of B momentum carried away by the $\ell^+\ell^-$ pair depends on q^2, and in this case, there is a visible difference between control and rare modes. Readers can consult Appendix B for a visual of these differences.

requirements on the muons and are more efficient compared to the electron hardware trigger. To make matters worse, the electrons suffer from Bremsstrahlung radiation and even with a photon recovery procedure, the momentum and invariant mass resolution of the electron modes are significantly smeared compared to the muon modes. These differences are a source of systematic uncertainties because they do not cancel out in the $\frac{\epsilon(ee)}{\epsilon(\mu\mu)}$ efficiency ratio of $r_{J/\psi}$ (Eq. 4.6-4.7), R_K (Eq. 4.4) and R_{K^*} (Eq. 4.5).

Instead of measuring R_K and R_{K^*} as single ratios, this analysis measures R_K and R_{K^*} as double ratios, defined as

$$\begin{aligned}
R_K^{DR} &= \frac{R_K^{SR}}{r_{J/\psi}^{K}} \\
&= \frac{N(B^+\to K^+\mu^+\mu^-)}{N(B^+\to K^+J/\psi(\to\mu^+\mu^-))} \times \frac{N(B^+\to K^+J/\psi(\to e^+e^-))}{N(B^+\to K^+e^+e^-)} \\
&\times \frac{\epsilon(B^+\to K^+J/\psi(\to\mu^+\mu^-))}{\epsilon(B^+\to K^+\mu^+\mu^-)} \times \frac{\epsilon(B^+\to K^+e^+e^-)}{\epsilon(B^+\to K^+J/\psi(\to e^+e^-))},
\end{aligned} \tag{4.8}$$

$$\begin{aligned}
R_{K^*}^{DR} &= \frac{R_{K^*}^{SR}}{r_{J/\psi}^{K^*}} \\
&= \frac{N(B^0\to K^{*0}\mu^+\mu^-)}{N(B^0\to K^{*0}J/\psi(\to\mu^+\mu^-))} \times \frac{N(B^0\to K^{*0}J/\psi(\to e^+e^-))}{N(B^0\to K^{*0}e^+e^-)} \\
&\times \frac{\epsilon(B^0\to K^{*0}J/\psi(\to\mu^+\mu^-))}{\epsilon(B^0\to K^{*0}\mu^+\mu^-)} \times \frac{\epsilon(B^0\to K^{*0}e^+e^-)}{\epsilon(B^0\to K^{*0}J/\psi(\to e^+e^-))},
\end{aligned} \tag{4.9}$$

which uses the fact that $r_{J/\psi}$ is consistent with unity. Since rare and control modes with the same final state have similar kinematic distribution, the systematics between them largely cancel out in the efficiency ratio. Thus, double ratios of R_K and R_{K^*} are more robust when compared to the single ratios in Eq. 4.4 and 4.5. The double ratio procedure can be cross checked with another b to charmonium decay, the $B^+\to K^+\psi(2S)(\to\ell^+\ell^-)$ and $B^0\to K^{*0}\psi(2S)(\to\ell^+\ell^-)$ decay modes. This involves the measurement of $R_{\psi(2S)}$ double ratio, defined as

$$\begin{aligned}
R_{\psi(2S)}^{DR} &= \frac{R_K^{SR}}{r_{J/\psi}^{K}} \\
&= \frac{N(B\to K^{(*)}\psi(2S)(\to\mu^+\mu^-))}{N(B\to K^{(*)}J/\psi(\to\mu^+\mu^-))} \times \frac{N(B\to K^{(*)}J/\psi(\to e^+e^-))}{N(B\to K^{(*)}\psi(2S)(\to e^+e^-))} \\
&\times \frac{\epsilon(B\to K^{(*)}J/\psi(\to\mu^+\mu^-))}{\epsilon(B\to K^{(*)}\psi(2S)(\to\mu^+\mu^-))} \times \frac{\epsilon(B\to K^{(*)}\psi(2S)(\to e^+e^-))}{\epsilon(B\to K^{(*)}J/\psi(\to e^+e^-))},
\end{aligned} \tag{4.10}$$

and like $r_{J/\psi}$, $R_{\psi(2S)}$ is expected to be lepton flavour universal according to the PDG (2)

$$\frac{\Gamma(\psi(2S)(\mu\mu))}{\Gamma(\psi(2S)(ee))} = 1.00 \pm 0.08. \tag{4.11}$$

These measurements, however, use the same control mode data sample to calibrate the corrections and extract the control mode yields. This induces an intrinsic correlation between the efficiency estimation and control mode yield measurements, a systematic effect previous analysis accounted for by bootstrapping the control mode samples. Nevertheless, this analysis takes a different approach where correction maps from $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$ modes are ported to $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ simulation and vice versa. This minimises the correlation between efficiency calculation and yield measurement in the control modes but it necessitates a change in `L0` trigger strategy.

4.2 Data Sample

4.2.1 Real Data Samples

The datasets used in this analysis are collected during LHCb's Run 1 data-taking period, from 2011 to 2012, and Run 2 data-taking period, spanning 2015 to 2018. They are summarised in Tab. 4.1. This analysis then selects the interesting bunch crossings with the trigger strategies and stripping lines discussed in this section.

Year	$\int \mathcal{L}$ $\left[\text{fb}^{-1}\right]$	$\sqrt{s}$ [TeV]
2011	1.1	7
2012	2.1	8
2015	0.3	13
2016	1.7	13
2017	1.7	13
2018	2.2	13

Table 4.1 – The datasets used by this analysis, and their integrated luminosities and centre-of-mass collision energies.

4.2.2 Triggers

L0

Previous LHCb measurements of R_K and R_{K^*} trigger $\mu\mu$ modes only with the `L0Muon` trigger while ee modes are selected with `L0Electron` as the primary category, `L0Hadron` the secondary and `L0Global_TIS` the tertiary category [7]. This is illustrated in Fig. 4.1. This trigger strategy forbids porting corrections between $K^+\pi^-\ell^+\ell^-$ and $K^+\ell^+\ell^-$ final states because the `L0Hadron` trigger would behave differently between the $K\pi$ and K hadron systems. Instead, this analysis uses a common trigger definition for both muon and electron modes. The `L0Global_TIS` trigger is now the primary category while `L0Muon` and `L0Electron` triggers are the secondary categories, and the `L0Hadron` trigger is dropped. This updated trigger strategy is illustrated in Fig. 4.2.

Figure 4.1 – The previous `L0` trigger strategy of R_K and R_{K^*} analysis. The $\mu\mu$ modes (left) only use the `L0Muon` while the ee modes (right) use the `L0Electron` as primary, `L0Hadron` as secondary and `L0Global_TIS` as tertiary categories.

This updated trigger strategy simplifies the `L0` corrections by dropping the `L0Hadron` trigger category, which is especially hard to calibrate because the HCAL resolution is significantly worse than the ECAL. Furthermore, this strategy treats `L0Global_TIS` as the primary category for $\mu\mu$ and ee modes. This decision allows the `L0Global_TIS` calibrations from the larger $\mu\mu$ samples to be applied on the ee modes because `L0Global_TIS` is expected to be independent of differences between $\mu\mu$ and ee modes. This assumption is checked in Sec. 4.4.1. These benefits do come at the cost of losing about $5 - 15\%$ of electron mode events, shown in Tab. 4.2.

This definition of `L0` trigger strategy allows us to measure R_K and R_{K^*} in two exclusive categories,

---

[7] The R_K only measurement, running in parallel to this analysis, uses the previous strategy as well.

Figure 4.2 – The current `L0` trigger strategy of this analysis. The $\mu\mu$ (left) and ee (right) modes use the `L0Global_TIS` as primary category and the `L0Muon` or `L0Electron` trigger as secondary categories.

Final state	q^2 bin	Run 1	Run 2
$K^+e^+e^-$	low	15.26 %	12.45 %
	central	11.67 %	9.10 %
$K^+\pi^-e^+e^-$	low	10.43 %	8.73 %
	central	7.59 %	5.68 %

Table 4.2 – Approximate loss of dropping the `L0Hadron` trigger in ee modes. The efficiency loss of dropping `L0Hadron` in low-q^2 is higher than central-q^2 because the low-q^2 hadron system carries more momentum away from B decays. Also, the $K^+e^+e^-$ final states lose more events compared to $K^+\pi^-e^+e^-$ final states within the same q^2 range. This is because the single K^+ from B^+ decays have more momentum to trigger `L0Hadron` than the individual momentum of K^+ or π^- from B^0 decays.

`L0I` : Data samples trigger by `L0Global_TIS`.

`L0L` : Data samples trigger by `L0Muon` and `L0Electron`, but are exclusive of `L0Global_TIS`.

where we use the `L0I` and `L0L` labels for brevity. As we shall see in the `L0` corrections (Sec. 4.4.1), the `L0L` weights relies on first correcting the inclusive `L0Muon` and `L0Electron` triggers. Therefore, the integrated $r_{J/\psi}$ and $R_{\psi(2S)}$ cross checks are performed in three `L0` categories,

- inclusive `L0I`,
- exclusive `L0L` which is exclusive of `L0Global_TIS`,

- inclusive `L0L`, which are triggered inclusively by `L0Muon` or `L0Electron`.

For the remainder of this thesis, the exclusive and inclusive `L0L` categories will be differentiated explicitly.

HLT

As we have discussed in Chap. 3.3.3, this analysis uses the `HLT1` inclusive b and c trigger and the `HLT2` topological b triggers. The `HLT2` trigger lines in this analysis follows the same selection as the generic LHCb topological b trigger with an added requirement for one or two of its daughters to be identified as a lepton. Compared to the LHCb's previous R_{K^*} analysis, this analysis drops the four body topological triggers in B^0 modes so that the `HLT` corrections between B^+ and B^0 modes are compatible and portable among each other. The efficiency loss of dropping the four body triggers are at the sub-percent level, a negligible effect on statistical sensitivity.

4.2.3 Stripping

After the online triggers, LHCb applies another stage of loose offline selections known as 'stripping'. The stripping cuts run centrally on LHCb grid resources and serves two purposes. First, the offline stripping uses the best quality detector alignment and calibration (that were unavailable online in Run 1) to maximise the reconstruction quality. Compared to the `HLT2` reconstruction, stripping contains additional variables, such as isolation variables, that are computationally too expensive to calculate in the online triggers. Secondly, the stripping applies a more stringent cut than the online triggers and distributes the dataset in long term LHCb storage into dedicated stripping lines. This significantly reduces the size of the datasets an analyst has to study.

This analysis selects the signal and rare modes candidates with the `Bu2LLKmm` and `Bu2LLKee` stripping lines. The important features of these lines are

- PID requirements on daughters,
- daughters are significantly displaced from the primary vertex,
- the reconstructed B candidate invariant mass falls within ± 1500 MeV of its PDG mass, and

- a small difference B direction of flight and momentum vector, since partially reconstructed candidates will have missing momentum.

The selections in `Bu2LLKmm` and `Bu2LLKee` stripping lines can be found in Appendix C.

4.2.4 Simulation Sample

LHCb simulates pp collisions with PYTHIA (4; 5), particle decays with EVTGEN (6) and radiative processes with PHOTOS (7). Particle interactions with the LHCb dipole magnet and detector materials are simulated by GEANT4 (8; 9; 10). The charge and energy deposits in the LHCb subdetectors are digitised by a dedicated software called BOOLE (11), developed by the LHCb collaboration. Once the simulated events are digitised by BOOLE and triggered by an `L0` trigger emulation, the LHCb trigger and reconstruction software treats it just like real data, with a small caveat that trigger prescales are removed to enhance the statistics of simulated decay of interest. The LHCb simulation has an optional *filtering* process at the end which applies appropriate stripping cuts to the simulated events. The sole purpose of the filtering is to save disk space by removing simulated events that would fail the stripping cuts anyway [8]. The LHCb simulation uses different simulation parameters, such as the proton-proton collision center or detector description, in each data-taking year. Additionally, the simulated events in each year use a trigger configuration unique to that year [9]. These two reasons motivate the data-driven corrections (Sec. 4.4.1) to treat each simulated year separately, and the integrated $r_{J/\psi}$ (Chap. 4.7.1) checks the efficiency calibration on a per-year basis.

4.3 Selections

This analysis uses an assortment of selections, which can be broken down into

---

[8] Take for example as case where you need 10000 events to model a signal shape. If the full selection is 1% efficient you would need to generate 1 million simulated events. Now if the stripping is 30% efficient you can save most of this extra disk space by storing only 300,000 of these events, discarding the other 700,000.

[9] Throughout data-taking, LHCb can vary the trigger threshold, change trigger algorithms or even add new trigger lines. These changes can happen multiple times throughout a year. The LHCb simulation, however, uses the configurations that collected the most data for each year.

- generic cuts [10],
- exclusive background vetoes,
- MVA classifiers, and
- the HOP cut.

Although the MVA classifiers and the HOP cut falls under event selection, they rely on fully corrected simulation. Therefore, the analysis workflow is to

1. select events with the generic cuts and exclusive background vetoes,
2. correct the simulation to represent real data,
3. train and optimise MVA classifiers, and
4. optimise the HOP variable cut

For brevity the trigger, stripping, generic selections, and exclusive background cuts [11] are collectively called the 'pre-MVA selections'.

4.3.1 Generic Selections

The full list of generic cuts is summarised in Tab. 4.3.

The calibration fiducial cuts use the same selection as the LHCb-wide PID calibration samples, which aligns the phase space of the data sample to that of the PID calibration samples. The acceptance fiducial cut not only removes electrons outside the ECAL acceptance, it also vetoes events falling into an ECAL region with a known hole, a region poorly modelled in simulation. With respect to the stripping cuts (Tab. C.1), the K^{*0} invariant mass cut is further tightened to remove non-resonant $K^+\pi^-$ daughters. Since the B^0 selections apply a K^{*0} $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ cut, this requirement is mirrored onto the B^+ modes hadron system with a K^+ $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$ requirement. A pair of clone tracks can mimic a signal decay if they are reconstructed as a signal candidate. The clone tracks mainly come from redundancies in the LHCb long track reconstruction [12] and they typically share the same VELO track segments. Because of this overlap, the VELO segments

---

[10] Generically applied to most, if not all, real and simulated events

[11] Basically every cut before the MVA.

[12] Recall that we have two long track reconstructions, one matches VELO tracks to the downstream tracker while the other extrapolates track segments in the tracker back into the VELO

of clone tracks are highly collinear. Thus, we require all reconstructed daughter tracks to have an angular separation of at least 0.5 mrad from each other.

The PID selections suppress misidentified daughter candidates, which are particularly dangerous when the invariant mass of misidentified B decays overlaps with the signal. The selections employ both the likelihood based and neural network based PID variables. The likelihood based variables $\texttt{DLL}(K-\pi)$ and $\texttt{DLL}(e-\pi)$ quantifies the change in RICH reconstruction likelihoods when the kaon and electron mass candidates assume the pion mass. The neural network based variables

Type		Requirement
Fiducial, Calibration	Multiplicity	`nSPDHits` $< 600(450)$ RUN 1 (RUN 2)
	all tracks	$\chi^2_{track}/ndf < 3$ $\chi^2_{IP} > 9$ `hasRICH==1` `GhostProb` < 0.4
	K, π	$p_{\mathrm{T}} > 250\,\mathrm{MeV}/c$ $p > 2000\,\mathrm{MeV}/c$ `InAccMuon==1`
	e	$p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$ $p > 3000\,\mathrm{MeV}/c$ `hasCalo==1`

	μ	$p_{\mathrm{T}} > 800\,\mathrm{MeV}/c$
		$p > 3000\,\mathrm{MeV}/c$
		`InAccMuon==1`
		`IsMuon==1`

Fiducial, Acceptance	e	$!(x_{ECAL} < 363.6\,\mathrm{mm}\ \&\&\ y_{ECAL} < 282.6\,\mathrm{mm})$ $\texttt{region}_{ECAL} \geq 0$
K^{*0}	K^{*0}	$\lvert m(K\pi) - m^{PDG}_{K^{*0}}\rvert < 100\,\mathrm{MeV}$ $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$
	K (B^+ modes only)	$p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$
Clones	all tracks	$\theta(\ell_{1,2}, h) > 0.5\,\mathrm{mrad}$ $\theta(\ell_1, \ell_2) > 0.5\,\mathrm{mrad}$
PID	K	$\texttt{DLL}(K-\pi) > 0$ $\texttt{ProbNNk} \cdot (1 - \texttt{ProbNNp}) > 0.05$
	π	$\texttt{ProbNNpi} \cdot (1 - \texttt{ProbNNk}) \cdot (1 - \texttt{ProbNNp}) > 0.1$
	μ	$\texttt{ProbNNmu} > 0.2$
	e	$\texttt{DLL}(e-\pi) > 2$ $\texttt{ProbNNe} > 0.2$

Table 4.3 – The generic selections used by this analysis.

`ProbNNk`, `ProbNNpi`, `ProbNNp`, `ProbNNmu` and `ProbNNe` can be interpreted as the probability the track is a kaon, pion, proton, muon and electron respectively. Despite the PID selections, some misidentified backgrounds will survive the generic selections. Thus, further vetoes are designed exclusively to remove these backgrounds if possible. Otherwise, the backgrounds which survive the full selection are modelled in the mass fits.

For the J/ψ and $\psi(2S)$ resonant modes, this analysis uses the q^2 cuts defined in Tab. 4.4. Notice that the ee modes have a much larger q^2 windows. This is because bremsstrahlung radiation significantly smears the $m(ee)$ distribution, and a looser q^2 cut is required to select resonant ee mode events with reasonable efficiency. A drawback of the wider q^2 window is the $J/\psi \to ee$ modes have a higher contamination from rare mode decays compared to the $J/\psi \to \mu\mu$, which can enhance the $J/\psi \to ee$ mode yield and bias the measurement of $r_{J/\psi}$.

q^2 bin	$\mathbf{J/\psi}$	$\psi\mathbf{(2S)}$
ee	$6 < m(ee) < 11\,\mathrm{GeV}^2$	$11 < m(ee) < 15\,\mathrm{GeV}^2$
$\mu\mu$	$\lvert m(\mu\mu) - m_{J/\psi}^{PDG}\rvert < 100\,\mathrm{MeV}$	$\lvert m(\mu\mu) - m_{\psi(2S)}^{PDG}\rvert < 100\,\mathrm{MeV}$

Table 4.4 – The q^2 bin definitions for the resonant modes.

The size of the rare mode contamination in $J/\psi \to ee$ can be estimated with the branching fractions of $B \to K^{(*)} J/\psi\,(\to e^+e^-)$ decays (2) and the differential branching fraction measurements of $B^0 \to K^{*0}\mu^+\mu^-$ (12) and $B^+ \to K^+\mu^+\mu^-$ (13) decays. Although the branching fraction of $B \to K^{(*)}\mu\mu$ around the $q^2 = J/\psi$ resonance are impossible to measure, a linear extrapolation from neighbouring bins can give an approximate estimate of its size. In order to extrapolate these branching fraction measurements from the rare $\mu\mu$ modes to the rare ee modes, a LFUV of $R_X = 0.7$ is assumed [13]. The efficiencies between rare ee and J/ψ mode are assumed to be the same to simplify this calculation. Put together, these calculations estimate a 0.13 % contamination in $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ and a 0.43 % contamination in $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$. Since these sub-percent bias are smaller than equivalent systematics in $r_{J/\psi}$ measurements, no systematic uncertainty is considered for rare ee mode contamination.

---

[13] This assumption is in line with the experimental measurements for R_{K^*} reported by LHCb and it is conservative for the R_K measurement by LHCb, summarised in Chap. 2.4.3. Note that $R_X = 0.7$ is a conservative assumption, as a higher value of R_X, $\frac{\mathcal{B}(B\to X\mu\mu)}{\mathcal{B}(B\to Xee)}$, equates to less rare ee mode contamination.

4.3.2 Exclusive Background Selections

Decays of b hadrons can be misconstructed as $B \to K^{(*)}\ell^+\ell^-$ candidates by daughter misidentification, partial reconstruction or over-reconstruction. Since partially reconstructed backgrounds lose particles in the reconstruction, their invariant mass curve lies below the signal region. Because these backgrounds have similar kinematics to the rare or control modes, it is difficult to efficiently veto them. Instead, these backgrounds are modelled in the mass fit. The exceptions to this are the semileptonic $B \to D$ cascades which, as we shall see, can be vetoed efficiently. Mis-identified backgrounds are particularly dangerous because their daughters, although misidentified, are fully reconstructed and these backgrounds tend to lie under the signal peak region. The same argument applies to overly reconstructed backgrounds where a random particle is added to the candidate B decay. These two types of background are either vetoed by exclusive selections or accounted for in the invariant mass fits.

In this section, some variables are computed with a `DecayTreeFitter` (14) approach. `DecayTreeFitter` reconstructs a decay tree by fitting a χ^2 defined from the full decay tree and external constraints, and iteratively minimises the χ^2 with a Kalman filter (15) [14]. The reconstructed invariant masses in this analysis are all reconstructed by the `DecayTreeFitter` with vertex constraints, i.e. the daughters are constrained to originate from the same point in space. Additionally, because the resonant mode leptons decay from a charmonium resonance, the B masses of these decays can be reconstructed with a J/ψ or $\psi(2S)$ invariant mass constraint, which are denoted as $m^{DTF}_{J/\psi}$ and $m^{DTF}_{\psi(2S)}$.

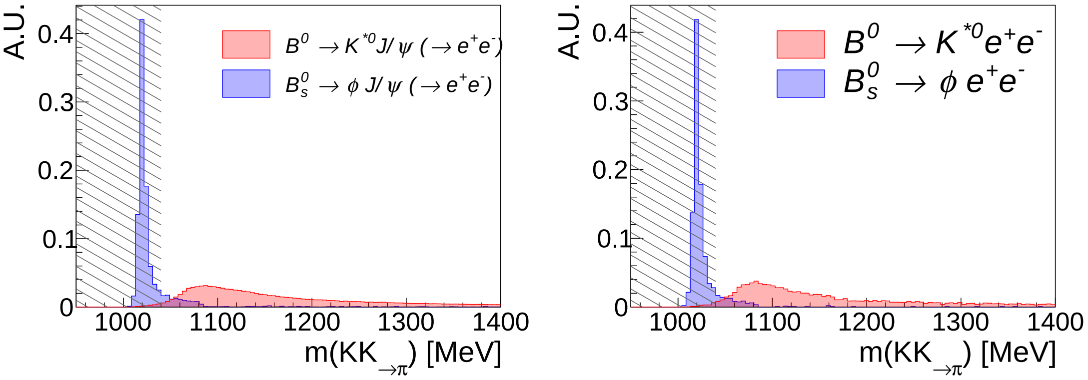
Most of the backgrounds in this section are due to particle misidentification and these backgrounds are vetoed by changing the mass hypothesis of the misidentified particle. The notation for mass hypothesis is better explained by an example, so let's consider a $\phi \to KK$ decay misidentified as a $K^{*0} \to K\pi$ candidate. The reconstructed ϕ mass veto needs to swap the π candidate mass hypothesis to the K mass, which will be denoted as $m(KK_{\to\pi})$.

---

[14] The other approach, 'leaf-by-leaf', decay tree fit takes a bottom-up approach, where the most downstream vertices are first reconstructed and the information propagated upwards until the full decay tree is reconstructed. The disadvantage of this method is that upstream information, like known invariant masses of intermediate particles, is not available downstream.

Backgrounds in $K^{*0}\ell\ell$ final states

$B^0 \to \phi\ell^+\ell^-$ backgrounds are due to $K \to \pi$ misidentification. Although the cut around the $m_{K^{*0}}$ invariant mass removes a significant amount of this background, we apply a further $m(KK_{\to\pi}) < 1040\,\mathrm{MeV}$ cut if the pion candidate does not have a high pion probability, $\pi_{\texttt{ProbNNpi}} < 0.8$. Notice in Fig. 4.3 that although this cut fully removes the sharp $\phi \to KK$ peak, there are some remaining events above 1040 MeV due to higher-mass KK contributions. The surviving backgrounds are modelled in the mass fits.



Background Rejection : 89.40 ± 0.51 %
Signal Efficiency : 98.77 ± 0.02 %

Background Rejection : 90.53 ± 1.09 %
Signal Efficiency : 98.52 ± 0.08 %

Figure 4.3 – The $\phi \to KK$ veto region (shaded) in simulated signal (red) and background (blue). The background on the left is $B_s^0 \to \phi(J/\psi \to ee)$ decay while the control mode $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ decay is the signal. On the right is $B_s^0 \to \phi ee$ decay as the background and $B^0 \to K^{*0}e^+e^-$ as the signal.

Semileptonic cascade decays, such as

- $B^0 \to (D^{*-} \to (\overline{D}^0 \to K^+\pi^-)\pi^-)\ell^+\nu_\ell$,
- $B^0 \to (\overline{D}^0 \to K^+\pi^-)\pi^-\ell^+\nu_\ell$,
- $B^0 \to (D^- \to (K^{*0} \to K^+\pi^-)\pi^-)\ell^+\nu_\ell$,

decays can be misidentified as $B^0 \to K^{*0}\ell^+\ell^-$ candidates due to $\pi \to \ell$ misidentification in the D decay chain. Because the neutrino carries missing momentum, semileptonic backgrounds are strongly suppressed by `DIRA` cuts on the B^0 candidate, defined as the angle between the B direction vector (the direction from its production, or primary, vertex to the decay, or secondary, vertex) and B momentum vector (by summing the momenta of its daughters). Despite the `DIRA` cut, this decay still leaks into our data sample due to its large branching fraction. This background is further suppressed with a PID requirement of `ProbNNmu > 0.8` or `ProbNNe > 0.8` within a 30 MeV window around the D^0 (D^-) mass in the $m(K\pi_{\to\ell})(m(K\pi_{\to\ell}))$. When calculating the D masses, bremsstrahlung corrections are removed from the electron momentum (the resulting momentum is called the track momentum). In Fig. 4.4, we can see that the D peak is highly suppressed while the signal retention is more than 98%. There is a residual D peak in the $\mu\mu$ modes, mainly due to $\pi \to \mu\nu$ decays in-flight.

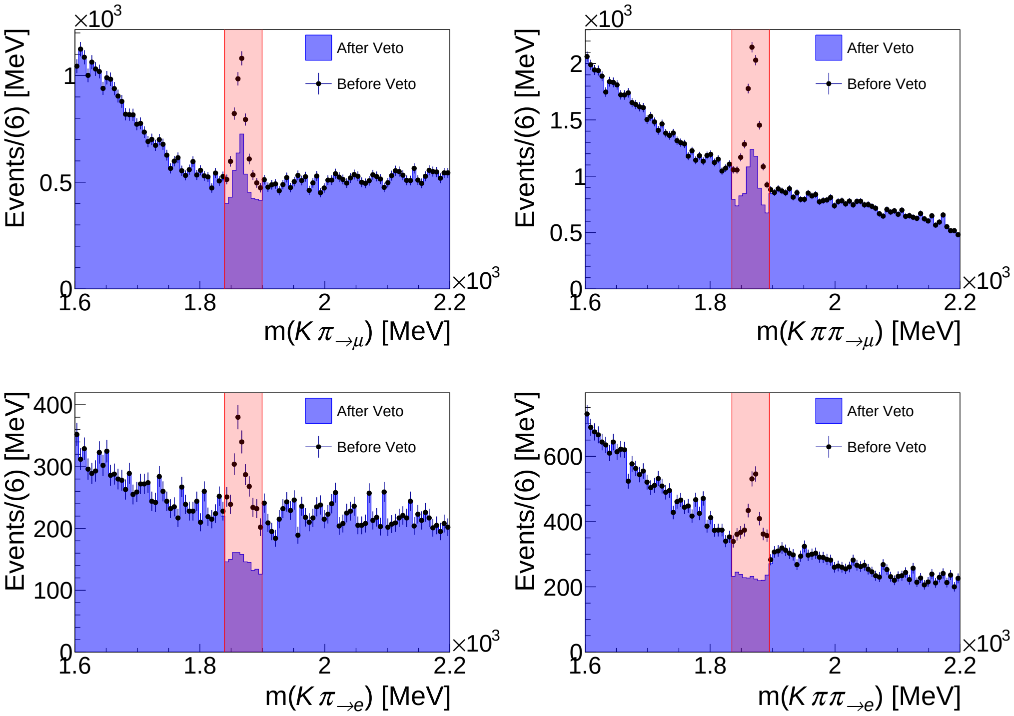
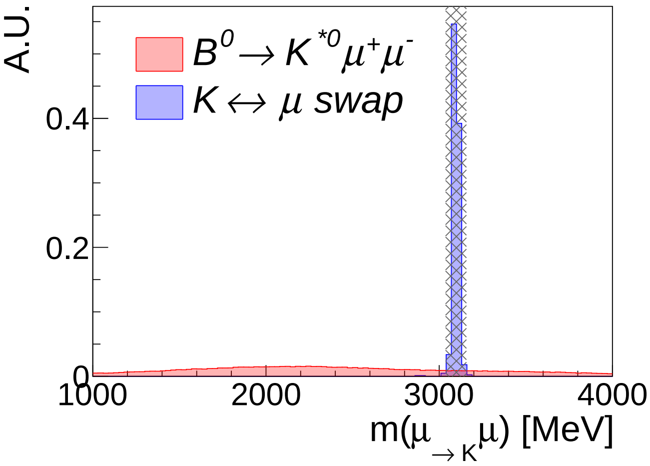


Figure 4.4 – Effect of the semileptonic cascade veto around the D mass (red) on Run 2 data in central-q^2 for $\mu\mu$ (top) and ee (bottom) modes. The distribution of $m(K\pi_{\to\ell})$ (left) and $m(K\pi\pi_{\to\ell})$ (right) has a D peak before (black marker) the veto, which is suppressed after (blue fill) the semileptonic cascade cuts.

***h* ↔ *ℓ* swaps** are caused by a double misidentification in the $B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-)$ or $B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$ daughters. Since one of the ℓ daughters is misidentified, this background no longer peaks in the J/ψ mass window and it can contaminate the rare modes. For muon candidates with `ProbNNmu` < 0.8, we reconstruct $m(\mu_{\to K}\mu)$ and $m(\mu\mu_{\to\pi})$ and veto events around the J/ψ and $\psi(2S)$ resonance. Electrons, however suffer from bremsstrahlung radiation which smears its mass resolution and render such a J/ψ or $\psi(2S)$ veto [15] inefficient. Rather, we reconstruct the $m(K_{\to e}\pi e_{\to K} e)$ and $m(K\pi_{\to e} e e_{\to\pi})$ with a J/ψ or $\psi(2S)$ invariant mass constraint on the dilepton system [16] and apply a veto around the $m^{DTF}_{J/\psi,\psi(2S)}(B^0)$ mass. These vetoes are shown in Fig. 4.5 and Fig. 4.6.

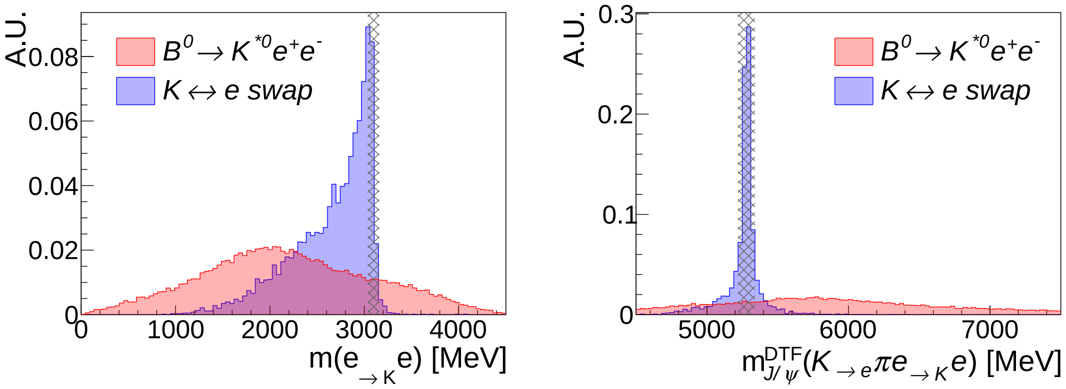


Background Rejection : 98.90 ± 0.40 %
Signal Efficiency : 96.16 ± 0.12 %

Figure 4.5 – The veto (shaded) against $K \leftrightarrow \mu$ swaps cuts $\pm 60\,\mathrm{MeV}$ around the J/ψ mass for muon candidates with low muon probabilities. The effect of this is evaluated on $B^0 \to K^{*0}\mu^+\mu^-$ signal (red) in the central-q^2 and $B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ $K \leftrightarrow \mu$ swap background (blue).

---

[15] With the analogous `ProbNNe` < 0.8 requirement as well.
[16] Using a `DecayTreeFitter`



Background Rejection : 18.40 ± 3.45 %
Signal Efficiency : 93.89 ± 0.19 %

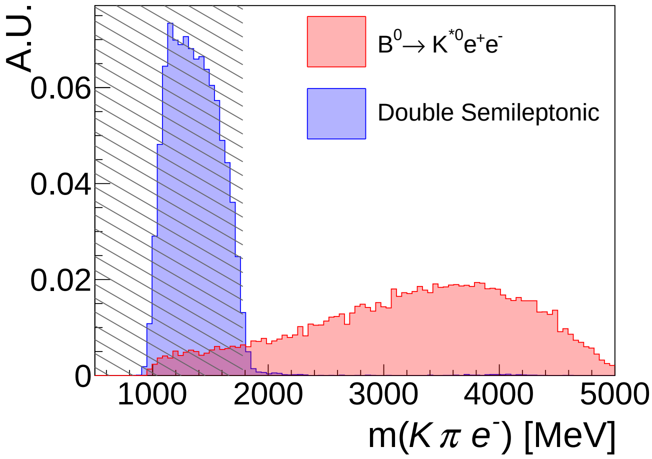
Background Rejection : $85.71^{+7.33}_{-7.41}$ %
Signal Efficiency : 97.98 ± 0.10 %

Figure 4.6 – A similar veto against $K \leftrightarrow e$ swaps around J/ψ mass (shaded area, left plot) is inefficient due to Bremsstrahlung radiation. Instead, we apply a veto around the B^0 mass in the J/ψ constrained $m^{DTF}_{J/\psi}(K_{\to e}\pi e_{\to K}e)$ reconstructed masses (shaded area, right plot). This veto is compared with $K \leftrightarrow e$ swapped $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ backgrounds (blue) against $B^0 \to K^{*0}e^+e^-$ (red) signal in the central-q^2 .

$\mathbf{B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)\ell^-\overline{\nu}_\ell)\ell^+\nu_\ell}$ double semileptonic cascade final state contains the same charged daughters as the B^0 modes. Hence, this background can easily survive the PID and K^{*0} selection in Tab. 4.3. Coupled with its large branching fraction, this background has a sizeable contribution at the low $m(K^+\pi^-\ell^+\ell^-)$ region due to the missing neutrinos. Nevertheless, the intermediate D^- decays semileptonically and a $m(K^+\pi^-\ell^-) > 1780\,\text{MeV}$ veto can suppress this background, as shown in Fig. 4.7.

$\boldsymbol{B^+ \to K^+\ell^+\ell^-}$ decays, can be over-reconstructed as $B^0 \to K^{*0}\ell^+\ell^-$ candidates when a slow π is added from the underlying event. A $m(K^+\ell^+\ell^-) < 5100\,\text{MeV}$ veto, which is approximately the mass difference between a B^0 and π, suppresses these events while preserving $B^0 \to K^{*0}\ell^+\ell^-$ signal events. This veto also account for the $B^+ \to K^+\ell^+\ell^-$ kaon misidentified as pion candidate, by cutting on $\texttt{max}\Big(m(K^+{}_{\to\pi^+}\ell^+\ell^-), m(K^+\ell^+\ell^-)\Big) < 5100\,\text{MeV}$

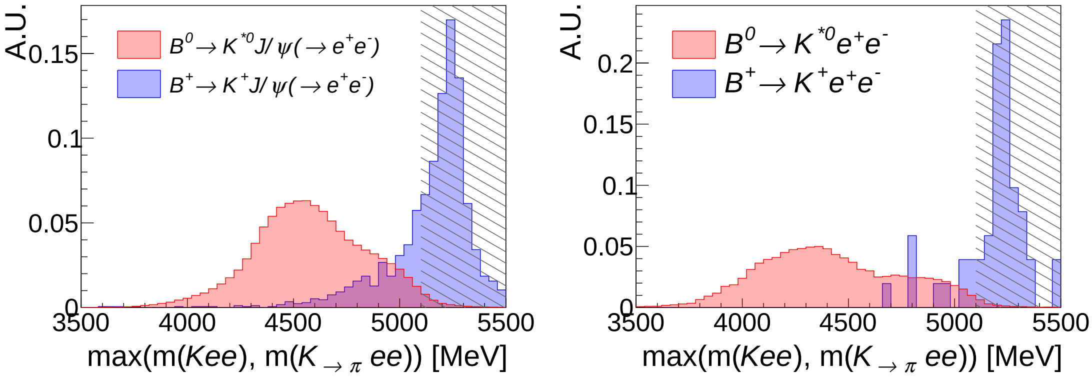
The effect of this veto is shown in Fig. 4.8.



Background Rejection : 97.28 ± 0.35 %

Signal Efficiency : 93.87 ± 0.16 %

Figure 4.7 – The $m(K^+\pi^-\ell^-) > 1780\,\text{MeV}$ veto (shaded) applied to $B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)\ell^-\overline{\nu}_\ell)\ell^+\nu_\ell$ backgrounds (blue) and $B^0 \to K^{*0}e^+e^-$ signals (red) in the central-q^2 bin.



Background Rejection : 71.29 ± 1.60 %

Signal Efficiency : 98.20 ± 0.03 %

Background Rejection : $75.00^{+7.59}_{-7.61}$ %

Signal Efficiency : 98.27 ± 0.09 %

Figure 4.8 – Left: The $\mathtt{max}\Big(m(K^+{}_{\to\pi^+}\ell^+\ell^-), m(K^+\ell^+\ell^-)\Big) < 5100\,\text{MeV}$ veto (shaded) applied against simulated $B^+ \to K^+J/\psi(\to e^+e^-)$ backgrounds (blue) against simulated $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ signal (red). The same distribution is shown on the right, but with simulated $B^+ \to K^+e^+e^-$ backgrounds against simulated $B^0 \to K^{*0}e^+e^-$ signal in the central-q^2 region.

Background	Requirement
lBs $\to \phi\ \ell^+\ell^-$	$!\Big(m(KK_{\to\pi}) < 1040\,\text{MeV}\ \&\&\ \pi_{\texttt{ProbNNpi}} < 0.8\Big)$
$B^0 \to (D^0 \to K\pi)\pi^-\ell^+\nu$	$!\Big(\lvert m(K^+\ell^-{}_{\to\pi}) - m^{PDG}(D^0)\rvert < 30\,\text{MeV}\ \&\&\ \ell_{\texttt{ProbNNl}} < 0.8\Big)$
$B^0 \to (D^- \to K\pi\pi)\ell^+\nu$	$!\Big(\lvert m(K^+\pi^+\ell^-{}_{\to\pi}) - m^{PDG}(D^-)\rvert < 30\,\text{MeV}\ \&\&\ \ell_{\texttt{ProbNNl}} < 0.8\Big)$
$h \leftrightarrow \mu$ swap	$!\Big(\lvert m(\mu_{\to h}\mu) - m^{PDG}(J/\psi, \psi(2S))\rvert < 60\,\text{MeV}\ \&\&\ \mu_{\texttt{ProbNNmu}} < 0.8\Big)$
$h \leftrightarrow e$ swap	$!\Big(\lvert m^{DTF}_{J/\psi,\psi(2S)}(h_{\to e}he_{\to h}e) - m^{PDG}(B^0)\rvert < 60\,\text{MeV}\ \&\&\ e_{\texttt{ProbNNe}} < 0.8\Big)$
$B^0 \to (D^- \to K^{*0}\ell\nu)\ell\nu$	$m(K\pi\ell^-) < 1780\,\text{MeV}$ (low and central q^2 only)
$B^+ \to K^+\ell^+\ell^-$	$max(m(K\ell\ell), m(K_{\to\pi}\ell\ell)) < 5100\,\text{MeV}$

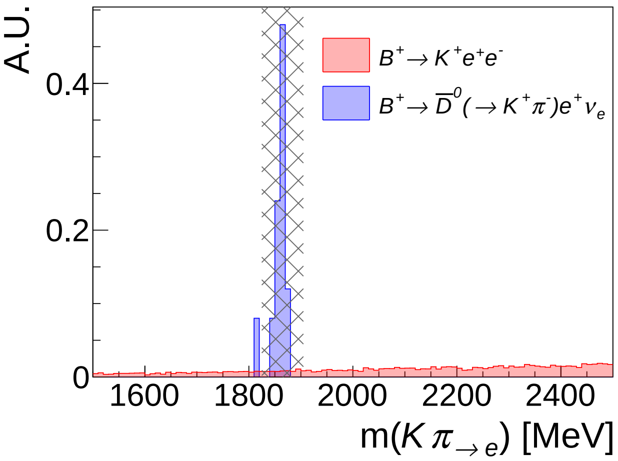
Table 4.5 – Summary of the exclusive background selection requirements in B^0 modes.

Backgrounds in $K^+\ell^+\ell^-$ final states

$\mathbf{B^+ \to \overline{D}^0(\to K^+\pi^-)\ell^+\nu}$ decays are suppressed with a tight `ProbNNl` > 0.8 PID requirement on lepton candidates within a 40 MeV window around the D^0 mass in the $m(K\pi_{\to e})$ spectrum. This selection is showin in Fig. 4.9.

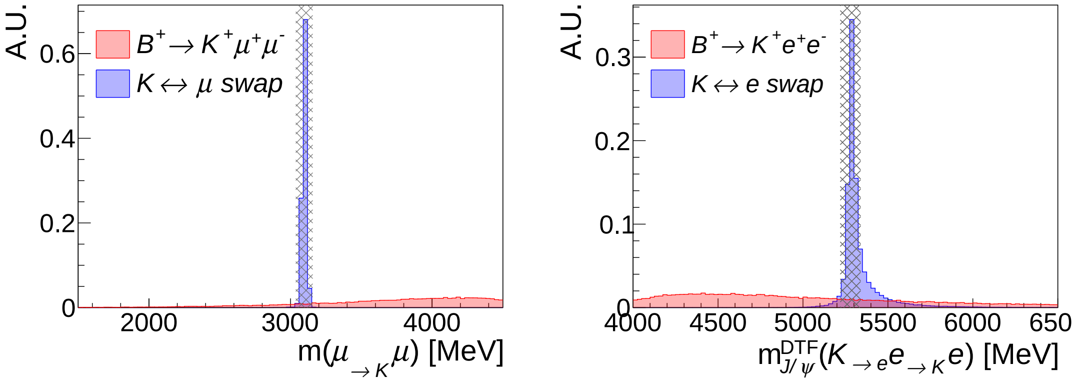
$\mathbf{K \leftrightarrow \ell}$ swap backgrounds are similar to $h \leftrightarrow \ell$ swap backgrounds in reconstructed $K^{*0}\ell\ell$ candidates. We apply analogous vetoes against lepton candidates with `ProbNNl` < 0.8, one around the J/ψ or $\psi(2S)$ masses in reconstructed $m(\mu\mu_{\to K})$ for the muon modes and another around the B^+ mass in reconstructed $m^{DTF}_{J/\psi,\psi(2S)}(K_{\to e}ee_{\to K})$ for the electron modes. The vetoes are shown in Fig. 4.10.

$\mathbf{B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu}$ **and** $\mathbf{B^+ \to (\overline{D}^0 \to K^+\ell^-\nu)\pi^+}$ backgrounds are suppressed with a $m(K\ell) < 1885\,\text{MeV}$ veto against the D^0. This is similar to the $B^0 \to D^-(\to K^{*0}(\to K^+\pi^-)\ell^-\overline{\nu}_\ell)\ell^+\nu_\ell$ background veto in $K^{*0}\ell\ell$ candidates. The veto is shown in Fig. 4.11.



Background Rejection : $91.86^{+3.51}_{-3.53}$ %
Signal Efficiency : 98.38 ± 0.17 %

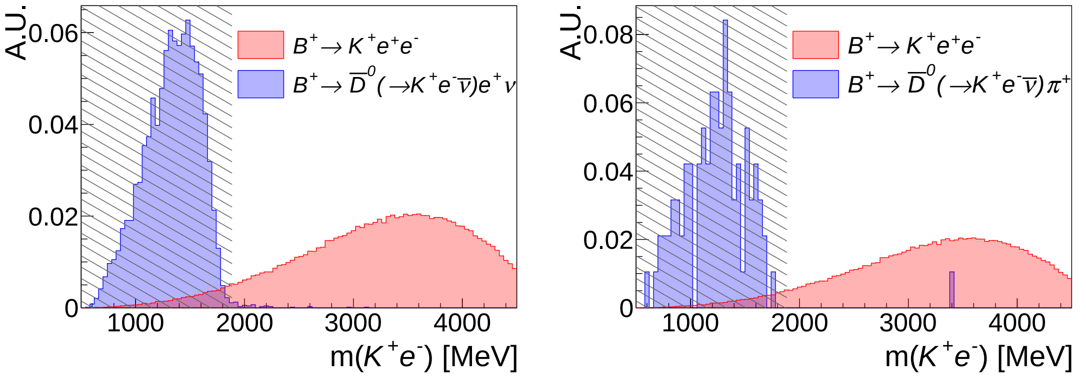
Figure 4.9 – A tight PID selection is appplied against $B^+ \to (\overline{D}^0 \to K^+\pi^-)e^+\nu_e$ backgrounds (blue) a $\pm 40\,\text{MeV}$ region (shaded) around the D^0 mass in reconstructed $m(K\pi_{\to e})$. These events are simulated in the central-q^2 region and simulated $B^+ \to K^+e^+e^-$ events are used as the signal (red).



Background Rejection : 57.14 ± 4.41 %
Signal Efficiency : 99.88 ± 0.01 %

Background Rejection : $83.33^{+10.14}_{-10.35}$ %
Signal Efficiency : 99.24 ± 0.06 %

Figure 4.10 – On the left, a veto around the J/ψ mass (red) suppress simulated $B^+ \to K^+J/\psi\,(\to \mu^+\mu^-)$ backgrounds (black) due to $K \leftrightarrow \mu$ but retains $B^+ \to K^+\mu^+\mu^-$ signals (blue) in the central-q^2 region. On the right, a veto around B^+ mass in reconstructed $m^{DTF}_{J/\psi}(K_{\to e}ee_{\to K}$ (red) is used instead for the electron modes, where $B^+ \to K^+J/\psi\,(\to e^+e^-)$ backgrounds (black) are vetoed against $B^+ \to K^+e^+e^-$ signals (blue) in the central-q^2 region.



Background Rejection : $96.55^{+1.17}_{-1.18}$ %
Signal Efficiency : 95.61 ± 0.15 %

Background Rejection : $95.83^{+3.42}_{-3.53}$ %
Signal Efficiency : 95.61 ± 0.15 %

Figure 4.11 – The $m(K\ell) < 1885\,\text{MeV}$ cut (shaded) vetoes $B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu$ (left) and $B^+ \to (\overline{D}^0 \to K^+\ell^-\nu)\pi^+$ (right) backgrounds (blue). The signal in both plots are simulated $B^+ \to K^+e^+e^-$ (red) decays in the central-q^2 region.

Background	**Requirement**
$B^+ \to (\overline{D}^0 \to K^+\pi^-)\ell^+\nu$	$!\Big(\lvert m(K^+\pi^-{}_{\to\ell^-}) - m^{PDG}(D^0)\rvert < 40\,\text{MeV} \;\&\&\; \ell_{\texttt{ProbNNl}} < 0.8\Big)$
$K \leftrightarrow \mu$ swap	$!\Big(\lvert m(\mu_{\to K}\mu) - m^{PDG}(J/\psi, \psi(2S))\rvert < 60\,\text{MeV} \;\&\&\; \mu_{\texttt{ProbNNmu}} < 0.8\Big)$
$K \leftrightarrow e$ swap	$!\Big(\lvert m^{DTF}_{J/\psi,\psi(2S)}(K_{\to e}e_{\to K}e) - m^{PDG}(B^+)\rvert < 60\,\text{MeV} \;\&\&\; e_{\texttt{ProbNNe}} < 0.8\Big)$
$B^+ \to (\overline{D}^0 \to K^+\ell^-\overline{\nu})\ell^+\nu$	$m(K\ell^-) < 1885\,\text{MeV}$ (low and central q^2 only)

Table 4.6 – Summary of the exclusive background selection requirements in B^+ modes.

4.3.3 Multivariate Classifiers

Although the specific vetoes described above can efficiently remove misidentified backgrounds or partially reconstructed semileptonic cascade backgrounds, a multivariate (MVA) classifier is able to further suppress combinatorial and partially reconstructed backgrounds. This analysis uses the `CatBoost` [17] algorithm (16) in conjunction with the Reproducible Experiment Platfrom (17).

`CatBoost` is a gradient boosted decision tree (18) algorithm. The decision tree classifier begins with a single node (root) containing the full dataset. The dataset is branched into exclusive nodes by a simple cut. This process iterates, until the decision tree nodes arrive at the terminal nodes (leaves), which will be the categorial predictions of the algorithm. Gradient boosting involves building an ensemble of decision trees in an iterative process. At every iteration, a new tree is added to the ensemble in order to minimise the loss function of the algorithm [18]. After a number of iterations, the `CatBoost` training will have an ensemble of decision trees, and the gradient boosted prediction is the weighted summed of predictions from individual trees in the ensemble. A schematic of this is shown in Fig. 4.12.

Combinatorial background classifiers (MVA_{COMB}) are trained for both muon and electron modes while partially-reconstructed background classifiers ($\text{MVA}_{\text{PRECO}}$)

---

[17] **Cat**egory **boost**ing.

[18] Essentially, the tree added at each step tries to compensate for the residuals, defined as the difference between classifier prediction and true value. Drawing an analogy to likelihood minimization, likelihood minimization fits converge to the optimal solution by iteratively optimizing its floating parameters while gradient boosting (in decision trees) converge by iteratively growing a new tree and adding it to the ensemble.

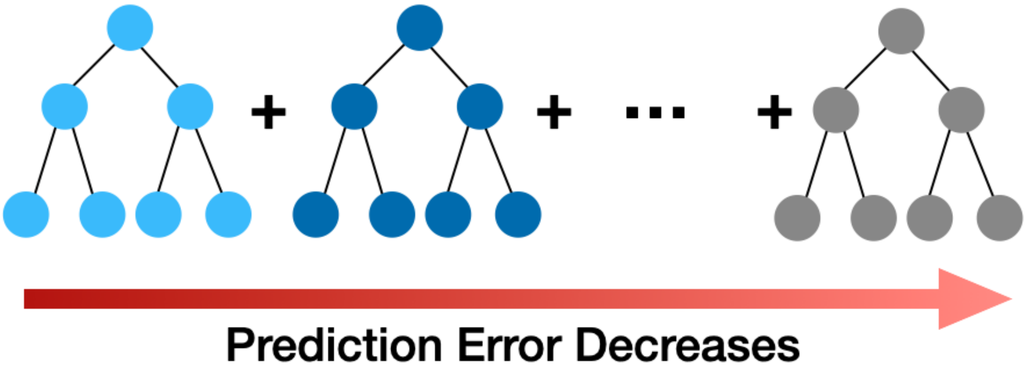


Figure 4.12 – Schematic of how decision trees are added linearly in gradient boosting. The training builds the ensemble tree-by-tree, and each added tree reduces the prediction errors of the ensemble.

are only trained for the electron modes. Separate classifiers are trained for $\mu\mu$ and ee modes, for reconstructed B^+ and B^0 decays, and in three separate data-taking periods (Run 1, 15+16, 17+18). This leads to a total of 12 distinct combinatorial classifiers and 6 additional partially-reconstructed background classifiers for the electron modes. Both MVA_{Comb} and $\text{MVA}_{\text{PReco}}$ use simulated $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ events as their signal proxies. Background proxies depend on the MVA and B mode in question, and we shall discuss this later. Regardless, these signal and background proxies are required to pass the pre-MVA selections we have discussed so far.

Due to limited statistics, the MVA training combines low-q^2 and central-q^2 bins into the same training sample. It is only later during the classifier response cut optimisation stage that each q^2 bins are treated separately. Furthermore, the training phase downsamples the signal sample to have equal amounts of signal and background events. The downsampling avoids an imbalanced training sample, which can bias the classifier performance on the smaller category [19].

The full dataset is split into exclusive sets of training and test sets. MVA classifiers are meant to learn a general set of features from their training sample and extrapolate these learned features to unseen data. Furthermore, seeing test

---

[19] Consider the simple case were a classifier is asked to optimise its accuracy but your training sample is 99% signal and 1% background. If the classifer always classifies the samples as ‘signal’, it would still be 99% accurate. Real-world cases of imbalanced samples are of course more complex but the idea is the same.

set examples during training would bias the MVA predictions during test. This analysis splits and resamples the dataset using a *k-fold* cross-validation approach, where $k = 10$. The *k-fold* method first splits the dataset into 10 exclusive subsets. The training and test then resamples the dataset, where each resampling trains on 9/10 of the full dataset and the remaining 1/10 of the dataset is used as test. This dataset is resampled 10 times, until the test set has cycled through the 10 exclusive subsets.

$\mathbf{MVA_{Comb}}$

The MVA_{COMB} classifier uses events from the upper mass sideband in real data as background proxies. This is defined with a $m(B) > 5400\,\text{MeV}$ cut in the $\mu\mu$ modes and a tighter $m(B) > 5600\,\text{MeV}$ cut in the electron modes because of bremsstrahlung smearing. A consequence of the tighter cuts in ee modes is a significantly reduced training size. In order to enhance the background proxy statistics, the background proxy selection relaxes the K^{*0} mass window from $\pm 100\,\text{MeV}$ around the peak (see Tab. 4.3) to $\pm 200\,\text{MeV}$.

The MVA_{COMB} chooses its input variables via a pruning strategy. We first select a large set of variables with noticeable differences between signal and background distributions. Then, we iteratively train a classifier using the full set of variables, rank them according to importance, and remove the least important variable. This is repeated until the area under the ROC curve changes by more than 1%. In the end, 23 variables remain for the B^0 modes while 16 varibles remain for the B^+ modes. The MVA_{COMB} input variables are listed in Tab. 4.7.

Post-training, the ROC curves are checked for classifier performance. The ROC curves between the ten *k-fold*s of $15 + 16$ MVA_{COMB} classifiers are shown in Fig. 4.13. In the plots, the ROC curves between different *k-fold*s are consistent with one another. The areas under the ROC curves are all above 0.98, regardless of *k-fold*, lepton mode or B mode, which suggests very good MVA_{COMB} performance. The classifier performance is compared between training and test set in Fig. 4.14. This comparison checks for signs of overtraining and verifies that the classifier is able to extrapolate its learned features into unseen data. From the plots, the performance between the training and test set agrees very well, and no sign of overtraining is found.

Particle	**Variables**
B^0, B^+	$p_{\mathrm{T}},\ \chi^2_{\text{IP\_OWNPV}},\ \chi^2_{\text{FD\_OWNPV}},\ \chi^2_{\text{vtx}}/\text{ndf},\ \chi^2_{\text{DTF}}/\text{ndf},\ \text{DIRA}$
K^{*0} (B^0 mode)	$p_{\mathrm{T}},\ \chi^2_{\text{IP\_OWNPV}},\ \chi^2_{\text{FD\_OWNPV}},\ \chi^2_{\text{vtx}}/\text{ndf},\ \text{DIRA}$
K (B^+ mode)	$p_{\mathrm{T}},\ \chi^2_{\text{IP\_OWNPV}}$
$\ell^+\ell^-$	$p_{\mathrm{T}},\ \chi^2_{\text{IP\_OWNPV}},\ \chi^2_{\text{FD\_OWNPV}},\ \chi^2_{\text{vtx}}/\text{ndf},\ \text{DIRA}$
h	$min, max(p_{T,K},\ p_{T,\pi}),\ \ min, max\ (\chi^2_{\text{IP\_OWNPV},K},\ \chi^2_{\text{IP\_OWNPV},\pi})$
ℓ	$min, max(p_{T,\ell^+},\ p_{T,\ell^-}),\ \ min(\chi^2_{\text{IP\_OWNPV},\ell^+},\ \chi^2_{\text{IP\_OWNPV},\ell^-})$

Table 4.7 – Summary of the input variables to MVA_{Comb} for the $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$ training. The only differences between the B^0 and B^+ modes are from the variables of their hadron system.

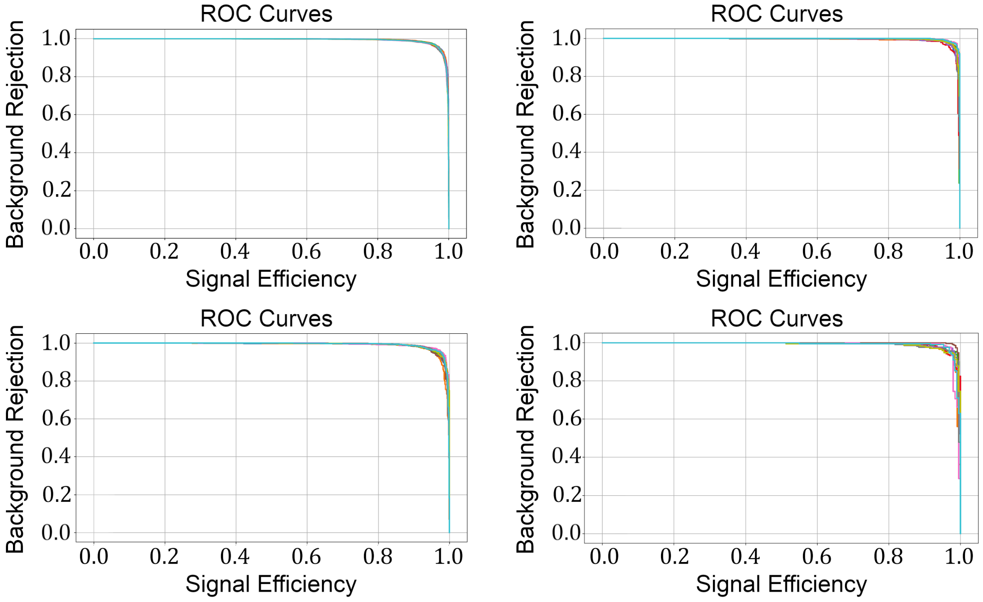


Figure 4.13 – 15 + 16 MVA_{Comb} ROC curves for $\mu\mu$ (top) and ee (bottom) final states in B^+ (left) and B^0 (right) modes. The ten differently coloured curves represents the ten different *k-folds*. Figures courtesy of a collaborator in this analysis.

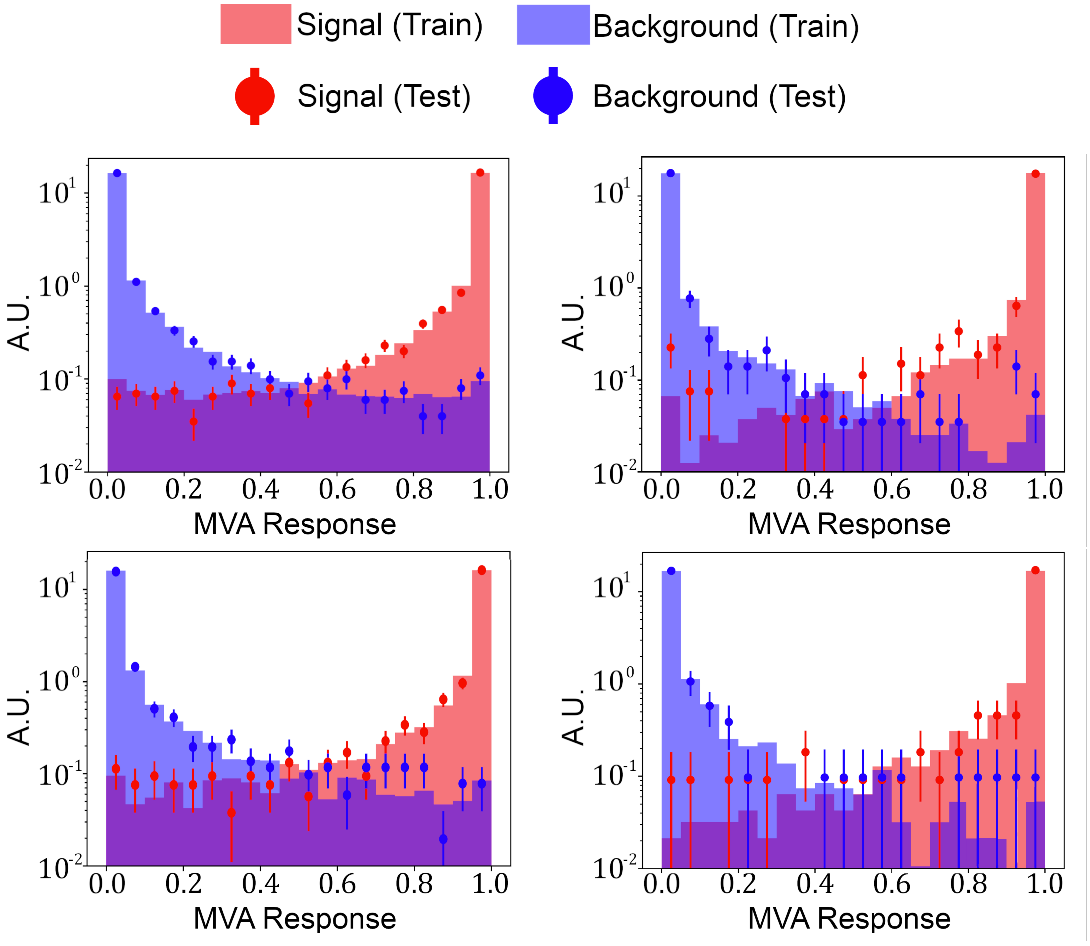


Figure 4.14 – Comparison of $\mathrm{MVA}_{\text{Comb}}$ response distribution between training and validation sample, for $\mu\mu$ (top) and ee (bottom) final states in B^+ (left) and B^0 (right) modes. The $\mathrm{MVA}_{\text{Comb}}$ performance between training (filled histogram) and validation (coloured marker) sets agrees for both signal (red) and background (blue) proxies. The ee modes have less statistics, hence the larger statistical uncertainties. Figures courtesy of a collaborator in this analysis.

Regarding classifier inputs, the most important variable for B^0 modes is the $\chi^2/ndof$ of the decay tree fit to B^0 invariant mass, followed by the B^0 candidate p_{T}. Likewise, the most important variables for B^+ modes is the B^+ invariant mass decay tree fit $\chi^2/ndof$. This is followed by the kaon candidate χ^2_{IP} for the $\mu\mu$ mode and the B^+ candidate p_{T} for the ee mode.

MVA$_{\mathbf{PReco}}$

The MVA$_{\mathrm{PReco}}$ uses simulated decays of $B^0 \to K^{*0}e^+e^-$ as background proxies of $B^+ \to K^+e^+e^-$ candidates, and simulated $B^+ \to K\pi\pi e^+e^-$ as background proxies of $B^0 \to K^{*0}e^+e^-$ candidates. These partially reconstructed decays only lose a pion, and their reconstructed B mass distribution overlaps with the signal events. This problem is exacerbated by the resolution smearing due to bremsstrahlung radiation. Therefore, it is important that these backgrounds are suppressed by the MVA$_{\mathrm{PReco}}$ classifier.

Just like MVA$_{\mathrm{Comb}}$, MVA$_{\mathrm{PReco}}$ prunes the input variables iteratively, and the final 14 variables used in the B^0 and B^+ modes classifiers are summarised in Tab. 4.8. The isolation variables compute reconstructed quantities within a 0.5 mrad cone around the particle of interest. 'MULT' varibles denote the multiplicity within the isolation cone excluding the particle of interest. 'SPT' variables denote the scalar sum of track p_{T} within the cone of interest excluding the particle of interest. 'IT' variables denote how much p_{T} within the cone is attributed to the particle of interest.

Table 4.8 – Summary of the input variables to MVA$_{\mathrm{PReco}}$ for B^0 trainings

Particle	**Variables**
B^0	$\chi^2_{\mathrm{IP\_OWNPV}}$, $\chi^2_{\mathrm{DTF}}/\mathrm{ndf}$, $\chi^2_{\mathrm{vtx}}/\mathrm{ndf}$, DIRA, vertex isolation one-track-χ^2, vertex isolation one-track-mass
K^{*0}	DIRA, $\chi^2_{\mathrm{vtx}}/\mathrm{ndf}$, $\chi^2_{\mathrm{IP\_OWNPV}}$
$\ell^+\ell^-$	DIRA, $\chi^2_{\mathrm{IP\_OWNPV}}$, $\chi^2_{\mathrm{FD\_OWNPV}}$
h	$min(p_{T,K},\ p_{T,\pi})$
l	charged-cone isolation min(MULT l^+, MULT l^-)

Table 4.9 – Summary of the input variables to $\mathrm{MVA}_{\mathrm{PReco}}$ for B^+ trainings

Particle	**Variables**
B^+	p_{T}, $\chi^2_{\mathrm{IP\_OWNPV}}$, DIRA, vertex isolation one-track-χ^2, vertex isolation one-track-mass
$\ell^+\ell^-$	DIRA, $\chi^2_{\mathrm{IP\_OWNPV}}$, $\chi^2_{\mathrm{FD\_OWNPV}}$
K	p_{T}
l	$min(\chi^2_{\mathrm{IP\_OWNPV},\ell^+}, \chi^2_{\mathrm{IP\_OWNPV},\ell^-})$, charged-cone isolation min, max(APT l^+, APT l^-) charged-cone isolation max(MULT l^+, MULT l^-) charged-cone isolation min(IT l^+, IT l^-)

The 15 + 16 ROC curves for B^0 and B^+ $\mathrm{MVA}_{\mathrm{PReco}}$ classifiers are shown in Fig. 4.15. Compared to $\mathrm{MVA}_{\mathrm{Comb}}$, the area under ROC curve is significantly reduced, due to the fact that partialy reconstructed backgrounds and signals have similar phase space distribution. In Fig. 4.16, the classifier performances between training and test sets show no signs of overtraining. For both B^0 and B^+ classifiers, the B vertex isolation one-track-mass is the most discriminating variable followed by B vertex isolation one-track-χ^2.

These B vertex isolation algorithm iteratively loops over all reconstructed charged tracks in the underlying event, and computes the changes in B vertex reconstruction by adding a single underlying event track to the B vertex. The B vertex isolation one-track-χ^2 variable is then defined as the smallest vertex fit χ^2, out of all the B vertex reconstruction with an added underlying event track. Calculation of the B vertex isolation one-track-mass variable first selects the single track that changes the B vertex fit χ^2 the least [20]. It then recomputes the invariant mass of the B vertex by adding the momentum of said track to the B daughter candidates.

---

[20] Note the subtle difference that the track which results in the smallest vertex fit χ^2 can be different from the track that causes the least change vertex fit χ^2.

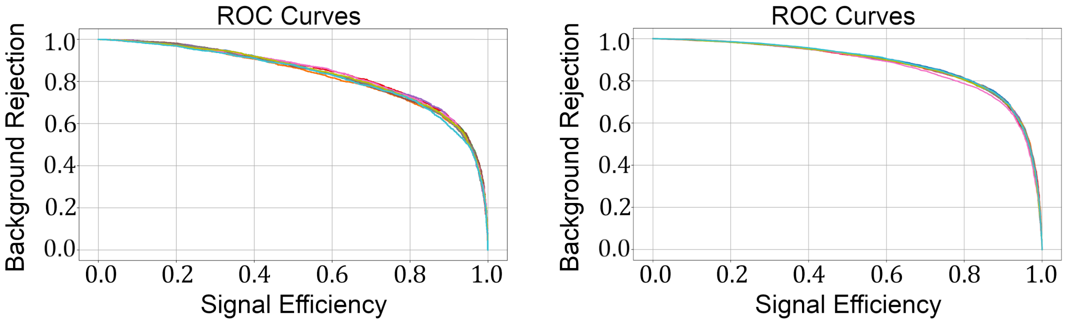


Figure 4.15 – $\mathrm{MVA}_{\mathrm{PReco}}$ ROC curves for ee final states in B^+ (left) and B^0 (right) modes. The ten differently coloured curves represents the ten different *k-folds*. Figures courtesy of a collaborator in this analysis.

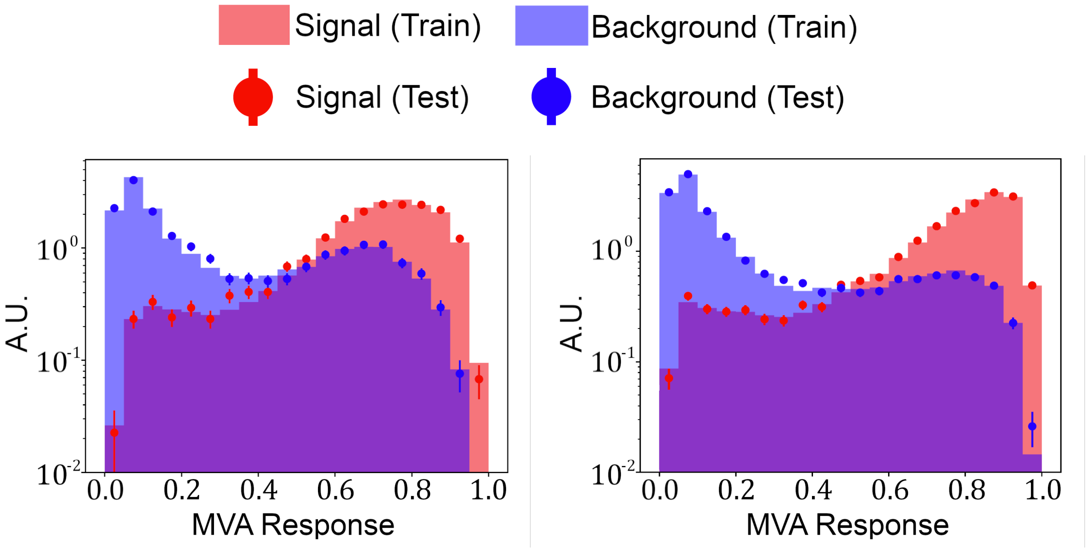


Figure 4.16 – Comparison of $\mathrm{MVA}_{\mathrm{PReco}}$ response distribution between training and validation sets for B^+ (left) and B^0 (right) modes. The $\mathrm{MVA}_{\mathrm{PReco}}$ performance between training (filled histogram) and validation (coloured marker) sets agrees for both signal (red) and background (blue) proxies. Figures courtesy of a collaborator in this analysis.

MVA Response Cut Optimisation

The MVA response cut is a trade-off between maximising background rejection and minimising signal loss. In order to find the optimal MVA working point, this analysis maximises a figure of merit, defined as the signal significance,

$$\frac{N_S}{\sqrt{N_S + N_B}}. \tag{4.12}$$

N_S and N_B is the number of signal and background, respectively, which survives the MVA response cut and falls into the signal B mass window. This signal window is $\pm 50\,\mathrm{MeV}$ around the B meson mass for $\mu\mu$ modes and $5150 - 5350\,\mathrm{MeV}$ for the ee modes. The larger window in ee modes accounts for bremsstrahlung radiation.

The calculation of N_S first involves estimating an expected signal yield after the pre-MVA selections, which is

$$N_{expected} = N_{control} \times \frac{\mathcal{B}_S}{\mathcal{B}_{control}} \times \frac{\epsilon_S}{\epsilon_{control}} \tag{4.13}$$

where the subscript $_{control}$ denotes the control mode with the same B mother and $\ell^+\ell^-$ final state as the signal mode. $N_{control}$ is determined from a fit to data. ϵ_S and $\epsilon_{control}$ are the efficiencies of the signal and control modes to survive their pre-MVA selections, which are determined from simulation. Finally, $\mathcal{B}_S$ and $\mathcal{B}(B \to K^{(*)} J/\psi (\to \ell\ell))$ are the branching fractions of the signal and control modes respectively. Note that $\mathcal{B}_S$ for the rare modes is different in the low-q^2 and central-q^2 regions. The number of expected signal events pre-MVA, is then multiplied by the MVA cut efficiency to get N_S,

$$N_S = N_{expected} \times \frac{N_{MC|MVA\ cut}}{N_{MC}} \tag{4.14}$$

where N_{MC} and $N_{MC|MVA\ cut}$ are the number of simulated events before and after the MVA response cut respectively.

Determination of N_B involve fits to real data, which are configured differently between q^2 bins and $\ell^+\ell^-$ modes:

- **Low- and central $-$ $\mathbf{q^2}$ bins $\mu\mu$ modes** optimise $\mathrm{MVA_{COMB}}$ in 1D. They determine N_B by fitting the combinatorial background with an exponential function in the lower and upper mass sidebands of reconstrcuted B mass.

- **J/ψ $\rightarrow \mu\mu$ modes** optimise MVA$_{\text{Comb}}$ in 1D. These modes fit the lower and upper mass sidebands of J/ψ constrained B mass and model the combinatorial background with an exponential background to determine N_B. However, the lower sideband contains partially reconstructed background [21] which is modelled with a `RooKeysPdf` kernel density estimator (KDE). The `RooKeysPdf` shape is built from an inclusive simulation of $B \rightarrow X J/\psi (\rightarrow \mu\mu)$ decays, which considers B^0, B^+ and B_s^0 decays. More details about this inclusive background is given in Chap. 4.5.6.

- **ψ(2S) $\rightarrow \mu\mu$ modes** optimise MVA$_{\text{Comb}}$ in 1D. They determine N_B by fitting an exponential function to the combinatorial background in the upper mass sidebands of $\psi(2S)$ constrained B mass.

- **low $-$ $\mathbf{q^2}$ee modes** optimise MVA$_{\text{Comb}}$ and MVA$_{\text{PReco}}$ in 2D. The fits to determine N_B consider the full mass fit window because the signal and partially reconstructed background overlaps in the B mass distribution. The signal is modelled with a double sided crystal ball (DSCB) function, the combinatorial with an exponential and the partially reconstructed is modelled by a `RooKeysPdf`, which is built from simulated $B^0 \rightarrow K^{*0} e^+ e^-$ decays in B^+ mode and simulated $B^+ \rightarrow K\pi\pi e^+ e^-$ decays in B^0 mode.

- **central $-$ $\mathbf{q^2}$ee** modes optimise MVA$_{\text{Comb}}$ and MVA$_{\text{PReco}}$ in 2D. Similar to low-q^2 ee modes, the fits to N_B consider the full mass fit window. It uses the considers the same signal and background component as the low-q^2 ee modes and an additional control mode leakage background, due to $J/\psi \rightarrow ee$ events leaking into the central-q^2 cut. The leakage is modelled with a `RooKeysPdf` built from simulated control mode $J/\psi \rightarrow ee$ decays.

- **J/ψ $\rightarrow$ ee modes** optimise MVA$_{\text{Comb}}$ and MVA$_{\text{PReco}}$ individually in two steps of 1D scans. Unlike the $J/\psi \rightarrow \mu\mu$ modes, the fits to determine N_B considers the full reconstructed B mass spectrum because the signal and partially reconstructed background overlaps. Hence, it considers the same backgrounds as $J/\psi \rightarrow \mu\mu$ modes, but the signal is modelled with using a linear combination of three DSCB functions.

- **ψ(2S) $\rightarrow$ ee modes** optimise MVA$_{\text{Comb}}$ in 1D. Similar to $\psi(2S) \rightarrow \mu\mu$ mode fits, N_B is determined by fitting an exponential function to the combinatorial

---

[21] We only optimise for MVA$_{\text{Comb}}$. The partially reconstructed background yield is a nuisance parameter.

background in the upper mass sidebands of $\psi(2S)$ constrained B mass.

The MVA training stage separates the dataset into 12 distinct groups, according to B^0 and B^+ modes, $\mu\mu$ and ee modes, and three different data-taking periods (Run 1, $15+16$, $17+18$). The MVA response cut optimisation further splits the dataset into four q^2 bins, which gives a total of 48 distinct optimisation. Fig. 4.17 shows an example of a 1D MVA_{Comb} optimisation and an example of the 2D MVA_{Comb}-$\text{MVA}_{\text{PReco}}$ optimisation. The optimal cuts for each category is given in Tab. 4.10.

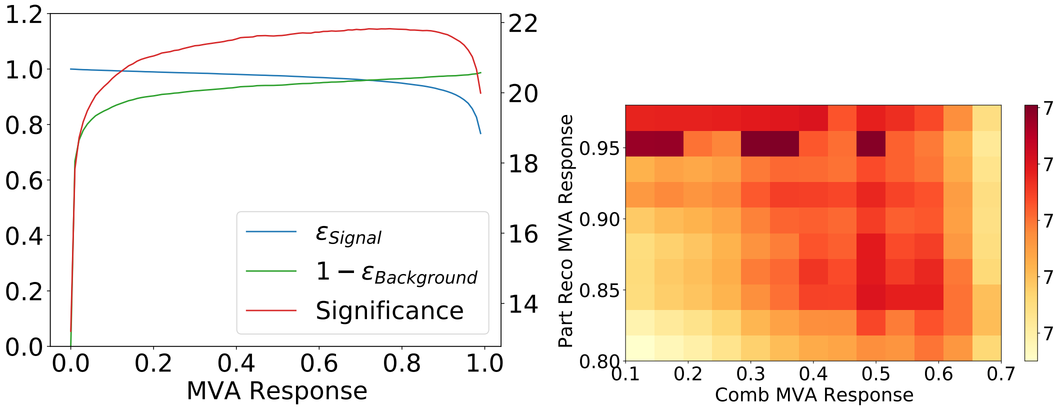


Figure 4.17 – Left: The 1D MVA_{Comb} response scan to 15+16 $B^0 \to K^{*0}\mu^+\mu^-$ candidates in the central-q^2 region, showing the evolution of significance (red), signal efficiency (blue) and background rejection (green) along the MVA_{Comb} response. Right: The 2D MVA_{Comb}-$\text{MVA}_{\text{PReco}}$ response scan to $15+16$ $B^0 \to K^{*0}e^+e^-$ candidate significance in the central-q^2 region.

<table>
<tr><th colspan="3">Optimisation</th><th>Run 1</th><th>15 + 16</th><th>17 + 18</th></tr>
<tr><td rowspan="8">B^+</td><td rowspan="4">ee</td><td>low-q^2</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.90$ && $\text{MVA}_{\text{PReco}} > 0.40$</td></tr>
<tr><td>central-q^2</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.90$ && $\text{MVA}_{\text{PReco}} > 0.40$</td></tr>
<tr><td>J/ψ</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.10$ && $\text{MVA}_{\text{PReco}} > 0.05$</td></tr>
<tr><td>$\psi(2S)$</td><td>$\text{MVA}_{\text{Comb}} > 0.47$</td><td>$\text{MVA}_{\text{Comb}} > 0.36$</td><td>$\text{MVA}_{\text{Comb}} > 0.59$</td></tr>
<tr><td rowspan="4">$\mu\mu$</td><td>low-q^2</td><td>$\text{MVA}_{\text{Comb}} > 0.70$</td><td>$\text{MVA}_{\text{Comb}} > 0.85$</td><td>$\text{MVA}_{\text{Comb}} > 0.85$</td></tr>
<tr><td>central-q^2</td><td>$\text{MVA}_{\text{Comb}} > 0.70$</td><td>$\text{MVA}_{\text{Comb}} > 0.80$</td><td>$\text{MVA}_{\text{Comb}} > 0.80$</td></tr>
<tr><td>J/ψ</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.05$</td></tr>
<tr><td>$\psi(2S)$</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.05$</td></tr>
<tr><td rowspan="8">B^+</td><td rowspan="4">ee</td><td>low-q^2</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.50$ && $\text{MVA}_{\text{PReco}} > 0.50$</td></tr>
<tr><td>central-q^2</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.90$ && $\text{MVA}_{\text{PReco}} > 0.40$</td></tr>
<tr><td>J/ψ</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.20$ && $\text{MVA}_{\text{PReco}} > 0.05$</td></tr>
<tr><td>$\psi(2S)$</td><td>$\text{MVA}_{\text{Comb}} > 0.30$</td><td>$\text{MVA}_{\text{Comb}} > 0.55$</td><td>$\text{MVA}_{\text{Comb}} > 0.67$</td></tr>
<tr><td rowspan="4">$\mu\mu$</td><td>low-q^2</td><td>$\text{MVA}_{\text{Comb}} > 0.29$</td><td>$\text{MVA}_{\text{Comb}} > 0.54$</td><td>$\text{MVA}_{\text{Comb}} > 0.55$</td></tr>
<tr><td>central-q^2</td><td>$\text{MVA}_{\text{Comb}} > 0.63$</td><td>$\text{MVA}_{\text{Comb}} > 0.77$</td><td>$\text{MVA}_{\text{Comb}} > 0.64$</td></tr>
<tr><td>J/ψ</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.05$</td></tr>
<tr><td>$\psi(2S)$</td><td colspan="3">$\text{MVA}_{\text{Comb}} > 0.05$</td></tr>
</table>

Table 4.10 – Summary of the optimal MVA_{Comb} and $\text{MVA}_{\text{PReco}}$ cuts determined by the MVA response optimisation.

4.3.4 HOP cut optimisation

The HOP variable [22] is a pseudo correction for the momentum loss in $b \to s\,ee$ electrons via bremsstrahlung radiation. A schematic of the relevant variables is presented in Fig. 4.18.

In the figure above, the B decays into an electron pair, X_e, and a hadronic system, Y_h. The B flight direction is reconstructed by vertex reconstruction while the B momentum vector, $\overrightarrow{P}(B)$, is a vector sum of the momentum from its final state particles, calculated from track reconstruction. The momentum of final

---

[22] Developed by Martino Borsato, Marie-Helene Schune, Francesco Polci. A document internal to LHCb members exists but no public documents are available for citation.

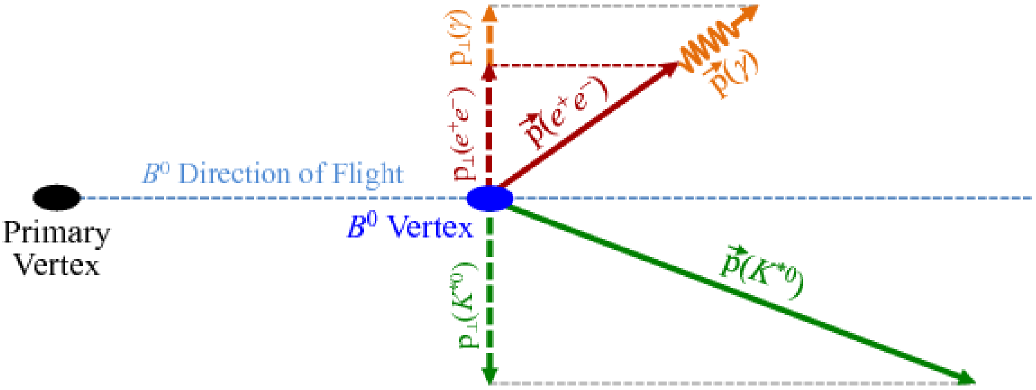


Figure 4.18 – Schematic of $B \rightarrow Y_h X_e$ decays, highlighting the relevant quantities. Figure extracted from Ref. (19).

state particles can be grouped into the di-electron momentum $\overrightarrow{P}(X_e)$ and hadron system momentum $\overrightarrow{P}(Y_h)$. The momentum components transverse to the B flight direction are labelled $\overrightarrow{p_\mathrm{T}}(X_e)$ and $\overrightarrow{p_\mathrm{T}}(Y_h)$ respectively.

By conservation of momentum, the di-electron and hadron system momentums has to balance out in the direction transverse to B flight, i.e. $\overrightarrow{p_\mathrm{T}}(X_e) = -\overrightarrow{p_\mathrm{T}}(Y_h)$. However, the measured value of $\overrightarrow{p_\mathrm{T}}(X_e)$ are significantly smeared due to bremsstrahlung radiation. Despite that, bremsstrahlung radiation in electrons are driven by material interaction with the LHCb detector and the photons emitted are collinear to the electrons. Hence, a correction ratio

$$\alpha_{HOP} = \frac{p_\mathrm{T}(Y_h)}{p_\mathrm{T}(X_e)} \tag{4.15}$$

calculated by balancing the di-electron and hadron system transverse momenta can be used to correct the total momentum of the di-electron system

$$\overrightarrow{P^{corr}}(X_e) = \alpha_{HOP} \times \overrightarrow{P}(X_e) \tag{4.16}$$

due to the collinearity between electrons and their bremsstrahlung photons.

Then, a HOP corrected B mass, m_{HOP}, can be calculated using $\overrightarrow{P^{corr}}(X_e)$. Unfortunately, the quality of m_{HOP} depends on the resolution on the B flight direction. As a result, m_{HOP} is correlated to the vertex reconstruction resolution and the angle between the electron and hadron system, $\theta(X_e)$ in Fig. 4.18. For this reason, electron momenta with bremsstrahlung recovery are favoured over $\overrightarrow{P^{corr}}$

to reconstruct the B mass. Yet, m_{HOP} is a strongly discriminating variable that can suppress partially-reconstructed background. This is because missing particles from the hadron system unbalance the measurement of α_{HOP}. As a result, partially reconstructed backgrounds tend towards lower values of m_{HOP} compared to real signal decays, which can be seen in Fig. 4.19.

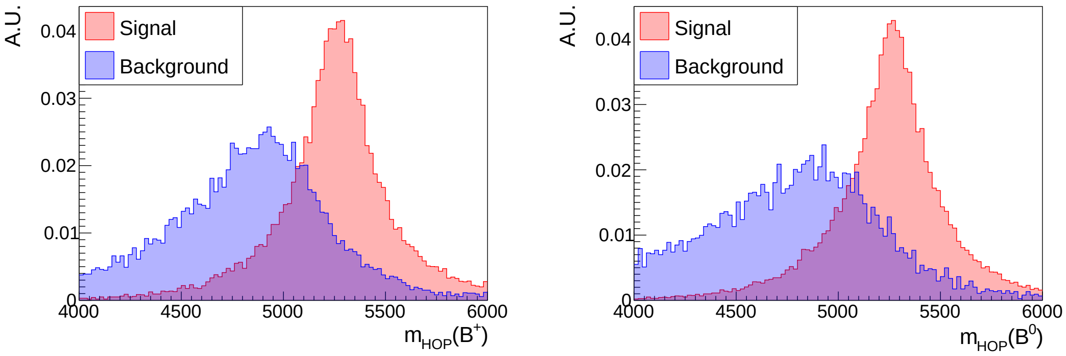


Figure 4.19 – m_{HOP} distribution in signal (red) and partially reconstructed backgrounds (blue) in the central-q^2 bins. The rare B^+ mode (left) signals are simulated $B^+ \to K^+e^+e^-$ decays while its partially reconstructed backgrounds are simulated $B^0 \to K^{*0}e^+e^-$ decays. The rare B^0 mode (right) signals are simulated $B^0 \to K^{*0}e^+e^-$ decays while its partially reconstructed backgrounds are simulated $B^+ \to K\pi\pi e^+e^-$ decays.

The m_{HOP} cut optimisation is similar to the MVA response cut optimisation. Different m_{HOP} cut values are scanned in 1D to profile the signal significance. Two example scans are shown in Fig. 4.20. In these plots, there is a significance plateau around the $4700 - 4900\,\text{MeV}$ range. Because the mass fits perform poorly when the are insufficient background events, this analysis opted for conservative cuts on the significance plateau, which are

$$m_{HOP} < 4800 \;\text{ for } low - q^2,$$
$$m_{HOP} < 4700 \;\text{ for } central - q^2.$$

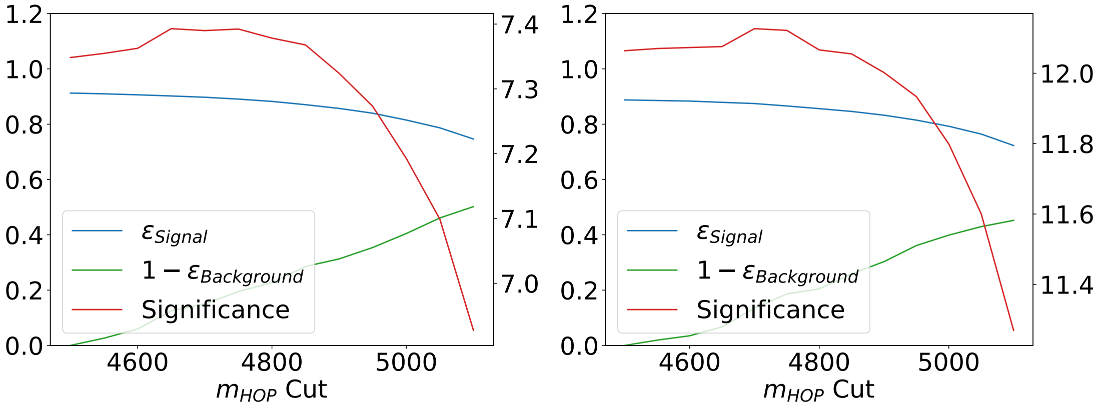


Figure 4.20 – 1D scans of m_{HOP} cuts on $B^0 \to K^{*0}e^+e^-$ (left) and $B^+ \to K^+e^+e^-$ (right) decays in the central-q^2 region, showing the evolution of significance (red), signal efficiency (blue), and background rejection (green) of across m_{HOP} cut values.

4.4 Corrections and Efficiencies

As foreshadowed in the analysis strategy discussion, correcting simulation to represent real data is an important part of this analysis. The corrections are vital for the efficiencies, which will bias the measurement of lepton flavour universality if calibrated poorly. Because LHCb simulation uses a different configuration for each data-taking year, the corrections will be calculated and applied on a per-year basis.

4.4.1 Corrections to Simulation

The corrections calculate weights for the simulated events, and account for different sources of simulation mismodelling:

- $\mathbf{w}_{\text{PID}}$ corrects the PID response.
- $\mathbf{w}_{\mathbf{TRK}}$ corrects the lepton track reconstruction efficiency.
- $\mathbf{w}_{\mathbf{B\&MULT}}$ corrects the B kinematics and underlying event multiplicity.
- $\mathbf{w}_{\text{L0}}$ corrects the `L0` hardware trigger response.
- $\mathbf{w}_{\texttt{HLT}}$ corrects the `HLT` software trigger response.

- $\mathbf{w}_{\text{reco}}$ corrects reconstructed variables of the B candidate, such as $IP\ \chi^2$ and vertex χ^2.
- $\mathbf{w}_{\mathbf{q^2}}$ corrects the q^2 selection efficiency due to mismodelling of bremsstrahlung radiation in ee modes.

w_{PID} and w_{TRK} are calculated using collaboration-wide calibration samples. For the rest, this analysis calculates the corrections with the tree-level control modes. As discussed in analysis strategy (Sec. 4.1), a correlation is induced if the same control mode sample is used to calculate efficiency and extract signal yields. To combat this, the analysis exchanges the correction weight maps between B^+ and B^0 modes.

The corrections above affect one another, hence the full, nominal, correction proceeds sequentially as a chain, illustrated in Fig. 4.21. In other words, the corrections up to the n^{th} step must be applied on simulated samples before the $(n+1)^{th}$ correction step proceeds. Notice, however, that $w_{B\&MULT}$ corrections follows the generator chain instead to arrive at its weights.

Since $w_{B,MULT}$ corrections account for mismodelling effects in the mother B kinematics and underlying event multiplicity, they are expected to be portable between ee and $\mu\mu$ modes. Therefore, $w_{B,MULT}$ are calculated with $B \to K^{(*)} J/\psi\,(\to \mu^+\mu^-)$ decays in the inclusive `L0Muon` category to maximise statistics, and the $w_{B,MULT}$ weights are ported to every candidates in $\mu\mu$ and ee modes including those in `L0I` and `L0L` exclusive categories. However, this strategy to calculate $w_{B,MULT}$ weights only in the inclusive `L0Muon` category requires trigger effects to be accounted for first in $w_{B,MULT}$ corrections, which is not possible in the nominal chain. Hence, the generator chain, as illustrated in Fig. 4.21, first corrects `L0` and `HLT` efficiencies instead before measuring the $w_{B,MULT}$ corrections.

The sequential nature of the corrections are mirrored in this section, which follows the nominal chain flow except for $w_{B\&MULT}$. Since $w_{B\&MULT}$ comes after trigger corrections, it is discussed after `L0` and `HLT` trigger corrections.

$w_{\mathbf{PID}}$ Correction

A unique w_{PID} weight is assigned to each final state particle. The w_{PID} weights calibrate the effect of the PID selections in Tab. 4.3, and w_{PID} can be interpreted as 'the probability this track will pass a PID selection cut, given this track's true particle species'. Effectively, this means pion candidates that are true pions use the correctly identified pion maps while pion candidates that are true kaons use the kaon-to-pion misidentification maps.

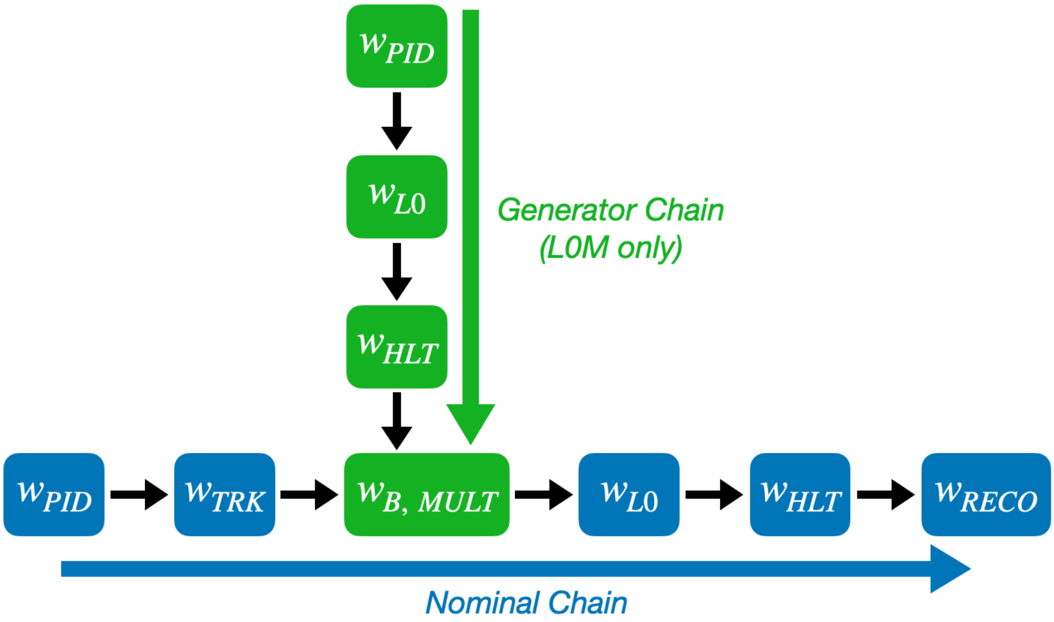


Figure 4.21 – The nominal (blue) and generator (green) correction chains. The generator chain corrects the w_{PID}, w_{L0} and $w_{\texttt{HLT}}$ efficiencies first before accessing the B kinematics and underlying event multiplicity distributions.

Following our discussion about the LHCb detector (Chap. 3.3), the RICH PID performance and muon identification [23] is momentum dependant. Detector elements have higher resolution and experience higher particle flux as they get closer to the beampipe, which induces an η dependence in PID performance. Also, the performance of event reconstruction degrades as the overall detector occupancy increase. For these reasons, w_{PID} corrections are 3D maps that encode PID efficiencies as a function of track momentum, p, track pseudorapidity, η, and a multiplicity proxy, which is the number of tracks reconstructed by LHCb for the event, n_{Tracks}.

The w_{PID} correction maps use the collaboration-wide calibration samples, which have dedicated stripping and Turbo lines in Run 1 and Run 2 respectively. The selections in these stripping and Turbo lines are replicated by the generic selections, presented under the 'Fiducial cuts, calibration' row in Tab. 4.3. The list of calibration samples for each particle species is presented in Tab. 4.11.

---

[23] Recall that muons are identified by their penetration in the muon chambers, and the penetration power increases with particle momentum.

Particle	Sample Decay
K/π	$D^0 \to K^-\pi^+$ from $D^{*+} \to D^0\pi^+$
μ	$J/\psi \to \mu^+\mu^-$ from B decays (Run 1)
μ	prompt $J/\psi \to \mu^+\mu^-$ (Run 2)
e	$J/\psi \to e^+e^-$ from B^+ decays

Table 4.11 – PID calibration samples.

Notice that all the calibration samples in Tab. 4.11 decay into a pair of daughters. This feature is essential to build the 3D w_{PID} correction maps with a tag-and-probe approach. The w_{PID} tag-and-probe selects events by applying strict selection criteria on the tag daughter. The other daughter, being the probe particle, has no PID requirement. The probe particles then evaluate the efficiency of the PID requirements, and the backgrounds are subtracted using an $_s\mathcal{P}lot$ technique.

This method works sufficiently well for the kaons, pions, and muons but not for the electrons. The electron calibration sample has insufficient statistics for a reliable $_s\mathcal{P}lot$ subtraction and the invariant mass reconstructed with electrons correlate with electron momentum and pseudorapidity, a correlation which breaks the validity of $_s\mathcal{P}lot$ background subtraction.

Rather than $_s\mathcal{P}lot$ background subtraction, the electron w_{PID} maps are built using a fit-and-count approach. This approach builds two 3D maps, one for probe electrons with zero recovered bremsstrahlung photon (0γ) and one for probe electrons with at least one recovered photon ($> 1\gamma$). The dataset in each bin are split into probe electrons that fail the electron PID requirements and probe electrons that pass it. A fit to B^+ invariant mass in the failed and passed samples then extracts a signal yield, which is used to calculate an efficiency,

$$\epsilon_{PIDe} = w_{PIDe} = \frac{N_{signal}(passed)}{N_{signal}(failed) + N_{signal}(passed)},$$

that is the w_{PID} weight for the bin. An example fit is shown in Fig. 4.22.

When the w_{PID} weights are accessed from the maps, a linear interpolation between bin centers is used to estimate the w_{PID} values. The method of accessing the w_{PID} weights are explored as a systematic uncertainty in Sec. 4.6.2.

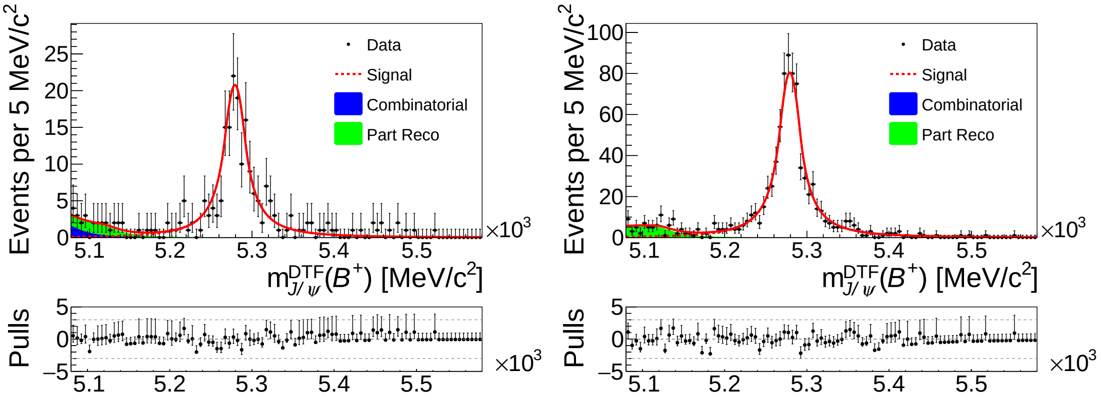


Figure 4.22 – Fits to reconstructed $m(B^+)$ with J/ψ mass constraint, in a $> 1\gamma$ bin. The datasets with probe electrons that failed (left) and passed (right) the PID selection uses a model with a DSCB signal (red dashed), an exponential combinatorial background (blue) and a `RooKeysPdf` partially reconstructed background (green).

$w_{\mathbf{TRK}}$ Corrections

w_{TRK} corrections accounts for mismodelling in electron and muon tracking efficiencies. The K and π tracking efficiencies cancel in the lepton universality ratios, and no w_{TRK} corrections are considered for them. For the leptons, w_{TRK} corrections measure the tracking efficiencies in data and simulation with a tag-and-probe approach described in Ref. (20).

This method measures the probability of reconstructing a lepton as a *long* track given that its *VELO* track is reconstructed. The tag-and-probe method uses $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ decays and the lepton pair are used as the tag and probe particles. The tag lepton is required to be reconstructed as a long track. The tracking efficiency is then measured as the number of probe electrons, reconstructed in the VELO, that pass and fail the long track reconstruction. This involves a fit, in the pass and fail probes, to the $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$ invariant mass reconstructed with a J/ψ mass constraint, since the failed probes provide no momentum information.

The tag-and-probe efficiencies are measured in bins of p_{T}, η and ϕ, which are used to define the per-lepton w_{TRK} weights, defined as

$$w_{TRK}(\ell) = \frac{\epsilon(long|velo)^{data}}{\epsilon(long|velo)^{MC}}. \tag{4.17}$$

The per-event w_{TRK} then assumes the per-lepton w_{TRK} correction factorises,

$$w_{\mathrm{TRK}}(B) = w_{\mathrm{TRK}}(\ell^{+}) \cdot w_{\mathrm{TRK}}(\ell^{-}). \tag{4.18}$$

$w_{\mathtt{L0}}$ Corrections

$w_{\mathtt{L0}}$ are mathematically defined as the efficiency ratios between data and simulation,

$$w_{\mathrm{L0}}^{\mathtt{L0I}} = \frac{\epsilon_{\mathtt{L0I}}^{data}}{\epsilon_{\mathtt{L0I}}^{MC}}, \tag{4.19}$$

$$w_{\mathrm{L0}}^{\mathtt{L0L}} = \frac{\epsilon_{\mathtt{L0L}}^{data}(B)}{\epsilon_{\mathtt{L0L}}^{MC}(B)}, \tag{4.20}$$

$$w_{\mathrm{L0}}^{\mathtt{L0L!}} = w_{\mathtt{L0}}^{\mathtt{L0L}} \cdot \frac{1 - \epsilon_{\mathtt{L0I}}^{data}}{1 - \epsilon_{\mathtt{L0I}}^{MC}}, \tag{4.21}$$

where `L0L` denotes the inclusive `L0L` category (`L0L == 1`) and `L0L`! denotes the exclusive `L0L` category (`!L0I && L0L`).

The `L0I` efficiencies in this section are measured as per-event efficiencies since `L0I` triggers on the underlying event. In contrast, the `L0Muon` and `L0Electron` trigger on individual leptons, thus the per-event `L0L` efficiencies are derived from per-lepton `L0` trigger efficiencies. The derivation of per-event `L0L` efficiencies, $\epsilon_{\mathtt{L0L}}(B)$, assumes that the measured per-particle `L0L` efficiencies, $\epsilon_{\mathtt{L0L}}(\ell)$, are uncorrelated

$$\epsilon_{\mathtt{L0L}}(B) = \epsilon_{\mathtt{L0L}}(\ell^{+}) + \epsilon_{\mathtt{L0L}}(\ell^{-}) - \epsilon_{\mathtt{L0L}}(\ell^{+}) \cdot \epsilon_{\mathtt{L0L}}(\ell^{-}). \tag{4.22}$$

which in probabilistic terms, can be interpreted as $P(A\&B) = P(A) + P(B) - P(A) \cdot P(B)$. This assumption then expresses the per-event `L0L` efficiencies of Eq. 4.20 and Eq. 4.21 in terms of per-lepton efficiencies as

$$w_{\mathrm{L0}}^{\mathtt{L0L}} = \frac{\epsilon_{\mathtt{L0L}}^{data}(\ell^{+}) + \epsilon_{\mathtt{L0L}}^{data}(\ell^{-}) - \epsilon_{\mathtt{L0L}}^{data}(\ell^{+}) \cdot \epsilon_{\mathtt{L0L}}^{data}(\ell^{-})}{\epsilon_{\mathtt{L0L}}^{MC}(\ell^{+}) + \epsilon_{\mathtt{L0L}}^{MC}(\ell^{-}) - \epsilon_{\mathtt{L0L}}^{MC}(\ell^{+}) \cdot \epsilon_{\mathtt{L0L}}^{MC}(\ell^{-})}, \tag{4.23}$$

$$w_{\mathrm{L0}}^{\mathtt{L0L!}} = \frac{\epsilon_{\mathtt{L0L}}^{data}(\ell^{+}) + \epsilon_{\mathtt{L0L}}^{data}(\ell^{-}) - \epsilon_{\mathtt{L0L}}^{data}(\ell^{+}) \cdot \epsilon_{\mathtt{L0L}}^{data}(\ell^{-})}{\epsilon_{\mathtt{L0L}}^{MC}(\ell^{+}) + \epsilon_{\mathtt{L0L}}^{MC}(\ell^{-}) - \epsilon_{\mathtt{L0L}}^{MC}(\ell^{+}) \cdot \epsilon_{\mathtt{L0L}}^{MC}(\ell^{-})} \cdot \frac{1 - \epsilon_{\mathtt{L0I}}^{data}}{1 - \epsilon_{\mathtt{L0I}}^{MC}}. \tag{4.24}$$

Now, measuring the true efficiencies in real data is an impossible task. Instead, the efficiencies in Eq. 4.19, 4.23 and 4.24 are actually TISTOS efficiencies, measured as [24]

$$\begin{aligned}
\epsilon_{\texttt{L0I}}^{data} &= \left(\frac{N_{TIS\&TOS}}{N_{TOS}}\right)^{data}, \qquad & \epsilon_{\texttt{L0L}}^{data}(\ell) &= \left(\frac{N_{TIS\&TOS}}{N_{TIS}}\right)^{data}, \\
\epsilon_{\texttt{L0I}}^{MC} &= \left(\frac{N_{TIS\&TOS}}{N_{TOS}}\right)^{MC}, \qquad & \epsilon_{\texttt{L0L}}^{MC}(\ell) &= \left(\frac{N_{TIS\&TOS}}{N_{TIS}}\right)^{MC}.
\end{aligned}$$

For the remainder the w_{L0} section, `L0` efficiencies implicitly refer to the TISTOS efficiencies. Furthermore, the TISTOS categories are referred as tag (TIS) and probe (TOS) categories instead to avoid ambiguities between the `L0I` trigger [25] and TIS [26] as a tag category.

The w_{L0} corrections use simulated samples and real data that has passed the pre-MVA selections, except for the `L0` requirements. An additional cut ±45 MeV around the J/ψ constrained B mass is applied to further reduced background contamination in real data.

The `L0Muon` hardware trigger fires when a muon passes the p_{T} threshold. This, together with the fact that muon station granularity increases closer to the beam pipe, motivates the measurement of per-lepton `L0Muon` efficiency as a function of $p_{\mathrm{T}}(\mu)$ in three η regions [27]. The boundaries of the three η bins are $[0, 2.75, 3.25, 10]$. The corrections are measured using real and simulated $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ and $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ events. The TISTOS methods uses three tag categories, which are

- **TIS tag** events, triggered by `L0` hadron or `L0` electron candidates from the underlying event.
- **Hadron tag** events, triggered by a hadron from the signal B candidate.
- **Lepton tag** events, triggered by the other muon from the intermediate $J/\psi \to \mu\mu$ decay.

---

[24] The reader can find a summary of the TISTOS method in Chap. 3.3.3 and a TISTOS method study documented in Ref. (21).

[25] `L0I` is technically a **T**rigger **I**ndependent of **S**ignal.

[26] Triggered by a particle not part of the signal candidate.

[27] Remember, the `L0Muon` corrections are calculated per final state μ.

The `L0Muon` efficiency is calculated in bins of p_{T} but it is modelled as a linear combination of an error and a Gompertz function

$$f_1 \cdot \left(1 + erf\left(\frac{x-t}{\sqrt{2}\cdot\sigma_1}\right)\right) + f_2 \cdot s^{exp(\frac{-(x-t)}{\sqrt{2}\cdot\sigma_2})} + a, \tag{4.25}$$

where $f_{1,2}$ denotes the normalisation fractions, $\sigma_{1,2}$ the resolutions, t the trigger threshold, s the skewness of the Gompertz function and a the constant effect. The data and simulation efficiencies can be seen in Fig. 4.23. From the plots, the overall efficiency curve resembles an error function but the asymmetry at the turn on curves is better modelled with contributions from a Gompertz function. The data/simulation ratio in Fig. 4.24 shows a good agreement between the three different tags. The TIS category measurement is used because it has the highest statistic. Also in Fig. 4.24, the TIS tag curves between B^0 and B^+ mode shows good agreement, which validates the portability of `L0Muon` efficiencies between B^+ and B^0 modes.

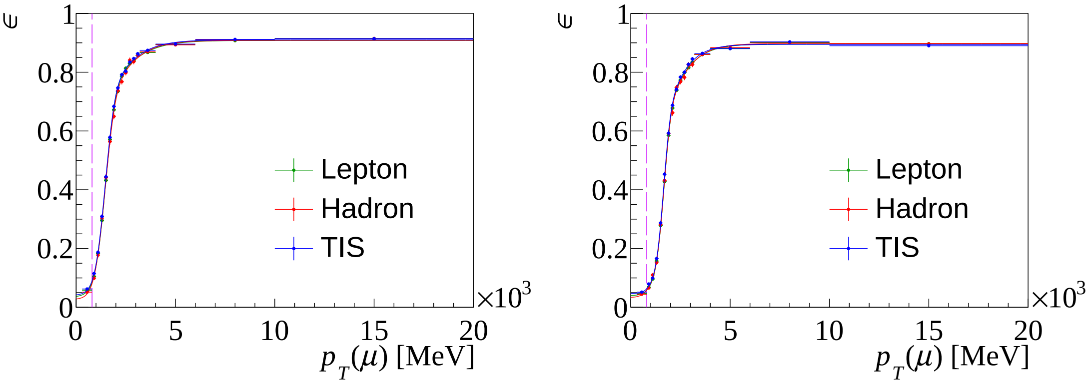


Figure 4.23 – The TISTOS efficiency of `L0Muon` trigger as a function of $p_{\mathrm{T}}(\mu)$, evaluated in 2016 B^+ modes with real data (left) and simulation (right). The curves above are from $2.75 < \eta < 3.25$ bins but they are representative of the other η bins.

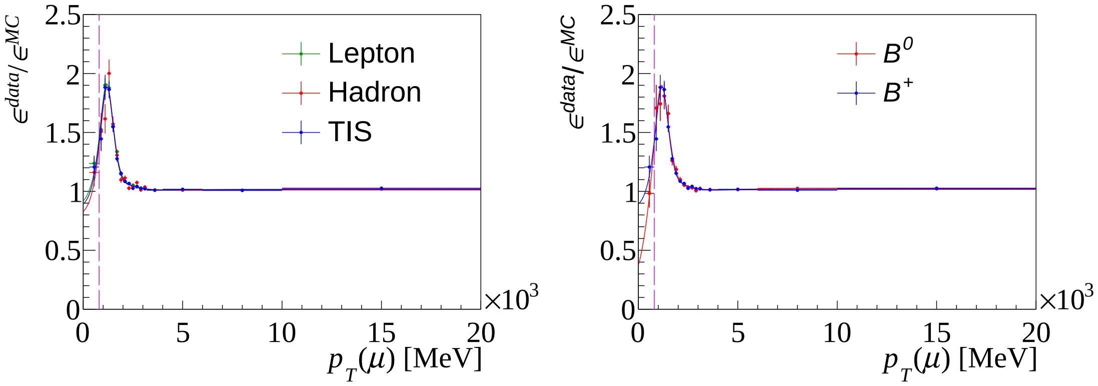


Figure 4.24 – Left: The `L0Muon` $\frac{\epsilon^{data}_{\text{TISTOS}}}{\epsilon^{MC}_{\text{TISTOS}}}$ ratio from the two curves in Fig. 4.23. The data/simulation ratio curves agree between the three different tags. Right: Comparison of `L0Muon` $\frac{\epsilon^{data}_{\text{TISTOS}}}{\epsilon^{MC}_{\text{TISTOS}}}$ curves between 2016 TIS tagged B^0 and B^+ modes.

The `L0Electron` hardware trigger fires on high E_{T} electrons and the ECAL has three different regions with increasing granularity towards the beampipe. Thus, the per-lepton `L0Electron` efficiencies are parameterised as a function of E_{T}, in the three different ECAL regions. The `L0Electron` efficiencies are measured using real and simulated $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ and $B^+ \to K^+ J/\psi(\to e^+e^-)$ events. The `L0Electron` efficiency measurement uses two tag categories, which are

- **TIS tag** events, triggered by `L0Hadron` or `L0Muon` candidates from the underlying event.
- **Hadron tag** events, triggered by a hadron from the signal B candidate.

Similar to the `L0Muon` efficiencies, `L0Electron` efficiencies are calculated in bins of E_{T} and modelled with the function in Eq. 4.25. The efficiency curves are shown in Fig. 4.25. The data/simulation efficiency ratios in Fig. 4.26 show a good agreement between TIS and hadron tags. Similarly, the TIS tag is chosen because it has a larger statistics. In the right of Fig. 4.26, the data/simulation efficiency ratio shows a good agreement between B^+ and B^0 mode.

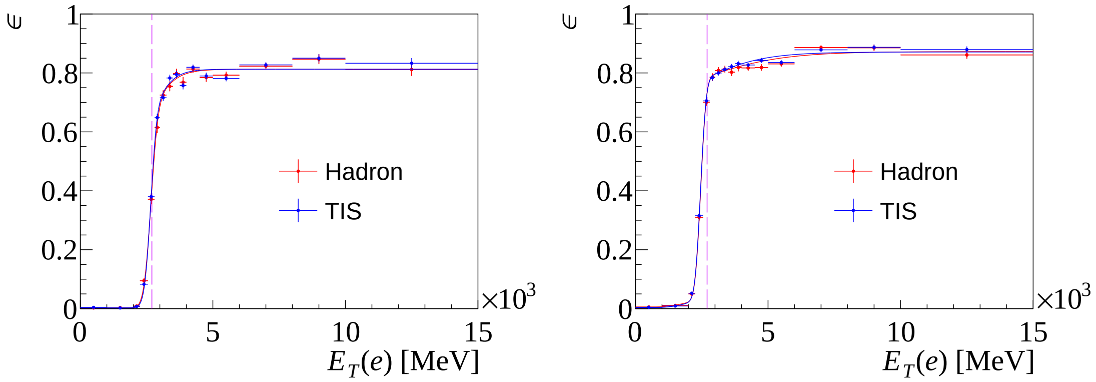


Figure 4.25 – The TISTOS efficiency of `L0Electron` trigger as a function of $p_{\mathrm{T}}(\mu)$, evaluated in 2016 B^+ modes with real data (left) and simulation (right). The curves above are from innermost ECAL region but they are representative of the other η bins.

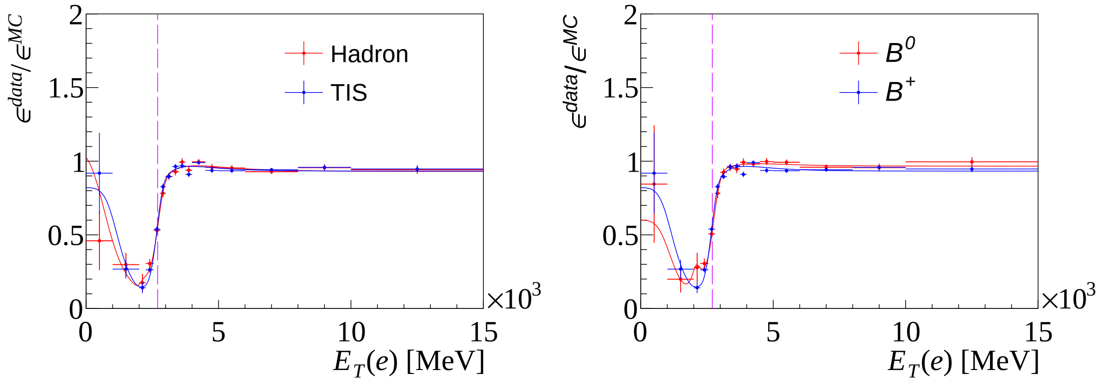


Figure 4.26 – Left: The `L0Electron` $\frac{\epsilon^{data}_{\mathrm{TISTOS}}}{\epsilon^{MC}_{\mathrm{TISTOS}}}$ ratio from the two curves in Fig. 4.25. The data/simulation ratio curves agree between the three different tags. Right: Comparison of `L0Electron` $\frac{\epsilon^{data}_{\mathrm{TISTOS}}}{\epsilon^{MC}_{\mathrm{TISTOS}}}$ curves between 2016 TIS tagged B^0 and B^+ modes.

The `L0I` category is triggered by non-signal particles in the underlying event. Recall that in LHCb, b quarks are produced in $b\bar{b}$ pairs. Hence, `L0I` are typically triggered by daughters from the opposite b hadron. Because the $b\bar{b}$ pair are correlated, the signal B kinematics would strongly correlate with the opposite B kinematics. This motivates the parameterisation of `L0I` efficiencies as a function of signal B p_{T}. Additionally, the `L0` trigger fires more often at higher multiplicity. Thus, we evaluate the `L0I` efficiencies in 6 different bins of `nTracks` but this is reduced to 4 bins in 2015 due to limited statistics. The efficiencies are measured using real and simulated $B^0 \to K^{*0} J/\psi\,(\to \ell^+\ell^-)$ and $B^+ \to K^+ J/\psi\,(\to \ell^+\ell^-)$ events that has passed the pre-MVA selection. The `L0I` TISTOS uses 3 tag categories, which are

- **Lepton tag** events, triggered by any lepton from the signal decay.
- **Hadron tag** events, triggered by any hadron from the signal decay.
- **Combined tag** events, triggered by any hadron or lepton from the signal decay.

The `L0I` TISTOS efficiencies are parameterised as a Gompertz function, which can be seen in Fig. 4.27. The data/simulation efficiency ratio, equivalent to the $w_{\mathtt{L0}}^{\mathtt{L0I}}$ correction, in Fig. 4.28 shows a good agreement between the three tags, and the **combined** tag is used as the default for its high statistics. Also in Fig. 4.28, the $w_{\mathtt{L0}}^{\mathtt{L0I}}$ corrections between B^+ and B^0 mode shows very good agreement.

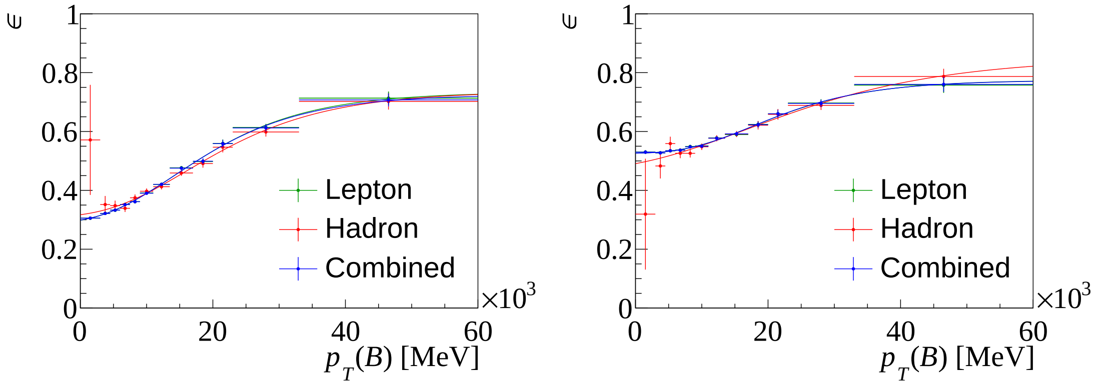


Figure 4.27 – The TISTOS efficiency of `L0I` trigger as a function of $p_{\mathrm{T}}(B^+)$, evaluated in 2016 $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ decays with real data (left) and simulation (right). The curves above are from innermost ECAL region but they are representative of the other η bins.

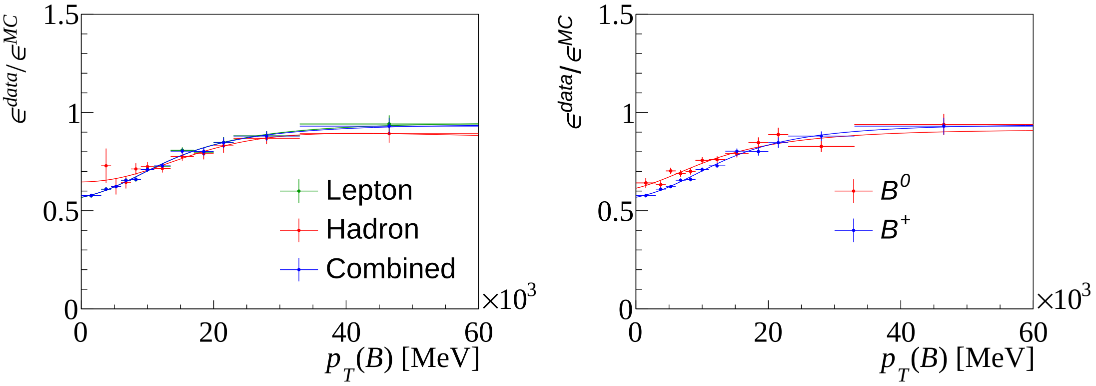


Figure 4.28 – Left: The `L0I` $\frac{\epsilon^{data}_{\text{TISTOS}}}{\epsilon^{data}_{\text{TISTOS}}}$ ratio from the two curves in Fig. 4.27. The data/simulation ratio curves agree between the three different tags. Right: Comparison of `L0I` $\frac{\epsilon^{data}_{\text{TISTOS}}}{\epsilon^{data}_{\text{TISTOS}}}$ curves between 2018 TIS tagged B^0 and B^+ modes.

Recall that `L0I` is the primary trigger in this analysis and that it triggers independently of the signal candidate. For these two reason, the $w_{\texttt{L0}}^{\texttt{L0I}}$ is expected bo be portable from the high statistics $\mu\mu$ modes to ee modes. This is checked in Fig. 4.29, which shows good agreement between $\mu\mu$ and ee modes.

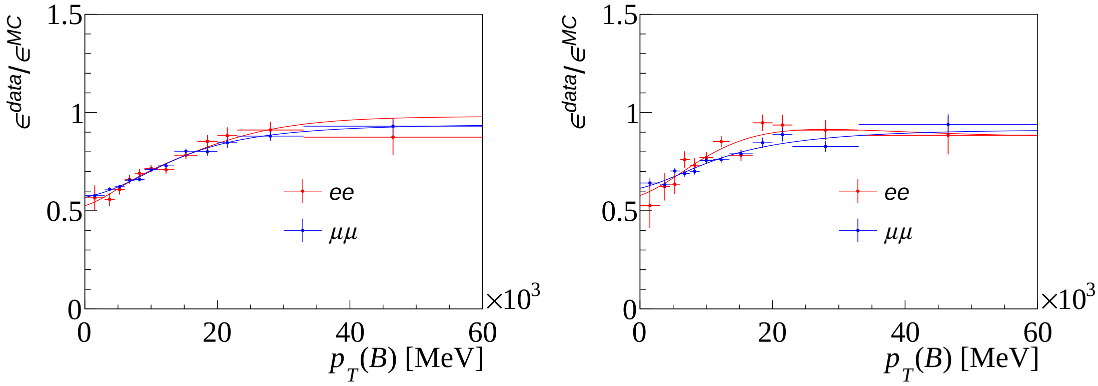


Figure 4.29 – The data/simulation ratio of `L0I` efficiency curves between electrons and muons in B^+ (left) and B^0 (right) modes in 2016, which shows good agreement between $\mu\mu$ and ee modes.

$w_{\texttt{HLT}}$ Corrections

w_{HLT} corrects the efficiencies of simulated events to pass both HLT1 and HLT2 triggers. It uses the same TISTOS method in `L0`, where the $w_{\texttt{HLT}}$ is defined as

$$w_{\texttt{HLT}} = \frac{\epsilon^{data}}{\epsilon^{MC}} = \frac{\left(\frac{N_{\text{TISTOS}}}{N_{\text{TIS}}}\right)^{data}}{\left(\frac{N_{\text{TISTOS}}}{N_{\text{TIS}}}\right)^{MC}}. \tag{4.26}$$

TOS refers to events where signal candidates that passes the `HLT1` and `HLT2` selections for this analysis. The TIS definition differs between data and simulation. For data, the underlying events in TIS candidates are selected from fully inclusive `HLT1` and fully inclusive `HLT2` triggers. In simulation, the fully inclusive `HLT` triggers do not reflect real data due to the fact that simulated events do not apply any trigger prescale. Since the trigger lines which are heavily prescaled in data tend to have a high rate, these trigger lines will accept a significant fractiom of simulated events. This in turn bias the simulated efficiencies of underlying events passing the inclusive `OR` of `HLT` trigger lines. Instead, the TIS candidates in simulated events are defined as:

- `HLT1` TIS is an inclusive `OR` of `HLT1` track triggers and `HLT1` muon triggers.
- `HLT2` TIS is an inclusive `OR` of topological triggers and detached muon triggers.

The specific trigger lines to define the `HLT1` TIS and `HLT2` TIS in simulation can be found in Appendix D.1.

$w_{\texttt{HLT}}$ correction maps are calculated using $B \to K^{(*)} J/\psi\,(\to \ell\ell)$ candidates, separately for the `L0I` and `L0L` inclusive candidates [28]. Because the $w_{\texttt{HLT}}$ corrections in `L0Muon` (`L0Electron`) inclusive is compatible with the corrections calculated from `L0Muon` (`L0Electron`) exclusive, the `L0L` exclusive categories will use the corrections from the high statistics `L0L` inclusive instead.

The `HLT` efficiencies are calculated as a function of `nTracks`, a multiplicity proxy, shown in Fig. 4.30.

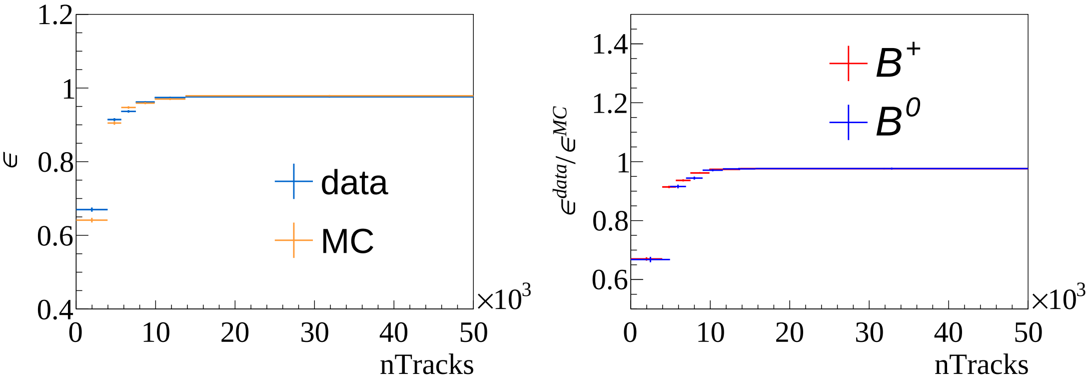


Figure 4.30 – Left: The HLT TISTOS efficiency as a function of `nTracks` in 2016 real data (blue) and simulation (orange), evaluated in 2016 `L0Muon` inclusive events of B^+ modes. Right: The data/simulation efficiency ratios shows a good agreement between B^+ (red) and B^0 (blue) modes in 2016 `L0Muon` inclusive.

$w_{B,MULT}$ Corrections

$w_{B,MULT}$ corrections account for the B kinematics mismodelling, in B p_{T} and η, by PYTHIA and event multiplicity mismodelling by the full event simulation [29]. The $w_{B,MULT}$ weights are calculated in the inclusive `L0Muon` category with a Gradient

---

[28] The correction chain is sequential and different `L0` triggers affects `HLT` trigger efficiencies differently.

[29] Sources of multiplicity effects are PYTHIA, EvtGen, GEANT4 and Boole.

Boosted **Reweighter** (GBReweighter). GBReweighter is an ensemble of gradient boosted regression trees, anologous to how a BDT is an ensemble of gradient boosted classifier trees. An individual tree in GBReweighter splits the dataset into $i = 1, 2 ... N$ regions (leaves) by greedily maximising a symmetrized χ^2, defined as [30]

$$\chi^2 = \sum_{i}^{N} \frac{(w_i^{MC} - w_i^{data})^2}{w_i^{MC} + w_i^{data}}. \tag{4.27}$$

Essentially, regions with high data-simulation discrepancy have a large contribution to χ^2, and these regions are prioritised by GBReweighter.

The gradient boosting in GBReweighter iteratively builds an ensemble of regression trees by repeating the cycle:

1. build a shallow tree which maximises the symmetrised χ^2,
2. compute predictions in each region i, $pred = \log \frac{w_i^{MC}}{w_i^{data}}$,
3. reweight each event in the MC distribution according $w^{MC} = e^{\sum_j pred_j}$, where $pred$ is summed over all j trees in the ensemble [31].

Thus, the gradient boosting builds an ensemble, tree-by-tree, that iteratively minimises the residual differences between data and simulation in each step.

Like the `CatBoost` algorithm trained as the MVA classifier, the training and test sets of GBReweighter must be exclusive of each other, in order to avoid regression biases. The training and application of $w_{B,MULT}$ uses the same *k-fold* resampling method, and the datasets are resampled $k = 4$ times to train 4 different regression tree.

In Fig. 4.31, the distribution of B kinematics, after the full correction chain, shows a good agreement between data and simulation. The same applies to `nTracks` distributions in Fig. 4.32 but a residual discrepancy remains in `nSPDHits`, a multiplicity variable which is not reweighted for. The effect of residual `nSPDHits` mismodelling is studied as a systematic uncertainty Sec. 4.6.2 and the `nSPDHits` variable is tested in the differential $r_{J/\psi}$ cross check of Sec. 4.7.2.

---

[30] Each depth of the regression tree splits the dataset by only considering one input variable. Greedy optimisation essentially means the regression tree chooses a 1D split at each depth that optimises the value of χ^2.

[31] Note that although two events belong in the region in tree j^{th}, this does not guarantee they belong in the same region in the $j + 1^{\text{th}}$ tree.

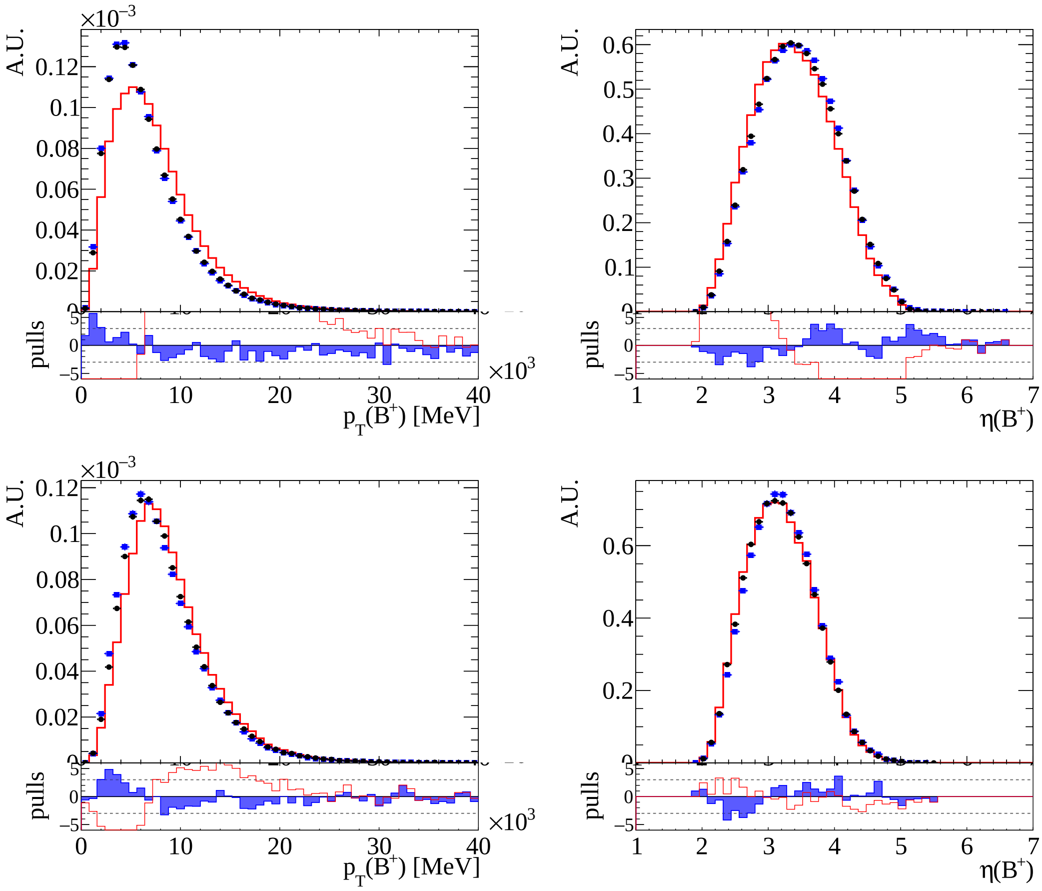


Figure 4.31 – Distributions of B^+ p_T (left) and η (right) in real data (black marker), uncorrected simulation (red histogram) and fully corrected simulation (blue marker) in $\mu\mu$ (top) and ee (bottom) modes.

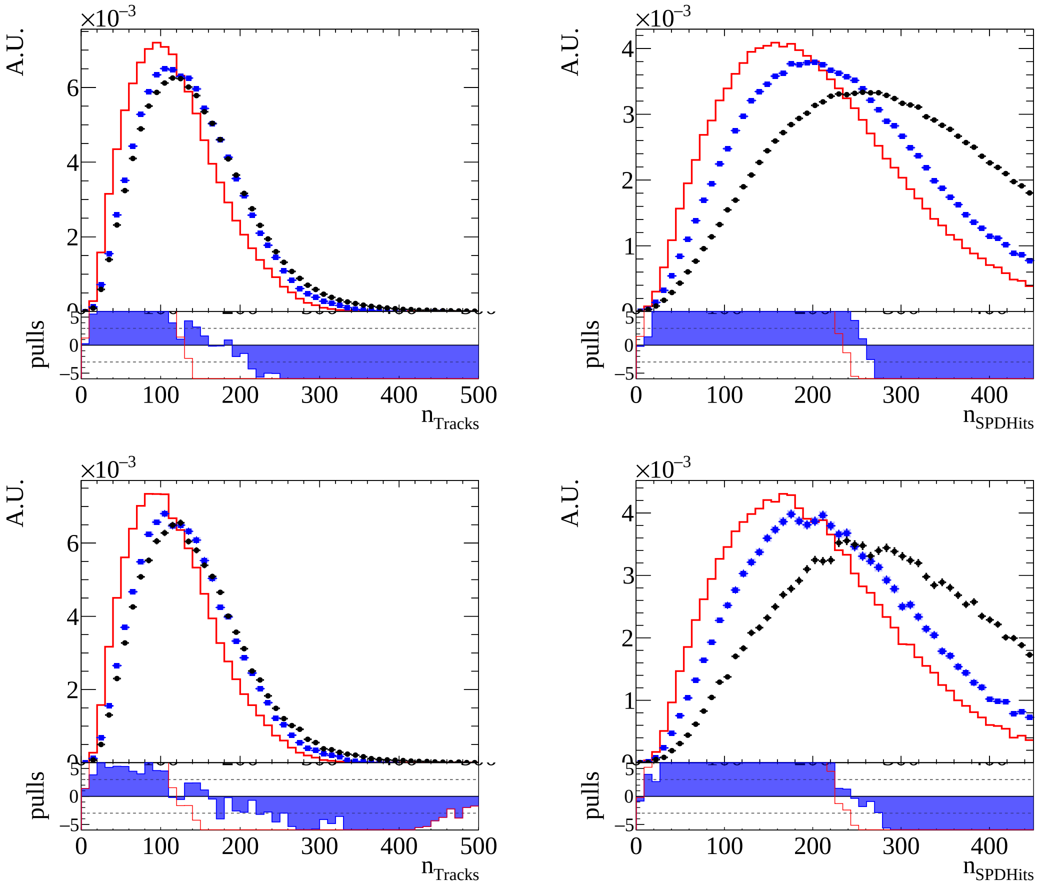


Figure 4.32 – Distributions of `nTracks` (left) and `nSPDHits` (right) in real data (black marker), uncorrected simulation (red histogram) and fully corrected simulation (blue marker) in B^+ $\mu\mu$ (top) and ee (bottom) modes. Although the `nSPDHits` distributions improved after event reweighting, a residual discrepancy remains.

w_{reco} Corrections

w_{reco} accounts for differences in reconstruction effects between simulation and real data. w_{reco} corrections reweight the vertex reconstruction quality, χ^2_{vertex}, and the impact parameter χ^2, χ^2_{IP}, of simulated B candidates to the distributions observed in real data. The w_{reco} weights are calculated by the GBReweighter regression tree algorithm, similar to the $w_{B,MULT}$ corrections. Likewise, w_{reco} weights are trained with a *k-fold* approach, with a $k = 4$ resampling.

The distribution of χ^2_{vertex} and χ^2_{IP} before and after w_{reco} corrections are shown in Fig. 4.33. These variables shows a good agreement between data and simulation after w_{reco} corrections.

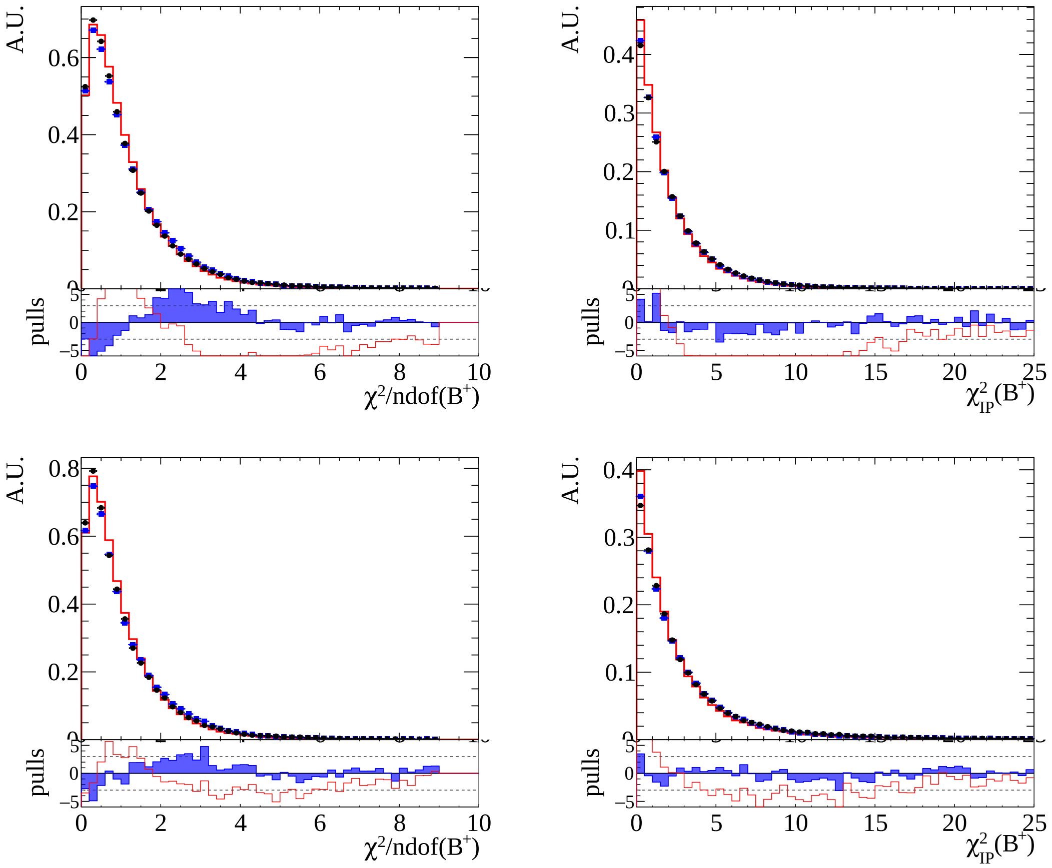


Figure 4.33 – Distributions of B^+ χ^2_{vertex} (left) and χ^2_{IP} (right) in real data (black marker), uncorrected simulation (red histogram) and fully corrected simulation (blue marker) in $\mu\mu$ (top) and ee (bottom) modes.

q^2 Smearing

w_{q^2} corrects the q^2 resolution in simulation. This correction is vital because R_K and R_{K^*} are measured in different bins of q^2, and a q^2 mismodelling in simulation would bias the efficiency of the q^2 cuts. Although the effect of q^2 resolution mismodelling is negligible in $\mu\mu$ modes, it can have sizeable contribution in ee modes.

The q^2 mismodelling of ee modes is largely caused by bremsstrahlung radiation. Simulated bremsstrahlung photons are highly dependent on the material budget in simulation because most of the bremsstrahlung radiation are emitted via material interactions. Additionally, the bremsstrahlung recovery procedure depends on the ECAL response, which is correlated to detector occupancy. Since material budget and detector occupancy are both badly modelled by simulation, the simulated electron momentum and $m(ee)$ resolution do not reflect real data.

In order to quantify the q^2 resolution mismodelling, we have to measure the q^2 resolution in both simulation and data. For this, the $J/\psi \to ee$ control modes are perfect because the true value of q^2 is known [32]. The $J/\psi \to ee$ datasets are selected with the pre-MVA selections without any q^2 requirement, and an additional $5200 < m_{J/\psi}^{DTF}(B) < 5680\,\mathrm{MeV}/c^2$ cut vetoes the partially reconstructed and combinatorial backgrounds. The datasets are further binned by how many bremsstrahlung photon(s) are recovered for the di-electron system,

- **0**γ no bremsstralung photon is recovered,
- **1**γ one bremsstrahlung photon is recovered for the ee system,
- **2**γ two or more bremstrahlung photons are recovered for the ee system,

due to the fact that the q^2 resolution function depends on how many bremsstrahlung photons are recovered.

The q^2 resolution, in simulation and real data, is parameterised as a DSCB function. First, a DSCB function is fitted to the q^2 spectrum in simulated $B \to K^{(*)} J/\psi\,(\to e^+e^-)$ decays. The parameters of this DSCB function is then fixed, and the shape is ported to real data. The data/simulation differences in the resolution function is parameterised as

$$\mu^{data} = \mu^{MC} + \Delta\mu, \tag{4.28}$$

$$\sigma^{data} = \sigma^{MC} \times s_{\sigma}, \tag{4.29}$$

$$\alpha_{high}^{data} = \alpha_{high}^{MC} \times b_{\alpha} \text{ (}1\gamma \text{ and } 2\gamma \text{ only)}, \tag{4.30}$$

---

[32] The J/ψ invariant mass!

where μ denotes the mean of the Gaussian core, σ the width of the Gaussian core, and α_{high} the boundary between the Gaussian core and high mass tail. Only $\Delta\mu$, s_σ and b_α are allowed to float in the data fit. The combinatorial background in real data is accounted for with an exponential function. Example fits are shown in Appendix D.2.

With the resolution function paramerised in both data and simulation, the resolution of simulated events are smeared to represent real data. The smearing formula is written as

$$q^{2,smeared} = q^{2,true} + s_\sigma \cdot (q^{2,reco} - q^{2,true}) + \Delta\mu + (1 - s_\sigma) \cdot (\mu^{MC} - m(J/\psi)^{PDG}), \quad (4.31)$$

which is then used to compute the efficiencies of the q^2 cuts. The measured values of $\Delta\mu$ and s_σ are given in Appendix D.3.

4.4.2 Efficiencies

Now that the simulated samples are fully corrected, we can compute an unbiased total efficiency ϵ_{total}, defined as

$$\epsilon_{total} = \epsilon_{geo} \cdot \epsilon_{flt|geo} \cdot \epsilon_{sel|geo,flt}. \quad (4.32)$$

ϵ_{geo} is the efficiency to pass the generator-level geometric cut. This cut is applied by LHCb simulation after Pythia and EvtGen simulations, and its definition is 'all signal decay daughters are required to be $10 < \theta < 400\,\text{mrad}$'. This selection ensures that events with daughters outside the LHCb acceptance are not simulated by post-EvtGen processes. This saves significant CPU resources especially since the computationally intensive Geant4 simulation is avoided. Furthermore, these events fall outside the LHCb acceptance and their simulation mismodellings are difficult to calibrate with data-driven methods. The geometric cut efficiencies are provided, with the simulated sample, by LHCb-centralised simulation jobs, and the geometric efficiencies typically range between $10 - 15\%$.

$\epsilon_{flt|geo}$ is the filtering efficiency, given an event has passed the geometric cuts. As mentioned previously, the filtering is an optional process at the end of a simulation job, which applies a stripping cut to simulated samples to remove events that are likely discarded by analysts anyway. If a simulation job opts for filtering, the filtering efficiency is provided by the simulation job. Otherwise, the filtering efficiency is 1 for simulated samples that did not request filtering, which is the case for all the simulated rare and resonant mode decays of this analysis.

$\epsilon_{sel|geo,flt}$ is the full offline selection efficiency, given an event has passed the filtering and geometric cuts. This includes the stripping cuts, `L0`, `HLT1` and `HLT2` triggers, the generic selections in Tab. 4.3, the exclusive background vetoes, the MVA selection and the HOP cut. All the corrections we have discussed calibrate the simulation to represent real data, except for w_{PID} corrections, which are interpreted as the efficiency to pass the PID requirements. Hence, the measurement of $\epsilon_{sel|geo,flt}$ excludes the PID requirements in simulated samples since PID efficiencies are encoded in the values of w_{PID}. $\epsilon_{sel|geo,flt}$ is then defined as

$$\epsilon_{sel|geo,flt} = \frac{\sum_i^{pass} w^i \cdot w^i_{PID}}{\sum_i^{geo,flt} w^i} \tag{4.33}$$

where w^i denotes the full correction weights for event i, excluding w_{PID} weights, $\sum_i^{pass}$ sums over all events that passes the full selection excluding PID requirements and $\sum_i^{geo,flt}$ sums over all simulated events that has passed the filtering and geometric cuts.

Although this analysis blinds the rare mode efficiencies as the blinding strategy, the fully calibrated efficiencies of the control mode are unblinded. Their total efficiencies in each data-taking year are presented in Tab. 4.12.

These efficiencies and the corrections are cross-checked by the tests of integrated $r_{J/\psi}$ and the flatness test in Sec. 4.7. But first, we shall discuss the mass fits, which also rely on simulated efficiencies to constrain the normalisation of irreducible backgrounds.

Final State	Lepton	Year	Efficiencies$[10^{-3}]$		
			`L0I`	`L0L` exc	`L0L` inc
$K^+\ell^+\ell^-$	$\mu\mu$	2011	4.73 ± 0.05	13.6 ± 0.1	18.1 ± 0.2
		2012	4.59 ± 0.05	12.2 ± 0.1	16.4 ± 0.2
		2015	5.58 ± 0.06	11.9 ± 0.1	16.1 ± 0.2
		2016	6.27 ± 0.06	15.1 ± 0.2	21 ± 0.2
		2017	6.51 ± 0.07	16.4 ± 0.2	22.6 ± 0.2
		2018	5.91 ± 0.06	16.3 ± 0.2	21.7 ± 0.2
	ee	2011	1.75 ± 0.02	2.27 ± 0.02	3.23 ± 0.03
		2012	1.6 ± 0.02	1.55 ± 0.02	2.28 ± 0.02
		2015	2.18 ± 0.02	2.42 ± 0.02	3.43 ± 0.03
		2016	2.79 ± 0.03	3.21 ± 0.03	4.71 ± 0.05
		2017	2.91 ± 0.03	3.16 ± 0.03	4.61 ± 0.05
		2018	2.51 ± 0.03	2.97 ± 0.03	4.18 ± 0.04
$K^{*0}\ell\ell$	$\mu\mu$	2011	1.74 ± 0.02	4.61 ± 0.05	6.29 ± 0.06
		2012	1.69 ± 0.02	4.26 ± 0.04	5.86 ± 0.06
		2015	2.15 ± 0.02	4.34 ± 0.04	5.98 ± 0.06
		2016	2.46 ± 0.02	5.51 ± 0.06	7.83 ± 0.08
		2017	2.56 ± 0.03	5.9 ± 0.06	8.33 ± 0.08
		2018	2.24 ± 0.02	5.78 ± 0.06	7.84 ± 0.08
	ee	2011	0.644 ± 0.006	0.833 ± 0.008	1.22 ± 0.01
		2012	0.639 ± 0.006	0.64 ± 0.006	0.962 ± 0.01
		2015	0.881 ± 0.009	0.996 ± 0.01	1.45 ± 0.01
		2016	1.07 ± 0.01	1.22 ± 0.01	1.83 ± 0.02
		2017	1.12 ± 0.01	1.22 ± 0.01	1.84 ± 0.02
		2018	0.967 ± 0.01	1.15 ± 0.01	1.65 ± 0.02

Table 4.12 – The total efficiencies measured in the control mode. Note that these simulation samples are not filtered, hence their filtering efficiencies are 1.

4.5 Mass fits

The mass fits not only measure the yields relevant to the LFU ratios, it can also be configured to measure the LFU ratios directly. Recall that single-ratio $r_{J/\psi}$ and double-ratios are defined as

$$r_{J/\psi} = \frac{N^{\mu\mu}_{J/\psi}}{N^{ee}_{J/\psi}} \cdot \frac{\epsilon^{\mu\mu}_{J/\psi}}{\epsilon^{ee}_{J/\psi}}, \tag{4.34}$$

$$R_X^{DR} = \frac{N^{\mu\mu}}{N^{ee}} \cdot \frac{\epsilon^{\mu\mu}}{\epsilon^{ee}} \cdot \frac{1}{r_{J/\psi}}, \tag{4.35}$$

and the mass fits can measure $r_{J/\psi}$ and R_X^{DR} by reparameterising the electron mode yields with the equation above [33] and float the LFU ratio as a fit parameter. The efficiencies are treated as external Gaussian constraints by the mass fits.

These fits are one-dimensional extended likelihood fits to the reconstructed B^+ and B^0 invariant masses. The fits in the resonant modes, $B^0 \to K^{*0} J/\psi (\to \ell^+\ell^-)$, $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$, $B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$ and $B^+ \to K^+\psi(2S)(\to \ell^+\ell^-)$, are binned for reasons of speed. Fits to the rare modes, $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^+\ell^+\ell^-$, are unbinned due to their limited statistics.

The values of R_K and R_{K^*} are extracted with a simultaneous fit to low, central and J/ψ q^2, simultaneously in B^+ and B^0 modes. This fit, known as the rare mode fit, extracts the LFU ratios using Eq. 4.35. Additionally in this fit, the B invariant masses in J/ψ q^2 bin are reconstructed without J/ψ mass constraints because the J/ψ mode is used to constrain shape parameters of the rare modes. [34]

In contrast, the cross checks of $r_{J/\psi}$, integrated or differential, are only interested in the validity of the efficiency calibration. For this reason, the $r_{J/\psi}$ cross checks fit the B masses with J/ψ mass constraints, and treat B^+ and B^0 modes separately. Similarly, the $R_{\psi(2S)}$ double ratio cross check validates the portability of the efficiency calibration separately in B^+ and B^0 modes but $R_{\psi(2S)}$ are measured with a simultaneous fit to J/ψ and $\psi(2S)$ modes, and the B masses are reconstructed with their respective $c\bar{c}$ mass constraint.

---

[33] Doing so means the electron yields are no longer a fit parameter.

[34] As we shall soon see.

4.5.1 Mass Ranges

The B^+ and B^0 invariant mass windows depend on the q^2 bin and leptonic final state. Electron modes typically have wider mass windows due to bremsstrahlung radiation losses, unless the invariant mass is reconstructed with J/ψ or $\psi(2S)$ mass constraints on $m(ee)$.

$J/\psi \to \mu\mu$ mode ranges from 5100 MeV to 5900 MeV, both with and without the J/ψ mass constraint. The partially reconstructed background dominates the 5100 MeV-5150 MeV region and leaks into the signal region at 5150 MeV. A lower limit of 5100 MeV constrains this background sufficiently. The upper sideband ends at 5900 MeV to constrain the combinatorial background.

$\psi(2S) \to \mu\mu$ mode ranges from 5100 MeV to 5750 MeV. Like the $J/\psi \to \mu\mu$ modes, the 5100 MeV limit allows the mass fits to constrain the partially reconstructed background. The upper limit at 5750 MeV symmetrises the mass windows between $\psi(2S) \to \mu\mu$ and $\psi(2S) \to ee$ modes.

Rare $\mu\mu$ mode ranges from 5150 MeV to 5850 MeV. Within this range only the signals and combinatorial backgrounds are modelled. The upper sideband constrains the combinatorial backgrounds while partially reconstructed backgrounds are vetoed by the lower 5150 MeV limit.

$J/\psi \to ee$ mode has two ranges: without the J/ψ mass constraint it is 4400 MeV to 6200 MeV while the J/ψ mass constrained window is 5100 MeV to 5900 MeV. Without the J/ψ mass constraint the bremsstrahlung radiation causes the tail to extend down to 4600 MeV. However, the partially reconstructed backgrounds dominate this region and peak around 4600 MeV. Constraining both the partially reconstructed backgrounds and J/ψ mode necessitates a lower mass range of 4400 MeV. On the high mass side, opening the window up to 6200 MeV allows us to better model the combinatorial background. If we constrain the J/ψ mass, we use the same mass window as the muon J/ψ mode: 5100 MeV to 5900 MeV. Like the muons, 5100 MeV is enough to constrain partially reconstructed background while the upper 5900 MeV limit controls the combinatorial background shape.

$\psi(2S) \to ee$ mode ranges from 5100 MeV to 5750 MeV. Similar to the $\psi(2S) \to \mu\mu$ mode, the 5100 MeV lower limit is sufficient to constrain partially reconstructed backgrounds. At the upper mass sideband, the $\psi(2S) \rightarrow ee$ mode has a $B \rightarrow K^{(*)} J/\psi\,(\rightarrow e^+e^-)$ leakage around the 5700-5750 MeV range. Hence, the 5750 MeV limit allows the mass fits to constrain this background.

Rare *ee* modes range from 4600 MeV to 6200 MeV. Like the J/ψ mode, radiative losses in electron rare modes extends the lower tail to 4600 MeV. In contrast to J/ψ mode, the partially reconstructed backgrounds are more strongly suppressed due to the dedicated partially reconstructed background MVA and the HOP cut. The lower boundary at 4600 MeV is sufficient to control the partially reconstructed backgrounds while an upper limit of 6200 MeV is needed to model the combinatorial background.

4.5.2 Simulated Shapes

As we have discussed above, backgrounds are the main consideration on how the mass windows are defined. The shape of these backgrounds are difficult to model in data, hence most of them are modelled with a `RooKeysPdf` KDE built from fully corrected simulation.

For the signal shapes, the analytical PDFs are first fixed by fits to simulated signal decay. Like the fits to model q^2 resolution, the signal shapes are extracted from fits to uncorrected simulation and data-simulation differences in B mass resolution are parameterised as

$$\mu^{data} = \mu^{MC} + \Delta\mu, \tag{4.36}$$

$$\sigma^{data} = s_\sigma \cdot \sigma^{MC}, \tag{4.37}$$

where $\Delta\mu$ and s_σ are floated in the data fits.

These parameters, $\Delta\mu$ and s_σ, are expected to be portable between q^2. Therefore, the rare mode fit shares $\Delta\mu$ and s_σ between the q^2 bins. This significantly improves the uncertainties of $\Delta\mu$ and s_σ in the statistically limited low and central q^2 bins, by relying on the sizeable statistics in the J/ψ q^2 bin. This strategy however, forbids the J/ψ modes from using the J/ψ mass constrained reconstructed B mass.

For the proceeding sections, the plots of simulated signal and background shapes can be found in Appendix E.1 and E.2.

4.5.3 Rare and resonant mode signal PDFs

Reconstructed invariant masses of $J/\psi \to \mu\mu$ modes are described by a linear combination of a double-sided Hypatia (DSH) (22) and a single-sided Crystal Ball (CB) function. The signal invariant mass distributions in rare muon and $\psi(2S) \to \mu\mu$ modes are modelled with a DSH function. Fig. 4.34 shows example mass shapes used to model $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ and $B^0 \to K^{*0}\mu^+\mu^-$ decays.

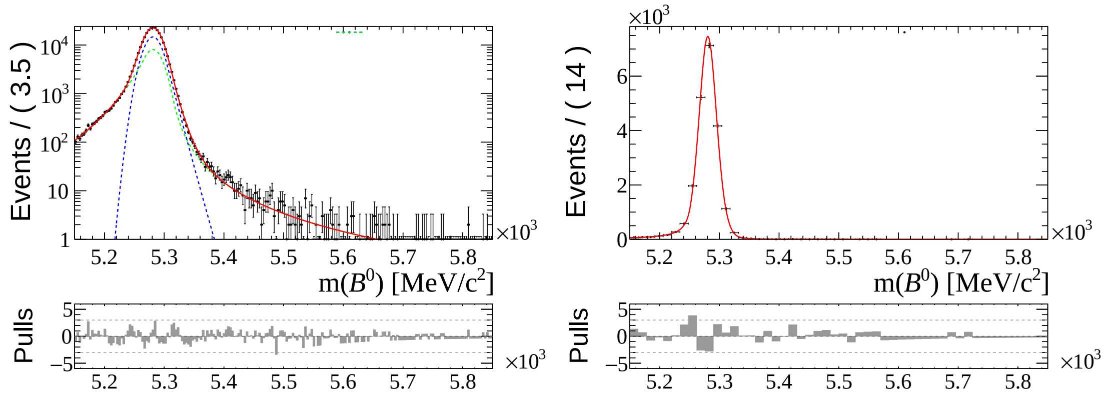


Figure 4.34 – Fits to simulated $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ (left) and central-q^2 $B^0 \to K^{*0}\mu^+\mu^-$ (right) decays. Note that the $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ shape is shown in logarithmic scale. The fits shown above use 15 + 16 simulation in the exclusive `LOL` category.

Analogous to the q^2 resolution model, the number of recovered photons strongly influence the invariant mass shapes in electron modes. Therefore, each of the three bremsstrahlung categories (0γ, 1γ, 2γ) has its own unique PDF. When the simulated shapes are ported to real data, the data fit combines the three bremsstrahlung category shapes. The three PDFs are added linearly by using their relative fractions calculated from fully corrected simulation.

In the rare ee modes, 0γ, 1γ and 2γ categories are described with a (DSCB) function. In the $J/\psi \to ee$ modes, without J/ψ mass constraints, the 0γ category is modelled with a DSCB while the 1γ and 2γ categories are modelled with the sum of a DSCB and a Gaussian with independent parameters.

For the electron resonant modes, the J/ψ and $\psi(2S)$ mass constraints correct, to first order, the mass resolution smeared by bremsstrahlung radiation. Residual effects on the resolution mostly come from the hadron system, and the mass-constrained B mass distributions are similar between electron and muon modes. Thus, mass-constrained ee modes in each bremsstrahlung category are modelled with the same function in the muon mode: the sum of a DSH and a CB function in J/ψ mode.

The means and widths of the Gaussian core in the simulated signals are given in Tab. 4.13 for the central and J/ψ q^2 bins. Although these are only for reconstructed B^0 decays in the $15+16$ `L0L` exclusive category, they are representative of the values for the other data-taking period, trigger categories, q^2 bins and B^+ modes. Of note, the widths decreases with the J/ψ mass constraint. Furthermore, the ee mode distributions, without J/ψ mass constraints, peak about $20-30$ MeV lower than the B^0 mass.

A comparison of the invariant mass distribution with and without a mass constraint on $m(ee)$ is shown in Fig. E.7-4.35, for the three bremsstrahlung categories. These figures show a significant improvement in the B mass resolution by constraining J/ψ mass.

Lepton	q^2 bin	bremsstrahlung	μ^{MC}	σ^{MC}
$\mu\mu$	J/ψ		5281.1	15.6
	J/ψ, DTF		5280.2	9.1
	central		5280.7	16.5
ee	J/ψ	0γ	5252.8	23.3
		1γ	5250.5	40.9
		2γ	5258.8	50.0
	J/ψ, DTF	0γ	5282.1	10.6
		1γ	5280.5	30.5
		2γ	5280.2	33.2
	central	0γ	5259.3	19.6
		1γ	5256.6	34.4
		2γ	5266.6	42.4

Table 4.13 – The means and width, in MeV, of the Gaussian-like core to simulated B^0 signals, in the $15+16$ `L0L` exclusive sample.

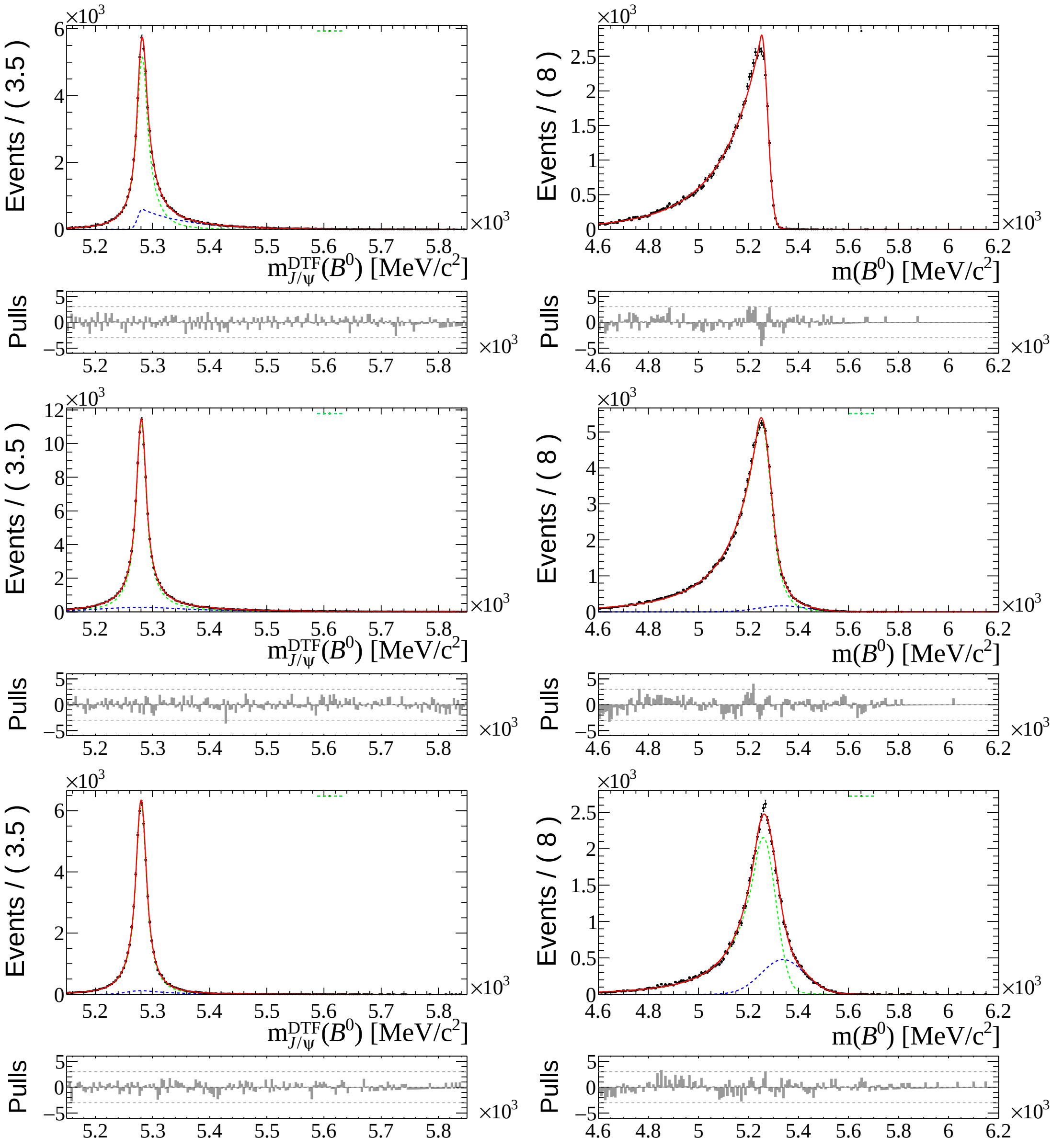


Figure 4.35 – Fits to simulated $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ in the 0γ (top), 1γ (middle) and 2γ (bottom) category to B^0 mass reconstructed with (left) and without (right) J/ψ mass constraint. The fits shown above use $15+16$ simulation in the exclusive `L0L` category.

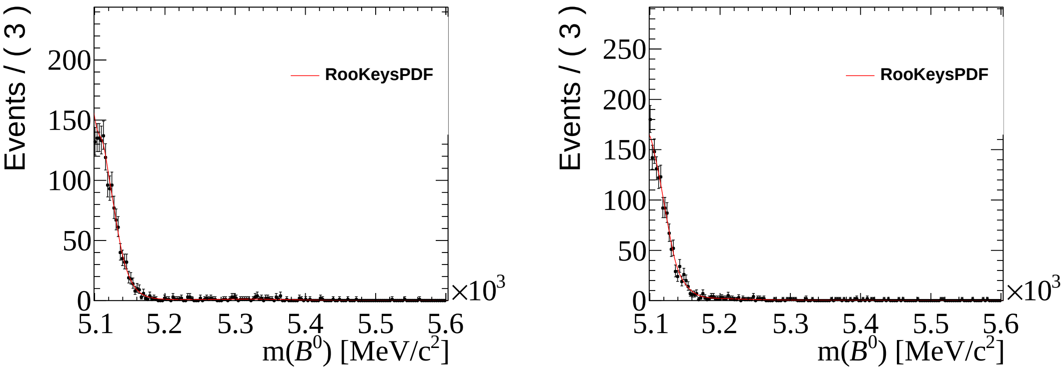
4.5.4 Background PDFs in B^0 modes

$B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ and $B^0 \to K^{*0}\psi(2S)(\to \mu^+\mu^-)$ data fits consider the following backgrounds:

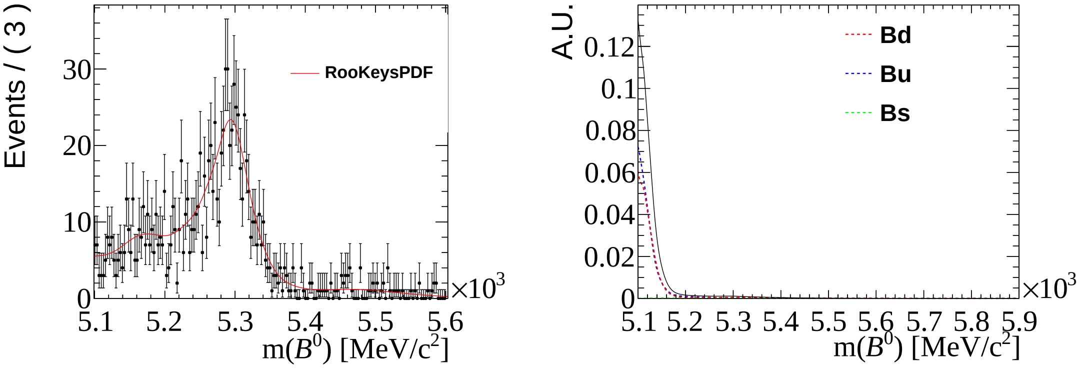
- *Combinatorial*: modelled with an exponential function where the slope is floating.
- *$\Lambda_b^0 \to pK J/\psi(\to \mu^+\mu^-)$ ($\Lambda_b^0 \to pK\psi(2S)(\to \mu^+\mu^-)$)*: modelled with a `RooKeysPdf` using simulated data corrected for the pK Dalitz plot, measured by Ref. (23).
- *$B_s^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ ($B_s^0 \to K^{*0}\psi(2S)(\to \mu^+\mu^-)$)*: modelled using the control mode signal PDF with the central value offset by $m_{B_s^0}^{PDG} - m_{B^0}^{PDG}$.
- *Partially Reconstructed $B \to X J/\psi(\to \mu\mu)$ (and $B \to X\psi(2S)(\to \mu\mu)$)*: modelled with a `RooKeysPdf` using inclusive, simulated samples of B^0, B^+ and B_s^0 decays. Sec. 4.5.6 will explain them in more detail. Fig. 4.36 illustrates the inclusive decays of B^0, B^+ and B_s^0, and their combined shape used in the data fit.

$B^0 \to K^{*0} J/\psi(\to e^+e^-)$ and $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ data fits consider the following backgrounds:

- *Combinatorial*: modelled with an exponential function where the slope is floating.
- *$\Lambda_b^0 \to pK J/\psi(\to e^+e^-)$ ($\Lambda_b^0 \to pK\psi(2S)(\to e^+e^-)$)*: modelled with a `RooKeysPdf` using simulated data corrected for the pK Dalitz plot, measured by Ref. (23).
- *$B_s^0 \to K^{*0} J/\psi(\to ee)$ ($B_s^0 \to K^{*0}\psi(2S)(\to ee)$)*: modelled using the control mode signal PDF but central value offset by $m_{B_s^0}^{PDG} - m_{B^0}^{PDG}$.
- *Partially Reconstructed $B \to X J/\psi(\to ee)$ (and $B \to X\psi(2S)(\to ee)$)*: modelled with a `RooKeysPdf` using inclusive Monte Carlo samples of B^0, B^+ and B_s^0 decay. In the unconstrained J/ψ mass fits, they are split in two: leptonic partially reconstructed backgrounds, where a higher $c\bar{c}$ resonance decays into the J/ψ and hadronic partially reconstructed backgrounds, which are exclusive of the leptonic partially reconstructed backgrounds.
- *Leakage $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ ($\psi(2S)$-q^2 only)*: modelled with a `RooKeysPdf` using simulated data which pass the $\psi(2S)$-q^2 selection.



(a) $B^0 \to X J/\psi\,(\to \mu^+\mu^-)$ (b) $B^+ \to X J/\psi\,(\to \mu^+\mu^-)$



(c) $B_s^0 \to X J/\psi\,(\to \mu^+\mu^-)$ (d) Combined

Figure 4.36 – Partially reconstructed $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on an inclusive simulation sample. The $B^0 \to X J/\psi\,(\to \mu^+\mu^-)$ (top left), $B^+ \to X J/\psi\,(\to \mu^+\mu^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to \mu^+\mu^-)$ (bottom left) shapes are added linearly according to their ratios (bottom right) when these shapes are used in the data fit. The figures here use simulated samples from $15+16$ in the exclusive `L0L` category.

$B^0 \to K^{*0}\mu^+\mu^-$ data fits only consider a combinatorial background, described by an exponential function with floating slope.

$B^0 \to K^{*0}e^+e^-$ data fits consider the following backgrounds:

- *Combinatorial*: modelled with an exponential function where the slope is floating;
- *Partially-reconstructed $B^+ \to K\pi\pi e^+e^-$*: modelled with a `RooKeysPdf` using simulated data containing decays of higher K^{*0} resonances, such as $K_1(1270)$ and $K_2^*(1430)$.
- *Leakage $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (central-q^2 only)*: modelled with a `RooKeysPdf` using simulated data which pass the central-q^2 selection.

4.5.5 Background PDFs in B^+ modes

$B^+ \to K^+J/\psi(\to \mu^+\mu^-)$ and $B^+ \to K^+\psi(2S)(\to \mu^+\mu^-)$ data fits consider the following backgrounds:

- *Combinatorial*: modelled with an exponential function where the slope is floating.
- *$B^+ \to \pi^+J/\psi(\to \mu^+\mu^-)$ (J/ψ only)*: modelled with a DSCB function, fitted to non-corrected simulation. Fig. 4.37 shows an example shape of this background component.
- *Partially Reconstructed (J/ψ only)* : modelled with a `RooKeysPdf` using inclusive, simulated samples of B^0, B^+ and B_s^0 decays. Sec. 4.5.6 will explain them in more detail.

$B^+ \to K^+J/\psi(\to e^+e^-)$ and $B^+ \to K^+\psi(2S)(\to e^+e^-)$ data fits consider the following backgrounds:

- *Combinatorial*: modelled with an exponential function where the slope is floating;
- *$B^+ \to \pi^+J/\psi(\to e^+e^-)$ (J/ψ only)*: modelled with a DSCB fucntion, fitted to non-corrected simulation. Fig. 4.37 shows an example shape of this background component.

- *Partially Reconstructed (J/ψ only)* : modelled with a `RooKeysPdf` using inclusive Monte Carlo samples of B^0, B^+ and B_s^0 decays. In the unconstrained J/ψ mass fits, they are split in two: leptonic partially reconstructed backgrounds, where a higher $c\bar{c}$ resonance decays into the J/ψ and hadronic partially reconstructed backgrounds, which are exclusive of the leptonic partially reconstructed backgrounds.

- *Leakage $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ ($\psi(2S)$-q^2 only)*: modelled with a `RooKeysPdf` using simulated data which pass the $\psi(2S)$-q^2 selection.

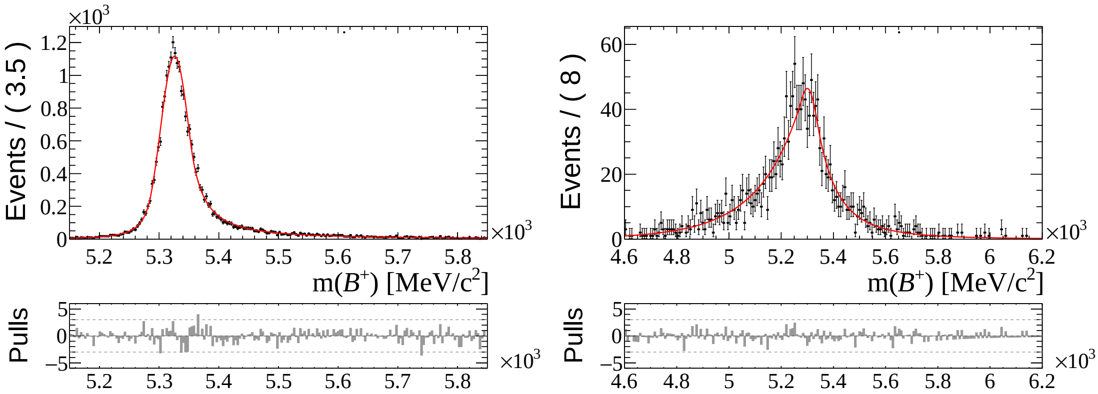


Figure 4.37 – Fits to simulated $\pi \to K$ mis-identified $B^+ \to \pi^+ J/\psi (\to \mu^+\mu^-)$ (left) and $B^+ \to \pi^+ J/\psi (\to e^+e^-)$ (right) backgrounds reconstructed as $B^+ \to K^+ J/\psi (\to \ell^+\ell^-)$. Modelled using 15 + 16 simulated samples which pass the exclusive `L0L` trigger.

$B^+ \to K^+\mu^+\mu^-$ data fits only consider the combinatorial background, described by an exponential function with floating slope.

$B^+ \to K^+e^+e^-$ data fits consider the following backgrounds:

- *Combinatorial*: modelled with an exponential function where the slope is floating;

- *Partially-reconstructed $B \to K^*ee$*: modelled with a `RooKeysPdf` using simulated $B^0 \to K^{*0}e^+e^-$ decays.

- *Partially-reconstructed $B^0 \to K^+\pi^-e^+e^-$*: modelled with a `RooKeysPdf` using phase-space $B^0 \to K^+\pi^-e^+e^-$ simulation.

- *Leakage from $B^+ \to K^+ J/\psi(\to e^+e^-)$ (central-q^2 only)*: modelled with a `RooKeysPdf` using simulated $B^+ \to K^+ J/\psi(\to e^+e^-)$ which pass the central-q^2 selection.

4.5.6 Background Normalisation Constraints

Although background normalisations are nuisance parameters, they are a source of fit instability. This is especially so for backgrounds with very low yields or backgrounds that peak underneath the signal region. At best, these backgrounds enlarge the signal mode uncertainties and at worse, the instability prevents MIGRAD from converging. Instead of floating their normalisation, the yields are parameterised as a function of the signal normalisations. These functions often include efficiencies of selecting and reconstructing a particular decay in the invariant mass window. The normalisation functions are channel-dependent.

Normalisation of B^0 resonant modes backgrounds

- *$\Lambda_b^0 \to pK J/\psi(\to \ell^+\ell^-)$ ($\Lambda_b^0 \to pK\psi(2S)(\to \ell^+\ell^-)$)*: Normalisation re-parameterised as a function of control mode yield, hadronisation fraction ratio, branching fraction ratio, and efficiency ratio. The value of $\frac{f_{\Lambda_b}}{f_d}$ is extracted from the p_{T}-averaged Λ_b^0 hadronisation fraction result in (24), which is 0.259 ± 0.018, and assuming isospin symmetry, $\frac{f_{\Lambda_b}}{f_d} = 2 \times (0.259 \pm 0.018) = 0.518 \pm 0.036$. Although this measurement is measured with 13 TeV proton-proton collisions, no such result exists for Run 1, and we assume the 13 TeV value for Run 1 as well.

$$N_{\Lambda_b^0}^{J/\psi} = \frac{f_{\Lambda_b}}{f_d} \cdot \frac{\mathcal{B}(\Lambda_b^0 \to pK J/\psi(\to \ell^+\ell^-))}{\mathcal{B}(B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-))} \cdot \frac{\varepsilon_{background}}{\varepsilon_{resonant}} \cdot N_{B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-)} \tag{4.38}$$

$$N_{\Lambda_b^0}^{\psi(2S)} = \frac{f_{\Lambda_b}}{f_d} \cdot \frac{\mathcal{B}(\Lambda_b^0 \to pK\psi(2S)(\to \ell^+\ell^-))}{\mathcal{B}(B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-))} \cdot \frac{\varepsilon_{background}}{\varepsilon_{resonant}} \cdot N_{B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)} \tag{4.39}$$

- *$B_s^0 \to K^{*0} J/\psi(\to \ell^+\ell^-)$ ($B_s^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$)*: Normalisation re-parameterised as a function of resonant mode yield, hadronisation fraction ratio (25; 24) and branching fraction ratio. The ratios are Gaussian constrained.

$$N_{B_s^0}^{J/\psi} = \frac{f_s}{f_d} \cdot \frac{\mathcal{B}(B_s^0 \to K^{*0} J/\psi(\to \ell^+\ell^-))}{\mathcal{B}(B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-))} \cdot N_{B^0 \to K^{*0} J/\psi(\to \ell^+\ell^-)} \tag{4.40}$$

$$N_{B_s^0}^{\psi(2S)} = \frac{f_s}{f_d} \cdot \frac{\mathcal{B}(B_s^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-))}{\mathcal{B}(B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-))} \cdot N_{B^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)} \tag{4.41}$$

Normalisation of $B^0 \to K^{*0}e^+e^-$ backgrounds

- *Leakage* $B^0 \to K^{*0}J/\psi\,(\to e^+e^-)$ *(central-*q^2 *only)*: Normalisation re-parameterised as a function of control mode yield in $J/\psi\,q^2$ and their efficiency ratio.

$$N_{Leakage} = \frac{\varepsilon^{central-q^2}_{B^0 \to K^{*0}J/\psi\,(\to e^+e^-)}}{\varepsilon^{J/\psi\,q^2}_{B^0 \to K^{*0}J/\psi\,(\to e^+e^-)}} \cdot N_{B^0 \to K^{*0}J/\psi\,(\to e^+e^-)} \tag{4.42}$$

Normalisation of B^+ resonant modes backgrounds

- $B^+ \to \pi^+ J/\psi\,(\to \ell^+\ell^-)$ *(*$B^+ \to \pi^+\psi(2S)(\to \ell^+\ell^-)$*)*: Normalisation re-parametrised as a function of control mode yield, branching fraction ratio and efficiency ratio. Branching fraction ratio is Gaussian constrained.

$$N_{B^+ \to \pi^+ J/\psi\,(\to \ell^+\ell^-)} = \frac{\mathcal{B}(B^+ \to \pi^+ J/\psi\,(\to \ell^+\ell^-))}{\mathcal{B}(B^+ \to K^+ J/\psi\,(\to \ell^+\ell^-))} \cdot \frac{\varepsilon_{background}}{\varepsilon_{control}} \cdot N_{B^+ \to K^+ J/\psi\,(\to \ell^+\ell^-)} \tag{4.43}$$

$$N_{B^+ \to \pi^+ \psi(2S)(\to \ell^+\ell^-)} = \frac{\mathcal{B}(B^+ \to \pi^+ \psi(2S)(\to \ell^+\ell^-))}{\mathcal{B}(B^+ \to K^+ \psi(2S)(\to \ell^+\ell^-))} \cdot \frac{\varepsilon_{background}}{\varepsilon_{control}} \cdot N_{B^+ \to K^+ \psi(2S)(\to \ell^+\ell^-)} \tag{4.44}$$

Normalisation of $B^+ \to K^+e^+e^-$ backgrounds

- *Leakage* $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ *(central-*q^2 *only)*: Normalisation re-parametrised as a function of control mode yield in $J/\psi\,q^2$ bin and their efficiency ratio.

$$N_{Leakage} = \frac{\varepsilon^{central-q^2}_{B^+ \to K^+ J/\psi\,(\to e^+e^-)}}{\varepsilon^{J/\psi\,q^2}_{B^+ \to K^+ J/\psi\,(\to e^+e^-)}} \cdot N_{B^+ \to K^+ J/\psi\,(\to e^+e^-)} \tag{4.45}$$

- *Partially-reconstructed* $B \to K^{(*)}ee$: This decay can be partially reconstructed as $B^+ \to K^+e^+e^-$ where the pion is missing. The normalisation is constrained to the $B^0 \to K^{*0}e^+e^-$ yield observed in data, an isospin conjugate extrapolation factor (explained in the next section), and the $B^0 \to K^{*0}e^+e^-$ efficiency ratio of being reconstructed in the B^+ and B^0 mass window. The efficiencies are Gaussian constrained.

$$N^{PartReco}_{B^0\to K^{*0}e^+e^-}(q^2) = \frac{\varepsilon^{PartReco}_{B^0\to K^{*0}e^+e^-}(q^2)}{\varepsilon^{Signal}_{B^0\to K^{*0}e^+e^-}(q^2)} \cdot f_{isospin} \cdot \mathcal{N}^{Signal}_{B^0\to K^{*0}e^+e^-}(q^2) \tag{4.46}$$

- *Partially-reconstructed* $B^0 \to K^+\pi^-e^+e^-$: The constraint above does not account for $B^0 \to K^+\pi^-e^+e^-$ decays outside the K^{*0} selection window. The yield of this background is fixed to the calculation in the next section.

Partially reconstructed backgrounds in low and central q^2 $K^+e^+e^-$ final state

$\underline{B^0 \to K^{*0}e^+e^-}$ As we have discussed above, the partially reconstructed $B^0 \to K^{*0}e^+e^-$ background constraint in $B^+ \to K^+e^+e^-$ mass window includes an isospin conjugate factor. This factor accounts for $B^+ \to K^{*+}e^+e^-$ decays and consists of two parts.

First, the isospin conjugate factor assumes the relative branching fraction of $B^0 \to K^{*0}\ell^+\ell^-$ and $B^+ \to K^{*+}\ell^+\ell^-$ respects isospin symmetry, supported by branching fraction measurements in their $\mu\mu$ modes by Ref. (12) and Ref. (13) respectively. The isospin symmetry assumption puts the ratio of their branching fraction at 1.0776. This factor is not unity due to difference between B^0 and B^+ lifetimes. Second, the branching fraction for $K^{*0} \to K^+\pi^-$ is 2/3 while the branching fraction for $K^{*+} \to K^+\pi^0$ is 1/3. Put together, the isospin extrapolation factor, $f_{isospin}$, is

$$\begin{aligned} f_{isospin} &= 1 + \frac{\tau(B^+)}{\tau(B^0)} \cdot \frac{\mathcal{B}(K^{*+} \to K^+\pi^0)}{\mathcal{B}(K^{*0} \to K^+\pi^-)} && (4.47) \\ &= 1 + 1.0776 \cdot \frac{1/3}{2/3} && (4.48) \\ &\simeq 1.54\ . && (4.49) \end{aligned}$$

Notice that because the relative K^* branching fraction modulates the size of $f_{isospin}$, even assigning a 10% relative uncertainty to the first part of the extrapolation only translate into a 3.5% relative on $f_{isospin}$. This is investigated as a systematic uncertainty in Sec. 4.6.3.

This extrapolation also accounts for s-wave $B^0 \to K^+\pi^-\ell^+\ell^-$ decays within the $\pm 100\,\text{MeV}$ K^{*0} mass window, since they contribute to the yields measured in $B^0 \to K^{*0}e^+e^-$ final states [35]. However in $K^+e^+e^-$ final states, no such protection exists against s-wave events outside the K^{*0} mass window. Thus, another calculation is needed to account for additional $B^0 \to K^+\pi^-e^+e^-$ backgrounds.

$\underline{B^0 \to K^+\pi^-e^+e^-}$ The efficiencies of s-wave $B^0 \to K^+\pi^-e^+e^-$ decays to be selected as $B^+ \to K^+e^+e^-$ candidates are computed using dedicated phase space $B^0 \to K^+\pi^-e^+e^-$ simulation. Their efficiencies are combined with the relevant beauty production cross-section and s-wave branching fractions to calculate the expected number of background events per unit luminosity.

A few considerations are made when using the s-wave branching fractions measured in Ref. (12). As already stated, the s-wave $\pm 100\,\text{MeV}$ around the K^{*0} mass is accounted for by the partially reconstructed $B^0 \to K^{*0}e^+e^-$ background in the data fit. Therefore, the s-wave branching fraction at $m(K\pi) < 1200\,\text{MeV}$, but outside $792 - 992\,\text{MeV}$, is $1.4 \cdot 10^{-8}$.

With their respective efficiencies and beauty production cross-section, this result in $1.3\,\text{fb}^{-1}$ and $2.5\,\text{fb}^{-1}$ events per unit luminosity at $7\,\text{TeV}$ and $13\,\text{TeV}$ respectively. This number then has to be scaled to for additional s-wave contributions above $m(K\pi) > 1200\,\text{MeV}$. This is done by noticing that because of the lower B^+ mass window of $4600\,\text{MeV}$, there is no efficiency to select s-wave events with a $m_{K\pi}$ above around $2400\,\text{MeV}$. We therefore extrapolate to this point, using again the efficiencies from phase space simulation, and assuming that the s-wave branching fraction below $1200\,\text{MeV}$ scales linearly to $2400\,\text{MeV}$. The expected s-wave contribution above $1200\,\text{MeV}$ is then $2.7\,\text{fb}^{-1}$ events per unit luminosity at $7\,\text{TeV}$ and $5.2\,\text{fb}^{-1}$ events per unit luminosity at $13\,\text{TeV}$.

Because the expected s-wave $B^0 \to K^+\pi^-e^+e^-$ yields use various extrapolations, they are assigned a 50% overall uncertainty. The associated systematic is measured in Sec. 4.6.3.

Partially Reconstructed Background Modelling in J/ψ and $\psi(2S)$-q^2

The partially reconstructed backgrounds in the J/ψ and $\psi(2S)$-q^2 bins contain an admixture of multiple decay modes. They are modelled with inclusive decays of B^0, B^+ and B_s^0 mesons, which are required to proceed via an intermediate $J/\psi \to \ell^+\ell^-$ or $\psi(2S) \to \ell^+\ell^-$ resonance.

---

[35] And also because no selections are applied on the K^{*0} helicity.

The individual PDFs of inclusive B^0, B^+ and B_s^0 simulations are combined linearly with normalisation factors, A_x,

$$g(B^0+B^++B_s^0) = \frac{1}{A_u+A_d+A_s}\cdot\Big(A_u\cdot g(B^+)+A_d\cdot g(B^0)+A_s\cdot g(B_s^0)\Big) \quad (4.50)$$

where the $g(B)$ denotes the individual PDFs of each B species, $g(B^0+B^++B_s^0)$ their total PDF and A_u, A_d, and A_s the normalisation factors of B^+, B^0 and B_s^0 respectively.

The normalisation factor of species x is computed by using the efficiency of the inclusive sample to pass the full selection, $\epsilon_x = \frac{N_{pass}(selection)}{N_{simulated}}$, the b hadronisation fraction of their respective species, f_x, and the total branching fraction of the inclusive decay channels, $\Sigma\mathcal{B}\ (B_x)$:

$$A_x = \epsilon_x * f_x * \Sigma\mathcal{B}(B_x). \quad (4.51)$$

Note that $f_{u,d} = 1$ for the B^0 and B^+ species while f_s assumes the values measured by Ref. (25; 24).

The total branching ratios of individual B decays can be computed by listing all the decay channels of their inclusive decay simulation, finding their branching fractions in PDG and summing them up. A tedious and error prone procedure. Instead, a dominant decay channel is chosen among the inclusive decay channels, and its PDG branching fraction is divided by the simulation fraction. For example, $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ has a fraction of 0.1850 in the inclusive B^0 simulation and its PDG branching fraction is 1.27×10^{-3} (2). Its total branching ratio is thus

$$\Sigma\mathcal{B}(B_d) = \frac{1.27\times10^{-3}}{0.1850} = 6.86\times10^{-3}. \quad (4.52)$$

4.5.7 Configurations of the Rare Mode Fits

Given the trigger definitions in Sec. 4.2.2, the rare mode fits are split into six exclusive samples:

i. Run 1 `L0I`,

ii. Run 1 `L0L` exclusive,

iii. $15+16$ `L0I`,

iv. $15+16$ `L0L` exclusive,

v. $17 + 18$ `L0I`,

vi. $17 + 18$ `L0L` exclusive.

The rare mode fits then defines four different configurations with these six orthogonal samples:

Nominal : R_K and R_{K^*} are measured simultaneously in all `L0` categories and data-taking periods.

`L0I` vs `L0L` exclusive : R_K and R_{K^*} are measured simultaneously in **two** exclusive datasets. One in the `L0I` category and one in the `L0L` exclusive category, measured simultaneously in the three data-taking periods.

Run 1 vs $15 + 16$ vs $17 + 18$: R_K and R_{K^*} are measured simultaneously in **three** exclusive datasets: Run 1, $15 + 16$ and $17 + 18$ data-taking periods. The measurements are done simultaneously in the two `L0` trigger categories.

Yields : The signal yields of B^0 and B^+ are measured in the **six** exclusive datasets.

The **nominal** configuration serves as the nominal result of R_K and R_{K^*}. The second, third and forth configurations serves as important cross check to ensure consistency between data taking periods and trigger categories.

4.5.8 Fits to Real Data

Since this analysis has blinded the efficiencies, the fit for **yields** configuration is the only test available to verify the convergence and stability of the fit models. Additionally, the results of **yields** rare mode fits are preliminary probes to access the statistical sensitivity of R_K and R_{K^*} measurements [36]. At the time of writing, another ongoing analysis blinds the signal yields in $17 + 18$ data samples. Hence, the $17 + 18$ dataset is excluded from the rare mode fits shown below [37]. The remaining four exclusive samples,

- Run 1 `L0I`,
- Run 1 `L0L` exclusive,

---

[36] Remember, these measurements are dominated by their statistical uncertainty.

[37] However, this blinding excludes the resonant modes, and the cross-checks to 2017 and 2018 datasets are not affected.

- $15 + 16$ `L0I`, and
- $15 + 16$ `L0L` exclusive,

are fitted independently. The rare mode fits in the $15 + 16$ `L0L` exclusive category are shown in Fig. 4.38- 4.43 while the rest can be found in Appendix E.3. The yields from the rare mode fits are tabulated in Table 4.14.

The simultaneous resonant mode fits between control and $\psi(2S)$ modes are shown in Figs. 4.44- 4.47, in the $15 + 16$ `L0L` exclusive sample.

final state	$q^2 bin$	lepton	year	L0I	L0L	Both Triggers
$K^{*0}l^+l^-$	low	ee	Run 1	27 ± 6	32 ± 6	58 ± 9
			15 + 16	30 ± 7	35 ± 7	66 ± 10
			Combined	57 ± 10	67 ± 10	124 ± 14
		$\mu\mu$	Run 1	87 ± 10	171 ± 14	258 ± 17
			15 + 16	101 ± 10	150 ± 13	252 ± 17
			Combined	189 ± 15	321 ± 19	509 ± 24
	central	ee	Run 1	48 ± 10	44 ± 8	92 ± 13
			15 + 16	52 ± 10	60 ± 10	112 ± 14
			Combined	100 ± 14	104 ± 13	204 ± 19
		$\mu\mu$	Run 1	117 ± 12	267 ± 18	384 ± 22
			15 + 16	158 ± 14	291 ± 18	450 ± 23
			Combined	276 ± 18	558 ± 26	834 ± 31
$K^+l^+l^-$	low	ee	Run 1	38 ± 8	45 ± 8	83 ± 11
			15 + 16	46 ± 9	41 ± 8	87 ± 12
			Combined	85 ± 12	86 ± 11	170 ± 16
		$\mu\mu$	Run 1	88 ± 11	238 ± 17	326 ± 20
			15 + 16	90 ± 10	231 ± 17	322 ± 20
			Combined	178 ± 15	470 ± 24	648 ± 28
	central	ee	Run 1	151 ± 17	144 ± 15	295 ± 23
			15 + 16	159 ± 18	170 ± 16	329 ± 24
			Combined	311 ± 25	313 ± 22	624 ± 33
		$\mu\mu$	Run 1	309 ± 20	889 ± 33	1199 ± 39
			15 + 16	344 ± 21	839 ± 32	1183 ± 38
			Combined	653 ± 29	1728 ± 46	2381 ± 55

Table 4.14 – Rare mode yields measured in the simultaneous fit to the rare and control modes. The yields in the four exclusive categories of Sec. 4.5.8 (Run 1 `L0I`, Run 1 `L0L` exclusive, $15 + 16$ `L0I`, and $15 + 16$ `L0L` exclusive) are measured independently. Note that `L0L` here denotes the exclusive `L0L` category.

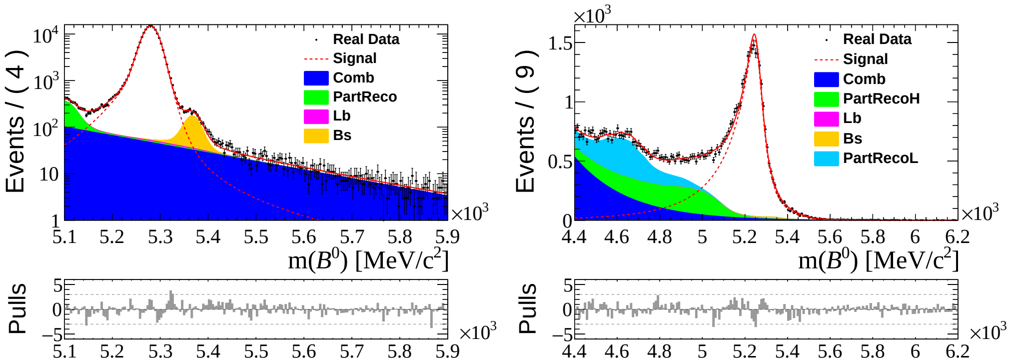


Figure 4.38 – Fits to reconstructed $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (right). Both models consist of a signal (dashed red line), a combinatorial background (blue), mis-identified $\Lambda_b^0 \to pKJ/\psi(\to \ell^+\ell^-)$ background (purple), and a $B_s^0 \to K^{*0}J/\psi(\to \ell^+\ell^-)$ background (orange) that decays into the same final state. The muon model has a partially reconstructed $B \to XJ/\psi(\to \mu\mu)$ background (green). The electron model splits the partially reconstructed $B \to XJ/\psi(\to ee)$ background into a hadronic (green) and leptonic (light blue) component.

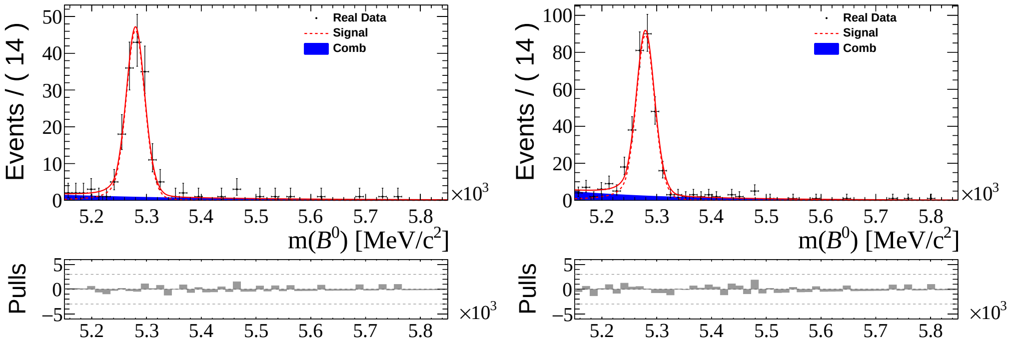


Figure 4.39 – Fits to $B^0 \to K^{*0}\mu^+\mu^-$ in the low (left) and central (right) q^2. The models consist of a signal (dashed red line) and a combinatorial background (blue).

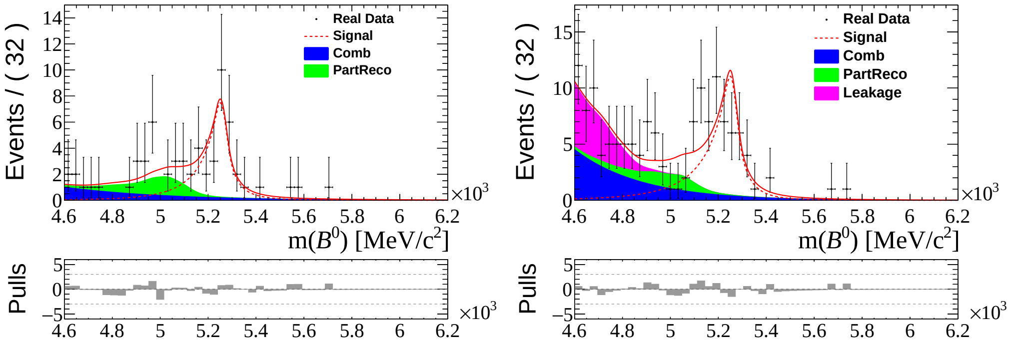


Figure 4.40 – Fits to $B^0 \to K^{*0}e^+e^-$ in the low (left) and central (right) q^2. Both models consist of a signal decay (dashed red line), partially reconstructed $B^+ \to K\pi\pi e^+e^-$ background (green) and a combinatorial background (blue). The central-q^2 also include a $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ leakage background (purple).

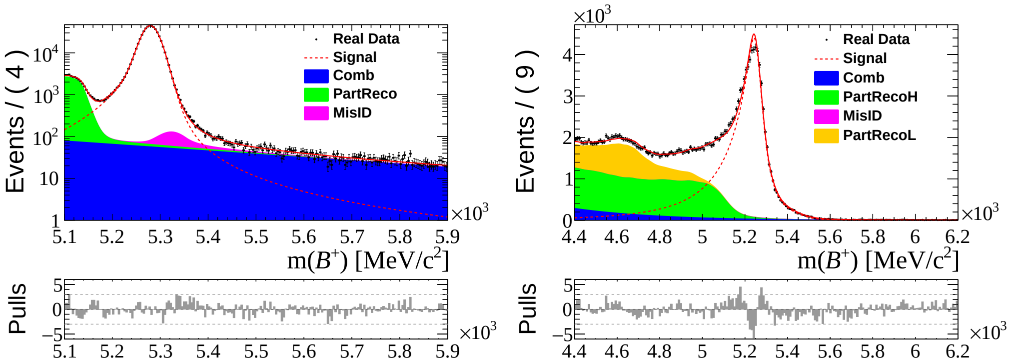


Figure 4.41 – Fits to reconstructed $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi (\to e^+e^-)$ (right). Both models consist of a signal (dashed red line), a combinatorial background (blue) and mis-identified $B^+ \to \pi^+ J/\psi (\to \ell^+\ell^-)$ background (purple). The muon model has a partially reconstructed $B \to X J/\psi (\to \mu\mu)$ background (green). The electron model splits the partially reconstructed $B \to X J/\psi (\to ee)$ background into a hadronic (green) and leptonic (light blue) component.

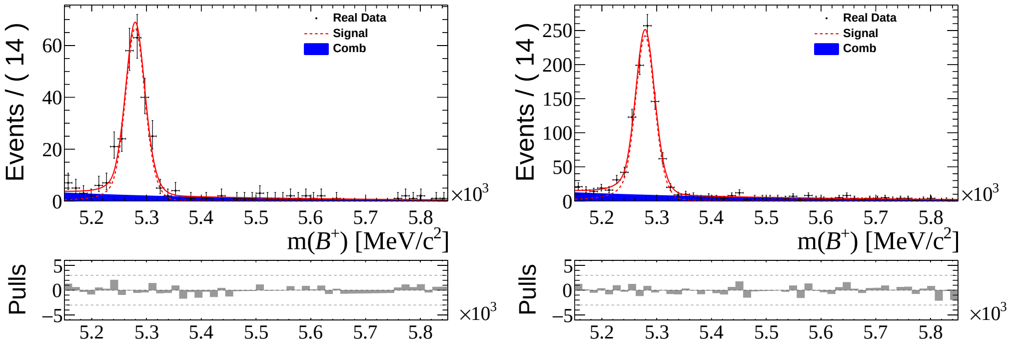


Figure 4.42 – Fits to $B^+ \to K^+\mu^+\mu^-$ in the low (left) and central (right) q^2. Both models consist of a signal (dashed red line) and a combinatorial background (blue).

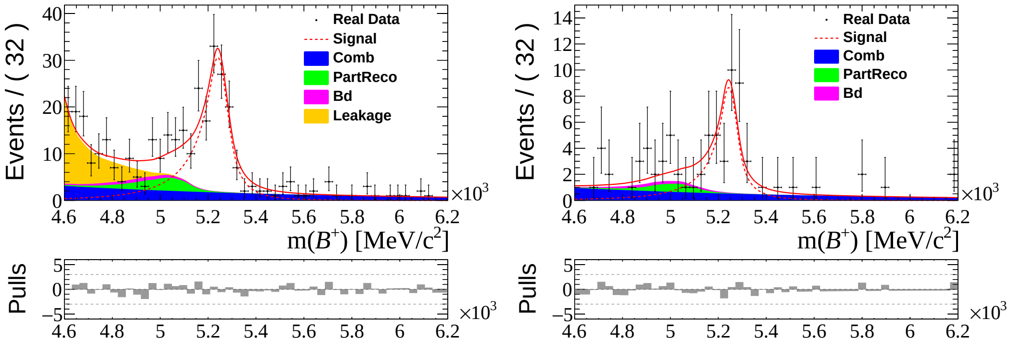


Figure 4.43 – Fits to $B^+ \to K^+e^+e^-$ in the low (left) and central (right) q^2. Both models consist of a signal decay (dashed red line), a combinatorial background (blue) , a partially reconstructed $B \to K^{(*)}ee$ background (green) and a partially reconstructed non-resonant $B^0 \to K^+\pi^-e^+e^-$ decays (purple) outside the K^{*0} mass window. The central-q^2 also include a $B^+ \to K^+ J/\psi (\to e^+e^-)$ leakage background (orange).

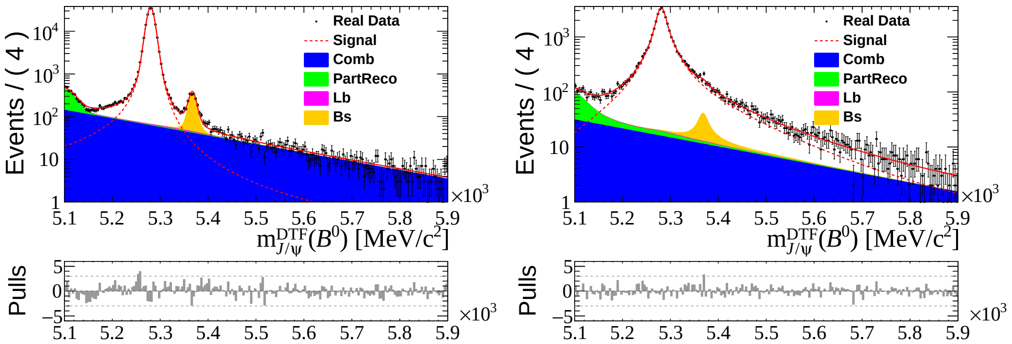


Figure 4.44 – Fits to data reconstructed as $B^0 \to K^{*0}J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0}J/\psi\,(\to e^+e^-)$ (right) with the J/ψ mass constraint. Both models consist of a signal (dashed red line), combinatorial background (blue), mis-identified $\Lambda_b^0 \to pK\,J/\psi\,(\to \ell^+\ell^-)$ baackgrounds (purple), $B_s^0 \to K^{*0}J/\psi\,(\to \ell^+\ell^-)$ background (light blue) with the same final state and a partially-reconstructed $B \to X\,J/\psi\,(\to \ell^+\ell^-)$ background (green) modelled with inclusive simulation samples. The fits above use datasets from $15+16$ `L0L` exclusive.

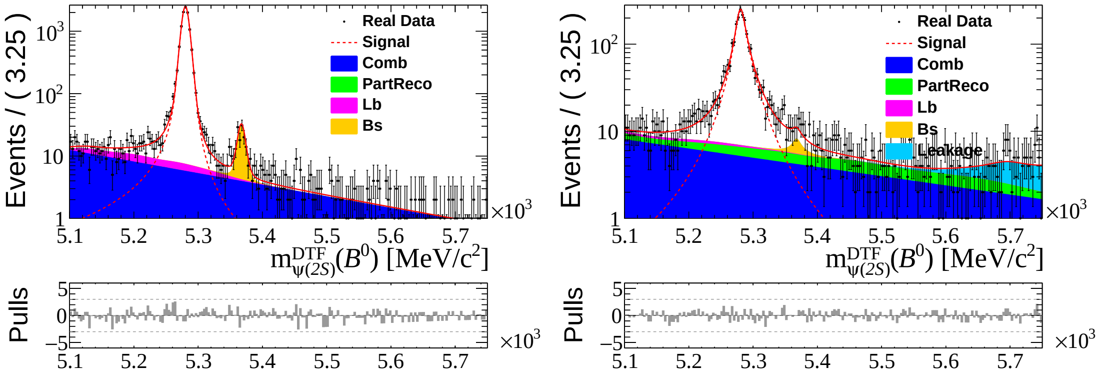


Figure 4.45 – Fits to data reconstructed as $B^0 \to K^{*0}\psi(2S)(\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ (right) with the $\psi(2S)$ mass constraint. Both models consist of a signal (dashed red line), a combinatorial background (blue), a mis-identified $\Lambda_b^0 \to pK\psi(2S)(\to \ell^+\ell^-)$ background (purple in $\mu\mu$, light blue in ee), $B_s^0 \to K^{*0}\psi(2S)(\to \ell^+\ell^-)$ background (light blue in $\mu\mu$, yellow in ee) that decays into the same final state and a partially-reconstructed background (green) modelled with inclusive simulation samples. Additionally, the ee modes contain a $B^0 \to K^{*0}J/\psi\,(\to e^+e^-)$ leakage background (purple). The fits above use datasets from $15+16$ `L0L` exclusive.

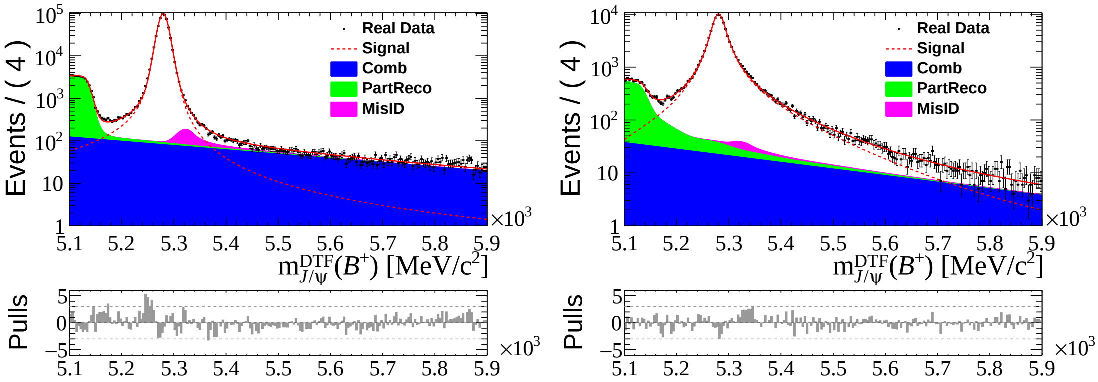


Figure 4.46 – Fits to data reconstructed as $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi(\to e^+e^-)$ (right) with the J/ψ mass constraint. Both models consist of a signal (dashed red line), a combinatorial background (blue), mis-identified $B^+ \to \pi^+ J/\psi(\to \ell^+\ell^-)$ background (purple) and a partially-reconstructed $B \to X J/\psi(\to \ell^+\ell^-)$ background (green) modelled with inclusive simulation samples. The fits above use datasets from 15 + 16 `L0L` exclusive.

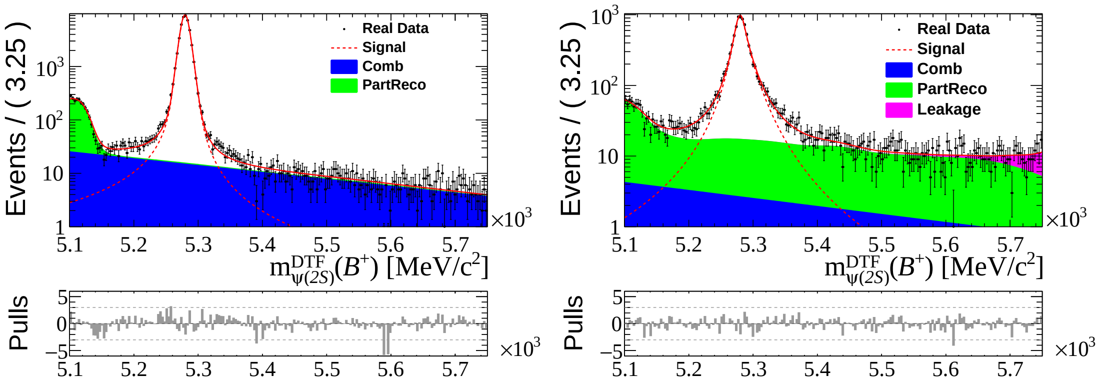


Figure 4.47 – Fits to data reconstructed as $B^+ \to K^+ \psi(2S)(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ \psi(2S)(\to e^+e^-)$ (right) with the $\psi(2S)$ mass constraint. Both models consist of a signal (dashed red line), a combinatorial background (blue), and a partially-reconstructed background (green) modelled with inclusive simulation samples. Additionally, the ee modes contain a $B^+ \to K^+ J/\psi(\to e^+e^-)$ leakage background (purple). The fits above use datasets from 15 + 16 `L0L` exclusive.

4.5.9 Fitter validation with pseudoexperiments

A standalone data-like pseudoexperiment generator was built using RooFit. The generator uses the converged fit model as truth seeds and it stores the output as ROOT nTuples, so that toy studies are repeatable. The tuples for each individual signal and background components are tagged and stored separately. By storing each component separately, the toy studies can vary the individual yields of each components and mix-and-match different background components. This allows the toy studies to investigate the effects of various background assumptions and to extrapolate the yields for sensitivity studies.

The toy fits feed pseudoexperiment data into the same fitting procedure as real data. Thus, the pseudoexperiment generate-and-fit procedure is used to validate the fit stability of the data fit and to detect the presence of intrinsic bias. Because efficiencies are blinded, only the **yields** configuration of the rare mode fit is validated with 1000 pseudoexperiment. A pseudoexperiment is classified as failed if either MIGRAD [38] or HESSE [39] failed in the minimization. For those that succeeds, the pull distributions of their fit parameters are fitted with a Gaussian. The Gaussian is expected to have zero mean and unit width for an unbiased fit.

The failure rates are reported in Tab. 4.15, which shows that the maximum failure rate of 5.8% is observed in the Run 1 `L0I` category. The failures in rare mode fits are mainly driven by the low background yields in rare *ee* modes, some of which are below 5.

Configuration	Splits	Failure Rate
Yields	Run 1 `L0I`	5.8%
	Run 1 `L0L` exclusive	5.4%
	$15+16$ `L0I`	1.5%
	$15+16$ `L0L` exclusive	3.1%

Table 4.15 – Failure rates of 1000 pseudoexperiments generated from fits to data and fitted with the same models.

Tab. 4.16-4.17 summarises the Gaussian distribution observed in signal yield parameters. The signal yields are unbiased except for the $B^+ \to K^+\mu^+\mu^-$ yields in Run 1 `L0I`, who show a roughly 10% bias from their truth values.

---

[38] Minima finder.

[39] A finite difference covariance matrix calculator.

final state	lepton	q^2 bin	year	`L0` trigger	$\mu \pm error_{\mu}$	$\sigma \pm error_{sigma}$
$K^+l^+l^-$	ee	J/ψ	Run 1	`L0I`	-0.017 ± 0.032	0.972 ± 0.024
				`L0L`	0.013 ± 0.033	1.016 ± 0.025
			15 + 16	`L0I`	-0.028 ± 0.035	1.060 ± 0.027
				`L0L`	0.046 ± 0.032	0.965 ± 0.023
		low	Run 1	`L0I`	-0.053 ± 0.032	0.982 ± 0.024
				`L0L`	0.041 ± 0.035	1.051 ± 0.027
			15 + 16	`L0I`	-0.040 ± 0.034	1.031 ± 0.025
				`L0L`	-0.015 ± 0.032	0.975 ± 0.024
		central	Run 1	`L0I`	-0.028 ± 0.033	1.007 ± 0.025
				`L0L`	-0.038 ± 0.032	0.963 ± 0.023
			15 + 16	`L0I`	0.015 ± 0.032	0.991 ± 0.024
				`L0L`	0.047 ± 0.033	1.005 ± 0.025
	$\mu\mu$	J/ψ	Run 1	`L0I`	0.044 ± 0.034	1.018 ± 0.025
				`L0L`	0.050 ± 0.034	1.024 ± 0.026
			15 + 16	`L0I`	-0.017 ± 0.033	1.001 ± 0.024
				`L0L`	0.042 ± 0.033	1.009 ± 0.025
		low	Run 1	`L0I`	-0.117 ± 0.035	1.047 ± 0.027
				`L0L`	-0.047 ± 0.034	1.019 ± 0.025
			15 + 16	`L0I`	-0.097 ± 0.033	1.024 ± 0.025
				`L0L`	0.000 ± 0.034	1.031 ± 0.026
		central	Run 1	`L0I`	-0.100 ± 0.033	1.001 ± 0.025
				`L0L`	0.032 ± 0.032	0.979 ± 0.024
			15 + 16	`L0I`	-0.037 ± 0.031	0.947 ± 0.022
				`L0L`	-0.005 ± 0.034	1.031 ± 0.026

Table 4.16 – Pulls of B^+ mode signal yields. The μ column contains the Gaussian means and σ column the Gaussian widths. A small bias, ≈ 10%, is observed in the $\mu\mu$ rare mode yields of Run 1 `L0I`.

final state	lepton	q^2 bin	year	`L0` trigger	$\mu \pm error_\mu$	$\sigma \pm error_{sigma}$
$K^{*0}l^+l^-$	ee	J/ψ	Run 1	`L0I`	-0.015 ± 0.033	0.994 ± 0.025
				`L0L`	0.004 ± 0.033	0.993 ± 0.024
			15 + 16	`L0I`	0.027 ± 0.034	1.045 ± 0.026
				`L0L`	-0.005 ± 0.034	1.035 ± 0.026
		low	Run 1	`L0I`	-0.007 ± 0.032	0.976 ± 0.024
				`L0L`	0.011 ± 0.034	1.034 ± 0.026
			15 + 16	`L0I`	-0.009 ± 0.031	0.963 ± 0.023
				`L0L`	0.013 ± 0.033	1.012 ± 0.025
		central	Run 1	`L0I`	0.016 ± 0.034	1.027 ± 0.026
				`L0L`	-0.080 ± 0.033	0.998 ± 0.025
			15 + 16	`L0I`	0.029 ± 0.031	0.961 ± 0.023
				`L0L`	-0.058 ± 0.034	1.028 ± 0.026
	$\mu\mu$	J/ψ	Run 1	`L0I`	0.026 ± 0.032	0.974 ± 0.024
				`L0L`	0.068 ± 0.033	1.000 ± 0.025
			15 + 16	`L0I`	0.051 ± 0.032	0.994 ± 0.024
				`L0L`	0.036 ± 0.031	0.954 ± 0.023
		low	Run 1	`L0I`	-0.017 ± 0.034	1.017 ± 0.025
				`L0L`	-0.001 ± 0.031	0.961 ± 0.023
			15 + 16	`L0I`	-0.037 ± 0.032	0.977 ± 0.024
				`L0L`	0.004 ± 0.032	0.969 ± 0.024
		central	Run 1	`L0I`	-0.041 ± 0.034	1.024 ± 0.026
				`L0L`	-0.033 ± 0.033	1.013 ± 0.025
			15 + 16	`L0I`	-0.028 ± 0.032	0.998 ± 0.024
				`L0L`	0.012 ± 0.034	1.023 ± 0.025

Table 4.17 – Pulls of B^0 mode signal yields. The μ column contains the Gaussian means and σ column the Gaussian widths.

4.6 Systematic Uncertainty

The systematic uncertainties due to limited size of calibration samples (Sec. 4.6.1) and efficiency corrections (Sec. 4.6.2) affects the efficiency estimation. The assumptions of the mass models can bias the yield measurements, and this systematic effect is probed in Sec. 4.6.3 with pseudoexperiments described in Sec. 4.5.9. If these systematics are properly accounted for, the residual systematics found by the flatness test of $r_{J/\psi}$ should be smaller than the total systematic uncertainty. If not, the non-flatness related systematics are accounted for by Sec. 4.6.4.

4.6.1 Systematic uncertainties due to the limited size of calibration samples

The calibration samples have a limited size. Although the associated uncertainty is of a statistical nature, it is a systematic uncertainty in the measurement of R_K and R_{K^*}. These systematic uncertainties, in the PID calibration samples and control mode samples, are evaluated with resampling methods.

PID calibration sample

The w_{PID} maps are resampled 100 times, where each bin resamples its efficiency by drawing from a Gaussian distribution. The mean of the Gaussian is the central value of the per-bin w_{PID} efficiency. The width of the Gaussian is the per-bin statistical uncertainty due to the size of the calibration sample falling into that bin. The resampling produces 100 variations of estimated efficiencies which is assigned as a systematic uncertainty.

Control modes

The systematic uncertainty due to limited statistics of the control mode is evaluated by bootstrapping the control sample, real and simulated, 100 times. The bootstrapping method assigns each event a weight, drawn from a Poisson distribution of mean 1. For each bootstrapped sample, the correction maps are recalibrated and efficiencies re-evaluated. The mass fits are not boostrapped because the corrections are swapped between B^0 and B^+ modes, and the correlation is negligible between the efficiency calibration and yield measurement.

At the time of writing, only the w_{L0} and $w_{\mathtt{HLT}}$ corrections are bootstrapped. Other corrections will be boostrapped as the analysis progresses. Examples of bootstrapped data/simulation efficiency ratios in `L0` corrections are shown in Fig. 4.48. The systematic uncertainties measured with bootstrapping w_{L0} and $w_{\mathtt{HLT}}$ are included in the integrated $r_{J/\psi}$ cross checks of Sec. 4.7.1.

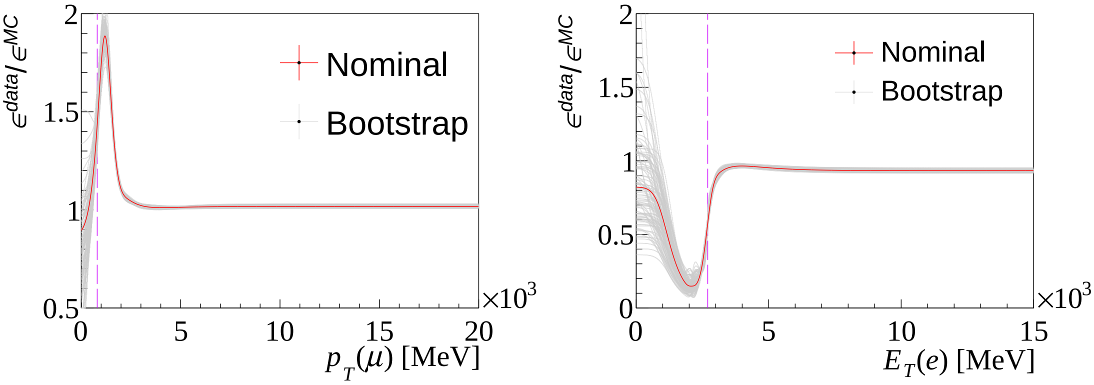


Figure 4.48 – The data/simulation efficiency ratios of inclusive `L0Muon` as a function of $p_{\mathrm{T}}(\mu)$ (left) and inclusive `L0Electron` as a function of E_T (right), evaluated with 2016 B^+ mode. The overlapping gray curves shows the statistical variation of the efficiency ratios between the 100 bootstrapped samples.

4.6.2 Systematic uncertainties due to simulation corrections and the efficiency measurement

These systematic uncertainties relate to the assumptions and methods used to calibrate and measure efficiencies. The associated systematics are evaluated by using an alternative approach, such as using different multiplicity proxy for the simulation reweighter. The difference between the alternative and nominal approach is taken as the systematic uncertainty.

Multiplicity Proxy

Although the GBReweighter used to reweight multiplicity in simulation shows a good data-simulation agreement in its multiplicity proxy `nTracks`, a significant difference between data and simulation remains in the `nSPDHits` distribution. Hence, the choice of multiplicity proxy is considered as a systematic uncertainty. Two alternative choices of multiplicity proxies are consider, the `nSPDHits` and `nPV`. GBReweighter is re-trained with these two alternatives, and the largest difference observed in the values of $r_{J/\psi}$ and rare mode efficiencies are assigned as the systematic uncertainties.

PID correction maps

There are three sources of systematic uncertainties with how PID efficiencies are evaluated [40]. First, the per-event PID efficiencies assume the per-track PID efficiencies are uncorrelated and factorisable. This assumption is not entirely true because the cherenkov rings between signal candidate tracks overlap in the RICH detectors. This systematic is evaluted using uncorrected simulation, by comparing the values of true efficiency, ϵ_{true}, to factorised efficiency, ϵ_{fact}, defined as

$$\epsilon_{true} = \frac{N(\text{Pass per-event PID cuts})}{N(\text{total})},$$

$$\epsilon_{fact} = \frac{\prod_i^m N_i(\text{Pass per-track PID cuts})}{\Big(N(\text{total})\Big)^m},$$

where i multiplies over m final state particles. Although the equation above uses uncorrected simulation [41], the difference between ϵ_{true} and ϵ_{fact} gives an estimate of how much the correlation between per-track w_{PID} affects per-event efficiencies. Therefore, this difference is assigned as a systematic uncertainty. The track independence systematics assigned to the values of $r_{J/\psi}$ due to the change in their efficiency ratios

$$\frac{\epsilon(B^+ \to K^+ J/\psi\,(\to e^+e^-))}{\epsilon(B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-))},$$

$$\frac{\epsilon(B^0 \to K^{*0} J/\psi\,(\to e^+e^-))}{\epsilon(B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-))},$$

are presented in Tab. 4.18.

---

[40] Beside the calibration sample statistics we have covered earlier.

[41] Remember that w_{PID} is the efficiency measured in real data. It is does not calibrate the simulated samples.

$r_{J/\psi}$	Year	Deviation (%)		
		`L0I` inc	`L0L` exc	`L0L` inc
$r_{J/\psi}^{K}$	11	0.4	0.3	0.4
	12	0.4	0.3	0.4
	15	1.0	0.7	1.0
	16	0.8	0.5	0.8
	17	0.7	0.5	0.7
	18	0.9	0.6	0.8
$r_{J/\psi}^{K^*}$	11	0.5	0.3	0.4
	12	0.5	0.3	0.5
	15	1.0	0.7	1.0
	16	0.9	0.6	0.8
	17	0.9	0.7	0.9
	18	0.9	0.5	0.8

Table 4.18 – The PID track independence systematics assigned, in percent, to the measured values of $r_{J/\psi}^{K}$ and $r_{J/\psi}^{K^*}$. The values quoted are the relative differences in control mode efficiency ratios, $\epsilon(ee)/\epsilon(\mu\mu)$, measured with and without the PID corrections.

Another source of w_{PID} systematics is related to how the w_{PID} efficiencies are accessed from the correction maps. Instead of using a linear interpolation between bin centers, w_{PID} weights are accessed by parameterising the w_{PID} maps with a KDE. Analogous to the track independence systematics, the difference in $\frac{\epsilon(ee)}{\epsilon(\mu\mu)}$ efficiency ratios are assigned as a systematic uncertainty on $r_{J/\psi}$. They are presented in Tab. 4.19 and the largest systematic shift observed is 1.1%.

The final source of systematic uncertainty is the ${}_s\mathcal{P}lot$ method used to subtract background contaminations in the K, π and μ efficiency maps [42]. The ${}_s\mathcal{P}lot$ systematic in K and π w_{PID} factorises in the lepton universality ratios, so their ${}_s\mathcal{P}lot$ related systematics are not considered. This systematic in the muon mode is assigned a 0.2% relative uncertainty, as per the suggestions by the working group responsible for maintaining LHCb's PID calibration samples.

---

[42] The electron fit-and-count method does not use ${}_s\mathcal{P}lot$

$r_{J/\psi}$	Year	Deviation (%)		
		`L0I` inc	`L0L` exc	`L0L` inc
$r^{K}_{J/\psi}$	11	1.1	1.0	1.1
	12	0.5	0.8	0.8
	15	0.2	0.2	0.2
	16	0.6	0.8	0.7
	17	0.3	0.5	0.4
	18	0.8	1.2	1.0
$r^{K^*}_{J/\psi}$	11	1.0	0.9	0.9
	12	0.5	0.8	0.7
	15	0.2	0.2	0.2
	16	0.6	0.8	0.7
	17	0.3	0.4	0.4
	18	0.7	1.1	1.0

Table 4.19 – The systematic uncertainty, in percent, assigned to the measured values of $r^{K}_{J/\psi}$ and $r^{K^*}_{J/\psi}$ due to way the PID calibration maps are accessed. The values here measured from the relative differences in control mode efficiency ratios, $\epsilon(ee)/\epsilon(\mu\mu)$.

Simulation form factors

The hadronic form-factors heavily affect the q^2 distribution in simulated $b \to s\,\ell^+\ell^-$ decays, which in turn modify the q^2 selection efficiencies. The simulated events are therefore reweighted with different form-factor models, and the resulting difference in simulated efficiencies are assigned as a systematic uncertainty.

4.6.3 Systematic uncertainty due to the mass fit

The systematic uncertainties due to the fit model are evaluated using toys fits performed with the pseudoexperiment setup described in Chap 4.5.9. Unless stated otherwise, the toy studies first generate toy nTuples with the nominal fit result as generator truths. These toy nTuples are fitted twice, once with the nominal model and once with an alternative fit model to obtain two values for our parameters of interest, the yields y^{nom} and y^{alt} respectively. The toy-by-toy difference distribution, of $y^{alt} - y^{nom}$, is fitted with a Gaussian. The mean and width of the Gaussian distribution is summed in quadrature and assigned as the systematic uncertainty. An example Gaussian distribution is shown on Fig. 4.49.

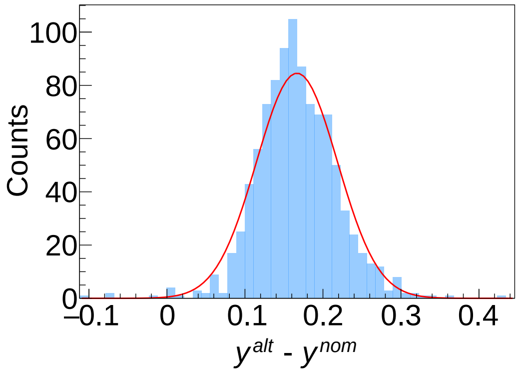


Figure 4.49 – Distribution of $N^{alt}(B^+ \to K^+e^+e^-) - N^{nom}(B^+ \to K^+e^+e^-)$ in central-q^2 15 + 16 `LOL` exclusive (blue). The alternative model constrains the background with different values than the nominal model. The toy-by-toy difference is fitted with a Gaussian (red). The Gaussian means and widths are added in quadrature, which gives an absolute bias of −0.18 events. With the nominal result of $N(B^+ \to K^+e^+e^-) = 170 \pm 16$, the relative systematic uncertainty is −0.10%.

Intrinsic fit bias

This systematic is intrinsic to the nominal fit model, and no alternative model is considered. As we have discussed in Sec. 4.5.9, this analysis generates and fit large samples of pseudoexperiments in order to verify fit stability and ascertain any intrinsic bias. A bias in the central values is observed when the pull mean is non-zero at $> 3\sigma$. The central value biases are used to correct the result of the data fit, and the biases are quoted as a systematic uncertainty on the relevant parameter. A bias observed in the uncertainties when their pull widths deviate from unity at $> 3\sigma$. The uncertainty biases are used to correct the covariance matrix in the data fit, but it is not assigned as a systematic uncertainty.

Signal Shape

This systematic is evaluated by using a different signal shape. The toy-by-toy variation between the nominal setup and the alternative signal shape is taken as the systematic uncertainty.

Background KDE shapes

The `RooKeysPdf` smoothing parameter is varied as an alternative background model. The toy-by-toy variation is taken as the systematic uncertainty.

Normalisation factors of inclusive partially-reconstructed background

The relative normalisation factor of the B^+, B^0 and B^0_s inclusive shapes (Sec. 4.5.6) is assigned a 10% uncertainty. These factors are varied within their uncertainty as an alternative model and the systematic uncertainty is evaluated on toy-by-toy variations between 1000 pseudoexperiments.

Uncertainty on the $K\pi\pi$ background

Because the resonant structure of $b \to K\pi\pi ee$ decay is unknown, the $m(K\pi\pi)$ spectrum is reweighted to the distributions observed in Ref. (26). The associated systematic uncertainty assumes the unweighted shape as an alternative model, and the systematic uncertainty is measured with 1000 pseudoexperiments.

Partially-reconstructed backgrounds in $B^+ \to K^+e^+e^-$ mass window

The two partially reconstructed backgrounds in $B^+ \to K^+e^+e^-$ mass fit windows rely on extrapolating branching fractions to calculate its yield (see Sec. 4.5.6). To calculate a systematic uncertainty, an alternative model is considered. This model varies the extrapolation factors are varied by $\pm 1\sigma$ of their associated uncertainty and the systematic uncertainty is extracted from the toy-by-toy difference of $N(B^+ \to K^+e^+e^-)$.

Although these are backgrounds in reconstructed $B^+ \to K^+e^+e^-$ mass windows, $B^0 \to K^{*0}e^+e^-$ signal yields are correlated to these extrapolation factors, via the signal yield constraint on partially reconstructed background yields in $B^+ \to K^+e^+e^-$. Therefore, the systematic effect of this extrapolation factor is also evaluated in the toy-by-toy difference of $N(B^0 \to K^{*0}e^+e^-)$.

Constrained backgrounds with $m_{K\pi}$ inside the K^* mass window

This background is constrained to the signal yield observed in $B^0 \to K^{*0}e^+e^-$ decays. The constraint relies on the relative branching fraction between isospin conjugates $B^+ \to K^{*+}e^+e^-$ and $B^0 \to K^{*0}e^+e^-$ decays. We assign a 10% uncertainty on the relative $B^+ \to K^{*+}e^+e^-$ branching fraction which propagates as a

3.5% uncertainty on the extrapolation factor. This systematic uncertainty is evaluated using 1000 pseudoexperiments and the uncertainty in $B^+ \to K^+e^+e^-$ and $B^0 \to K^{*0}e^+e^-$ signal yields are shown in Table 4.20.

q^2 bin	Data-taking period	`L0` Trigger	$B^+ \to K^+e^+e^-$	$B^0 \to K^{*0}e^+e^-$
low	Run 1	`L0I`	0.14 %	0.14 %
		`L0L`	0.14 %	0.12 %
	15 + 16	`L0I`	0.13 %	0.18 %
		`L0L`	0.12 %	0.14 %
central	Run 1	`L0I`	0.13 %	0.21 %
		`L0L`	0.12 %	0.25 %
	15 + 16	`L0I`	0.05 %	0.22 %
		`L0L`	0.10 %	0.26 %

Table 4.20 – The bias (in percentage) of $B^+ \to K^+e^+e^-$ and $B^0 \to K^{*0}e^+e^-$ signal yields between fits with nominal extrapolation factor and fits with isospin extrapolation factor varied by $\pm1\sigma$ of their uncertainties. The effect on $B^0 \to K^{*0}e^+e^-$ signal yields is slightly larger than on $B^+ \to K^+e^+e^-$ signal yields in central-q^2 . The overall systematic effect is at the sub-percent level.

Non-resonant backgrounds with $m_{K\pi}$ outside the K^* mass window

We assign the full calculation in Sec. 4.5.6 an overall 50% uncertainty on the expected yield. Table 4.21 shows the systematic effect of this background on $B^0 \to K^{*0}e^+e^-$ and $B^+ \to K^+e^+e^-$ signal yields, observed from 1000 fits to pseudoexperiments.

4.6.4 Residual Non-flatness of $r_{J/\psi}$

The values of $r_{J/\psi}$ should not vary as a function of any given variable. The effect of residual non-flatness as a systematic is quantified as a flatness parameter, d_f which will be given in Sec. 4.7.2 (Eq. 4.54 for the impatient). Should a residual non-flatness is observed, the simulated samples are re-weighted such that the differential $r_{J/\psi}$ is flat, and the observed change in R_K and R_{K^*} is taken as a systematic uncertainty.

q^2 bin	Data-taking period	`L0` Trigger	$B^+ \to K^+e^+e^-$	$B^0 \to K^{*0}e^+e^-$
low	Run 1	`L0I`	0.22 %	0.28 %
		`L0L`	0.58 %	0.40 %
	15 + 16	`L0I`	0.12 %	0.11 %
		`L0L`	0.37 %	0.32 %
central	Run 1	`L0I`	0.07 %	0.24 %
		`L0L`	0.12 %	0.55 %
	15 + 16	`L0I`	0.05 %	0.22 %
		`L0L`	0.11 %	0.46 %

Table 4.21 – The bias (in percentage) of $B^+ \to K^+e^+e^-$ and $B^0 \to K^{*0}e^+e^-$ signal yields between fits with nominal expected sWave yields and fits with expected sWave yields varied by $\pm 1\sigma$ of their uncertainties. The systematic effect on $B^0 \to K^{*0}e^+e^-$ compared to $B^+ \to K^+e^+e^-$ is smaller in low-q^2 but larger in central-q^2 . The overall systematic effect is at the sub-percent level.

4.6.5 Summary of Systematic Uncertainties

A summary of the systematic uncertainties is shown in Tab. 4.22 which includes all the systematic uncertainties considered for the measurement of R_K and R_{K^*}. Since this analysis is ongoing, some systematics are currently unavailable but wherever possible, an expected systematic uncertainty is assigned based on the published measurements of Ref. **??** and Ref. **??**.

Systematic Uncertainty		R_K [%]	R_{K^*} [%]
Calibration Sample Size		2.9	1.7
Efficiencies	Multiplicity Proxy	1.5	WIP
	PID track independence	1.0	1.0
	PID, interpolation	1.1	1.1
	PID, ${}_s\mathcal{P}lot$	0.2	0.2
	Hadronic form-factors	0.4	1.2
Fit Model	Intrinsic Fit Bias	WIP	WIP
	Signal Model	WIP	WIP
	Background KDE	WIP	WIP
	Inclusive $B \to XJ/\psi$ part reco	WIP	WIP
	$B \to K\pi\pi ee$	N/A	1.0
	$B \to K^*ee$ background isospin extrapolation	0.1	0.3
	s-wave $B \to K^+\pi^-e^+e^-$ background	0.6	0.6
Non Flatness		2.0	2.0
Total		4.2	3.5

Table 4.22 – Summary of systematic uncertainties and their preliminary values. The numbers in black has been evaluated by this analysis while the numbers in blue are extrapolated from previous analysis. The total values exclude systematics that are work in progress (WIP).

4.7 Cross checks

As we have discussed in Chap. 4.1, the measurement of integrated $r_{J/\psi}$ is a stringent closure test to ensure the efficiencies are calibrated correctly. Another cross check is the $r_{J/\psi}$ flatness test to probe for potential trends in differential $r_{J/\psi}$, which could be diluted in the integrated $r_{J/\psi}$ measurement. Lastly, the $R_{\psi(2S)}$ double ratio test checks that correction porting between q^2 regions do not bias the measurement of LFU ratios.

4.7.1 Integrated $r_{J/\psi}$

The values of $r_{J/\psi}$ are measured in each data-taking year, driven by the fact that the simulation are configured differently between each year. As we have explained in Chap. 4.2.2, the values of $r_{J/\psi}$ are probed in three `L0` trigger categories, which are

- inclusive `L0I`,
- exclusive `L0L` (`(!L0I) && L0L`),
- inclusive `L0L`.

Although the measurements of R_K and R_{K^*} only considers the first two categories, the exclusive `L0L` efficiency calibration depends on correctly calibrating the inclusive `L0L` efficiency (Eq. 4.20 for reference). Therefore, the resonant mode cross checks will also include the `L0L` inclusive category for completeness.

Recall back in Chap. 4.1, the correction maps between B^0 and B^+ modes are swapped. This avoids an irreducible correlation due to using the same control sample to calibrate efficiencies and measure yields. Although the swapping corrections between B modes serve as the nominal method, the values of $r_{J/\psi}$, and later $R_{\psi(2S)}$, are measured with both B^0 and B^+ weights. This redundancy ensures that, the values of $r_{J/\psi}$ and $R_{\psi(2S)}$ are compatible with and without swapping corrections between B modes, despite the intrinsic correlation without swapping correction maps.

The values of $r_{J/\psi}$ (B^0) and $r_{J/\psi}$ (B^+) are shown in Fig. 4.50 for the 15+16 data-taking periods. The rest can be found in Appendix F.1. The result of $r_{J/\psi}$ with B^+ and B^0 as the calibration sample are superimposed. In these plots, $r_{J/\psi}$ are mostly incompatible with unity. This is because the error bars are underestimated as they only include the statistical uncertainties reported by RooFit and the systematic uncertainties from bootstrapping the `L0` and `HLT` corrections. Other systematic uncertainties are unaccounted for at this stage of the analysis.

The measured values of $r_{J/\psi}$ (B^0) along the correction chain is shown in Fig. 4.51, in the $15 + 16$ `L0L` exclusive category. As we can see, $r_{J/\psi}$ approaches unity as more corrections are accounted for in the efficiencies. Note that the quoted values of $r_{J/\psi}$ (B^0) in Fig. 4.51 only include statistical uncertainties.

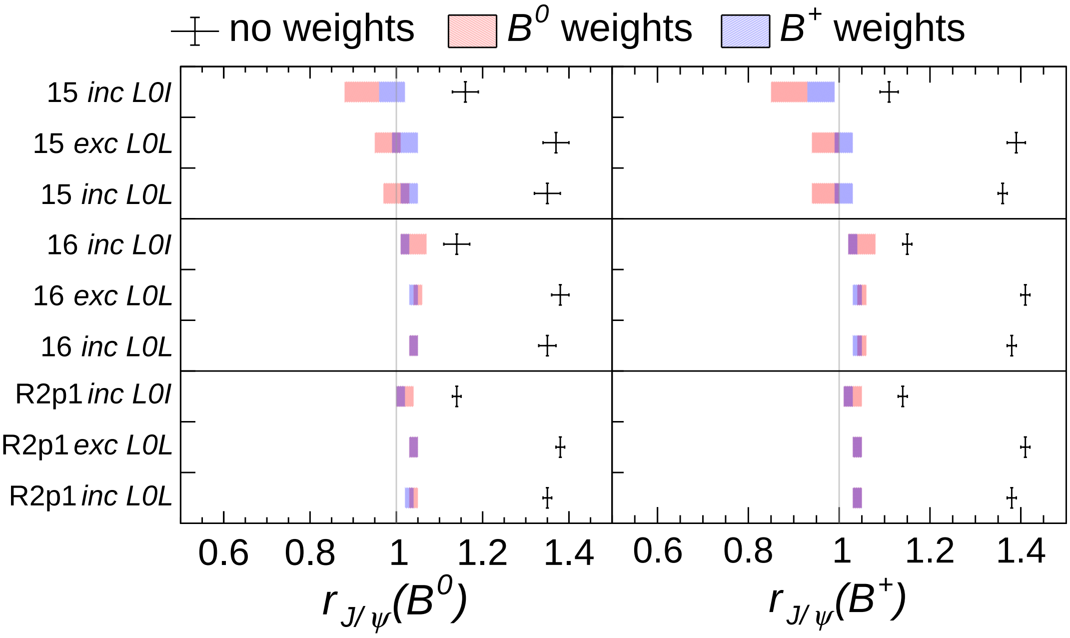


Figure 4.50 – The results of integrated $r_{J/\psi}$ measurements, in the $K^{*0}\ell\ell$ (left) and $K^+\ell^+\ell^-$ (right) final states. The efficiencies which uses B^0 (red) and B^+ (blue) control mode as the calibration samples show a good compatibility with one another. The 2015 and 2016 datasets are combined into the Run 2 part 1 (R2p1) result.

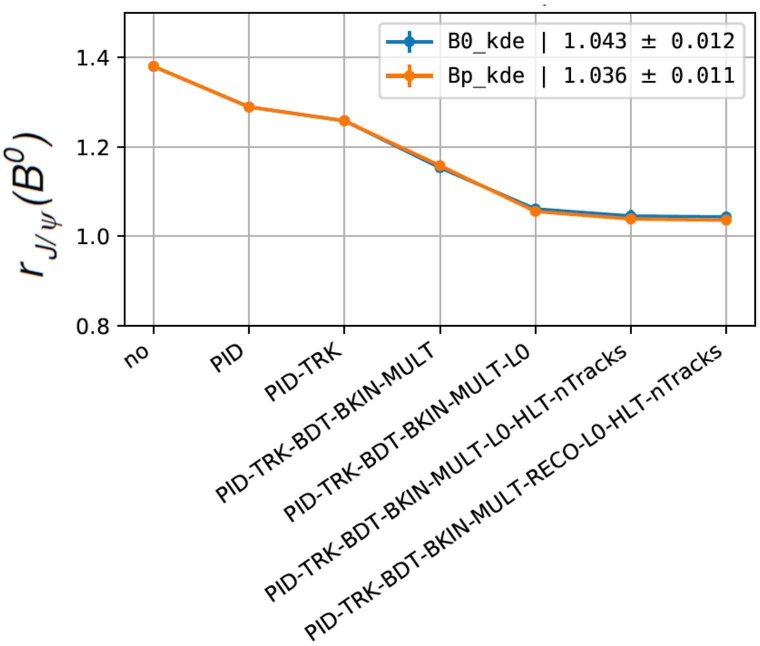


Figure 4.51 – The evolution of $r_{J/\psi}$ (B^0) across the correction chain, in the $15+16$ `L0L` exclusive category. The corrections with B^+ control modes are shown in orange and the corrections with B^0 control modes are shown blue.

4.7.2 Flatness of $r_{J/\psi}$

The $r_{J/\psi}$ flatness test checks for local deviations or trends in the differential values of $r_{J/\psi}$, which can be washed out by the integrated $r_{J/\psi}$ measurements. The flatness test uses a 1D iso-binning scheme. The iso-binning scheme selects events ±60 MeV around the J/ψ mass constrained B mass, and determine the cuts by mixing the electron and muon samples. The electron modes have smaller statistics, and the iso-binning scheme assigns a weight to electron events

$$w_e = \frac{N_{\mu,total}}{N_{e,total}} \tag{4.53}$$

such that the electron and muon samples have equal total weight.

The flatness test measures the differential $r_{J/\psi}$ across 8 bins of 20 variables in total. The full list of these variables can be found Appendix F.2.

These variables are composed of particle kinematics [43] and also event multiplicity related variables. Fig. 4.52 shows the differential $r_{J/\psi}$ (B^0) as a function of B^0 p_{T} and nSPDHits. In these plots, $r_{J/\psi}$ shows no dependency along B^0 p_{T} but there is a clear trend if $r_{J/\psi}$ is measured differentially along nSPDHits.

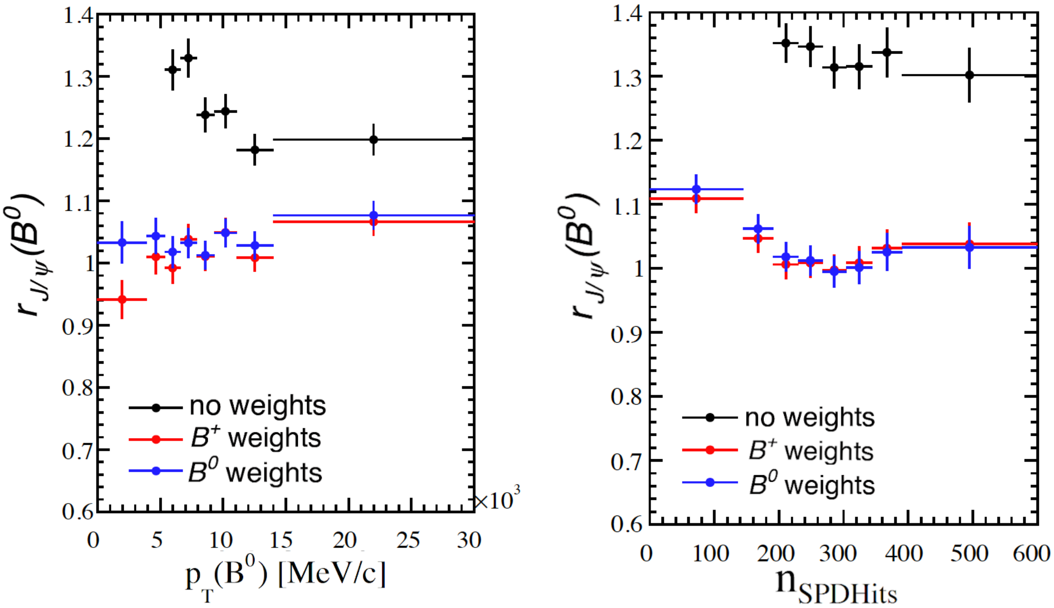


Figure 4.52 – The differential values of $r_{J/\psi}$ (B^0) as a function of B^0 p_{T} (left) and nSPDHits (right), in the 15 + 16 `L0L` exclusive datasets.

---

[43] Those of the reconstructed B and its daughter particles.

Any residual non-flatness, as we have observed in nSPDHit, can systematically affect the measured value of R_K and R_{K^*}. These effects can be quantified with a flatness parameter, d_f, defined as

$$d_f = \frac{\sum_i^8 \epsilon^i_{rare,\mu} \cdot \mathcal{Y}^i_\mu}{\sum_i^8 \epsilon^i_{rare,\mu}} \cdot \frac{\sum_i^8 \epsilon^i_{J/\psi,\mu}}{\sum_i^8 N^i_\mu} \Bigg/ \frac{\sum_i^8 \epsilon^i_{rare,e} \cdot \mathcal{Y}^i_e}{\sum_i^8 \epsilon^i_{rare,e}} \cdot \frac{\sum_i^8 \epsilon^i_{J/\psi,e}}{\sum_i^8 N^i_e} - 1, \tag{4.54}$$

where N^i_ℓ denotes the control mode yield measured in bin i and $\mathcal{Y}^i_e$ is the efficiency corrected control mode yield,

$$\mathcal{Y}^i_\ell = \frac{N^i_\ell}{\epsilon^i_{J/\psi,\ell}}. \tag{4.55}$$

Essentially, the d_f parameter calculates a pseudo double ratio, but the rare or $\psi(2S)$ mode yields are replaced with $\epsilon^i_{rare,\ell} \cdot \mathcal{Y}^i_\ell$ terms instead. Because $\mathcal{Y}^i_\ell$ is lepton flavour universal, d_f should be compatible with zero if the efficiencies are properly calibrated. In practice, d_f are required to be smaller than the total systematic uncertainty of R_K and R_{K^*}. Otherwise, it hints at additional systematic effects that do not cancel in the efficiency double ratio of Eq. 4.8 and 4.9.

4.7.3 Double ratio $R_{\psi(2S)}$

The cross-check of $R_{\psi(2S)}$ validates the portability of simulation corrections from the control modes to a different q^2 region. Similar to the cross-check of integrated $r_{J/\psi}$, the values of $R_{\psi(2S)}$ is checked in each data-taking year, in three `L0` trigger categories, and with correction weights from both B^+ and B^0 modes. The results of $R_{\psi(2S)}$ (B^0) are shown in Fig. 4.53 for the 15 + 16 data-taking period. The rest can be found in Appendix F.3. For $R_{\psi(2S)}$ (B^+), their results are not shown because at the time of writing, a bug in simulated $B^+ \to K^+\psi(2S)(\to e^+e^-)$ biases its efficiency measurement. Therefore, the result of $R_{\psi(2S)}$ (B^+) are currently unreliable.

Of note, the measurement of $R_{\psi(2S)}$ differs little between the unweighted efficiencies against fully weighted efficiencies. This is because $R_{\psi(2S)}$ is measured as a double-ratio. Fig. 4.54 shows the evolution of $R_{\psi(2S)}$ in the correction chain, which changes little between correction steps compared to the evolution of $r_{J/\psi}$. This speaks to the robustness of double-ratio measurements, where systematic effects embedded in the same final states cancel off.

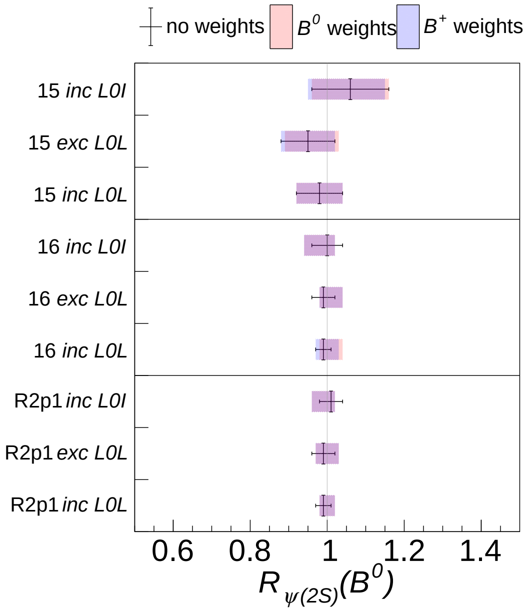


Figure 4.53 – The results of integrated $R_{\psi(2S)}$ (B^0) measurements. The efficiencies which uses B^0 (red) and B^+ (blue) control mode as the calibration samples show a good compatibility with one another. The 2015 and 2016 datasets are combined into the Run 2 part 1 (R2p1) result.

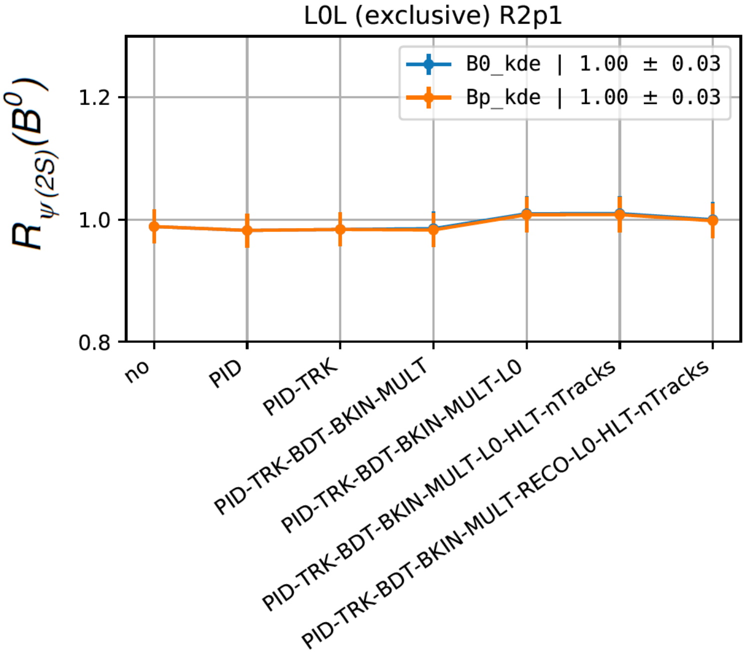


Figure 4.54 – The evolution of $R_{\psi(2S)}$ (B^0) across the correction chain, in the $15+16$ `L0L` exclusive category. The corrections with B^+ control modes are shown in orange and the corrections with B^0 control modes are shown blue. Notice that $R_{\psi(2S)}$ (B^0) varies much less between correction steps compared to $r_{J/\psi}$ (B^0).

4.8 Results and Conclusion

This analysis is in an advanced state but many systematic uncertainties remain to be evaluated before the measurement of R_K and R_{K^*} is unblinded. That being said, we can infer an expected statistical sensitivity from the measured yields. As we can see from the statistical sensitivities in Tab. 4.23, the electron modes, due to their lower yields, are the main drivers of the statistical reach in R_X measurements by LHCb.

final state	$q^2 bin$	lepton	year	`L0I`	`L0L`	Both Triggers
$K^{*0}l^+l^-$	low	ee	Run 1	23.6 %	20.3 %	15.4 %
			15 + 16	24.7 %	20.5 %	15.9 %
			Combined	17.2 %	14.4 %	11.1 %
		$\mu\mu$	Run 1	11.7 %	8.1 %	6.7 %
			15 + 16	10.2 %	8.7 %	6.6 %
			Combined	7.7 %	5.9 %	4.7 %
	central	ee	Run 1	20.4 %	18.0 %	13.7 %
			15 + 16	19.4 %	16.2 %	12.5 %
			Combined	14.1 %	12.0 %	9.2 %
		$\mu\mu$	Run 1	10.1 %	6.7 %	5.6 %
			15 + 16	8.6 %	6.3 %	5.1 %
			Combined	6.6 %	4.6 %	3.8 %
$K^+l^+l^-$	low	ee	Run 1	20.3 %	17.4 %	13.2 %
			15 + 16	19.3 %	19.2 %	13.7 %
			Combined	14.0 %	12.9 %	9.5 %
		$\mu\mu$	Run 1	12.4 %	7.1 %	6.2 %
			15 + 16	11.5 %	7.3 %	6.1 %
			Combined	8.4 %	5.1 %	4.3 %
	central	ee	Run 1	11.5 %	10.4 %	7.8 %
			15 + 16	11.0 %	9.7 %	7.3 %
			Combined	8.0 %	7.1 %	5.3 %
		$\mu\mu$	Run 1	6.6 %	3.7 %	3.2 %
			15 + 16	6.2 %	3.8 %	3.2 %
			Combined	4.5 %	2.7 %	2.3 %

Table 4.23 – The statistical sensitivity of this measurement, based on the signal yields in Tab. 4.14. Notice that ee mode sensitivities are at best half the sensitivity of corresponding $\mu\mu$ modes. The `L0L` column denotes the `L0L` exclusive trigger category.

The number of $B^+ \to K^+e^+e^-$ signal in central-q^2 measured by this analysis in the Run 1 and $15+16$ data-taking period is 624 ± 33. This analysis observed fewer events compared to LHCb's latest published result of R_K (27), which measured $N(B^+ \to K^+e^+e^-) = 766 \pm 48$. Despite losing events by dropping the `L0Hadron` trigger, this analysis is able to achieve a better statistical sensitivity of 5.3 % compared to the 6.3 % achieved by Ref. (27), approximately a 16 % improvement. This improvement can be attributed to the better background suppression provided by the MVA and HOP cuts (Chap. 4.3.3 and 4.3.4). Equally important to this sensitivity improvement is the constraints on partially reconstructed $B^0 \to K^{*0}e^+e^-$ and $B^0 \to K^+\pi^-e^+e^-$ backgrounds in the the mass fits to $B^+ \to K^+e^+e^-$ candidates (Chap. 4.5.6). These constaints add sub-percent level systematics to the yields measurements (Chap. 4.6.3), a negligible cost compared to the significant increase in sensitivity. In the low-q^2 , this analysis will provide the first measurement of R_K in the $0.1 < q^2 < 1.1\,\mathrm{GeV}^2$ region. This measurement can serve as an additional constraint to the global fits of $b \to s\,\ell^+\ell^-$ decays (Chap. 2.5.1) and as a complementary cross check of R_{K^*} measurements in this q^2 bin [44].

The improvement in sensitivity does not extend to the measurement of R_{K^*} with Run 1 data. The number of $B^0 \to K^{*0}e^+e^-$ numbers observed in Run 1 by this analysis and the previous measurement by LHCb are presented in Tab. 4.24. In the low-q^2 bin, these yields translate into statistical sensitivities of 13.4 % by Ref. (19) and 15.5 % by this analysis. For the central-q^2 bins, it is 12.7 % by Ref. (19) and 13 % by this analysis. The larger reduction in low-q^2 sensitivity is expected due to its higher `L0Hadron` efficiency loss, as we have discussed in Chap. 4.2.2.

q^2 bin	Ref. (19)	*This*
low	89 ± 12	58 ± 9
central	111 ± 14	92 ± 13

Table 4.24 – Number of $B^0 \to K^{*0}e^+e^-$ events observed in Run 1 by Ref. (19) and this analysis.

---

[44] Remember our discussion in Chap. 4.1 that photon pole related Wilson coefficients agrees with SM expectations.

Although the yields of $17+18$ rare modes are blinded due to a parallel measurement of R_K [45], they can be extrapolated with the yields of the unblinded $J/\psi \to \ell^+\ell^-$ control modes. The extrapolations in Tab. 4.25 assumes the $17+18$ rare mode yields scale from $15+16$ just like the control modes. The uncertainties quoted in this table assumes the $15+16$ uncertainties scale as $\sqrt{N}$ into $17+18$.

final state	lepton	$q^2 bin$	`L0I`	`L0L`	Both Triggers
$K^{*0}\ell\ell$	ee	low	61 ± 10	67 ± 10	128 ± 14
		central	106 ± 14	114 ± 14	220 ± 20
	$\mu\mu$	low	200 ± 14	310 ± 19	510 ± 24
		central	313 ± 20	602 ± 26	915 ± 33
$K^+\ell^+\ell^-$	ee	low	90 ± 13	79 ± 11	169 ± 17
		central	310 ± 25	328 ± 22	638 ± 33
	$\mu\mu$	low	172 ± 14	476 ± 24	648 ± 28
		central	656 ± 29	1729 ± 46	2385 ± 54

Table 4.25 – The extrapolated yields of rare modes in $17+18$. The `L0L` column denotes the `L0L` exclusive trigger category.

Combined with the yields observed in Tab. 4.14, the expected yields and sensitivity of this analysis with the full LHCb datasets from Run 1 and Run 2 are given in Tab. 4.26. Compared to the current published results of R_K and R_{K^*} by LHCb, the statistical sensitivities will improve in ee modes by about 41 % in central-q^2 R_K, 49 % in central-q^2 R_{K^*} and 41 % in low-q^2 R_{K^*}.

---

[45] The other analysis blinds the data yields, as opposed to the efficiency blinding strategy of this analysis.

final state	lepton	$q^2 bin$	Sensitivity
$K^{*0}\ell\ell$	ee	low	7.9 %
		central	6.5 %
	$\mu\mu$	low	3.3 %
		central	2.6 %
$K^+\ell^+\ell^-$	ee	low	6.9 %
		central	3.7 %
	$\mu\mu$	low	3.1 %
		central	1.6 %

Table 4.26 – The expected statistical sensitivities of the rare modes with the full LHCb Run 1 and Run 2 datasets, including the extrapolated 17 + 18 yields in Tab. 4.25.

4.9 Bibliography

[1] M. Bordone, G. Isidori, and A. Pattori, *On the standard model predictions for R_K and R_{K^*}*, The European Physical Journal C **76** (2016) .

[2] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020), no. 8 083C01.

[3] BESIII, M. Ablikim *et al.*, *Precision measurements of B[ψ(3686)→π+π-J/ψ] and B[J/ψ→l+l-]*, Phys. Rev. D **88** (2013), no. 3 032007, `arXiv:1307.1189`.

[4] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.

[5] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, `arXiv:hep-ph/0603175`.

[6] A. Ryd *et al.*, *EvtGen: A Monte Carlo Generator for B-Physics*, .

[7] E. Barberio, B. van Eijk, and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. **66** (1991) 115.

[8] S. Agostinelli *et al.*, *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003), no. 3 250 .

[9] J. Allison *et al.*, *Geant4 developments and applications*, IEEE Transactions on Nuclear Science **53** (2006), no. 1 270.

[10] J. Allison *et al.*, *Recent developments in geant4*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **835** (2016) 186 .

[11] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.

[12] LHCb, R. Aaij *et al.*, *Measurements of the S-wave fraction in* $B^0 \to K^+\pi^-\mu^+\mu^-$ *decays and the* $B^0 \to K^*(892)^0\mu^+\mu^-$ *differential branching fraction*, JHEP **11** (2016) 047, `arXiv:1606.04731`, [Erratum: JHEP 04, 142 (2017)].

[13] LHCb, R. Aaij *et al.*, *Differential branching fractions and isospin asymmetries of* $B \to K^{(*)}\mu^+\mu^-$ *decays*, JHEP **06** (2014) 133, `arXiv:1403.8044`.

[14] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. A **552** (2005) 566, `arXiv:physics/0503191`.

[15] R. Kalman, *A new approach to linear filtering and prediction problems" transaction of the asme journal of basic*, 1960.

[16] A. V. Dorogush, V. Ershov, and A. Gulin, *Catboost: gradient boosting with categorical features support*, ArXiv **abs/1810.11363** (2018).

[17] T. Likhomanenko *et al.*, *Reproducible Experiment Platform*, J. Phys. Conf. Ser. **664** (2015), no. 5 052022, `arXiv:1510.00624`.

[18] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, 01, 2017. doi: 10.1201/9781315139470.

[19] LHCb, R. Aaij *et al.*, *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, JHEP **08** (2017) 055, `arXiv:1705.05802`.

[20] LHCb, R. Aaij *et al.*, *Measurement of the electron reconstruction efficiency at LHCb*, `arXiv:1909.02957`.

[21] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at lhcb*, Tech. Rep. LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039, CERN, Geneva, May, 2014.

[22] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. A **764** (2014) 150, `arXiv:1312.5000`.

[23] LHCb, R. Aaij *et al.*, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001, `arXiv:1507.03414`.

[24] LHCb, R. Aaij *et al.*, *Measurement of b hadron fractions in 13 TeV pp collisions*, Phys. Rev. D **100** (2019), no. 3 031102, `arXiv:1902.06794`.

[25] LHCb, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 001, `arXiv:1301.5286`.

[26] LHCb, R. Aaij *et al.*, *First observations of the rare decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$*, JHEP **10** (2014) 064, `arXiv:1408.1137`.

[27] LHCb, R. Aaij *et al.*, *Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **122** (2019), no. 19 191801, `arXiv:1903.09252`.

Chapter 5

The Fault in Our Fitter: Binned Fits with RooFit

In the measurement of R_K and R_{K^*}, the mass fits (Chap. 4.5) to the low statistics rare modes are unbinned while the control mode fits are binned for computing efficiency. This is because RooFit's time complexity scales linearly as $\mathcal{O}(N_{bins})$ and $\mathcal{O}(N_{events})$ [1] for the binned and unbinned fits respectively. Since the high statistics control modes are binned such that $N_{bins} \ll N_{events}$, using a binned fit is orders of magnitude faster than an unbinned fit.

This speedup would be free lunch if not for the fact that the binned fits consistently produce biased toy pulls. Interestingly, unbinned fits to the control modes produce unbiased pulls. In fact, RooFit has a known bias when fitting binned datasets. [2] Public discussions between analysts and ROOT developers point to RooFit's definition of binned likelihood as the source of this bias.

In order to understand this bias and correct it, the author undertook an independent study. The source code of this study is publicly available on gitlab [3] and this study only depends on the GNU Scientific Library (GSL) (1) and ROOT, both open sourced libraries. The pseudorexperiments in this study are generated by RooFit and neither LHCb data nor LHCb simulation are used in this study. This design ensures that the results of this study is reproducible by any HEP analyst, even those outside the LHCb collaboration [4].

---

[1] In simpler terms, how long does it take an algorithm to run as a function of N_{bins} or N_{events}.

[2] This is acknowledged by the ROOT developers in a JIRA task and a ROOT forum post.

[3] https://gitlab.cern.ch/dtou/binned-minimization.

[4] Provided they have the neccessary computing resources.

5.1 RooFit and Likelihood Maximisation

RooFit (2) is the experimental high energy physics (HEP) community's de facto framework for likelihood or χ^2 based fit. RooFit's success owes to the robustness and functionality of RooFit's design, and also to its ease-of-use. Despite that, RooFit's definition of binned likelihood intrinsically biases its binned fits.

5.1.1 A Historical Introduction to RooFit

RooFit's origin can be traced to the $\sin(2\beta)$ measurement by BaBar [5], which involves a complicated fit to the B decay time. This complexity was beyond the simple data modelling functionalities provided by ROOT's base fitter class, the `TF1`. In the face of this limitation, RooFit rose as an alternative to the `TF1` fitter.

RooFit's data modelling can handle complicated simultaneous unbinned fits that were previously difficult to implement with vanilla ROOT. Since RooFit has to tackle increasingly complicated models, the developers also implemented a function cache to optimise likelihood and χ^2 computations. Furthermore, RooFit normalises the fit models automatically, and together with its user friendly interface contributed to its ease of use. These features were an upgrade compared to `TF1`'s offerings.

That being said, RooFit was designed as an extension to ROOT's functionality. Once RooFit has defined a data model, it offloads the minimisation to ROOT's MINUIT routine. Also, many of RooFit's classes either inherit or depend on ROOT classes. RooFit's dataset classes use ROOT's `TTree` as a storage backend by default. Another example is RooFit's data visualisation functionalities, which heavily depend on ROOT's graph plotting features.

After RooFit was integrated to ROOT distribution, it emerged as the most popular fitting framework among the experimental HEP community. RooFit's ascension was underpinned by RooFit's ease-of-use, optimised function caches and ability to handle complex models. Additionally, the experimental HEP community has further extended the RooFit library with the RooStats subpackage (3), bringing additional statistical tools to RooFit. Even today, RooFit is undergoing continuous development with dedicated resources by CERN's ROOT team.

---

[5] This is one of the golden measurements of B-factories.

5.1.2 Likelihoods

Most HEP analysis, including the analysis of this thesis, involves a maximum likelihood fit to estimate the parameters of interest. For conciseness, we shall briefly discuss the mathematics behind RooFit's maximum likelihood fits. For a detailed overview of maximum likelihood estimation, with an experimental HEP flavour, the reader is referred to PDG's review on statistics (4).

The maximum likelihood fit to unbinned data estimates the parameters, $\boldsymbol{\theta}$, of a model by maximising the likelihood, $\mathcal{L}$, defined as

$$\mathcal{L} = \prod_{i}^{N} f(x_i; \boldsymbol{\theta}), \tag{5.1}$$

where $f(x_i; \boldsymbol{\theta})$ is the probability density function (PDF) of the model, and the likelihood is a product over N observations x_i. Essentially, $f(x_i; \boldsymbol{\theta})$ is the 'per-event probability to observe x_i, given the model f and parameters $\boldsymbol{\theta}$'.

Commonly, experimental HEP analyses are interested in the normalisation of the dataset, such as fits where the signal yield is a parameter of interest. In such cases, an extended term is added to the likelihood,

$$\mathcal{L} = \frac{\mu^N}{N!} e^{\mu} \prod_{i}^{N} f(x_i; \boldsymbol{\theta}), \tag{5.2}$$

where μ is the expected number of events predicted by the model. The extended term, $\frac{\mu^N}{N!} e^{\mu}$, is basically the Poisson probability to observe N events given the model expects μ events.

In practice, likelihood maximisation runs into two numerical issues. First, a maximisation problem involves finding the optimal solution on a concave curve, which is numerically unstable [6]. Second, the product of per-event likelihoods at larger N would produce increasingly smaller $\mathcal{L}$. Eventually, the $\mathcal{L}$ encounters the floating point underflow, where the value of $\mathcal{L}$ is smaller than the absolute lower limit of floating point precision.

Therefore, maximum likelihood fits avoid these numerical issues by optimising the negative log likelihood instead, written as

$$-\log \mathcal{L} = -\mu - N \log \mu + \log N! - \sum_{i}^{N} \log f(x_i; \boldsymbol{\theta}). \tag{5.3}$$

---

[6] An analogy to physics can be found in energy potentials. An equilibrium at the maxima is an unstable equilibrium but the equilibrium at the minima is a stable equilibrium.

Because the negative logarithm is a one-to-one transformation, minimising $-\log\mathcal{L}$ is equivalent to maximising $\mathcal{L}$. On the numerical issues, the logarithm solves the underflow problem by summing individual per-event log likelihood. On the other hand, the sign change inverts the likelihood maxima into a minima, and the numerical optimisation now search for the solution on a convex curve instead.

Notice in Eq. 5.3 that the $\log N!$ term stays constant throughout the minimisation procedure. This means $\log N!$ only affects the depth of the minima, but not the optimum values of $\boldsymbol{\theta}$ [7]. RooFit therefore drops this term, and define the unbinned negative log likelihood as

$$-\log\mathcal{L} = -\mu - N\log\mu - \sum_i^N \log f(x_i;\boldsymbol{\theta}) \tag{5.4}$$

which is minimised by MINUIT.

The binned likelihood is instead

$$\mathcal{L} = \prod_i^{N_{bins}} \frac{\mu_i^{n_i}(\boldsymbol{\theta})}{n_i!} e^{\mu_i(\boldsymbol{\theta})} \tag{5.5}$$

where n_i is the number of events observed in bin i, and $\mu_i(\boldsymbol{\theta})$ the expected number of events in bin i, predicted by the model given parameters $\boldsymbol{\theta}$. In other words, the per-bin likelihood is the probability, assuming Poisson statistics, to observe n_i events given the model expects μ_i events.

Similarly, binned fits optimise the negative log likelihood instead,

$$-\log\mathcal{L} = \sum_i^{N_{bins}} -n_i\log\mu_i(\boldsymbol{\theta}) + \log n_i! - \mu_i(\boldsymbol{\theta}), \tag{5.6}$$

$$= -\mu(\boldsymbol{\theta}) - \sum_i^{N_{bins}} n_i\log\mu_i(\boldsymbol{\theta}) + \cancel{\sum_i^{N_{bins}}\log} n_i!, \tag{5.7}$$

where $\mu(\boldsymbol{\theta})$ is the expected total number of events. Likewise, RooFit drops the $\log n_i!$ terms because they remain constant throughout the minimisation.

---

[7] Even in analysis where a few different models are tested, $\log N!$ only depends on data and it does not affect the $\Delta\mathcal{L}$ comparison between different models.

5.1.3 Binned vs Unbinned Time Complexity

In RooFit, minimising these negative log likelihoods is the responsibility of MINUIT. MINUIT (5) itself is composed of various subroutines, but for the purpose of our discussions we are only interested in MIGRAD and HESSE. MIGRAD is a minima finding algorithm while HESSE is a covariance matrix calculator. These two are commonly used together, MIGRAD after HESSE, by experimental HEP fits to find the best fit to $\boldsymbol{\theta}$ and to estimate the uncertainties and correlations of $\boldsymbol{\theta}$.

At each minimisation step, MIGRAD estimates the first derivatives of the parameters, $\partial_{\boldsymbol{\theta}}$ [8], by finite differences. MIGRAD then uses the Davidon-Fletcher-Powell (6) algorithm to approximate the covariance matrix, $\boldsymbol{G}$, and update the parameters as

$$\Delta\boldsymbol{\theta} = \boldsymbol{G}^{-1}\partial_{\boldsymbol{\theta}}. \tag{5.8}$$

Once the minima is found, HESSE uses the second order finite differences to calculate the covariance matrix.

Both MIGRAD and HESSE make heavy use of finite difference calculations, and their time complexities are bounded by the likelihood calculations. Now, recall that the unbinned (Eq. 5.3) and binned (Eq. 5.7) likelihoods sum over N_{events} and N_{bins} respectively. This means that the unbinned and binned likelihoods time complexities scale as $\mathcal{O}(N_{events})$ and $\mathcal{O}(N_{bins})$ respectively. Assuming that $\log f(x_i;\boldsymbol{\theta})$ and $\log \mu_i(\boldsymbol{\theta})$ calculation times take roughly the same order of magnitude, the binned fit speed up with respect to unbinned fit scale as $\mathcal{O}(N_{events}/N_{bins})$.

Since binned fits are invoked with large datasets, it is usually the case that $N_{bins} \ll N_{events}$. Due to the $\mathcal{O}(N_{events}/N_{bins})$ scaling, binned fits will be significantly faster than unbinned fits. This is the case for the control mode fits in the analysis of R_K and R_{K^*}.

5.1.4 RooFit's Binned Likelihood

RooFit's definition of binned likelihood substitutes μ_i of Eq. 5.7 with the PDF value at the bin center, x_i^{center}:

$$-\log\mathcal{L} = -\mu(\boldsymbol{\theta}) - \sum_{i}^{N_{bins}} n_i \log f(x_i^{center};\boldsymbol{\theta}). \tag{5.9}$$

---

[8] First derivative w.r.t. negative log likelihood, $\partial_{\boldsymbol{\theta}} = \frac{\partial - \log\mathcal{L}}{\partial\boldsymbol{\theta}}$.

This is mathematically equivalent to calculating the per-bin expected population (μ_i) with the midpoint rule integral [9]. The midpoint rule approximation of μ_i works well when the model is linear within the bin. However, this approximation breaks down when the model has a non-linear behaviour within a bin. Unfortunately, most experimental HEP analyses fit highly non-linear models, such as the Gaussian-like distributions that are commonly used to model signal peaks.

The midpoint rule approximation is hypothesised as the reason binned fits produce biased pulls in the measurement of R_K and R_{K^*}. This hypothesis is tested with an independent study decoupled from the R_K and R_{K^*} analysis framework.

5.2 The Independent Study

The purpose of this independent study is to verify the midpoint rule approximation as the culprit of binned fit biases. Decoupling this study from the R_K and R_{K^*} framework would remove the framework overhead and isolate this study from possible bugs in the framework. Furthermore, this study will fit large numbers of pseudoexperiments, which motivates the use of a relatively simple model to decrease turnover time.

5.2.1 Strategy

The baseline model consists of a signal and a combinatorial background. The signal is a DSCB function, with a Gaussian core of mean $\mu_{gauss} = 5428.65$ MeV and width $\sigma_{gauss} = 20$ MeV, values which are similar to those of a B invariant mass fit. The combinatorial background is an exponential function with a slope of $b_{slope} = -2 \times 10^{-3}$ MeV^{-1}. The baseline model will have 1 million signal events and 150 thousand background events within a mass window of $5000 - 6000$ MeV. A plot of the typical toy generated by this model is shown in Fig. 5.1 [10]. In the following texts, a ^{truth} superscript denotes the pseudoexperiment generator truth values while anything without this tag denotes the values in the toy fit.

---

[9] The reader might have noticed that we have dropped the bin width from the likelihood. This is because the bin widths, w_i, can be taken out of the logarithm into $n_i \log w_i$ terms that are constant during the minimisation.

[10] An attentive reader might have noticed we did not discuss the DSCB tail parameters. This omission is deliberate because the tail parameters are fixed during the toy fit and their values have little significance in our discussions.

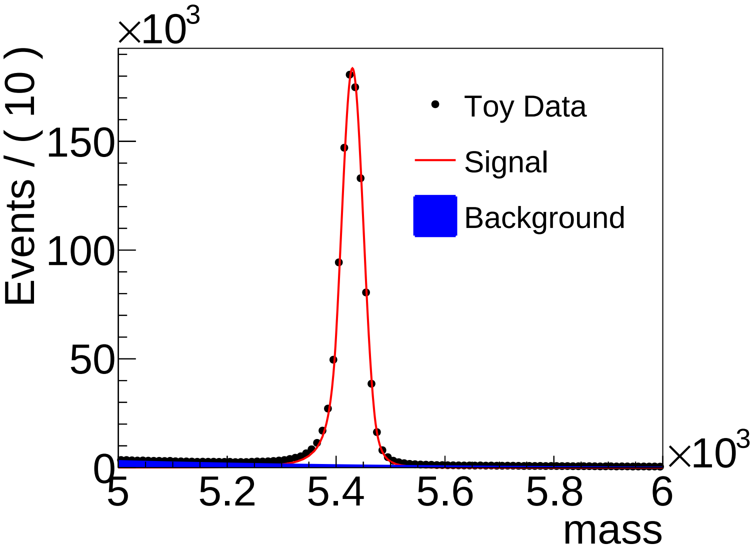


Figure 5.1 – An example pseudoexperiment generated by the baseline mode. The toy dataset (black) is generated by a DSCB signal (red) and combinatorial background (blue) model.

If the midpoint rule approximation is the cause of binned fit bias, this approximation is expected to degrade if the PDF varies more within a bin. By this assumption, the bias on σ_{gauss} should increase as σ_{gauss}^{truth} decreases. This is because a decreasing σ_{gauss} would produce narrower signal peak, and increase the per-bin PDF variance [11] around the signal peak. Another experiment is to measure how the bias varies with the number of bins in the fit, N_{bins}. As the number of bin decreases, the bin width will increase, causing the midpoint rule approximation to deviate further from the true value of μ_i. We shall study the central value bias in σ_{gauss} as a proxy for the fit bias in these experiments.

As a control, the same experiment with σ_{gauss}^{truth} will be run on RooFit's unbinned fit, since the unbinned fits produce unbiased results mass fits of R_K and R_{K^*} measurement. Furthermore, the simple one dimensional model allows cross checks to be done with the binned `TF1` fitter. `TF1` also has an 'integral' option which the ROOT developers suggest will produce unbiased binned fit results. Therefore, the `TF1` fits with and without the 'integral' option will act as a cross check of RooFit results.

---

[11] Per-bin variance refers to how much the PDF varies within a bin.

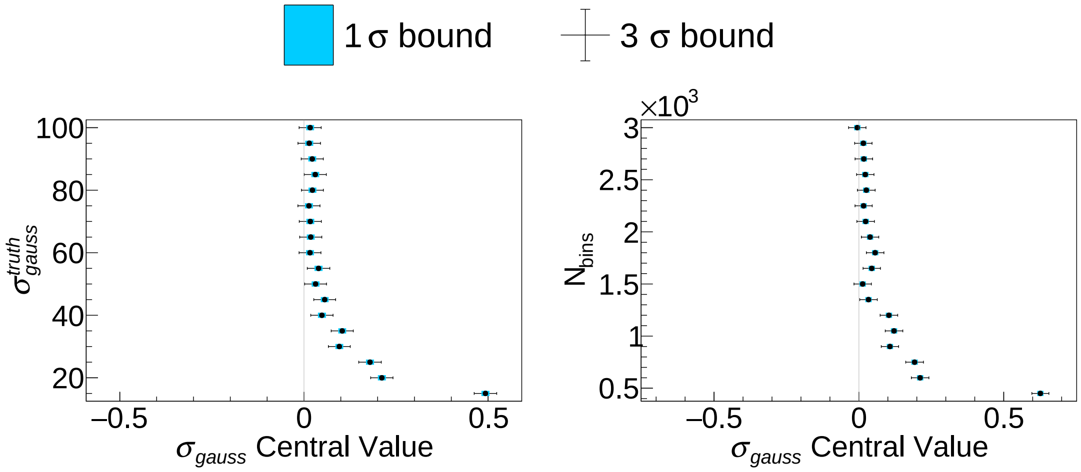


Figure 5.2 – The evolution of central value bias in σ_{gauss} along σ_{gauss}^{truth} (left) and N_{bins} (right).

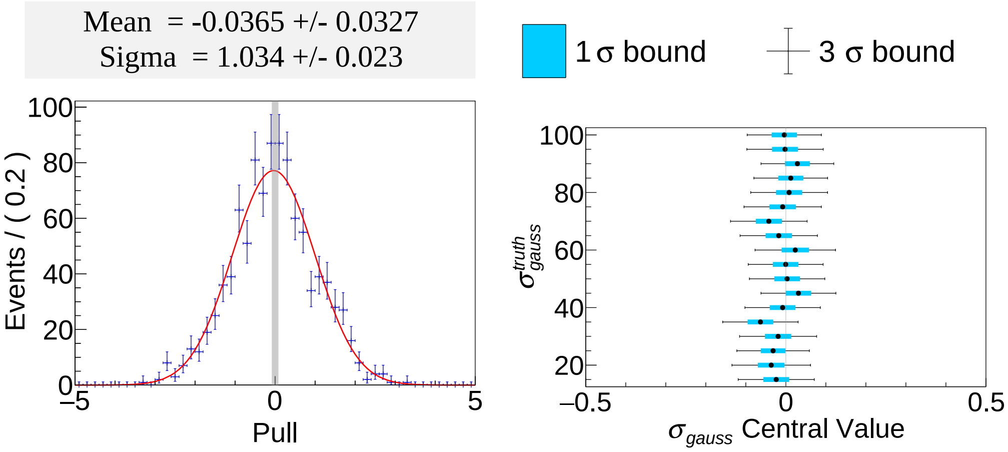


Figure 5.3 – On the left, the pull plot of σ_{gauss} from unbinned fits to 1000 pseudo-experiments generated by the baseline model. On the right, the central values of σ_{gauss} pulls in the unbinned fit as a function of σ_{gauss}^{truth}.

5.2.2 Results

The central value biases of σ_{gauss} as a function of σ_{gauss}^{truth} and N_{bins} are tested on 10000 pseudoexperiments. The results are shown in Fig. 5.2. As we can see, the pull bias in σ_{gauss} increases as σ_{gauss}^{truth} decreases and as N_{bins} decreases. These results support our earlier predictions of how the bias will vary in binned fits. As for the control, the unbinned fits are tested on 1000 pseudoexperiment only for computational reasons. As expected, the unbinned fits produce an unbiased results regardless of σ_{gauss}^{truth}, shown in Fig. 5.3.

At this point, the reader might wonder 'Why isn't this problem more widely known?' [12]. Possible explanations are either analysts overkill this problem with unbinned fits [13] or that typical analyses do not have the statistics required to observe this problem. The latter can be argued by the fact that binned fit biases are absolute biases, and the observed pull biases are amplified by the decreasing relative statistical uncertainty in larger samples. This argument is observed in the distribution of σ_{gauss} bias as a function of $N^{truth}(signal)$ in Fig. 5.4. More interestingly, the bias is well modelled as $A\sqrt{N(signal)}$. This proofs that the absolute bias remains unchanged with respect to $N(signal)^{truth}$, but it is the $\frac{1}{\sqrt{N(signal)}}$ uncertainty scaling that amplifies the pull bias in σ_{gauss}. Therefore, the binned fit bias will become apparent as analysts fit increasingly larger datasets.

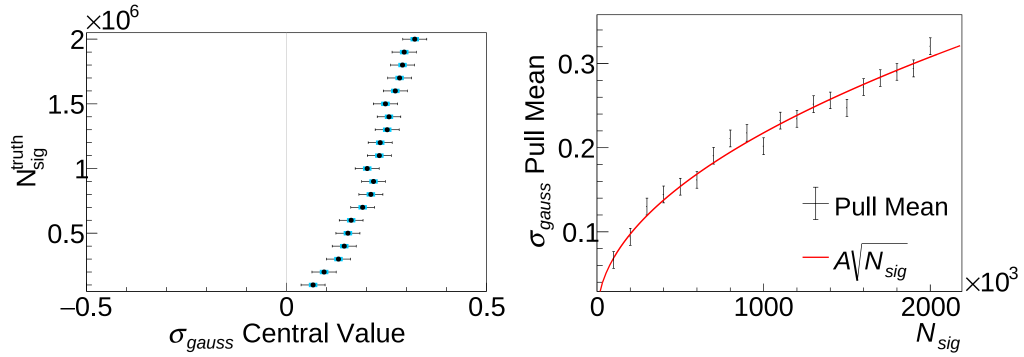


Figure 5.4 – On the left, the pull bias in σ_{gauss} as a function of $N^{truth}(signal)$. On the right, a fit to the distribution σ_{gauss} bias against $N^{truth}(signal)$ (black), modelled as $A\sqrt{N(signal)}$ (red).

---

[12] Some initially refused to believe RooFit's binned fit is inherently biased. This included the author's supervisor. The author had to argue with his supervisor for a week or two and show him the results here to convince his supervisor otherwise.

[13] And waste lots of CPU cycles.

5.2.3 `TF1` Cross check

Now that we have established that binned fits with RooFit is biased, the same setup is cross checked with `TF1`. Fig. 5.5 shows the bias in σ_{gauss} as a function of σ_{gauss}^{truth} in `TF1`. The default `TF1` binned fit replicates the bias in RooFit's binned fit. However, `TF1` fits with 'integral' enabled produce unbiased results. This is because the 'integral' method numerically integrates the PDF over a single bin to estimate the value of μ_i. Compared to the midpoint rule approximation, the per-bin integral is able to calculate an accurate value of per-bin expected population, μ_i. This indicates that a per-bin numerical integration will unbias RooFit's binned fit.

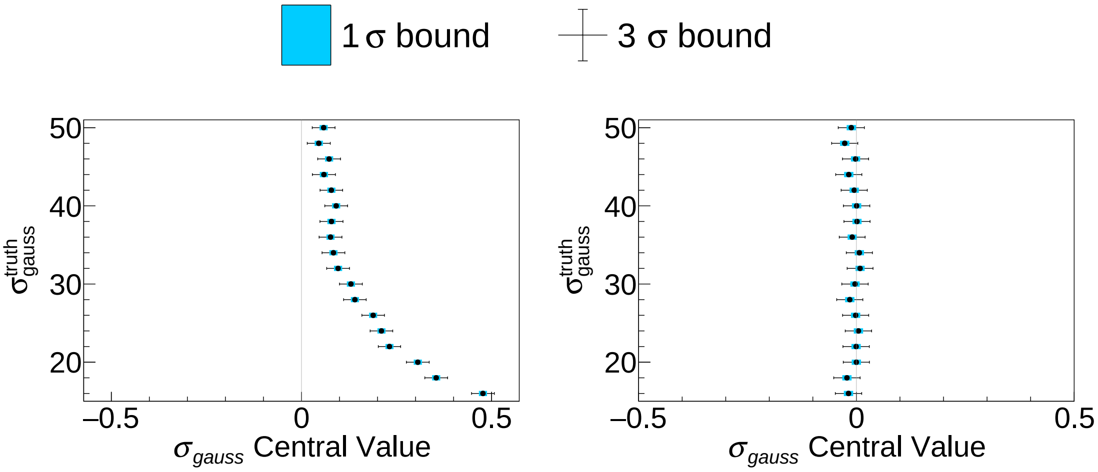


Figure 5.5 – The variation of σ_{gauss} pull mean as a function of σ_{gauss}^{truth} in `TF1` binned fits without (left) and with (right) the 'integral' option.

5.3 Numerical Integration in RooFit's Binned Likelihood

Inspired by the implementation of `TF1` per-bin integral, the solution to RooFit's binned likelihood will rely on GSL's numerical integration library. The choice of numerical integration algorithm will be a topic of discussion in this section. This solution replaces the midpoint rule with a numerical integration instead to aprroximate the value of $\mu_i(\theta)$ (Eq. 5.3). The implementation of numerical integration in RooFit inherits the `RooNLLVar` likelihood class and overloads the function that computes negative log likelihood. This ensures the numerical integration extension is seemlessly incorporated into RooFit.

5.3.1 Choice of Integral Algorithms

The GSL numerical integration library offers a huge selection of algorithms (7) to choose from. Nevertheless, all these algorithms first evaluate the function at multiple points along the integral range and then use a numerical integration rule to calculate the integral. These algorithms can be broadly classified into fixed points integration or adaptive integration. As their name suggests, fixed points integrations require the user to specify how many, and sometimes which, points to sample the function along the integration range. On the other hand, the adaptive integrals iteratively samples the functions at more points until it converges, which happens when the error estimate is below a user defined tolerance.

The choice between fixed points and adaptive integral boils down to two considerations. First, the adaptive integral is more robust with a user defined error tolerance. This feature allows an analyst to define her tolerance such that the integration error estimates are negligible compared to the uncertainties in parameter of interests. Also, we have discussed earlier that the binned fit bias is absolute. Hence, analyses with larger datasets can set lower tolerances, such that integration errors are negligible with respect to their statistical uncertainties.

Adaptive integrals have another advantage in their per-bin iterations. Most invariant mass windows contain a sharp signal peak which is highly non-linear but the accompanying backgrounds, especially the combinatorial, behave linearly towards the tails. Thus, the adaptive methods will spend more time around the high per-bin variance signal peaks while shaving time off the linear tails, without any cost to integration accuracy. In contrast, the fixed points methods use the same number of sampling in each bin. As a result, fixed points methods either waste CPU resources on the linear bins or sample insufficient points to accurately integrate high variance bins.

Among GSL's adaptive integral methods are algorithms capable to handle integrable singularity. Given HEP fits rarely, if ever involve singular functions, these algorithms are not considered for implementation with RooFit. This leaves two candidate algorithms, the Gauss-Kronrod (8) and Romberg (9) adaptive integrations. Both of them will be tested in the following sections.

5.3.2 Results

The tests of binned fit bias in Sec. 5.2.2, as a function of σ_{gauss}^{truth}, N_{bins}, and $N(signal)^{truth}$, is repeated for the Gauss-Kronrod and Romberg algorithms. Fig. 5.6 shows the result of both algorithms, and they both produce unbiased results.

The ultimate choice between Gauss-Kronrod or Romberg algorithms boiled down to their timing. In a benchmark on 10 pseudorexperiments, the Gauss-Kronrod based solution took 0.332 s while the Romberg based solution took 0.236 s. This result is rather expected because the Gauss Kronrod sampling scale as $15 \times n$ at the minimum [14], where n is the number of iterations. In constrast, the Romberg algorithm samples $2^n + 1$ points. Therefore, a Gauss-Kronrod based fit takes longer because it samples at least 15 points per bin, which takes the Romberg algorithm at least $n = 4$ iterations to sample.

---

[14] The Gauss Kronrod algorithms samples at least 15 points in each iteration, but the user can request 21, 31, 41, 51 or 61 points as well.

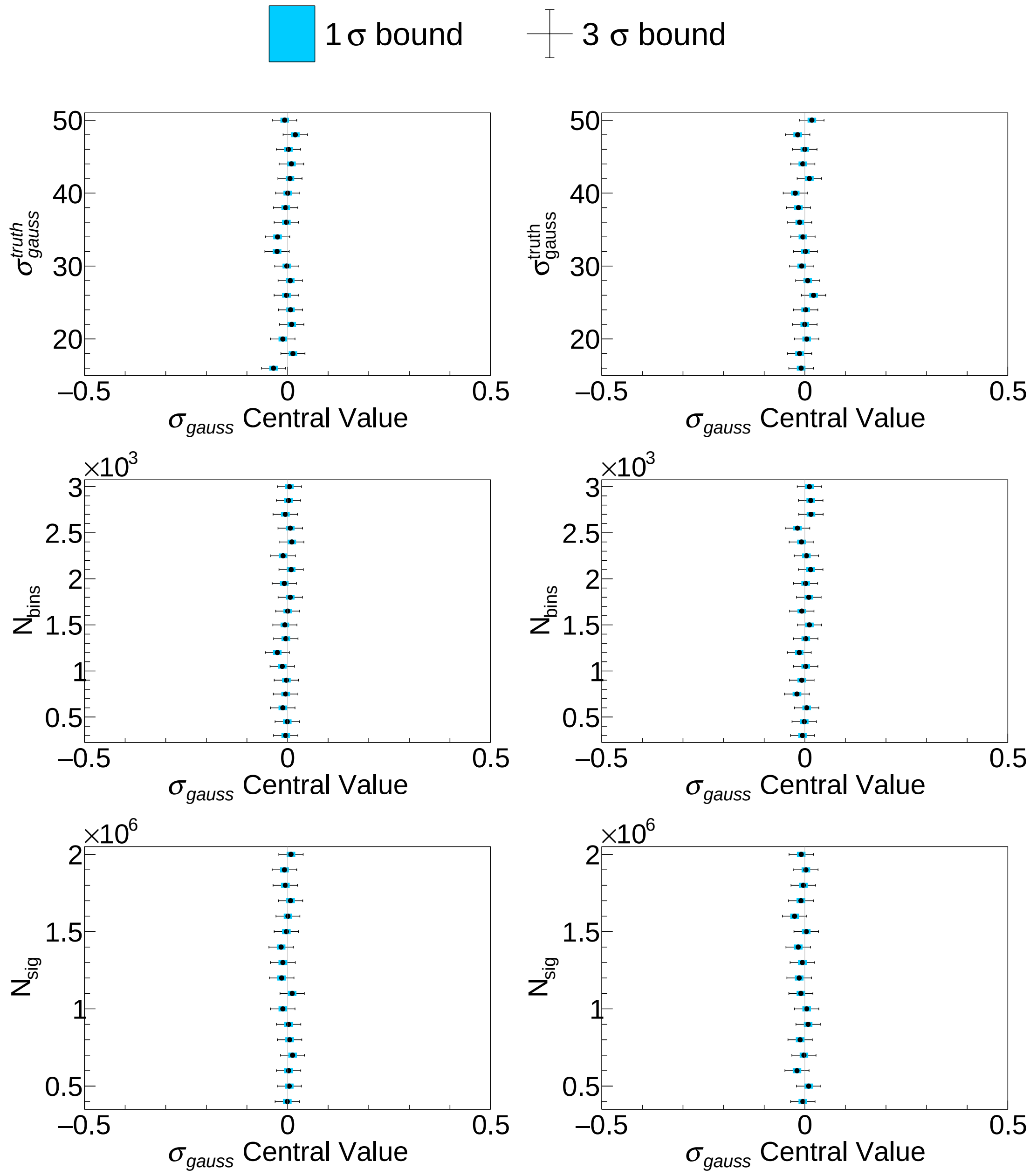


Figure 5.6 – The pull means of σ_{gauss} as a function of σ^{truth}_{gauss} (top), N_{bins} (middle), and $N(signal)^{truth}$ (bottom). The Gauss-Kronrod (left) and Romberg (right) algorithms are used as the numerical integrator.

5.4 Conclusion

The binned fits are significantly faster than unbinned fits for large datasets because the ratio of their time complexities scales as $\mathcal{O}(N_{bins}/N_{events})$. Despite this, RooFit's native binned likelihood fit is biased because it uses the midpoint rule approximation to estimate per-bin expected population.

The independent study has shown that the binned fit bias increases as the per-bin variance increases. This support that argument binned fit bias are caused by the midpoint rule integration. Furthermore, the binned fit bias is an absolute bias and larger datasets, with their smaller relative statistical errors, will enlarge the pull bias.

As a solution to this bias, RooFit's binned likelihood is redefined to use numerical integrations to calculate the per-bin expected population. Of the possible numerical integration algorithms offered by GSL, the Romberg algorithm was chosen because it is the fastest adaptive integral. This implementation has been shown to unbias the binned fit by RooFit.

5.5 Bibliography

[1] B. Gough, *GNU Scientific Library Reference Manual - Third Edition*, Network Theory Ltd., 3rd ed., 2009.

[2] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, `arXiv:physics/0306116`.

[3] L. Moneta *et al.*, *The RooStats Project*, PoS **ACAT2010** (2010) 057, `arXiv:1009.1003`.

[4] Particle Data Group, P. A. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020), no. 8 083C01.

[5] F. James, *MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1*, .

[6] R. Fletcher, *A new approach to variable metric algorithms*, The Computer Journal **13** (1970) 317, `arXiv:https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf`.

[7] R. Piessens *et al.*, *Quadpack: A Subroutine Package for Automatic Integration*, Computational Mathematics Series, Springer-Verlag, 1983.

[8] A. S. Kronrod, *Nodes and Weights of Quadrature Formulas: Sixteen-place Tables*, Consultants Bureau, 1966.

[9] W. Romberg, *Vereinfachte numerische integration*, Det Kongelige Norske Videnskabers Selskab Forhandlinger, Trondheim **28** (1955), no. 7 30–36.

Appendices

Appendix A

`HLT1` Lines Efficiencies

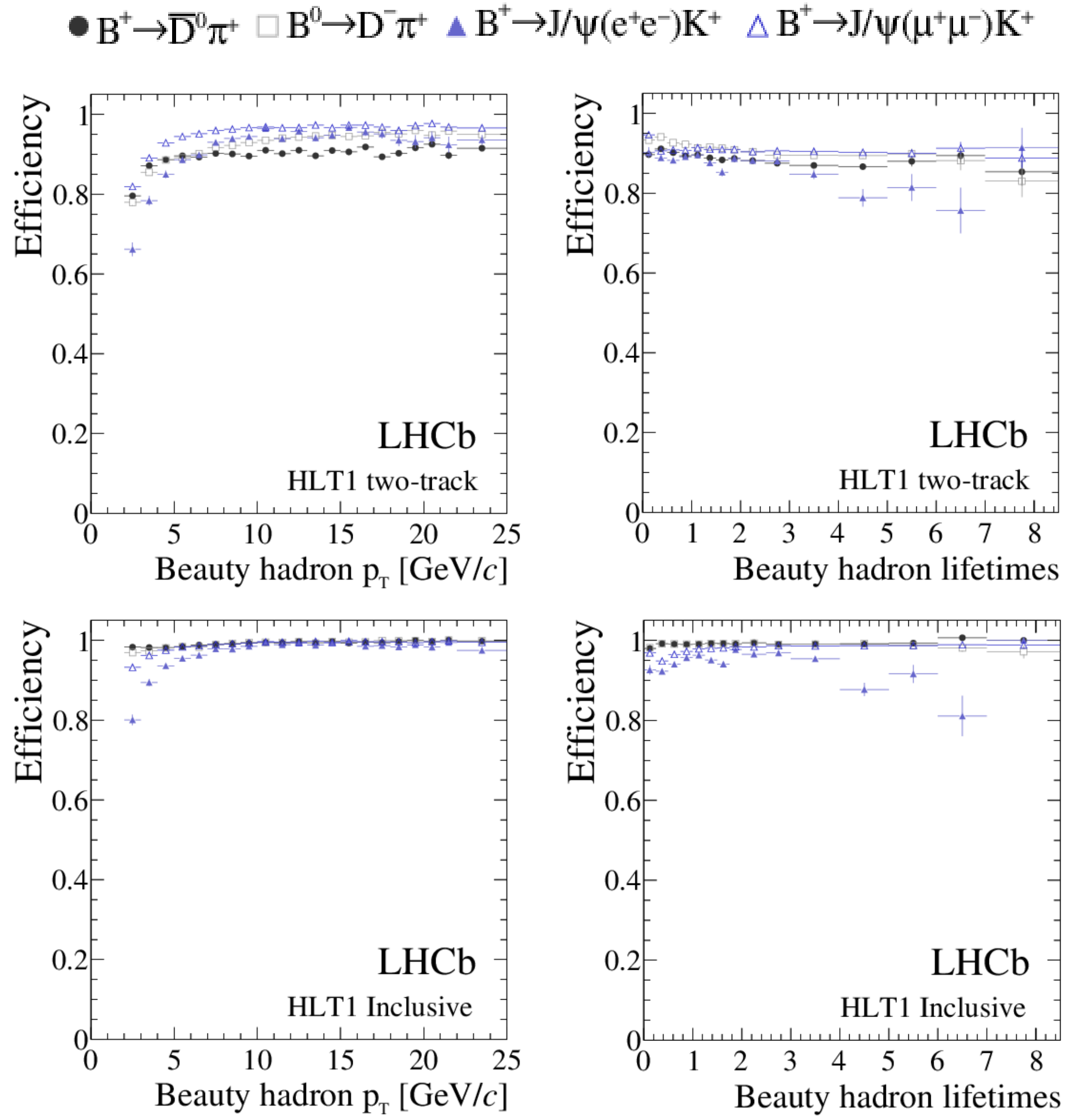


Figure A.1 – The Run 2 efficiencies of `HLT1TwoTrackMVA` (top), and the inclusive `OR` of `HLT1TrackMVA` and `HLT1TwoTrackMVA` (bottom) as a function of B p_{T} (left) and units of B lifetime (right). Figure extracted from Ref. (**?**).

Appendix B

B Momentum Fraction in Rare and Control modes

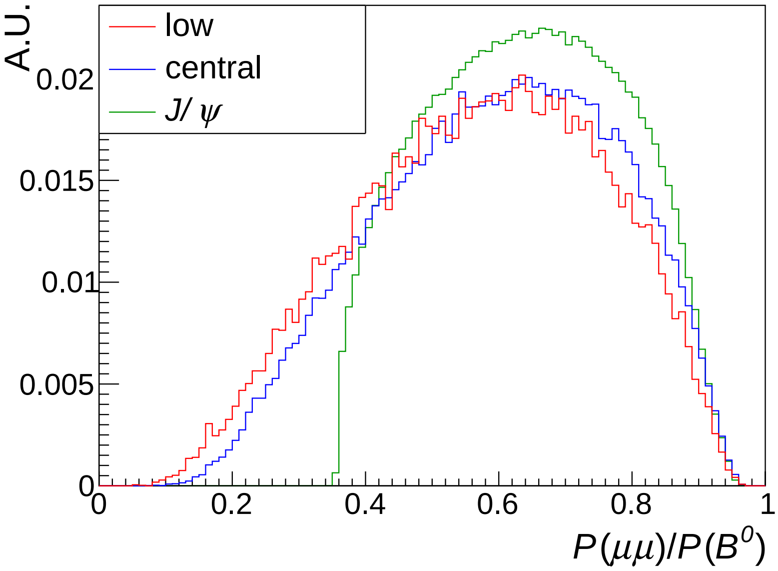


Figure B.1 – The simulated distribution of B^0 momentum fraction carried away by $\mu^+\mu^-$ pair, in the low (red), central (blue) and J/ψ (black) bins.

Appendix C

Stripping Cuts

The definitions for the variables below are

- `DIRA` (DIRection Angle) is the cosine of the angle between a particle's reconstructed momentum vector (reconstructed from tracking) and its direction-of-flight vector (reconstructed from vertexing),
- **Vertex fit $\boldsymbol{\chi^2}$/ndof**, encodes the quality of the vertex reconstruction fit,
- $\boldsymbol{\chi^2_{IP}}$**(primary)** is the difference in primary vertex reconstruction χ^2 with and without the track of interest, and high $\chi^2_{IP}(primary)$ are very likely tracks displaced from the primary vertex,
- **Primary-end vertex $\boldsymbol{\chi^2}$ separation** is the distance χ^2 between the B candidate decay vertex and its associated PV,
- **DLL**$_{\alpha\beta}$ is the change in RICH reconstruction log likelihood if mass hypothesis α is assigned to the candidate instead of mass hypothesis β,
- `isMuon` boolean is true if the particle penetrated the muon stations,
- `hasMuon` is true if the particle has Muon PID information.

Type	**Requirement**
Multiplicity	`nSPDHits` $< 600(450)$ Run 1 (Run 2)
B	$\lvert m - m_B^{\mathrm{PDG}} \rvert < 1500\,\mathrm{MeV}/c^2$
	`DIRA` > 0.9995
	$\chi^2_{\mathrm{IP}}(primary) < 25$
	end vertex fit $\chi^2/\mathrm{ndf} < 9$
	primary-end vertex χ^2 separation > 100
K^{*0} (R_{K^*})	$\lvert m - m_{K^{*0}}^{PDG} \rvert < 300\,\mathrm{MeV}/c^2$
	$p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$
	origin vertex fit $\chi^2/\mathrm{ndf} < 25$
K	`DLL`$_{K\pi} > -5$ (only data)
	$\chi^2_{\mathrm{IP}}(primary) > 9(4)$ R_{K^*} (R_K)
	$p_{\mathrm{T}} > 400\,\mathrm{MeV}/c$ (only for R_K)
π (R_{K^*})	$\chi^2_{\mathrm{IP}}(primary) > 9$
$\ell\ell$	$m < 5500\,\mathrm{MeV}/c^2$
	end vertex fit $\chi^2/\mathrm{ndf} < 9$
	primary-end vertex χ^2 separation > 16
μ	`isMuon` (only data), `hasMuon`
	$p_{\mathrm{T}} > 300\,\mathrm{MeV}/c$
	$\chi^2_{\mathrm{IP}}(primary) > 9$
e	`DLL`$_{e\pi} > 0$ (only data)
	$p_{\mathrm{T}} > 300\,\mathrm{MeV}/c$
	$\chi^2_{\mathrm{IP}}(primary) > 9$

Table C.1 – Summary of `Bu2LLKmm` and `Bu2LLKee` stripping cuts.

Appendix D

Corrections

D.1 `HLT` TIS Lines

`HLT` level	Run 1	15
HLT1	HLT1TrackAllL0_TIS HLT1Track{Muon,DiMuon}_TIS	HLT1Track{MVA,Muon}_TIS HLT1TwoTrack_TIS
HLT2	HLT2Topo{2,3,4}BodyBBDT_TIS HLT2TopoMu{2,3,4}BodyBBDT_TIS HLT2TopoE{2,3,4}BodyBBDT_TIS HLT2DiMuonDetached_TIS	HLT2Topo{2,3,4}Body_TIS HLT2TopoMu2,3,4Body_TIS Hlt2DiMuonDetachedHeavy_TIS

Table D.1 – The individual trigger lines used to compose the `HLT` inclusive TIS categories in Run 1 and 2015 simulation.

`HLT` level	16	17&18
HLT1	HLT1Track{MVA,Muon,MuonMVA}_TIS HLT1TwoTrack_TIS	HLT1Track{MVA,Muon,MuonMVA}_TIS HLT1TwoTrack_TIS
HLT2	HLT2Topo{2,3,4}Body_TIS HLT2TopoMu{2,3,4}Body_TIS HLT2TopoMuMu{2,3,4}Body_TIS Hlt2DiMuonDetachedHeavy_TIS HLT2TopoE{2,3,4}Body_TIS HLT2TopoEE{2,3,4}Body_TIS	HLT2Topo2,3,4,Body_TIS HLT2TopoMu2,3,4Body_TIS HLT2TopoMuMu2,3,4Body_TIS Hlt2DiMuonDetachedHeavy_TIS HLT2TopoE2,3,4Body_TIS HLT2TopoEE2,3,4Body_TIS

Table D.2 – The individual trigger lines used to compose the `HLT` inclusive TIS categories in 2016, 2017 and 2018 simulation.

D.2 Invariant mass fits to $m(ee)$

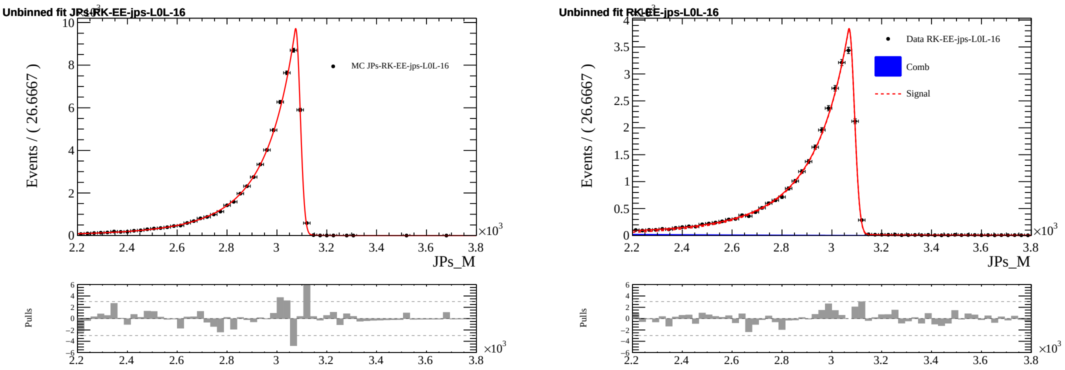


Figure D.1 – Fits to m_{ee} in simulated $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (left) and real data (right) of 0γ bremsstrahlung category.

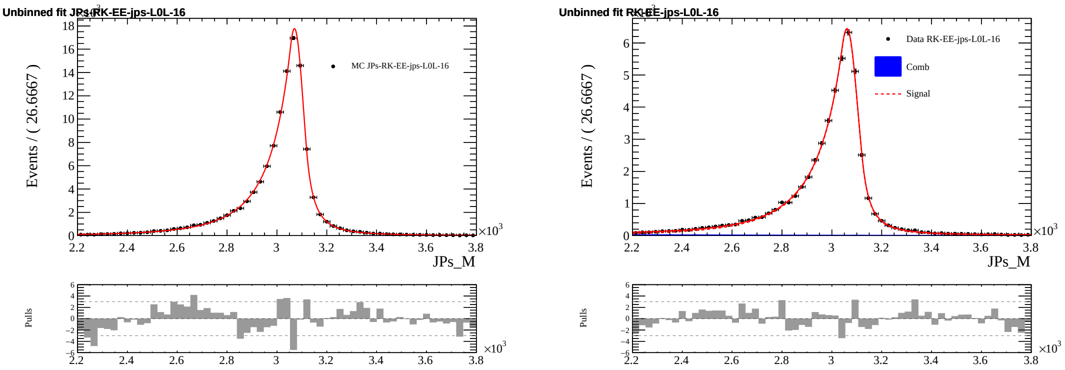


Figure D.2 – Fits to m_{ee} in simulated $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (left) and real data (right) of 1γ bremsstrahlung category.

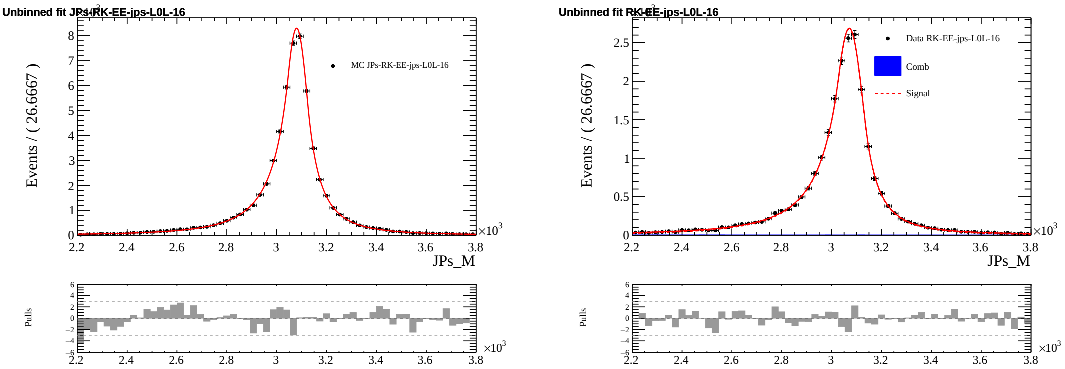


Figure D.3 – Fits to m_{ee} in simulated $B^+ \to K^+ J/\psi (\to e^+e^-)$ (left) and real data (right) of 2γ bremsstrahlung category.

D.3 Measurements of $\Delta\mu$ and s_σ

Table D.3 – Values of the sigma scale ,s_σ, mass shift, $\Delta\mu$ averaged over `L0I` and `L0L` exclusive trigger categories.

Parameter	Year	`nBrems`	B^+ mode	B^0 mode
s_σ	11	0γ	1.15 ± 0.03	1.14 ± 0.04
		1γ	1.16 ± 0.02	1.21 ± 0.04
		2γ	1.15 ± 0.03	1.14 ± 0.05
	12	0γ	1.13 ± 0.02	1.11 ± 0.03
		1γ	1.15 ± 0.01	1.16 ± 0.02
		2γ	1.14 ± 0.02	1.15 ± 0.04
	15	0γ	1.26 ± 0.03	1.20 ± 0.05
		1γ	1.18 ± 0.03	1.18 ± 0.04
		2γ	1.17 ± 0.03	1.14 ± 0.07
	16	0γ	1.20 ± 0.01	1.17 ± 0.02
		1γ	1.19 ± 0.01	1.21 ± 0.02
		2γ	1.18 ± 0.02	1.19 ± 0.03
	17	0γ	1.16 ± 0.01	1.15 ± 0.02
		1γ	1.16 ± 0.01	1.13 ± 0.02
		2γ	1.13 ± 0.01	1.12 ± 0.03
	18	0γ	1.16 ± 0.01	1.13 ± 0.02
		1γ	1.16 ± 0.01	1.17 ± 0.02
		2γ	1.12 ± 0.01	1.14 ± 0.02
$\Delta\mu$	11	0γ	-4.90 ± 0.89	-2.75 ± 1.44
		1γ	-1.80 ± 0.92	-1.07 ± 1.54
		2γ	2.54 ± 1.45	2.65 ± 2.36
	12	0γ	-2.18 ± 0.59	-2.04 ± 0.93
		1γ	0.06 ± 0.64	-0.39 ± 1.08
		2γ	5.78 ± 1.05	7.24 ± 1.81
	15	0γ	-7.99 ± 1.15	-6.60 ± 1.95
		1γ	-9.51 ± 1.24	-9.39 ± 2.01
		2γ	-10.92 ± 1.60	-13.62 ± 3.33
	16	0γ	-6.45 ± 0.45	-5.97 ± 0.75
		1γ	-11.54 ± 0.48	-12.68 ± 0.84
		2γ	-13.70 ± 0.79	-12.14 ± 1.36
	17	0γ	-6.37 ± 0.44	-5.36 ± 0.74
		1γ	-8.66 ± 0.46	-7.13 ± 0.78
		2γ	-9.24 ± 0.74	-7.64 ± 1.27
	18	0γ	-5.83 ± 0.41	-4.97 ± 0.65
		1γ	-7.22 ± 0.42	-5.90 ± 0.73
		2γ	-6.16 ± 0.69	-5.94 ± 1.16

Appendix E

Mass Fits

E.1 Shapes of Simulated Signal

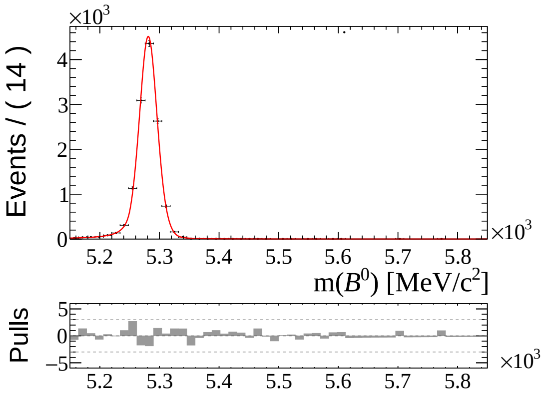


Figure E.1 – Fits to simulated $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ (left) and low-q^2 $B^0 \to K^{*0}\mu^+\mu^-$ (right) decays. Note that the $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ shape is shown in logarithmic scale. The fits shown above use $15+16$ simulation in the exclusive `L0L` category.

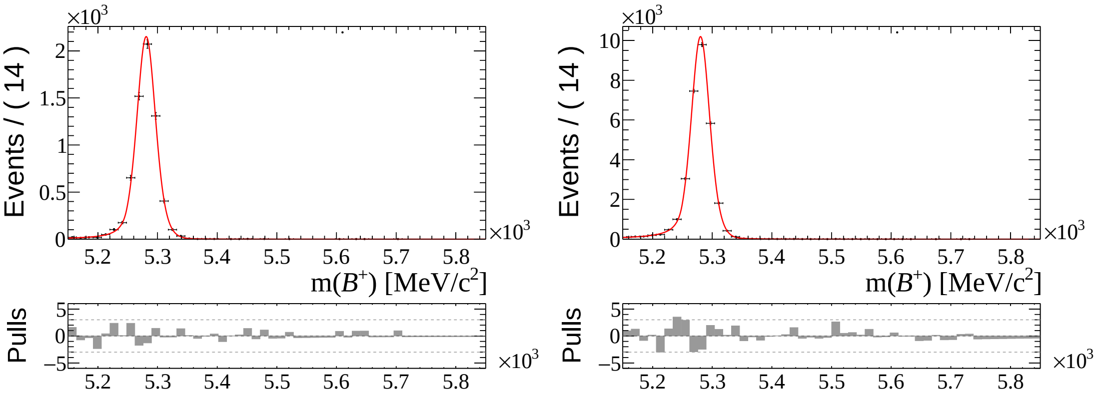
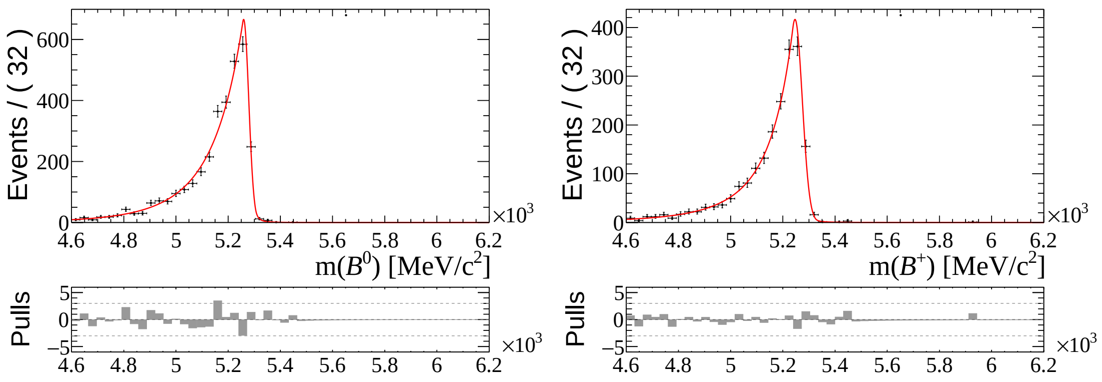
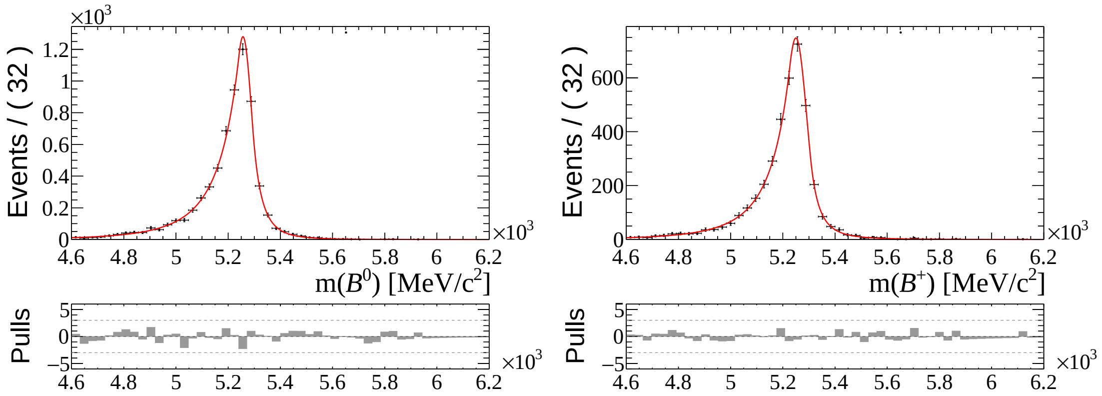


Figure E.2 – Fits to simulated $B^+ \to K^+\mu^+\mu^-$ signal are in the low-q^2 bin (left) and central-q^2 bin (right). The fits shown above uses $15 + 16$ simulation passing the muon `L0L` triggers.



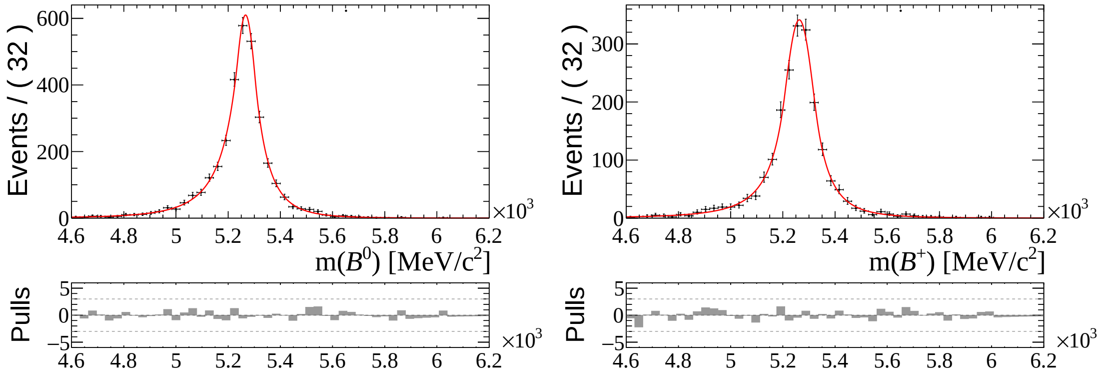
(a) **0γ** category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0}e^+e^-}$.

(b) **0γ** category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+e^+e^-}$.



(c) **1γ** category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0}e^+e^-}$.

(d) **1γ** category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+e^+e^-}$.



(e) **2γ** category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0}e^+e^-}$.

(f) **2γ** category, central-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+e^+e^-}$.

Figure E.3 – Simulated electron signal modes in central-q^2 bin. $B^0\to K^{*0}e^+e^-$ ($B^0\to K^{*0}e^+e^-$) are shown on the left(right), separated into **0γ** (top), **0γ** (middle) and **2γ** (bottom) bremsstrahlung categories. The fits shown above use **15 + 16** simulation in the exclusive `L0L` category.

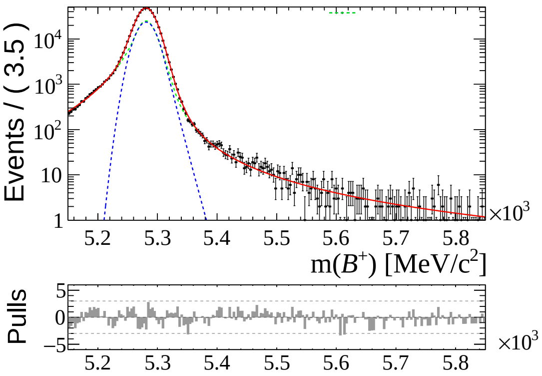


Figure E.4 – Fits to simulated $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ decays without a J/ψ mass constraint on $m(\mu^+\mu^-)$. The fits shown above use $15+16$ simulation in the exclusive `L0L` category.

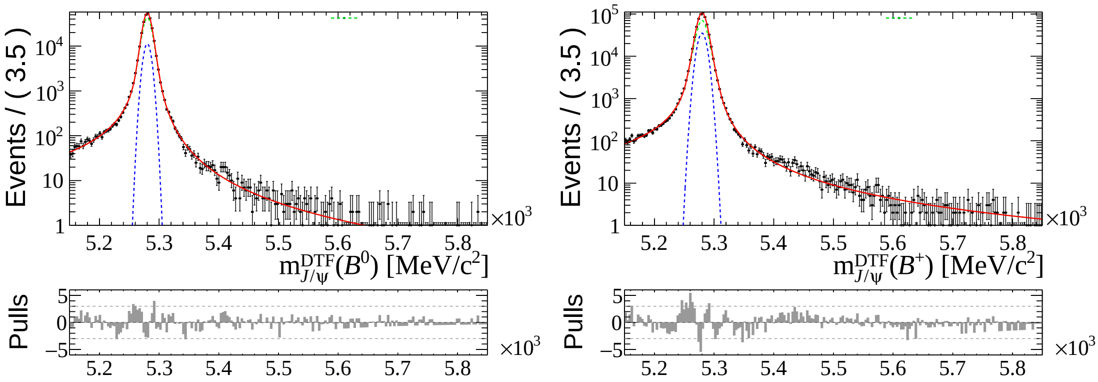
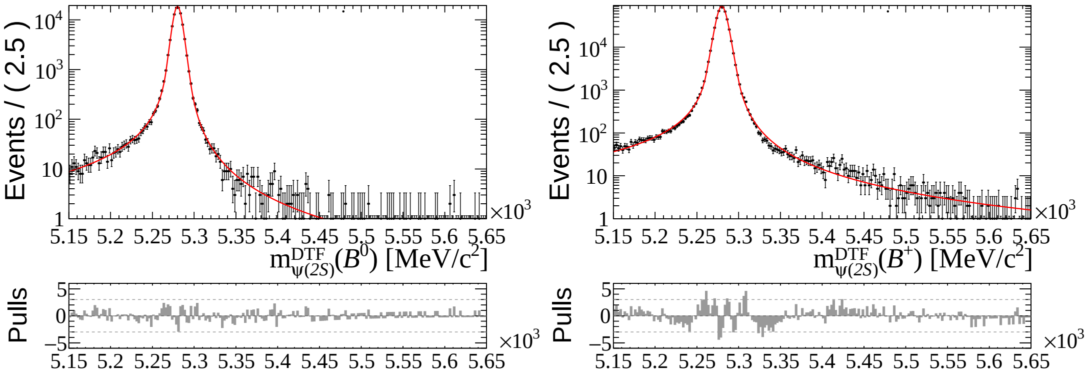


Figure E.5 – Fits to simulated $B^0 \to K^{*0} J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ (right) decays with a J/ψ mass constraint on $m(\mu^+\mu^-)$. The fits shown above use $15+16$ simulation in the exclusive `L0L` category.



(a) $B^0 \rightarrow K^{*0}\psi(2S)(\rightarrow \mu^+\mu^-)$ (b) $B^+ \rightarrow K^+\psi(2S)(\rightarrow \mu^+\mu^-)$

Figure E.6 – Fits to simulated muon $\psi(2S)$ resonant modes $B^0 \rightarrow K^{*0}\psi(2S)(\rightarrow \mu^+\mu^-)$ ($B^+ \rightarrow K^+\psi(2S)(\rightarrow \mu^+\mu^-)$) are shown on the left(right). The invariant mass is reconstructed with a DTF $\psi(2S)$ mass constraint on the dimuon system. The fits shown above uses $15 + 16$ simulation passing the muon `L0L` triggers.

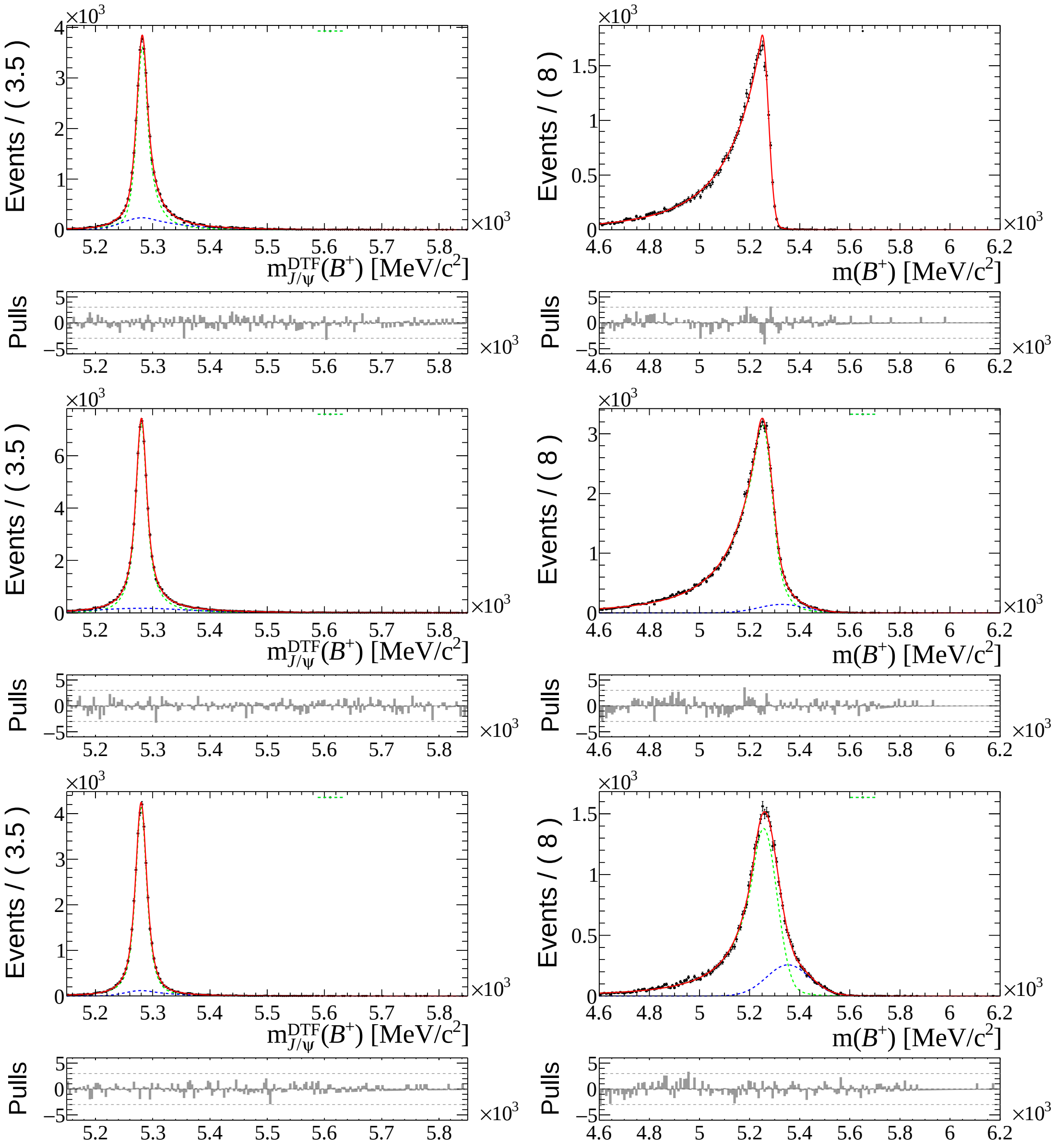
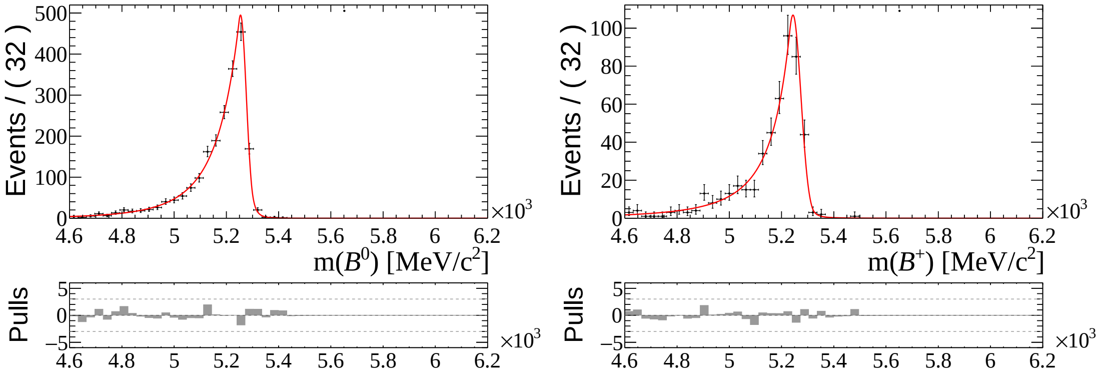
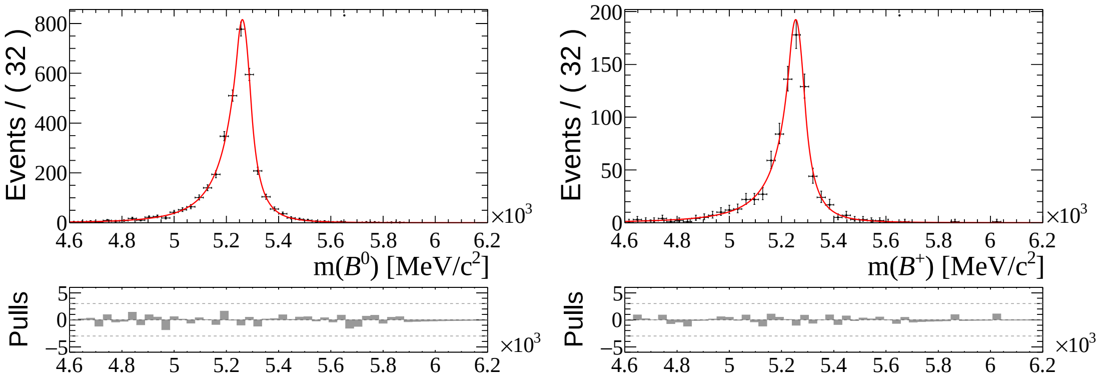


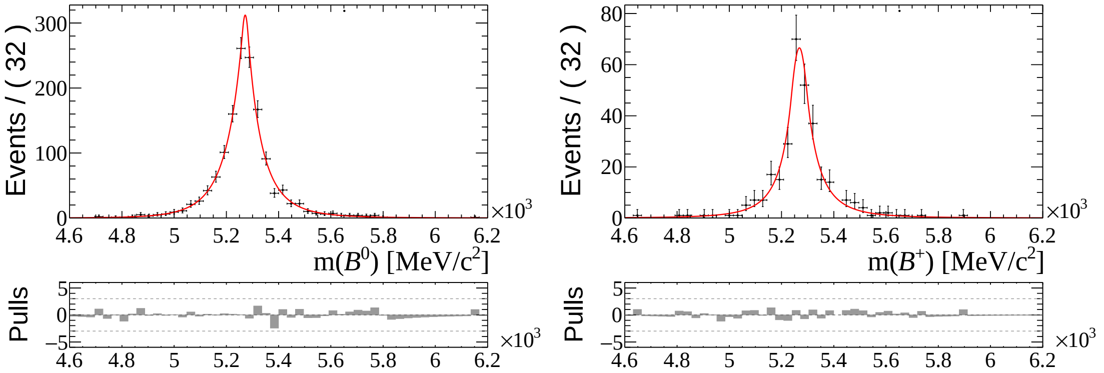
Figure E.7 – Fits to simulated $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ in the 0γ (top), 1γ (middle) and 2γ (bottom) category to B^0 mass reconstructed with (left) and without (right) J/ψ mass constraint. The fits shown above use $15+16$ simulation in the exclusive `L0L` category.



(a) $\mathbf{0\gamma}$ category, low-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0}e^+e^-}$ (b) $\mathbf{0\gamma}$ category, low-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+e^+e^-}$

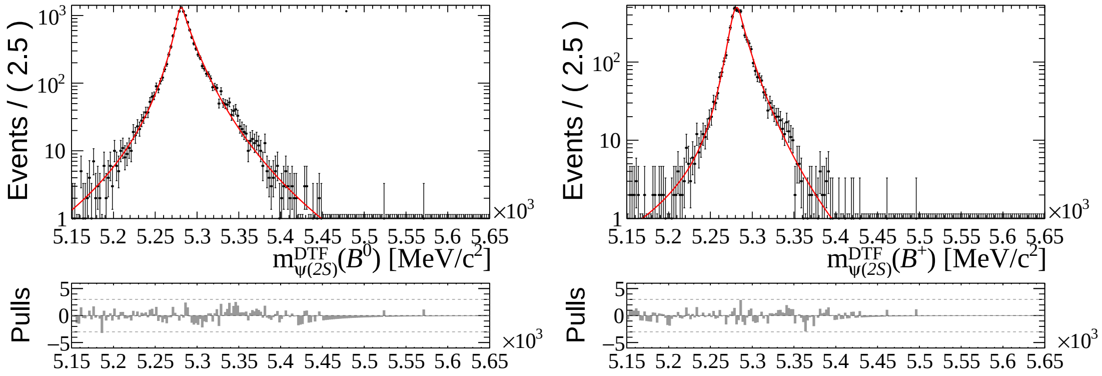


(c) $\mathbf{1\gamma}$ category, low-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0}e^+e^-}$ (d) $\mathbf{1\gamma}$ category, low-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+e^+e^-}$

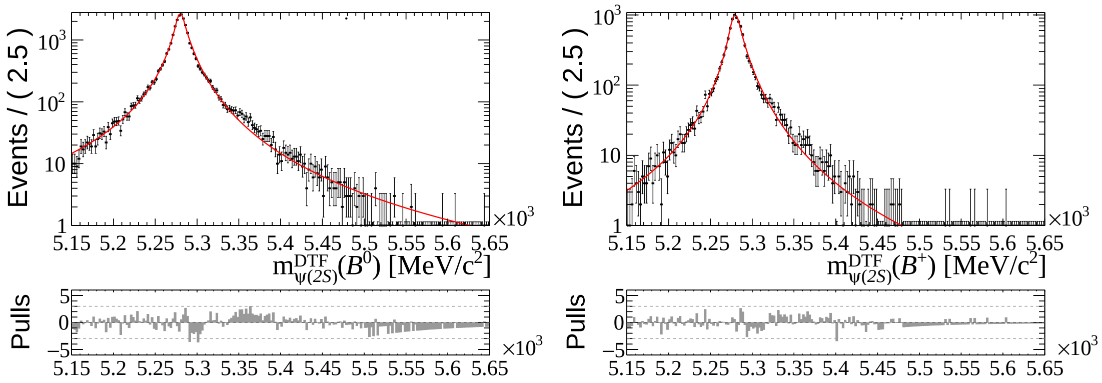


(e) $\mathbf{2\gamma}$ category, low-$\boldsymbol{q^2}$ $\boldsymbol{B^0 \to K^{*0}e^+e^-}$ (f) $\mathbf{2\gamma}$ category, low-$\boldsymbol{q^2}$ $\boldsymbol{B^+ \to K^+e^+e^-}$

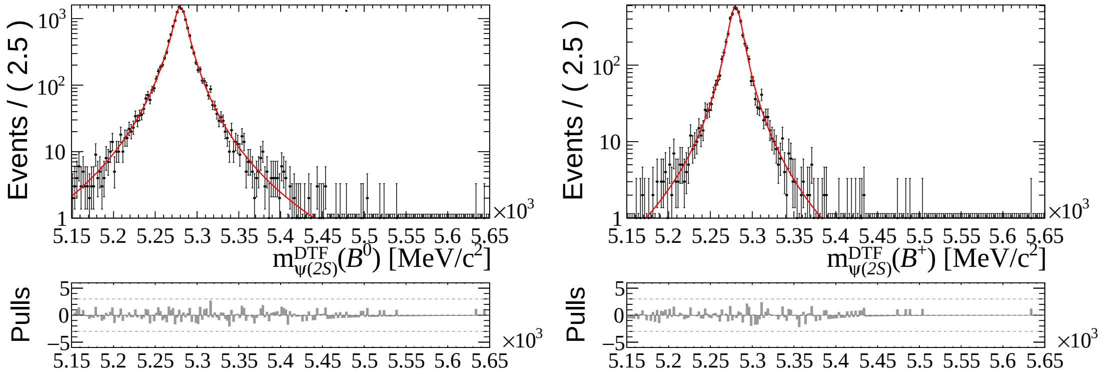
Figure E.8 – Simulated electron signal modes in low-q^2 bin. $B^0 \to K^{*0}e^+e^-$ ($B^0 \to K^{*0}e^+e^-$) are shown on the left(right), separated into $\mathbf{0\gamma}$ (top), $\mathbf{0\gamma}$ (middle) and $\mathbf{2\gamma}$ (bottom) bremsstrahlung categories. The fits shown above use **15 + 16** simulation in the exclusive `L0L` category.



(a) $\mathbf{0\gamma}$ category, $\boldsymbol{B^0 \to K^{*0}\psi(2S)(\to e^+e^-)}$.
(b) $\mathbf{0\gamma}$ category, $\boldsymbol{B^+ \to K^+\psi(2S)(\to e^+e^-)}$.



(c) $\mathbf{1\gamma}$ category, $\boldsymbol{B^0 \to K^{*0}\psi(2S)(\to e^+e^-)}$.
(d) $\mathbf{1\gamma}$ category, $\boldsymbol{B^+ \to K^+\psi(2S)(\to e^+e^-)}$.

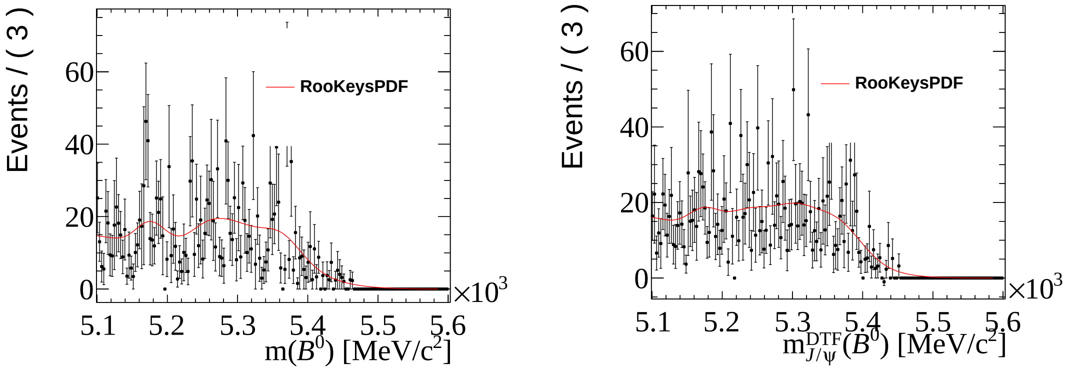


(e) $\mathbf{2\gamma}$ category, $\boldsymbol{B^0 \to K^{*0}\psi(2S)(\to e^+e^-)}$.
(f) $\mathbf{2\gamma}$ category, $\boldsymbol{B^+ \to K^+\psi(2S)(\to e^+e^-)}$.

Figure E.9 – Fits to simulated electron $\psi(2S)$ resonant modes $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ ($B^+ \to K^+\psi(2S)(\to e^+e^-)$) are shown on the left(right). A DTF $\psi(2S)$ mass constraint is applied on the dielectron system. The $\mathbf{0\gamma}$ (top), $\mathbf{1\gamma}$ (middle) and $\mathbf{2\gamma}$ (bottom) bremsstrahlung categories are modelled separately. The fits shown above use $\mathbf{15+16}$ simulation in the exclusive `L0L` category.

E.2 Shapes of Simulated Backgrounds

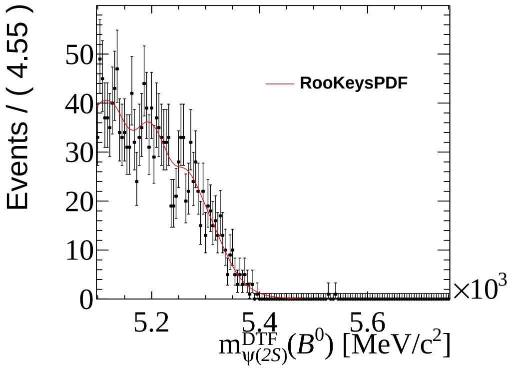
E.2.1 B^0 Mode Backgrounds



(a) $m(K\pi\mu\mu)$

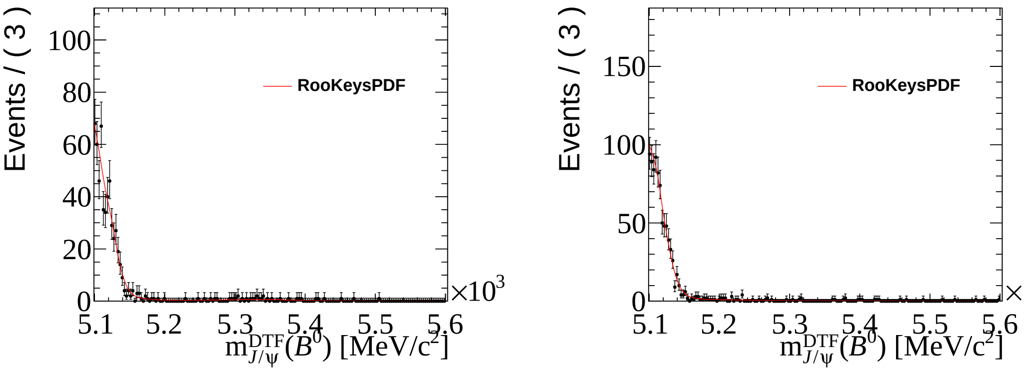
(b) $m(K\pi\mu\mu)^{J/\psi}_{DTF}$

$\Lambda_b^0 \to pKJ/\psi(\to \mu^+\mu^-)$

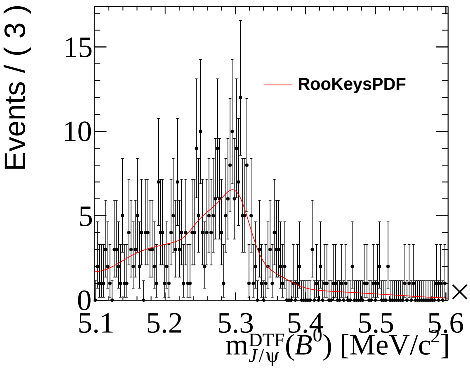


(c) $m(K\pi\mu\mu)^{\psi(2S)}_{DTF}$, $\Lambda_b^0 \to pK\psi(2S)(\to \mu^+\mu^-)$

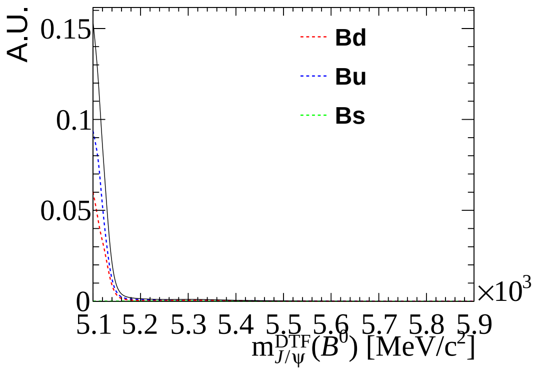
Figure E.10 – `RooKeysPdf` KDE modelled with simulated mis-identified $\Lambda_b^0 \to pKJ/\psi(\to \mu^+\mu^-)$ (top) and $\Lambda_b^0 \to pK\psi(2S)(\to \mu^+\mu^-)$ (bottom) backgrounds in $\mu\mu$ resonant modes. The $\Lambda_b^0 \to pKJ/\psi(\to \mu^+\mu^-)$ background on the top left(top right) is reconstructed without(with) J/ψ mass constraint. `LOL` 15+16.



(a) $B^0 \to X J/\psi (\to \mu^+\mu^-)$ (b) $B^+ \to X J/\psi (\to \mu^+\mu^-)$

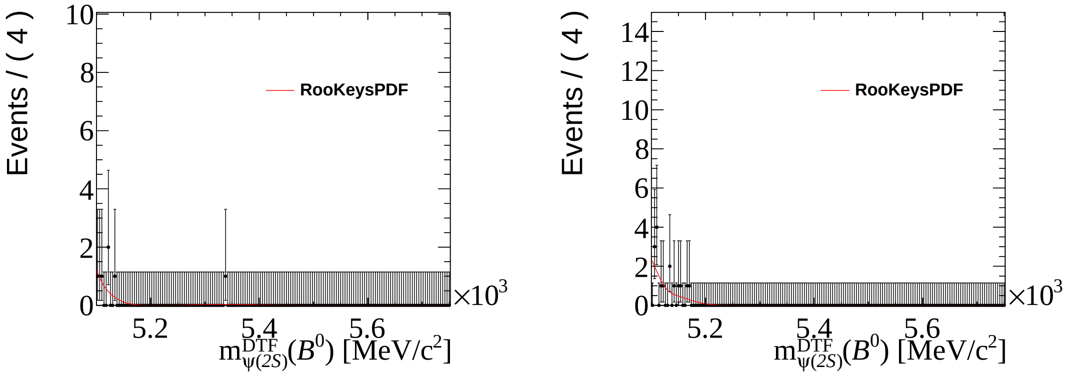


(c) $B_s^0 \to X J/\psi (\to \mu^+\mu^-)$



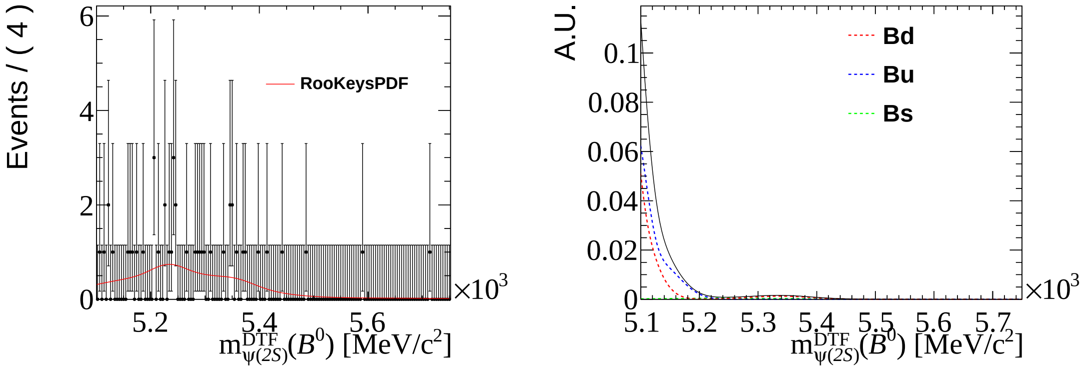
(d) $B^0 \to X J/\psi (\to \mu^+\mu^-)$, $B^+ \to X J/\psi (\to \mu^+\mu^-)$, and $B_s^0 \to X J/\psi (\to \mu^+\mu^-)$ Combined

Figure E.11 – Partially reconstructed J/ψ mass constrained $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 simulation passing `L0L` exclusive trigger. $B^0 \to X J/\psi (\to \mu^+\mu^-)$ (top left), $B^+ \to X J/\psi (\to \mu^+\mu^-)$ (top right) and $B_s^0 \to X J/\psi (\to \mu^+\mu^-)$ (bottom left) are added linearly according to their ratios (bottom right).



(a) $B^0 \rightarrow XJ/\psi(\rightarrow \mu^+\mu^-)$ and $B^0 \rightarrow X\psi(2S)(\rightarrow \mu^+\mu^-)$

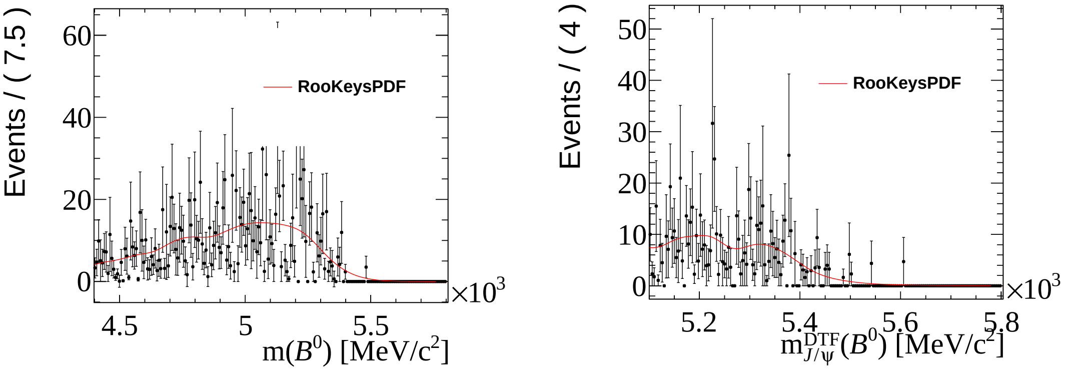
(b) $B^+ \rightarrow XJ/\psi(\rightarrow \mu^+\mu^-)$ and $B^+ \rightarrow X\psi(2S)(\rightarrow \mu^+\mu^-)$



(c) $B_s^0 \rightarrow XJ/\psi(\rightarrow \mu^+\mu^-)$ and $B_s^0 \rightarrow X\psi(2S)(\rightarrow \mu^+\mu^-)$

(d) Inclusive B^0, B^+ and B_s^0 decays combined

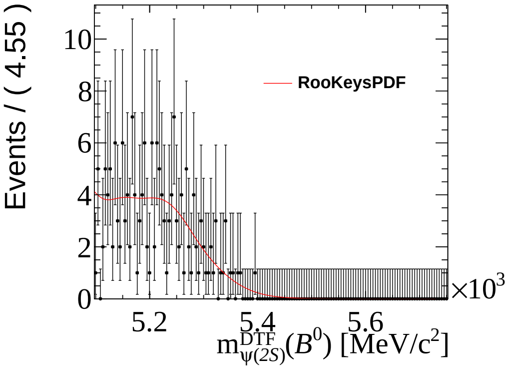
Figure E.12 – Partially reconstructed $\psi(2S)$ mass constrained $B^0 \rightarrow K^{*0}\psi(2S)(\rightarrow \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015+2016 simulation passing `L0L` exclusive trigger. Inclusive B^0 (top left), B^+ (top right) and B_s^0 (bottom left) decays are added linearly according to their ratios (bottom right).



(a) $m(K\pi ee)$

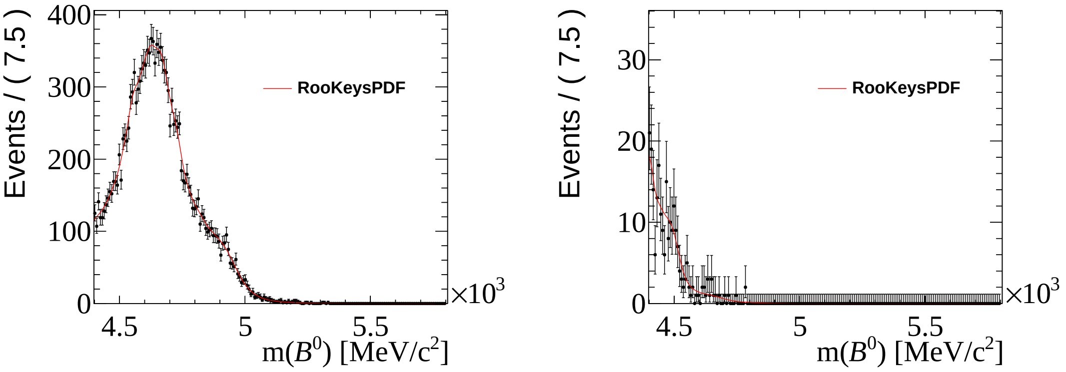
(b) $m(K\pi ee)^{J/\psi}_{DTF}$

$\Lambda_b^0 \to pKJ/\psi\,(\to e^+e^-)$

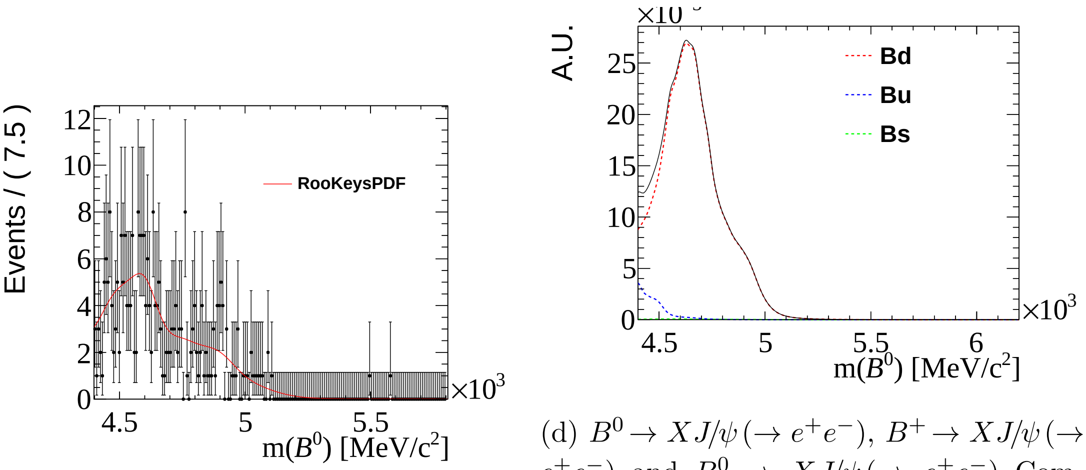


(c) $m(K\pi ee)^{\psi(2S)}_{DTF}$, $\Lambda_b^0 \to pK\psi(2S)(\to e^+e^-)$

Figure E.13 – `RooKeysPdf` KDE modelled with simulated mis-identified $\Lambda_b^0 \to pKJ/\psi\,(\to e^+e^-)$ (top) and $\Lambda_b^0 \to pK\psi(2S)(\to e^+e^-)$ (bottom) backgrounds in $\mu\mu$ resonant modes. The $\Lambda_b^0 \to pKJ/\psi\,(\to e^+e^-)$ background on the top left(top right) is reconstructed without(with) J/ψ mass constraint. `LOL` exclusive $15+16$.



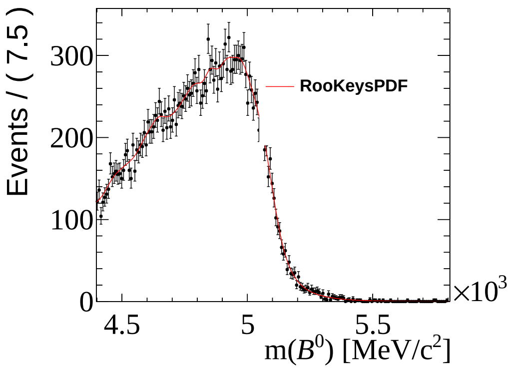
(a) $B^0 \to X J/\psi\,(\to e^+e^-)$ (b) $B^+ \to X J/\psi\,(\to e^+e^-)$



(c) $B_s^0 \to X J/\psi\,(\to e^+e^-)$

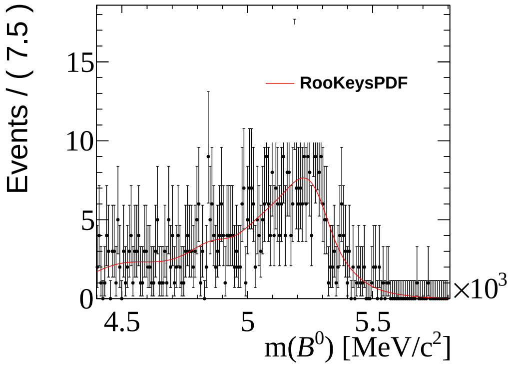
(d) $B^0 \to X J/\psi\,(\to e^+e^-)$, $B^+ \to X J/\psi\,(\to e^+e^-)$ and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ Combined

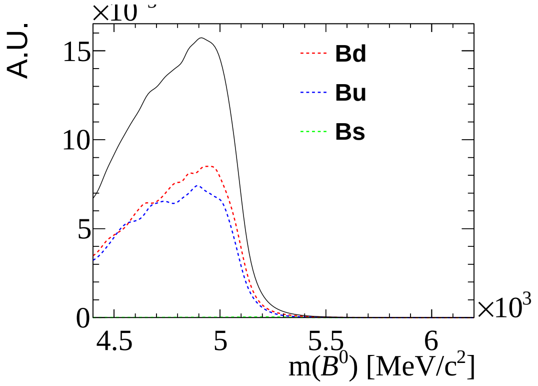
Figure E.14 – Leptonic partially reconstructed $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015+2016 Monte-Carlo passing `L0L` exclusive trigger. The B meson $\to J/\psi$ decay chain is truth matched to proceed via an intermediate $c\bar{c}$ resonance. $B^0 \to X J/\psi\,(\to e^+e^-)$ (top left), $B^+ \to X J/\psi\,(\to e^+e^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ (bottom left) are added linearly according to their ratios (bottom right).



(a) $B^0 \to X J/\psi\,(\to e^+e^-)$

(b) $B^+ \to X J/\psi\,(\to e^+e^-)$

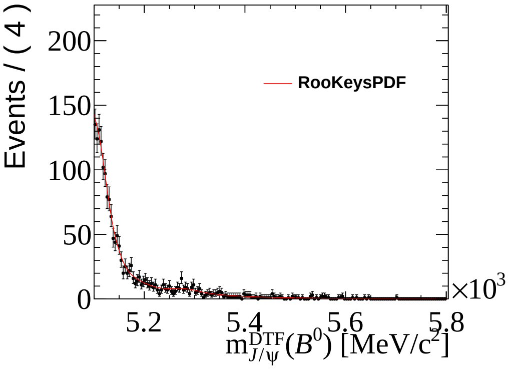




(c) $B_s^0 \to X J/\psi\,(\to e^+e^-)$

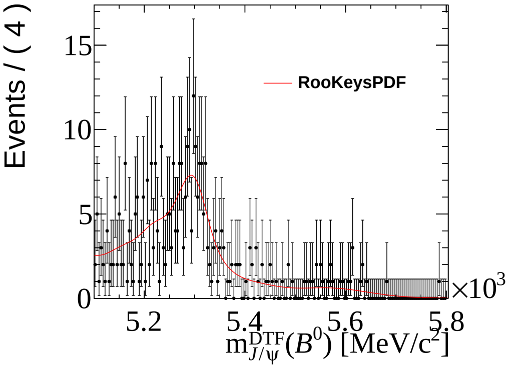
(d) $B^0 \to X J/\psi\,(\to e^+e^-)$, $B^+ \to X J/\psi\,(\to e^+e^-)$ and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ Combined

Figure E.15 – Hadronic partially reconstructed $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 simulation passing `L0L` exclusive trigger. $B^0 \to X J/\psi\,(\to e^+e^-)$ (top left), $B^+ \to X J/\psi\,(\to e^+e^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ (bottom left) are required to be the exclusive of the leptonic truth matching criteria and added linearly according to their ratios (bottom right).

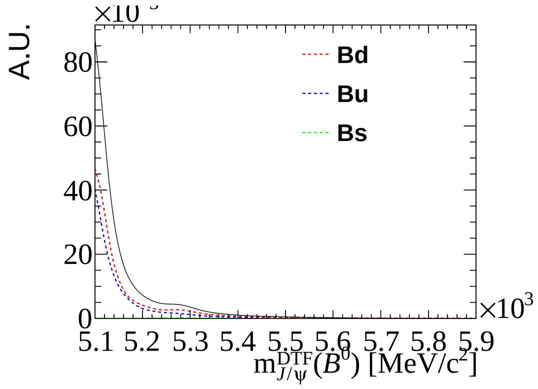


(a) $B^0 \to X J/\psi\,(\to e^+e^-)$

(b) $B^+ \to X J/\psi\,(\to e^+e^-)$

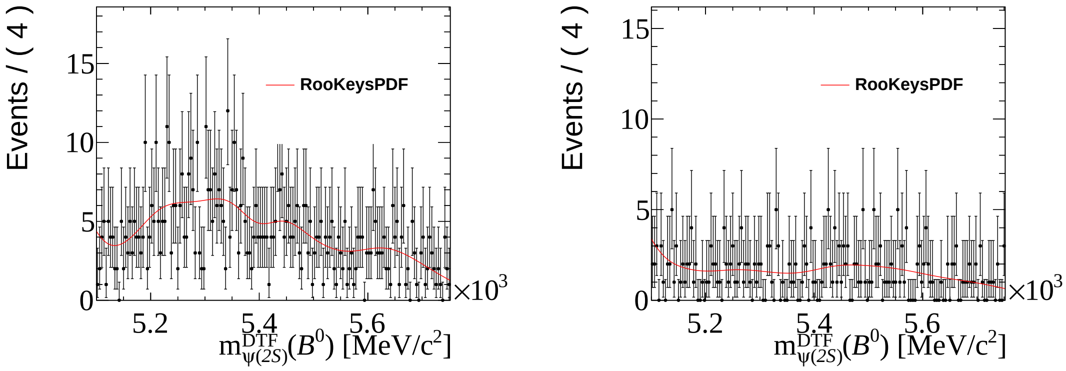


(c) $B_s^0 \to X J/\psi\,(\to e^+e^-)$



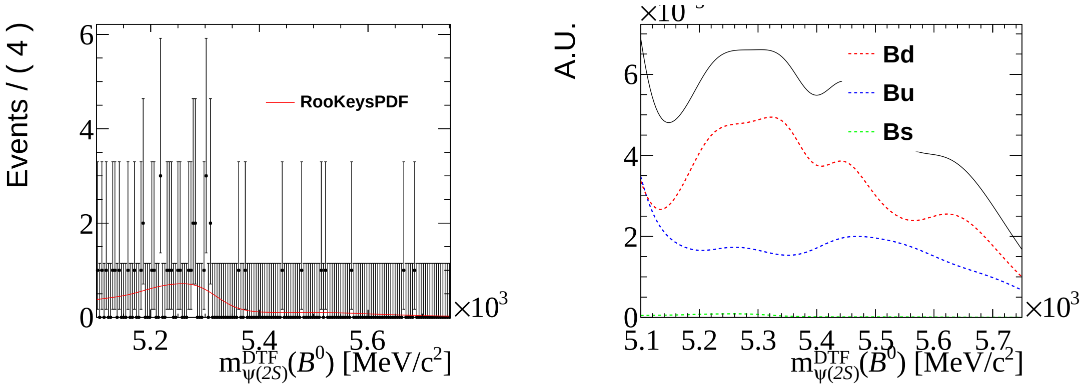
(d) $B^0 \to X J/\psi\,(\to e^+e^-)$, $B^+ \to X J/\psi\,(\to e^+e^-)$ and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ Combined

Figure E.16 – Partially reconstructed J/ψ mass constrained $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015+2016 Monte-Carlo passing `L0L` exclusive trigger. $B^0 \to X J/\psi\,(\to e^+e^-)$ (top left), $B^+ \to X J/\psi\,(\to e^+e^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ (bottom left) are added linearly according to their ratios (bottom right).



(a) $B^0 \to XJ/\psi(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$

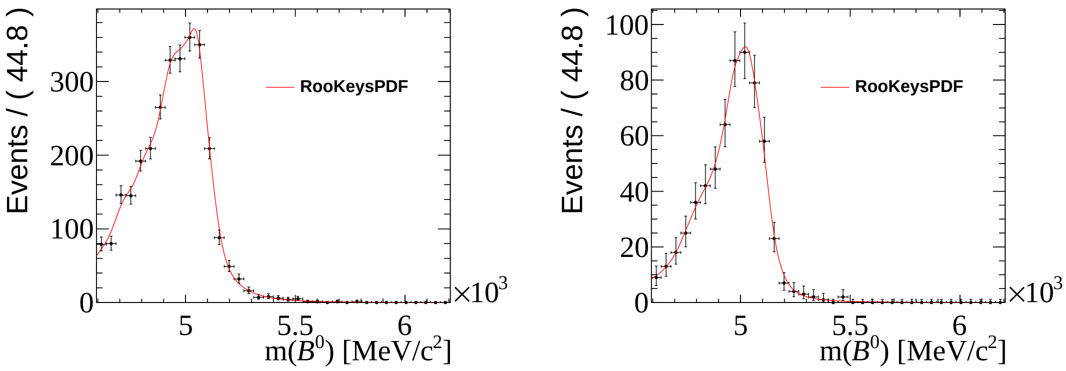
(b) $B^+ \to XJ/\psi(\to e^+e^-)$ and $B^+ \to X\psi(2S)(\to e^+e^-)$



(c) $B_s^0 \to XJ/\psi(\to e^+e^-)$ and $B_s^0 \to X\psi(2S)(\to e^+e^-)$

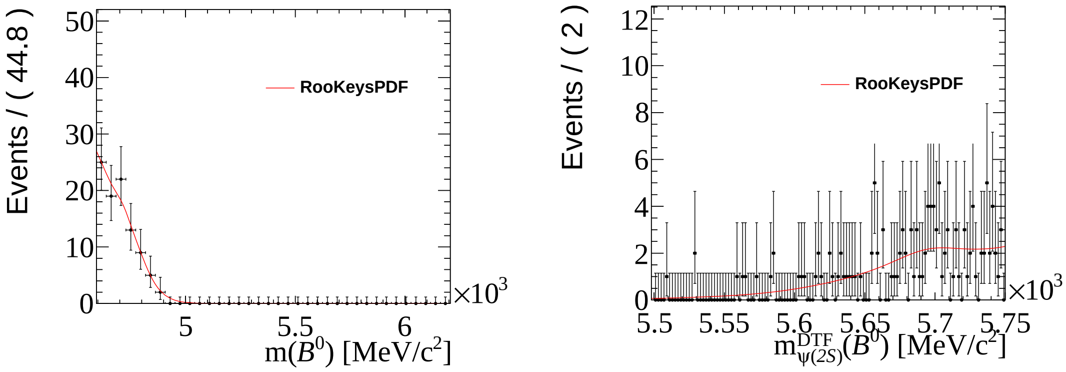
(d) Inclusive B^0, B^+ and B_s^0 decays combined

Figure E.17 – Partially reconstructed $\psi(2S)$ mass constrained $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 simulation passing `L0L` exclusive trigger. Inclusive B^0 (top left), B^+ (top right) and B_s^0 (bottom left) decays are added linearly according to their ratios (bottom right).



(a) low-q^2 (b) central-q^2

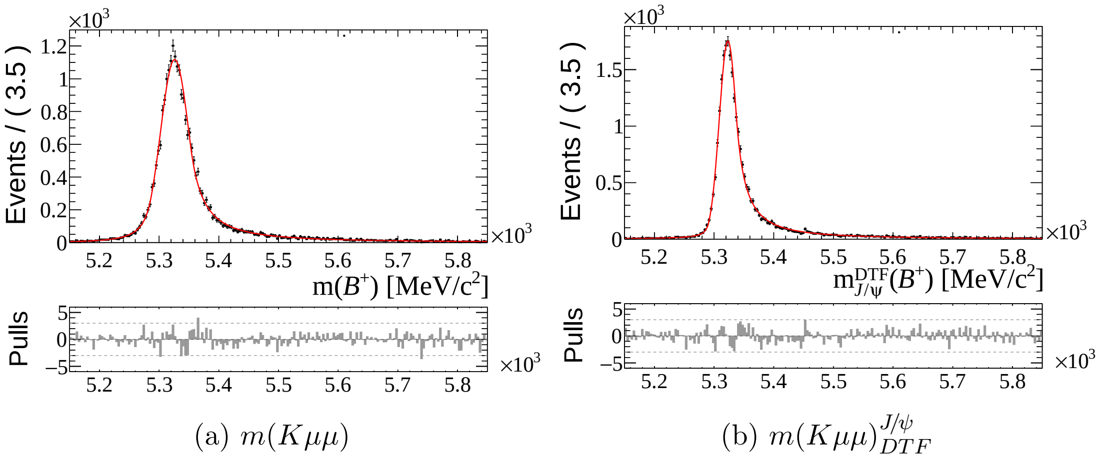
Figure E.18 – Partially reconstructed $B^+ \to K\pi\pi e^+e^-$ background, reconstructed as $B^0 \to K^{*0}e^+e^-$ in the low-q^2 (left) and central-q^2 (right). Modelled using `RooKeysPdf` KDE on 2015 + 2016 Monte-Carlo which passes the `L0L` exclusive trigger.



(a) central-q^2 (b) $\psi(2S)$ central-q^2 , $m(K\pi ee)^{\psi(2S)}_{DTF}$

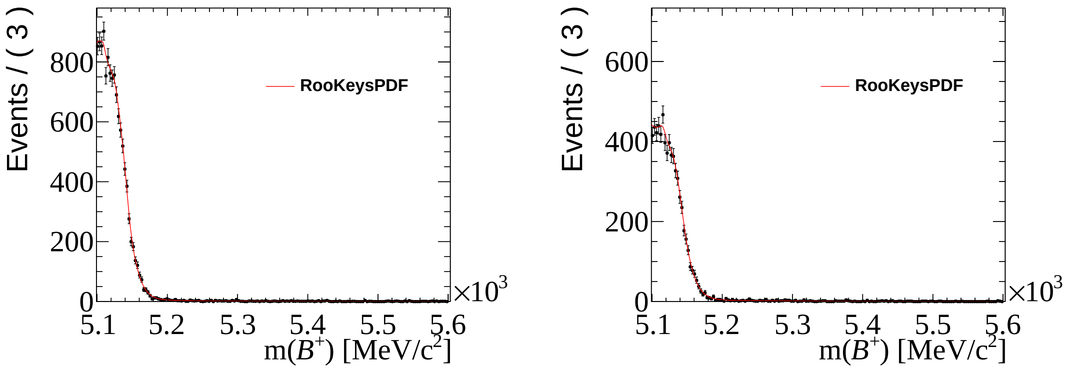
Figure E.19 – $B^0 \to K^{*0}J/\psi(\to e^+e^-)q^2$ bin leakage into the central-q^2 (left) and $\psi(2S)q^2$ (right) bin of $B^0 \to K^{*0}e^+e^-$ and $B^0 \to K^{*0}\psi(2S)(\to e^+e^-)$ due to bremsstrahlung losses. The background shape is $\psi(2S)q^2$ bin is reconstructed with a $\psi(2S)$ mass constraint which shifts the shape above the B^0 mass. Modelled using `RooKeysPdf` KDE on 2015 + 2016 Monte Carlo passing the `L0L` exclusive trigger.

E.2.2 B^+ Mode Backgrounds

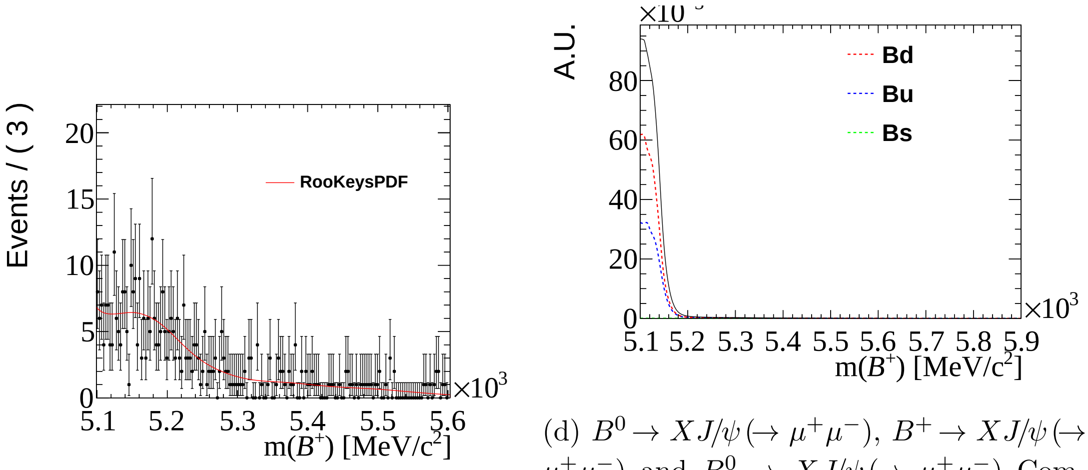


(a) $m(K\mu\mu)$ (b) $m(K\mu\mu)^{J/\psi}_{DTF}$

Figure E.20 – Fits to simulated $\pi \to K$ mis-identified $B^+ \to \pi^+ J/\psi(\to \mu^+\mu^-)$ backgrounds reconstructed as $B^+ \to K^+ J/\psi(\to \mu^+\mu^-)$. On the left(right) is the invariant mass reconstructed without(with) DTF constrained mass. Modelled using 2015 + 2016 Monte Carlo passing the `L0L` exclusive trigger.



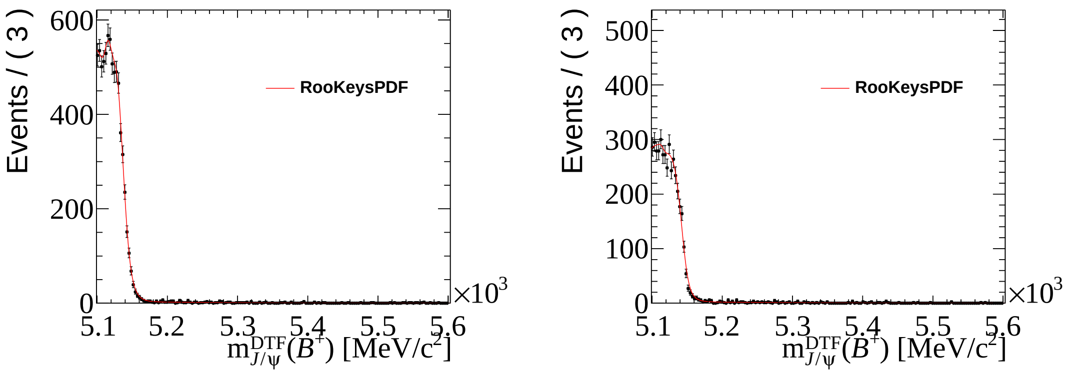
(a) $B^0 \to X J/\psi (\to \mu^+\mu^-)$ (b) $B^+ \to X J/\psi (\to \mu^+\mu^-)$



(c) $B_s^0 \to X J/\psi (\to \mu^+\mu^-)$

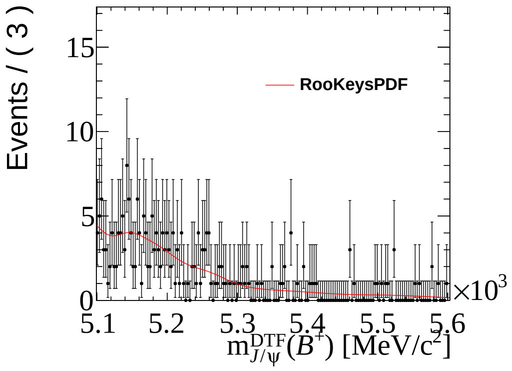
(d) $B^0 \to X J/\psi (\to \mu^+\mu^-)$, $B^+ \to X J/\psi (\to \mu^+\mu^-)$ and $B_s^0 \to X J/\psi (\to \mu^+\mu^-)$ Combined

Figure E.21 – Partially reconstructed $B^+ \to K^+ J/\psi (\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 Monte-Carlo passing `L0L` exclusive trigger. $B^0 \to X J/\psi (\to \mu^+\mu^-)$ (top left), $B^+ \to X J/\psi (\to \mu^+\mu^-)$ (top right) and $B_s^0 \to X J/\psi (\to \mu^+\mu^-)$ (bottom left) are added linearly according to their ratios (bottom right).

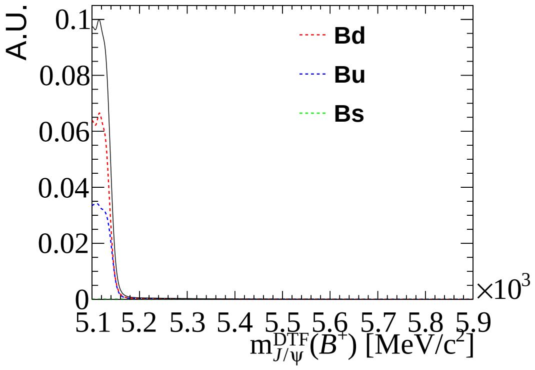


(a) $B^0 \to X J/\psi\,(\to \mu^+\mu^-)$

(b) $B^+ \to X J/\psi\,(\to \mu^+\mu^-)$

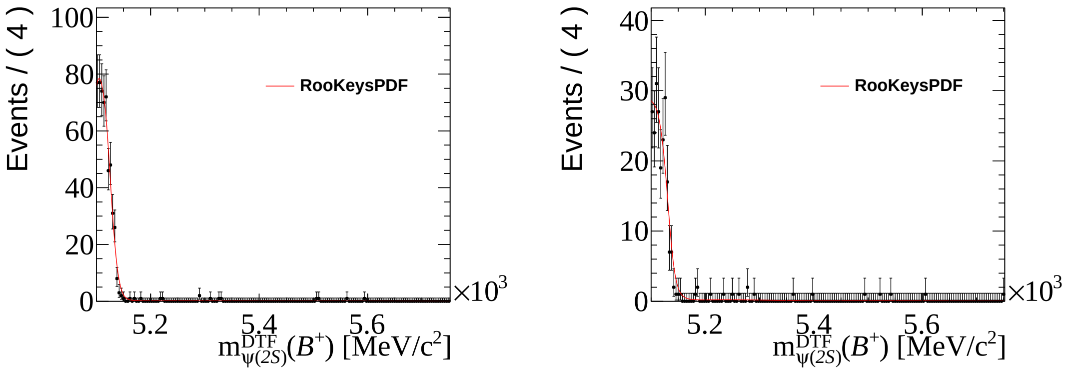


(c) $B_s^0 \to X J/\psi\,(\to \mu^+\mu^-)$



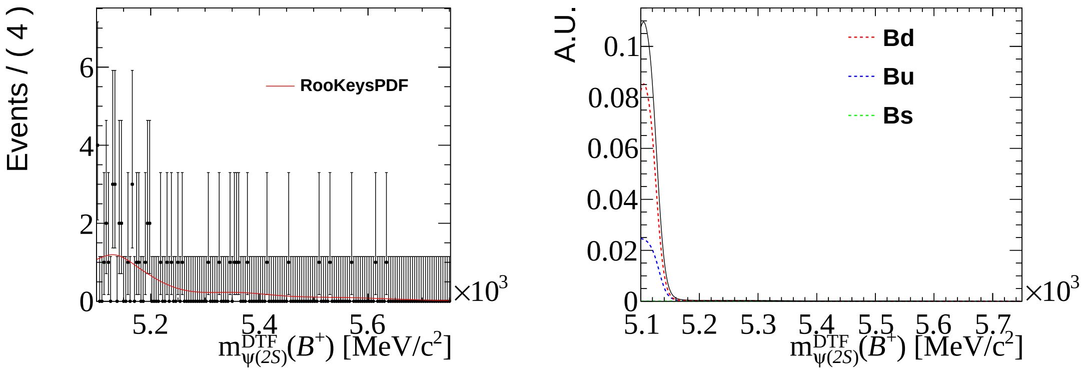
(d) $B^0 \to X J/\psi\,(\to \mu^+\mu^-)$, $B^+ \to X J/\psi\,(\to \mu^+\mu^-)$ and $B_s^0 \to X J/\psi\,(\to \mu^+\mu^-)$ Combined

Figure E.22 – Partially reconstructed J/ψ mass constrained $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 Monte-Carlo passing `L0L` exclusive trigger. $B^0 \to X J/\psi\,(\to \mu^+\mu^-)$ (top left), $B^+ \to X J/\psi\,(\to \mu^+\mu^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to \mu^+\mu^-)$ (bottom left) are added linearly according to their ratios (bottom right).



(a) $B^0 \to XJ/\psi(\to \mu^+\mu^-)$ and $B^0 \to X\psi(2S)(\to \mu^+\mu^-)$

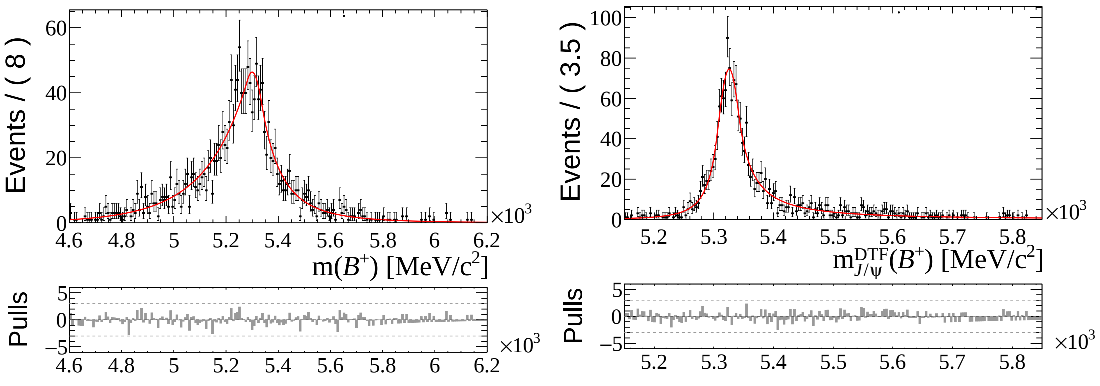
(b) $B^+ \to XJ/\psi(\to \mu^+\mu^-)$ and $B^+ \to X\psi(2S)(\to \mu^+\mu^-)$



(c) $B^0_s \to XJ/\psi(\to \mu^+\mu^-)$ and $B^0_s \to X\psi(2S)(\to \mu^+\mu^-)$

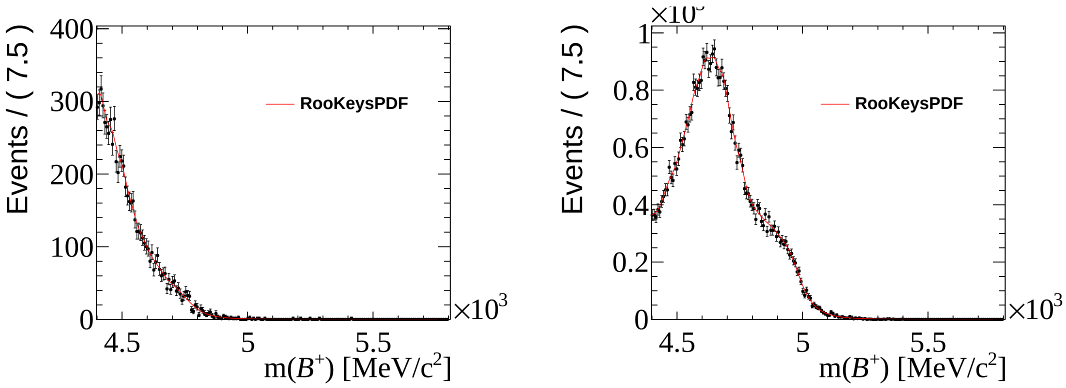
(d) Inclusive B^0, B^+ and B^0_s decays combined

Figure E.23 – Partially reconstructed $\psi(2S)$ mass constrained $B^+ \to K^+\psi(2S)(\to \mu^+\mu^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015+2016 simulation passing `L0L` exclusive trigger. Inclusive B^0 (top left), B^+ (top right) and B^0_s (bottom left) decays are added linearly according to their ratios (bottom right).



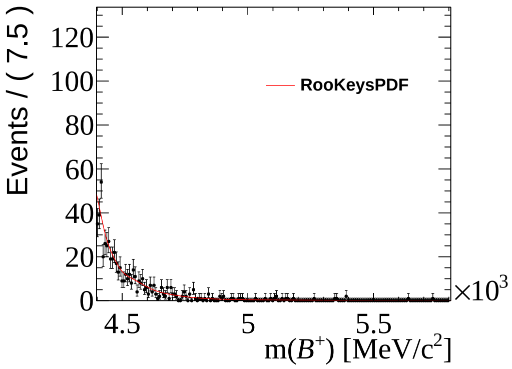
(a) $m(Kee)$ (b) $m(Kee)^{J/\psi}_{DTF}$

Figure E.24 – Fits to simulated $\pi \to K$ mis-identified $B^+ \to \pi^+ J/\psi(\to e^+e^-)$ backgrounds reconstructed as $B^+ \to K^+ J/\psi(\to e^+e^-)$. On the left(right) is the invariant mass reconstructed without(with) DTF constrained mass. Modelled using $2015 + 2016$ Monte Carlo passing the `L0L` exclusive trigger.

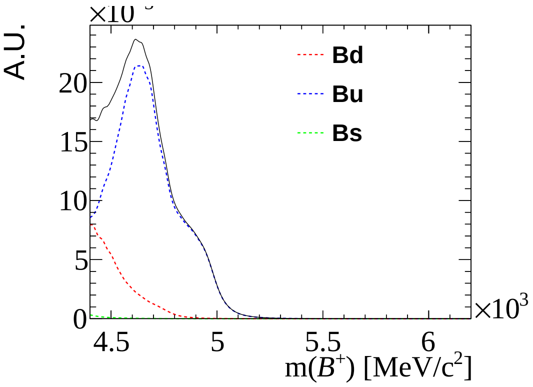


(a) $B^0 \to X J/\psi\,(\to e^+e^-)$

(b) $B^+ \to X J/\psi\,(\to e^+e^-)$

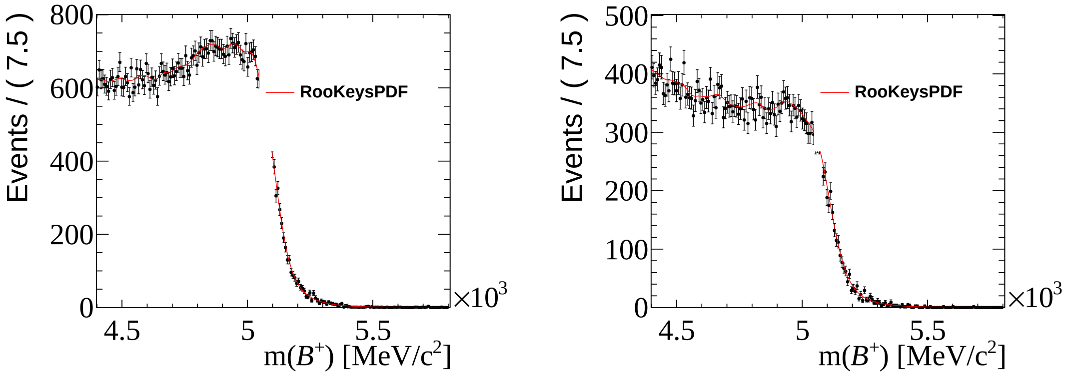


(c) $B_s^0 \to X J/\psi\,(\to e^+e^-)$



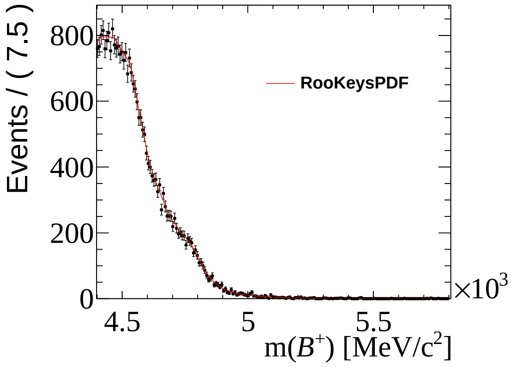
(d) $B^0 \to X J/\psi\,(\to e^+e^-)$, $B^+ \to X J/\psi\,(\to e^+e^-)$ and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ Combined

Figure E.25 – Leptonic partially reconstructed $B^0 \to K^{*0} J/\psi\,(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015+2016 Monte-Carlo passing `L0L` exclusive trigger. The B meson $\to J/\psi$ decay chain is truth matched to proceed via an intermediate $c\bar{c}$ resonance. $B^0 \to X J/\psi\,(\to e^+e^-)$ (top left), $B^+ \to X J/\psi\,(\to e^+e^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ (bottom left) are added linearly according to their ratios (bottom right).

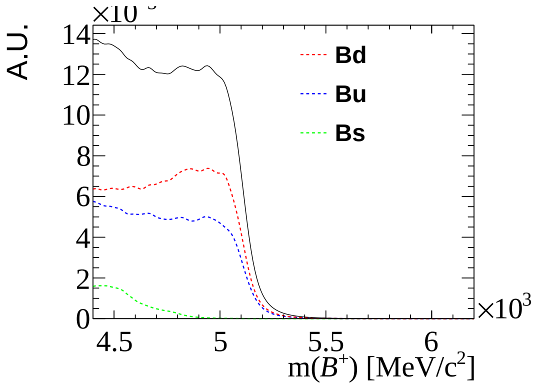


(a) $B^0 \to X J/\psi\, (\to e^+e^-)$

(b) $B^+ \to X J/\psi\, (\to e^+e^-)$

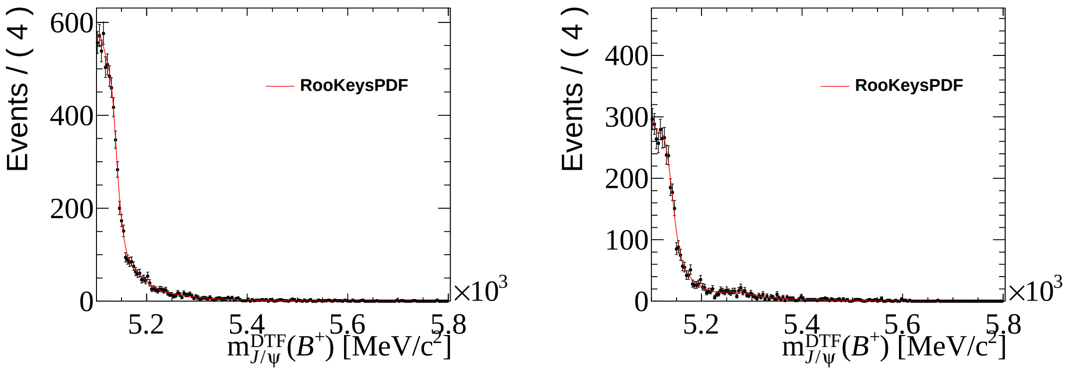


(c) $B_s^0 \to X J/\psi\, (\to e^+e^-)$



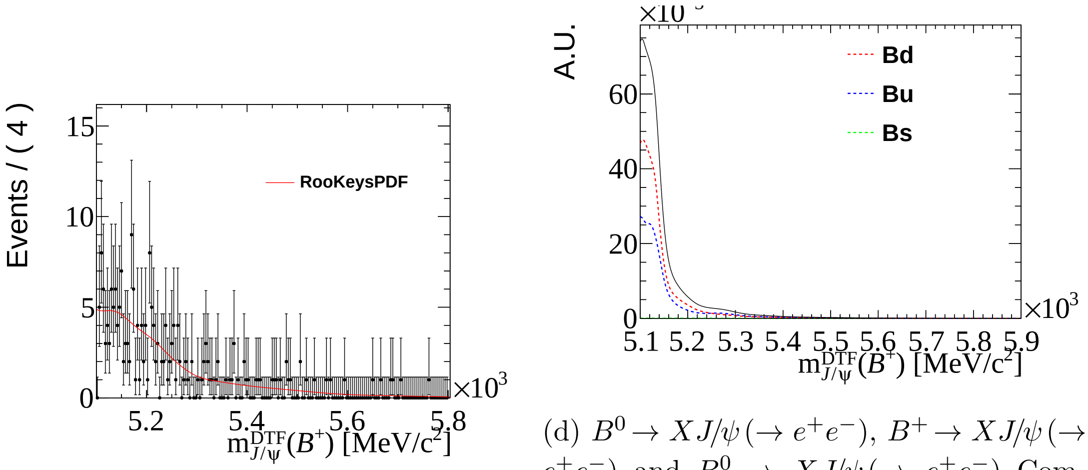
(d) $B^0 \to X J/\psi\, (\to e^+e^-)$, $B^+ \to X J/\psi\, (\to e^+e^-)$ and $B_s^0 \to X J/\psi\, (\to e^+e^-)$ Combined

Figure E.26 – Hadronic partially reconstructed $B^0 \to K^{*0} J/\psi\, (\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 simulation passing `L0L` exclusive trigger. $B^0 \to X J/\psi\, (\to e^+e^-)$ (top left), $B^+ \to X J/\psi\, (\to e^+e^-)$ (top right) and $B_s^0 \to X J/\psi\, (\to e^+e^-)$ (bottom left) are required to be the exclusive of the leptonic truth matching criteria and added linearly according to their ratios (bottom right).



(a) $B^0 \to X J/\psi\,(\to e^+e^-)$

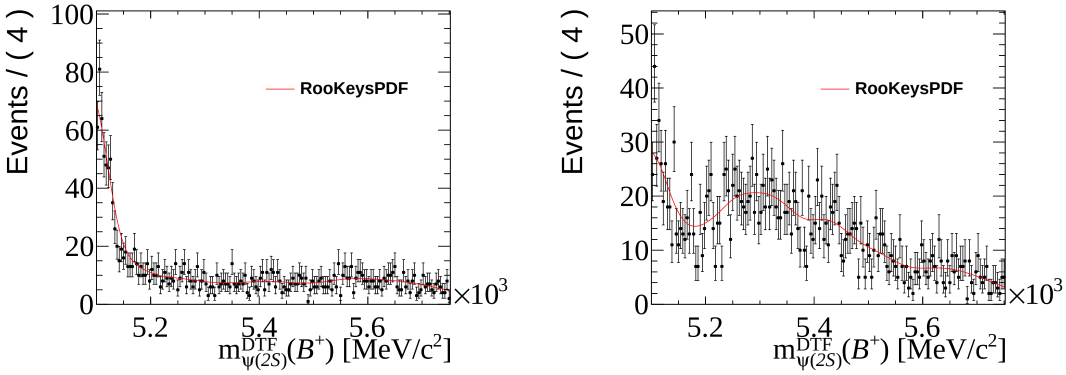
(b) $B^+ \to X J/\psi\,(\to e^+e^-)$



(c) $B_s^0 \to X J/\psi\,(\to e^+e^-)$

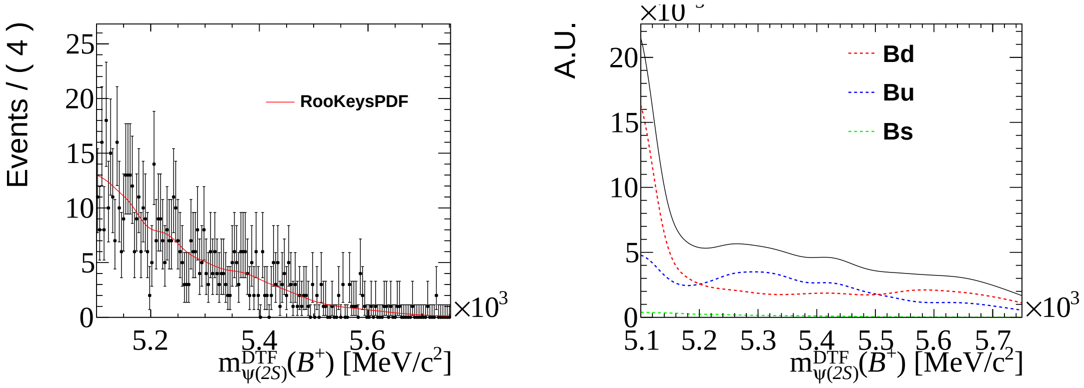
(d) $B^0 \to X J/\psi\,(\to e^+e^-)$, $B^+ \to X J/\psi\,(\to e^+e^-)$ and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ Combined

Figure E.27 – Partially reconstructed J/ψ mass constrained $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015+2016 Monte-Carlo passing `L0L` exclusive trigger. $B^0 \to X J/\psi\,(\to e^+e^-)$ (top left), $B^+ \to X J/\psi\,(\to e^+e^-)$ (top right) and $B_s^0 \to X J/\psi\,(\to e^+e^-)$ (bottom left) are added linearly according to their ratios (bottom right).



(a) $B^0 \to XJ/\psi(\to e^+e^-)$ and $B^0 \to X\psi(2S)(\to e^+e^-)$

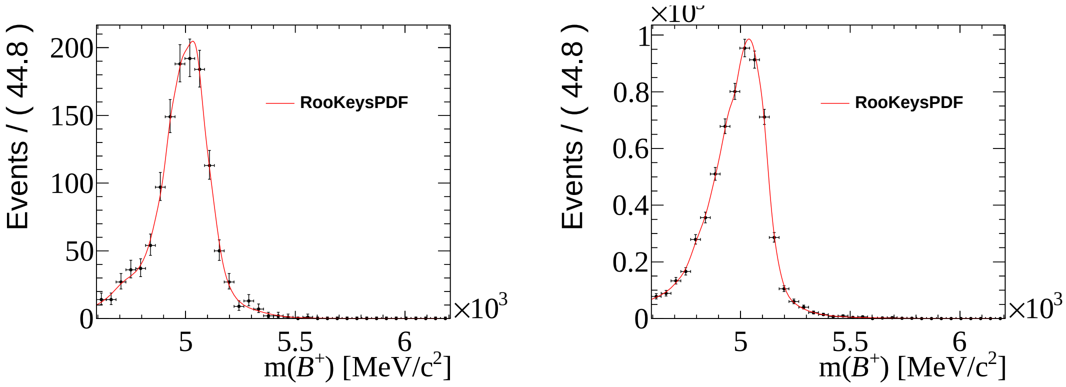
(b) $B^+ \to XJ/\psi(\to e^+e^-)$ and $B^+ \to X\psi(2S)(\to e^+e^-)$



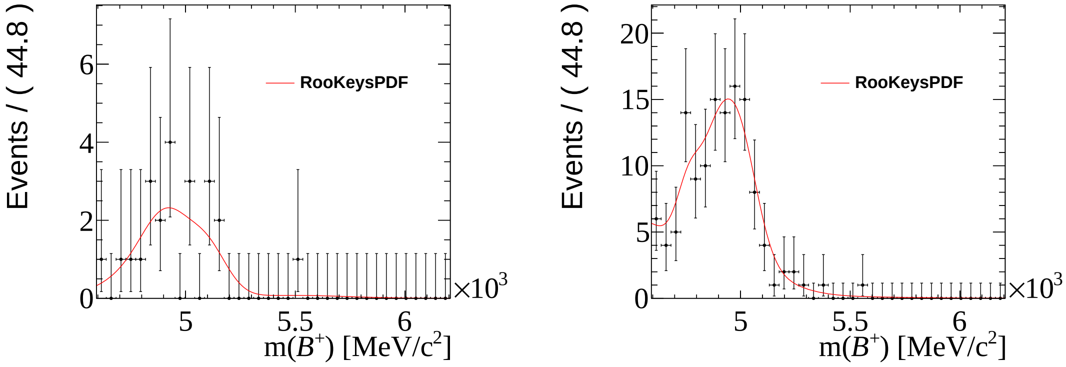
(c) $B^0_s \to XJ/\psi(\to e^+e^-)$ and $B^0_s \to X\psi(2S)(\to e^+e^-)$

(d) Inclusive B^0, B^+ and B^0_s decays combined

Figure E.28 – Partially reconstructed $\psi(2S)$ mass constrained $B^+ \to K^+\psi(2S)(\to e^+e^-)$ background modelled using `RooKeysPdf` KDE on cocktail 2015 + 2016 simulation passing `L0L` exclusive trigger. Inclusive B^0 (top left), B^+ (top right) and B^0_s (bottom left) decays are added linearly according to their ratios (bottom right).

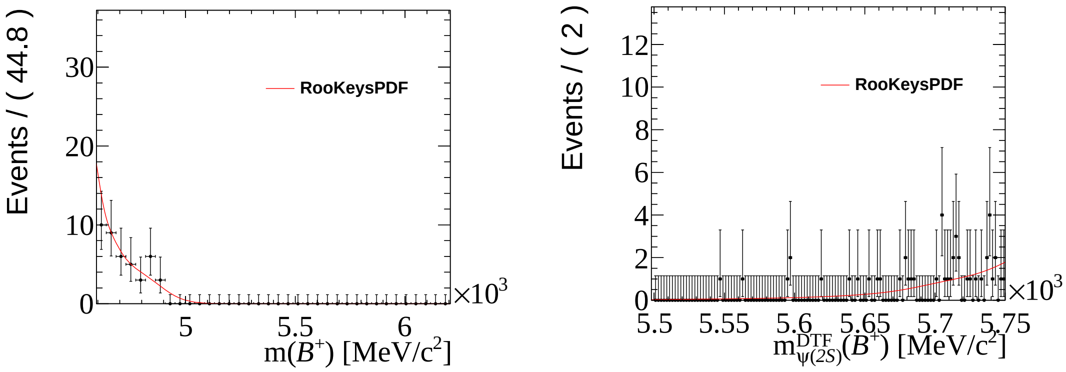


(a) low-q^2 (b) central-q^2



(c) low-q^2 (d) central-q^2

Figure E.29 – Partially reconstructed $B^0 \to K^{*0} e^+ e^-$ (top) and $B^0 \to K^+ \pi^- e^+ e^-$ (bottom) background, reconstructed as $B^+ \to K^+ e^+ e^-$ in the low-q^2 (left) and central-q^2 (right). Modelled using `RooKeysPdf` KDE on 2015 + 2016 simulation which passes the `L0L` exclusive trigger.



(a) central-q^2 (b) $\psi(2S)$ central-q^2 , $m(Kee)^{\psi(2S)}_{DTF}$

Figure E.30 – $B^+ \to K^+ J/\psi(\to e^+e^-)q^2$ bin leakage into central-q^2 (left) and $\psi(2S)q^2$ (right) bin of $B^+ \to K^+e^+e^-$ and $B^+ \to K^+\psi(2S)(\to e^+e^-)$ due to bremsstrahlung losses. The background shape in $\psi(2S)q^2$ bin is reconstructed with a $\psi(2S)$ mass constraint which shifts the shape above the B^+ mass. Modelled using `RooKeysPdf` KDE on $2015 + 2016$ Monte Carlo passing the `L0L` exclusive trigger.

E.3 Rare Mode Fits to Data

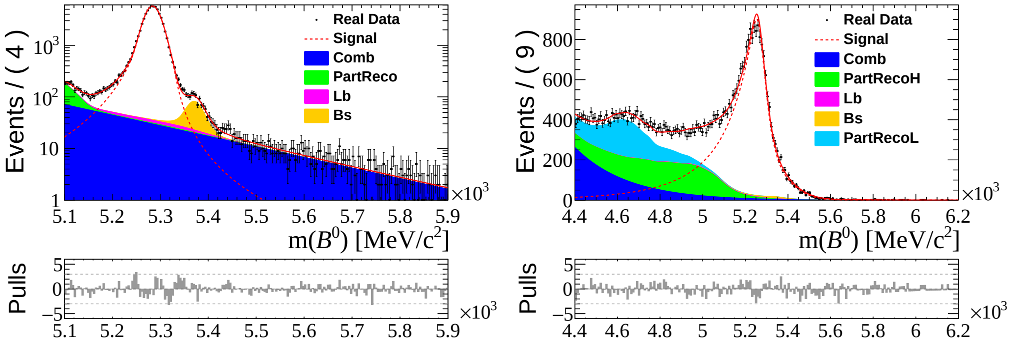


Figure E.31 – Fits to reconstructed $B^0 \to K^{*0} J/\psi (\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0} J/\psi (\to e^+e^-)$ (right). Run 1 `L0I`.

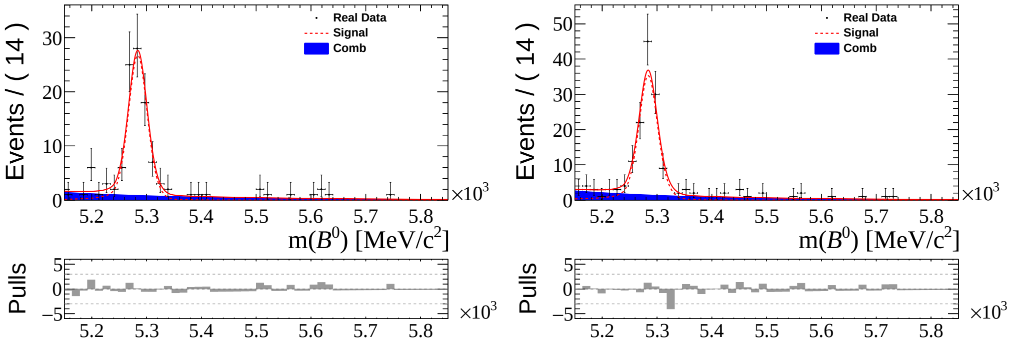


Figure E.32 – Fits to $B^0 \to K^{*0}\mu^+\mu^-$ in the low (left) and central (right) q^2. Run 1 `L0I`.

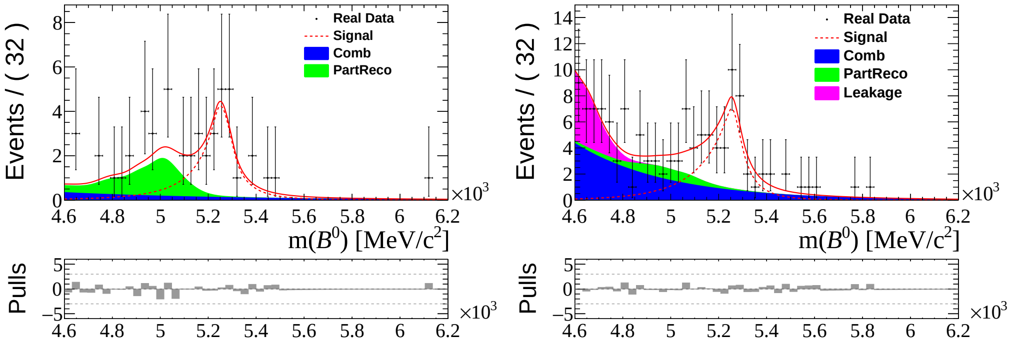


Figure E.33 – Fits to $B^0 \to K^{*0}e^+e^-$ in the low (left) and central (right) q^2. Run 1 `L0I`.

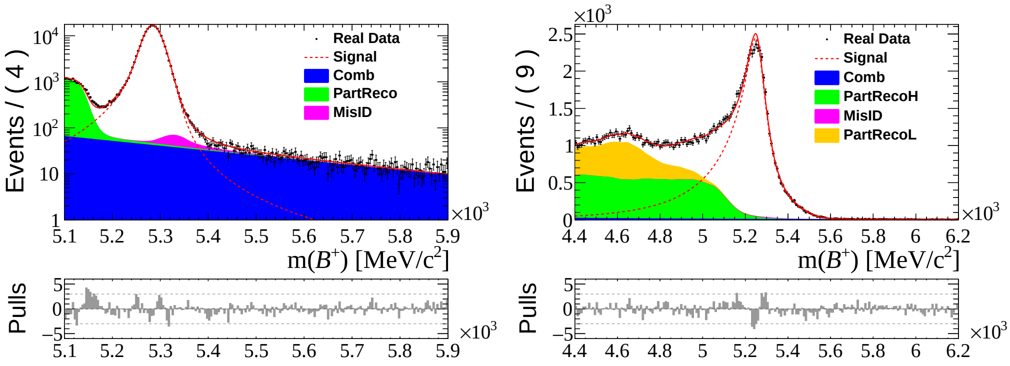


Figure E.34 – Fits to reconstructed $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (right). Run 1 `L0I`.

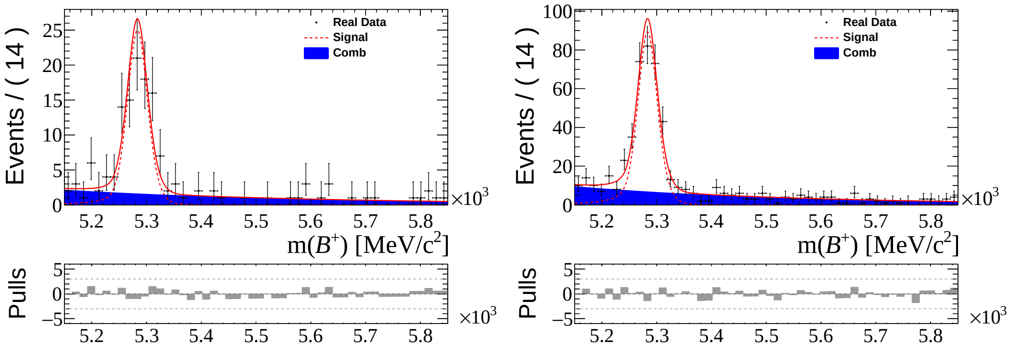


Figure E.35 – Fits to $B^+ \to K^+\mu^+\mu^-$ in the low (left) and central (right) q^2. Run 1 `L0I`.

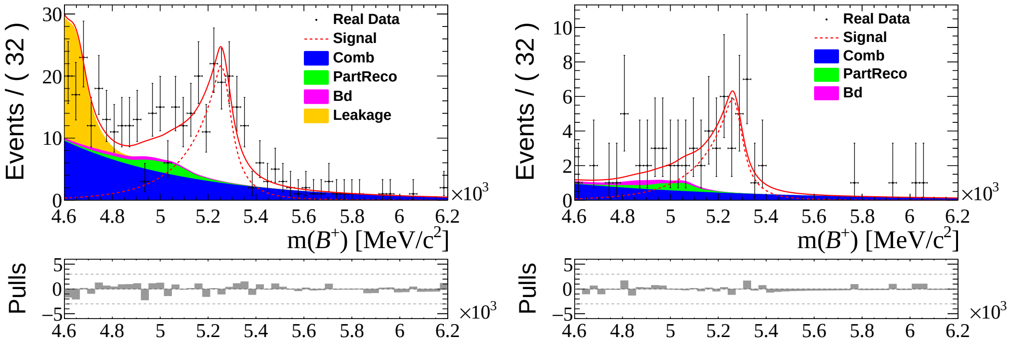


Figure E.36 – Fits to $B^+ \to K^+e^+e^-$ in the low (left) and central (right) q^2. Run 1 `L0I`.

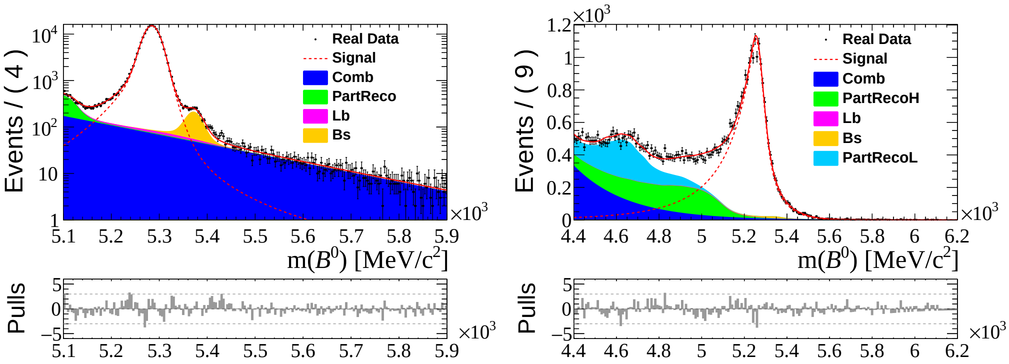


Figure E.37 – Fits to reconstructed $B^0 \to K^{*0} J/\psi(\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0} J/\psi(\to e^+e^-)$ (right). Run 1 `L0L` exclusive.

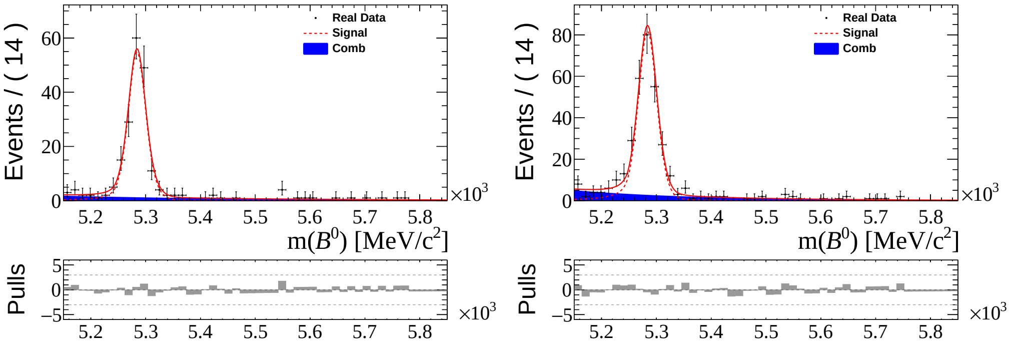


Figure E.38 – Fits to $B^0 \to K^{*0}\mu^+\mu^-$ in the low (left) and central (right) q^2. Run 1 `L0L` exclusive.

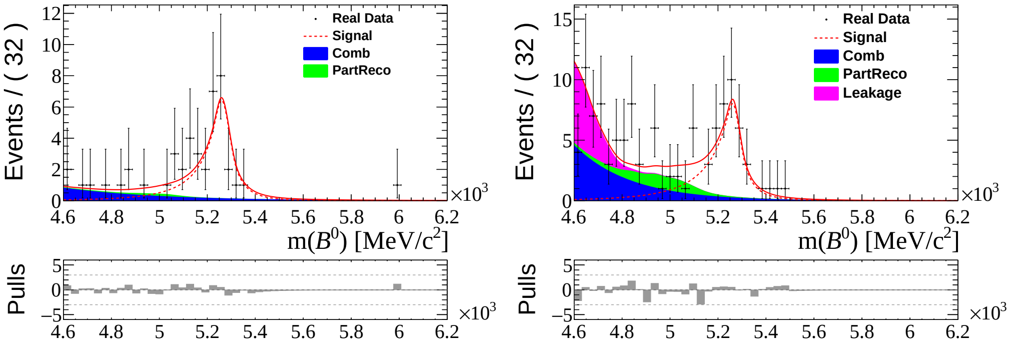


Figure E.39 – Fits to $B^0 \to K^{*0}e^+e^-$ in the low (left) and central (right) q^2. Run 1 `L0L` exclusive.

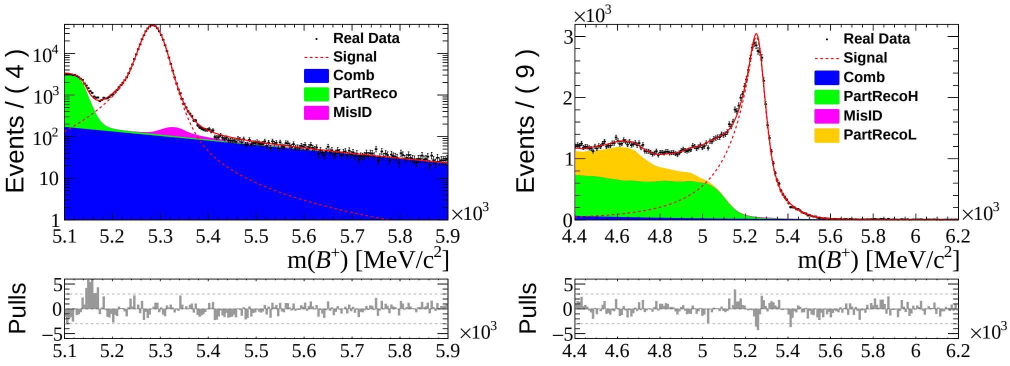


Figure E.40 – Fits to reconstructed $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (right). Run 1 `L0L` exclusive.

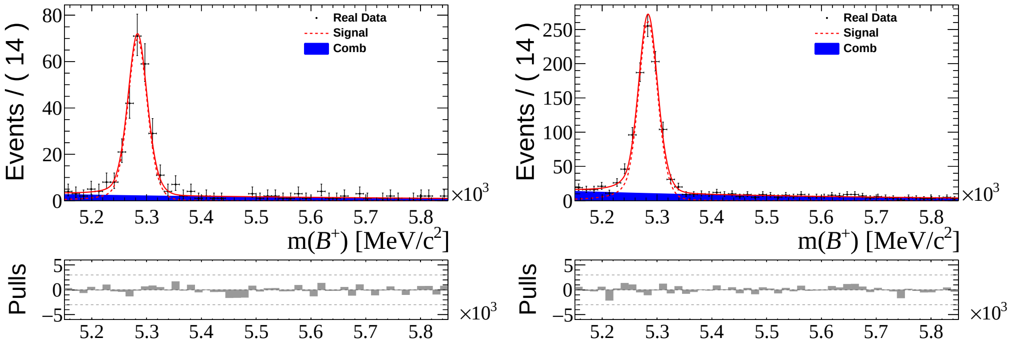


Figure E.41 – Fits to $B^+ \to K^+\mu^+\mu^-$ in the low (left) and central (right) q^2. Run 1 `L0L` exclusive.

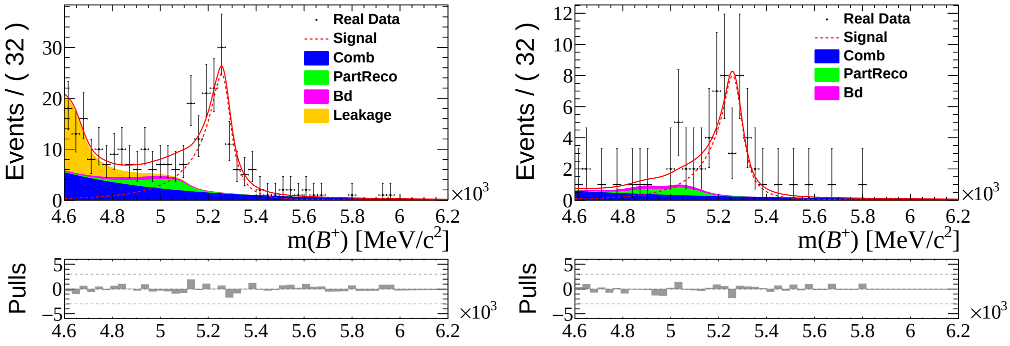


Figure E.42 – Fits to $B^+ \to K^+e^+e^-$ in the low (left) and central (right) q^2. Run 1 `L0L` exclusive.

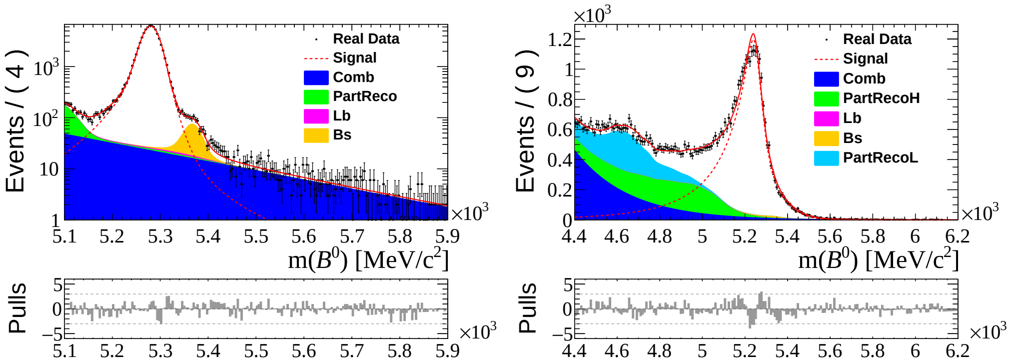


Figure E.43 – Fits to reconstructed $B^0 \to K^{*0}J/\psi(\to \mu^+\mu^-)$ (left) and $B^0 \to K^{*0}J/\psi(\to e^+e^-)$ (right). $15+16$ `LOI`.

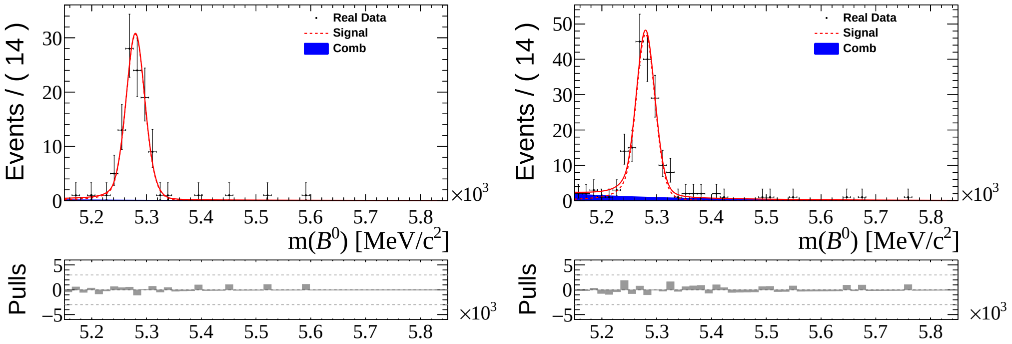


Figure E.44 – Fits to $B^0 \to K^{*0}\mu^+\mu^-$ in the low (left) and central (right) q^2. $15+16$ `LOI`.

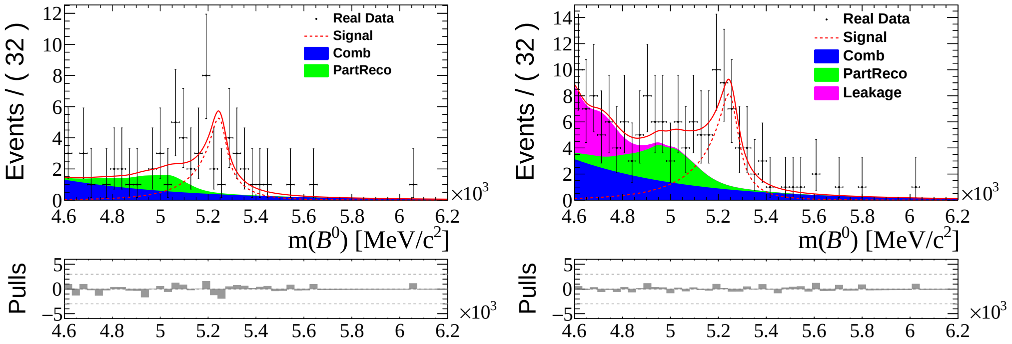


Figure E.45 – Fits to $B^0 \to K^{*0}e^+e^-$ in the low (left) and central (right) q^2. $15+16$ `LOI`.

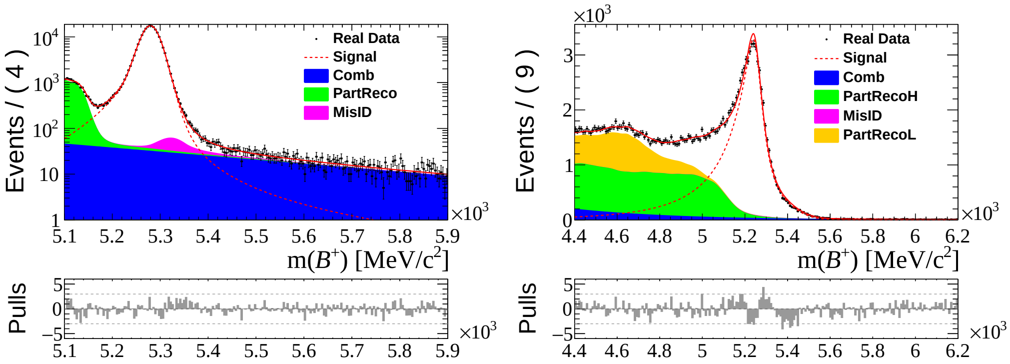


Figure E.46 – Fits to reconstructed $B^+ \to K^+ J/\psi\,(\to \mu^+\mu^-)$ (left) and $B^+ \to K^+ J/\psi\,(\to e^+e^-)$ (right). $15 + 16$ `LOI`.

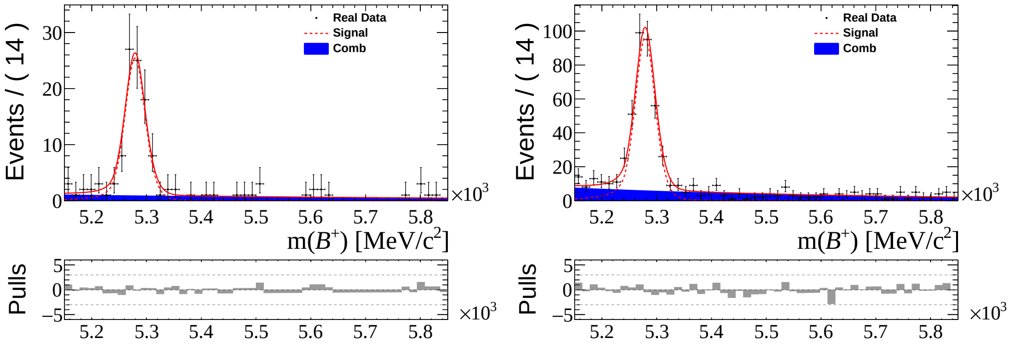


Figure E.47 – Fits to $B^+ \to K^+\mu^+\mu^-$ in the low (left) and central (right) q^2. $15 + 16$ `LOI`.

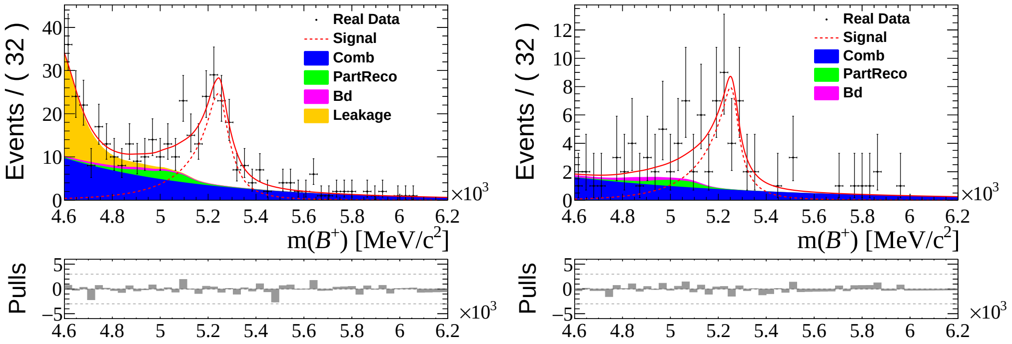


Figure E.48 – Fits to $B^+ \to K^+e^+e^-$ in the low (left) and central (right) q^2. $15+16$ `LOI`.

Appendix F

Cross Checks

F.1 Integrated $r_{J/\psi}$

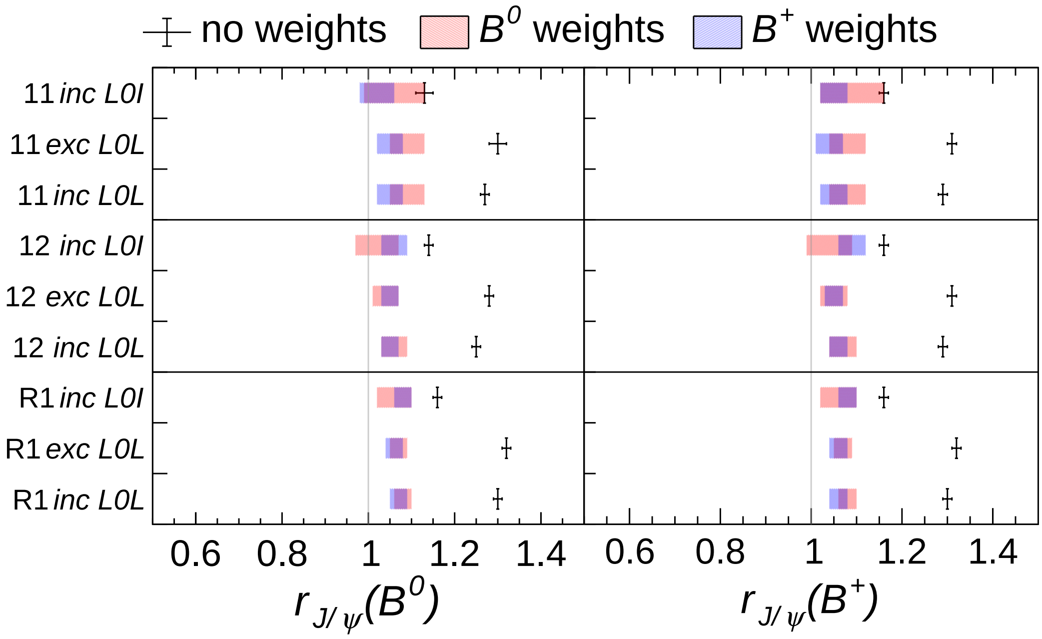


Figure F.1 – The results of integrated $r_{J/\psi}$ measurements, in the $K^{*0}\ell\ell$ (left) and $K^+\ell^+\ell^-$ (right) final states. The efficiencies which uses B^0 (red) and B^+ (blue) control mode as the calibration samples show a good compatibility with one another. The 2011 and 2012 datasets are combined into the Run 1 (R1) result.

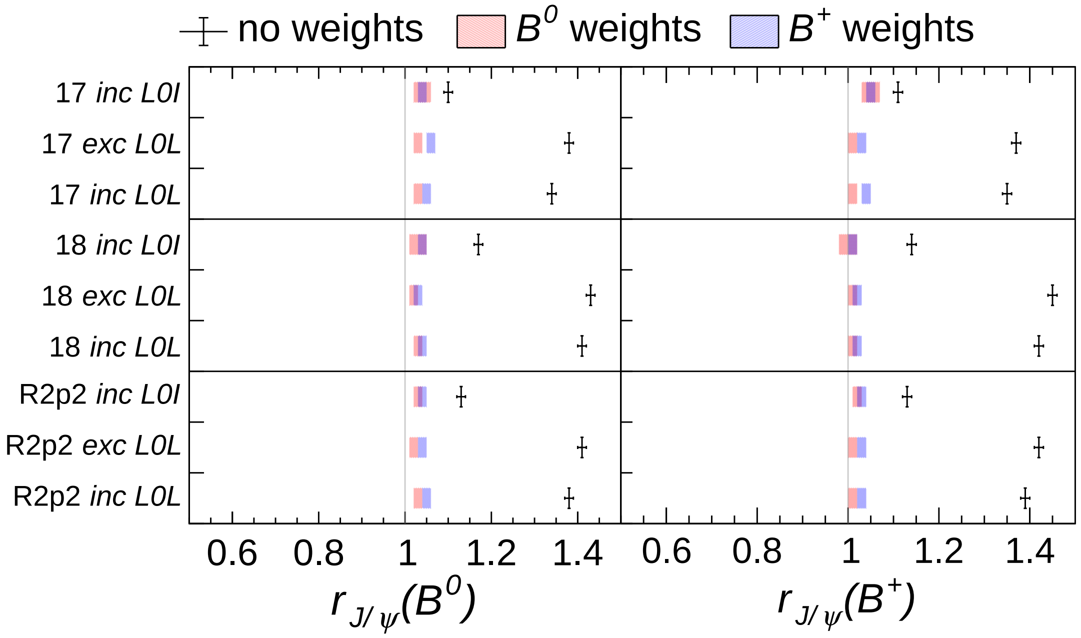


Figure F.2 – The results of integrated $r_{J/\psi}$ measurements, in the $K^{*0}\ell\ell$ (left) and $K^+\ell^+\ell^-$ (right) final states. The efficiencies which uses B^0 (red) and B^+ (blue) control mode as the calibration samples show a good compatibility with one another. The 2017 and 2018 datasets are combined into the Run 2 part 2 (R2p2) result.

F.2 Differential $r_{J/\psi}$

The flatness of $r_{J/\psi}$ is tested on the following variables:

- $p_{\mathrm{T}}(B)$,
- $p(B)$,
- $\eta(B)$,
- B FD χ^2,
- B `DIRA`,
- B VTX χ^2/ndf,
- $\tau(B)$,
- MVA response,
- $nTracks$,
- nSPDHits,
- $p_{\mathrm{T}}(J/\psi)$,
- $p(J/\psi)$,
- $\eta(J/\psi)$,
- J/ψ FD χ^2,

- min(η) $\ell^+\ell^-$,
- max(η) $\ell^+\ell^-$,
- min(IP χ^2) $\ell^+\ell^-$,
- $p_{\mathrm{T}}(K)$,
- $\eta(K)$,
- IP χ^2 (K).

F.3 $R_{\psi(2S)}$

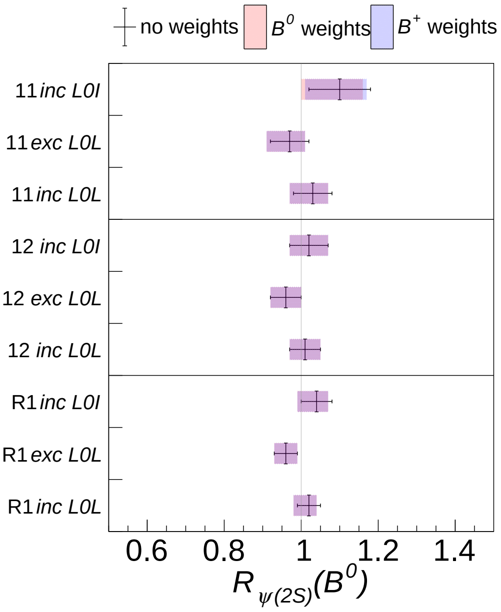


Figure F.3 – The results of integrated $R_{\psi(2S)}$ (B^0) measurements. The efficiencies which uses B^0 (red) and B^+ (blue) control mode as the calibration samples show a good compatibility with one another. The 2011 and 2012 datasets are combined into the Run 1 (R1) result.

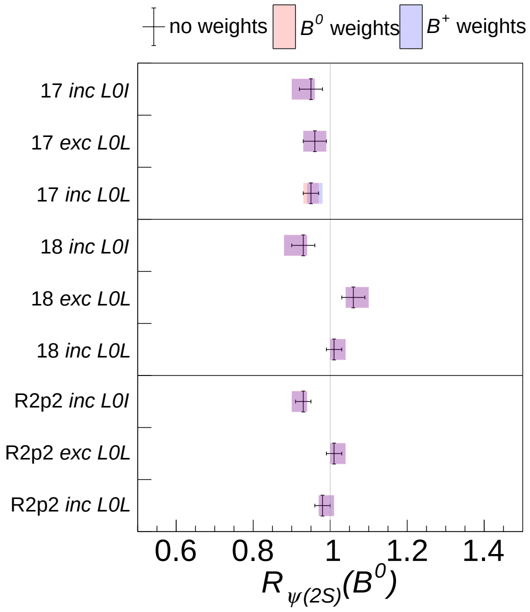


Figure F.4 – The results of integrated $R_{\psi(2S)}$ (B^0) measurements. The efficiencies which uses B^0 (red) and B^+ (blue) control mode as the calibration samples show a good compatibility with one another. The 2017 and 2018 datasets are combined into the Run 2 part 2 (R2p2) result.

Appendix G

List of figures

The list of figures provided to the author, courtesy of Ryan Bernard Calladine.

- 4.13
- 4.14
- 4.15
- 4.16