Search for lepton universality violation through $b \to s\ell^+\ell^-$ mediated rare decays at Belle II

A Thesis

Submitted to the

Tata Institute of Fundamental Research, Mumbai

for the degree of Doctor of Philosophy

in Subject Board of Physics

by
Soumen Halder

Department of High Energy Physics

Tata Institute of Fundamental Research

Mumbai

February 2024

Final Version Submitted in April 2024

Dedicated to my cherished ancestors and beloved family members, whose unwavering support and enduring legacy have guided my journey

DECLARATION

This thesis is a presentation of my original research work. Wherever contributions of other are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

The work was done under the guidance of Prof. Gagan Mohanty, at the Tata Institute of Fundamental Research, Mumbai.

Soumen Halder
16/04/2024
Soumen Halder

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Gagan Mohanty
Prof. Gagan Mohanty

Date: 16/4/2024

Acknowledgment

I am deeply grateful to Prof. Gagan Mohanty, my thesis guide, whose invaluable guidance, unwavering support, and insightful feedback have been instrumental throughout the research process. Prof. Mohanty's expertise and dedication played a pivotal role in shaping this thesis, and I am sincerely thankful for the enriching experience under his mentorship.

Heartfelt thanks are extended to Prof. Saurabh Sandilya, my senior collaborator, whose collaboration, valuable insights, and encouragement significantly enhanced the scope of my research. I appreciate the collaborative spirit that characterized our work together.

This thesis owes its existence to the continuous encouragement and support from Prof. Gagan Mohanty and Prof. Saurabh Sandilya. I am indebted to them for their unwavering guidance, encouragement, and mentorship.

Special appreciation goes to my junior, Mr. Rahul Tiwary, who served as a friend, philosopher, and guide. Our collaboration was invaluable, and his understanding of the subject greatly contributed to my work.

I extend my thanks to Prof. Elisa Manoni and Prof. Alexander Glazov, the Belle II EWP working group convenors, for their invaluable assistance and support in my analyses.

I would like to extend my deepest gratitude to Mr. Siddhartha Karmakar for his invaluable assistance throughout my research journey. His unwavering support and willingness to help have been instrumental in enhancing my understanding of various theoretical concepts. I am truly grateful for his friendship.

Acknowledgment is due to my academic elder brothers, Varghese Babu and Debashis Sahoo, for sharing their experiences. Engaging discussions with juniors, Mr. Sagar Hazra and Mr. Rishabh Mehta, enriched my perspective.

I extend my thanks to Belle II colleagues Dr. Niharika Rout, Dr. Lopamudra Nayak, Ms. Rashmi Dhamija, and Dr. Renu Garg for their helpful discussions and unwavering support.

Special gratitude is reserved for Ms. Medha Chakraborty, who inspired me to explore and appreciate one of the cultural heritages of Bharat, 'Hindustani classical music'.

Embarking on my journey as an integrated PhD student at TIFR in 2016, the initial 2.5 years of coursework were a mandatory phase. I extend my gratitude to my batchmates—Dr. Anupam Ray, Dr. Mursalin Islam, Dr. Rabisankar Samanta, Mr. Anirban Bhattacharjee, and Dr. Subhajit Sinha—who played a pivotal role in my academic growth. Engaging in enlightening discussions, particularly during assignment solving, their camaraderie made the coursework phase both intellectually stimulating and memorable.

I am grateful for the camaraderie and warmth extended by my hostel neighbors, Dr. Subhajit Ghosh, Dr. Uttiya Sarkar, and Mr. Sabir Ali. Despite being seniors, they treated me as a friend, and we shared beautiful moments, particularly during the challenging times of the COVID pandemic. Their friendship added a sense of community and support, making those moments not only memorable but also a source of strength.

My deep appreciation goes to my school friends, Dr. Bikramadiya Naskar and Mr. Tushar Sardar, who were consistently just a call away, demonstrating unwavering support despite their demanding responsibilities and life challenges.

Finally, heartfelt appreciation goes to my mother, father, and friend, Ms. Puja Sarkar, unsung heroes who provided constant support in coping with uncertainties and emotional challenges throughout this journey.

I express my gratitude to the members of TIFR and its department of High Energy Physics, who, knowingly and unknowingly, played a vital role in helping me complete this journey. While it's not possible to acknowledge everyone on this piece of paper, you are all woven into the fabric of this thesis.

Abstract

We report on the progress made at Belle II in probing lepton universality violation through rare B meson decays mediated by $b \to s\ell^+\ell^-$ transitions. Key measurements include branching fractions for $B^+ \to K^+\ell^+\ell^-$ decays, with ℓ representing an electron or a muon. Lepton universality, quantified by the observable R_K, is investigated across a broad range of the dilepton invariant-mass squared excluding charmonium resonance regions. Validation of lepton universality studies is performed through the measurement of absolute branching fractions for $B \to KJ/\psi(\ell^+\ell^-)$ decays, the lepton universality observable $R_K(J/\psi)$, and isospin asymmetries. The rediscovery of other decays involving $b \to s\ell^+\ell^-$ transitions is also discussed, particularly the branching fraction for $B \to K^*\ell^+\ell^-$ decays. Lastly, we discuss the performance of the silicon-strip vertex detector of Belle II during its commissioning phase.

Publications and public documents

1. Measurements of the branching fraction, isospin asymmetry, and lepton-universality ratio in $B \to J/\psi K$ decays at Belle II, F. Abudinén et al. (Belle II Collaboration), arXiv:2207.11275.

2. Measurement of the branching fraction for the decay $B \to K^*(892)\ell^+\ell^-$ at Belle II, F. Abudinén et al. (Belle II Collaboration), arXiv:2206.05946.

3. Study of $B^+ \to K^+\ell^+\ell^-$ at Belle II, F. Abudinén et al. (Belle II Collaboration), BELLE2-NOTE-PL-2021-005.

4. The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector, K. Adamczyk et al. (Belle II SVD Collaboration), JINST, **17** P11042 (2022).

Belle II Internal Note

1. Measurement of lepton flavour universality using $B \to K\ell^+\ell^-$ at Belle II. S. Halder, R. Tiwary, S. Sandilya, G. B. Mohanty [BELLE2-NOTE-PH-2022-068] (Review is ongoing)

2. Study of $B \to K\ell^+\ell^-$ decays at Belle II. S. Halder, R. Tiwary, S. Sandilya, G. B. Mohanty [BELLE2-NOTE-PH-2020-014]

3. Measurement of branching fraction, isospin asymmetry and $R_K(J/\psi)$ using the $B \to J/\psi K$ channel. S. Halder, R. Tiwary, S. Sandilya, G. B. Mohanty [BELLE2-NOTE-PH-2021-049]

4. Search for $B \to K^*\ell\ell$ decays, at Belle II. S. Choudhury, S. Halder, L. Nayak, S. Sandilya, G. B. Mohanty, A. Giri [BELLE2-NOTE-PH-2020-017]

5. Validation of the BDT-based lepton identification method. L. Nayak, S. Halder, G. B. Mohanty, A. Giri [BELLE2-NOTE-PH-2020-065]

6. Performance of Belle II Silicon Vertex Detector in Phase-2. S. Halder, G. B. Mohanty, A. Bozek, G. Casarosa [BELLE2-NOTE-TE-2019-026]

Conference Proceedings

1. Proceedings of the Conference on Flavour Physics and CP violation (FPCP) 2020 on "Results and prospects of radiative and electroweak penguin decays at Belle II", arXiv:2101.11573.
2. Proceedings of the 27th International Workshop on Vertex Detectors on "Spatial Resolution of the Belle II Silicon Vertex Detector" PoS (Vertex 2018) 054.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Studying beyond the Standard Model (BSM) is imperative for the advancement of our understanding of the fundamental constituents of the universe and the forces that govern their interactions. As such, the Standard Model (SM) of particle physics has been an incredibly successful theoretical framework for describing the behaviour of subatomic particles, incorporating the electromagnetic, weak, and strong nuclear forces. It has predicted the existence of various particles, such as the W and Z bosons, and culminated in the discovery of the Higgs boson in 2012, which confirmed the existence of an all-pervading Higgs field responsible for imparting mass to particles. However, this model is far from complete and leaves several profound questions unanswered.

One such question that motivates the study of BSM physics is the presence of dark matter and dark energy. These enigmatic substances constitute the majority of the universe's mass-energy content, yet they do not fit within the framework of the SM. Researchers seek to identify new particles or interactions that could explain the nature of dark matter and understand why dark energy causes the universe to expand at an accelerating rate. These discoveries could not only revolutionize our understanding of the cosmos but also have profound implications for our understanding of the universe's fate.

In addition to the fundamental issue mentioned earlier, there are two significant conceptual challenges within the SM that serve as compelling drivers for BSM research, particularly in the context of the Belle II experiment.

First, the SM grapples with the problem of mass hierarchy, which refers to the vast disparities in the masses of fundamental particles. This hierarchy is particularly evident in the masses of neutrinos compared to other particles, such as the top quark. While the SM successfully explains the masses of quarks and charged leptons through their Yukawa interactions with the Higgs field, it fails to account for the tiny masses of neutrinos. This intriguing discrepancy raises questions about the underlying mechanisms responsible for generating particle masses and hints at BSM physics. Some BSM theories, like the seesaw mechanism [10], suggest new particles and interactions that might help explain where neutrino masses come from and give us a better picture of the mass hierarchy as a whole.

Second, lepton universality, a fundamental tenet of the SM, postulates that the weak interactions for all three generations of charged leptons (electrons, muons, and taus) should be identical. But previous findings, mostly from the LHCb experiment, suggest that some rare decays of B mesons may not follow lepton universality. These observations challenge the integrity of the SM's predictions and underscore the need for further exploration into BSM theories. BSM models, like those with additional gauge bosons or leptoquarks, might be able to explain these findings, giving us new information about the symmetry and interactions that govern particle physics.

In the specific context of the Belle II experiment, the pursuit of BSM physics is intertwined with the aim of shedding light on these conceptual challenges. Belle II's precise measurements of rare decays and properties of B mesons provide fertile ground for probing BSM theories that address the mass hierarchy problem and lepton universality violations. By conducting meticulous experiments and analyzing the collected data, Belle II contributes significantly to the quest for new physics that extends beyond the limits of the SM, ultimately deepening our understanding of the universe's fundamental principles and phenomena.

This thesis is organized as follows: Chapter 2 introduces the motivation behind the lepton universality measurement within the SM framework. Chapter 3 provides an overview of the experimental setup, offering insights into the SuperKEKB accelerator and the Belle II detector. Chapters 4 to 6 detail the event reconstruction and background suppression techniques, including a step-by-step approach and the application of multivariate classifiers. In Chapter 7, the signal extraction procedure is elucidated, employing unbinned maximum-likelihood fitting. Chapter 8 presents a comprehensive study of control samples and the corresponding results. Chapter 9 focuses on the analysis of $B \to K\ell\ell$ events with a relatively smaller dataset. Chapter 10 delves into the intricate process of estimating systematic uncertainties. Chapter 13 offers the results and a discussion of the lepton universality measurement. Chapter 11 shifts the focus to the measurement of the branching fraction of $B \to K^*\ell\ell$, another rare decay process. Finally, Chapter 12 provides an in-depth examination of the SVD's performance during the commissioning phase.

Chapter 2

Standard Model and lepton universality

In this chapter, we delve into the intricacies of the SM, the bedrock of particle physics, and explore the theoretical motivations that drive the quest for precise measurements of lepton universality. We commence by discussing the core principles of the SM, shedding light on the fundamental forces and particles that constitute our universe.

As we venture deeper into our exploration, we introduce an essential theoretical tool known as the Operator Product Expansion (OPE). This mathematical framework plays a pivotal role in our approach to understand flavor physics and Effective Field Theory, enabling us to dissect complex phenomena in the realm of particle physics.

Furthermore, we review and analyse earlier measurements that have laid the foundation for our pursuit of lepton universality measurements. By studying the history of these measurements, we gain valuable insights into the progress and challenges of this scientific endeavour.

2.1 The Standard Model

2.1.1 Gauge symmetries and fields

The SM is a quantum field theory characterized by the invariance of its Lagrangian under a collection of local and global symmetries. The equations of motion are determined by the principle of least action. In this theoretical framework, fundamental particles are described as fields, each with a unique representation under the Lorentz symmetry, dictating their spin, as well as under various global symmetries defined by specific groups. These fields also transform consistently across spacetime, known as local or gauge symmetry. Gauge symmetry is of paramount importance, as it underpins the description of the fundamental forces and interactions among particles within the SM. Additionally, the Higgs field plays a crucial role in providing mass to particles through spontaneous symmetry breaking, completing the SM's theoretical structure.

In the SM, discrete symmetries, including parity (P), charge conjugation (C), and time reversal (T), play a fundamental role in understanding the behaviour of particles and their interactions. While P and C are not always conserved, their combination with T (CPT) is expected to be exact, providing essential constraints on the laws of physics. The study of discrete symmetries and their violations has led to crucial insights into the universe's matter-antimatter asymmetry and the behaviour of fundamental particles.

The gauge symmetry group of the SM is represented as

$$G_{\text{SM}} = SU(3)_C \times SU(2)_L \times U(1)_Y.$$

These three groups correspond to the three fundamental forces in the SM. The strong force is described by $SU(3)_C$, while specific combinations of $SU(2)_L$ and $U(1)_Y$ represent the weak and electromagnetic forces. The force carriers of these interactions are spin-one particles with adjoint representations under the respective gauge groups. For example, gluons mediate the strong interaction and are represented as octets under $SU(3)_C$, while being invariant under the other two gauge groups.

Matter particles in the SM have Lorentz group representations like $(0, \frac{1}{2})$ or $(\frac{1}{2}, 0)$, signifying their spin properties. These spin-half fields make up three fermion generations, with each generation comprising five representations under G_{SM}. These representations include

$$Q_{Li}(3,2)_{+\frac{1}{6}}, U_{Ri}(3,1)_{+\frac{2}{3}}, D_{Ri}(3,1)_{-\frac{1}{3}}, L_{Li}(1,2)_{-\frac{1}{2}}, E_{Ri}(1,1)_{-1},$$

where 'L/R' denotes left- or right-handed states, 'i' represents the three generations, and the numbers indicate their transformation properties under the gauge groups. As an example, $Q_{Li}(3,2)_{+\frac{1}{6}}$ are triplets under $SU(3)_C$, doublets under $SU(2)_L$, and carry the hypercharge $Y = +\frac{1}{6}$ There is a single scalar field, known as the Higgs field, which has representation

$$\phi(1,2)_{+\frac{1}{2}}.$$

As mentioned previously, the Lagrangian is invariant under all group transformations of the fields. The terms formed by these fields, which are also invariant under group transformations, contribute to the Lagrangian. We can categorize these terms in the Lagrangian into three parts:

$$\mathcal{L} = \mathcal{L}_{\text{Kin}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}.$$

The kinetic term $\mathcal{L}_{\text{Kin}}$ includes terms with derivatives. To maintain gauge invariance, the derivatives are replaced by their covariant counterparts

$$D^{\mu} = \partial^{\mu} + ig_s\, G^{\mu\,a}\, L^a + ig\, W^{\mu\,b}\, T^b + ig'\, B^{\mu}\, Y,$$

here $G^{\mu\,a}$ is octet gluon fields, $W^{\mu\,b}$ is triplet weak boson fields, and B^{μ} is singlet hypercharge boson field. L^a's are $SU(3)$ generators, T^b's are $SU(2)$ generators, Y are $U(1)_Y$ hypercharges and g's are the interaction strength. The terms containing only the derivatives of boson fields $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu$ are not relevant for the current discussion. The terms containing covariant derivatives of fields are shown below:

$$\mathcal{L}_{\text{Kin}} \supset \sum_{p=Q_{Li},U_{Ri},D_{Ri},L_{Li},E_{Li}} i\overline{\psi}_p \gamma_\mu D^\mu \psi_p + (D_\mu H)^\dagger (D_\mu H).$$

The Higgs term contains self interaction :

$$\mathcal{L}_{\text{Higgs}} = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.1}$$

Finally, the Yukawa terms contain interaction between the Higgs field and fermions

$$-\mathcal{L}_{\text{Yukawa}} = Y^E_{ij} \overline{L_{Li}}\, \phi\, E_{Rj} + Y^U_{ij}\, \overline{Q_{Li}}\, \tilde{\phi}\, U_{Rj} + Y^D_{ij}\, \overline{Q_{Li}}\, \phi\, D_{Rj} + \text{hermitian conjugate terms}, \tag{2.2}$$

here Y^E, Y^U and Y^D are three 3×3 matrices. So for any of the fields, be they matter fermions or gauge bosons, there are no mass terms given the symmetry structure. That's why the SM is called a chiral theory to begin with.

2.1.2 Spontaneous Symmetry Breaking (SSB)

In the SM Higgs potential, as expressed in Eq. 2.1, $\mu^2 < 0$, indicating that $\phi = 0$ is not the minimum of the potential. Instead, the vacuum state is degenerate, and choosing the physical vacuum state spontaneously breaks the symmetry. The scalar assumes the vacuum expectation value (VEV)

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix},$$

where $v = \sqrt{\frac{-\mu^2}{\lambda}}$ is a constant with a value of 246 GeV. This VEV leads to the spontaneous breaking of the gauge group, $G_{\mathrm{SM}} \to SU(3)_{\mathrm{C}} \times U(1)_{\mathrm{em}}$.

When introducing the Higgs VEV into the Lagrangian, the mass term appears for the three gauge bosons. Specifically, within the kinetic term $\mathcal{L}_{\mathrm{Kin}} \supset (D_\mu H)^\dagger (D_\mu H)$, three out of the four bosons from the $SU(2)_W \times U(1)_Y$ gauge group acquire mass, while the fourth one remains massless.

The field $|D_\mu H|^2$ can be expressed as

$$\begin{aligned} |D_\mu H|^2 &= \frac{g^2 v^2}{8}\left(\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2 + \left(\frac{g'}{g}B_\mu - W_\mu^3\right)^2\right) \\ &= \frac{1}{2}m_W^2 W_\mu^+ W_\mu^- + \frac{1}{2}m_Z^2 Z^\mu Z_\mu. \end{aligned}$$

This leads to the generation of mass terms for the $W^\pm$ and Z_μ bosons, with $m_W = \frac{1}{2}gv$ and $m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}$. The fields are redefined as follows: $W^\pm = W_\mu^1 \pm iW_\mu^2$, $Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}}$, and $A_\mu = \frac{g'B_\mu + gW_\mu^3}{\sqrt{g^2+g'^2}}$. While the Lagrangian contains mass terms for the $W^\pm$ and Z_μ fields, there is no mass term for the A_μ field. The associated conserved quantities for the corresponding gauge groups are referred to as charges. In this context, the remaining gauge symmetry, other than color, is the electromagnetic field, and the charge for this symmetry is known as electric charge, defined as $Q = T^3 + Y$, where T^3 represents the third generator of $SU(2)$. As a result, the final gauge group becomes $SU(3)_C \times U(1)_{\mathrm{em}}$, leading to the existence of two distinct charges: color and electromagnetic charge.

Let's decompose $SU(2)_W$ doublet before SSB of Q_{Li} and L_{Li} as

$$Q_{Li} = \begin{pmatrix} U_{Li} \\ D_{Li} \end{pmatrix}, \qquad L_{Li} = \begin{pmatrix} \nu_{Li} \\ E_{Li} \end{pmatrix}.$$

When we plug in the expression of Higgs VEV into $\mathcal{L}_{\text{Yukawa}}$, we get

$$\mathcal{L}_{\text{Yukawa}} = (M_d)_{ij}\overline{D_{Li}}D_{Rj} + (M_u)_{ij}\overline{U_{Li}}U_{Rj} + (M_e)_{ij}\overline{E_{Li}}E_{Rj},$$

where $M_x = \frac{v}{\sqrt{2}}Y_x$ with $x \in \{U, D, E\}$. It is important to note that there are no mass terms corresponding to ν_{Li}, which is made by construction to keep neutrinos massless. There are several ways to introduce mass and one possibility is introducing a ν_{Ri} spinor, but this is beyond the scope of this thesis. Here the M_x are not necessarily diagonal, and we can change the basis to make them diagonal by a biunitary transformation. For any matrix, there exist two unitary matrices such that it can be diagonalised by

$$V_{xL}M_xV_{xR}^{\dagger} = M_x^{\text{diag}},$$

where $M_U^{\text{diag}} = \text{diag}(m_u^2, m_c^2, m_t^2)$, $M_D^{\text{diag}} = \text{diag}(m_d^2, m_s^2, m_b^2)$, $M_E^{\text{diag}} = \text{diag}(m_e^2, m_\mu^2, m_\tau^2)$. So we have recognized a few SM parameters as masses of all the fundamental fermions in the SM.

We can transform the flavor basis of fields of quarks and charged leptons to the mass eigenstates basis by transforming,

$$x_{Li} = (V_{xL})_{ij}x_{Lj} \qquad x_{Ri} = (V_{xR})_{ij}x_{Rj}, \qquad \text{with } x \in \{U, D, E\}.$$

After transitioning to the mass basis, the charged current interactions take on the following form:

$$-\mathcal{L}_{\text{CC}} \supset \frac{g}{\sqrt{2}}\left(\overline{U_{Li}}\gamma^\mu(V_{\text{CKM}})_{ij}D_{Lj}W_\mu^+\right) + \frac{g}{\sqrt{2}}\left(\overline{\nu_{Li}}\gamma^\mu E_{Lj}W_\mu^+\right) + \text{h.c.} \tag{2.3}$$

Here, $V_{\text{CKM}} = V_{uL}V_{dL}^{\dagger}$ represents the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix for quarks. Notably, this matrix is non-diagonal, causing the $W^{\pm}$ bosons to couple with different quark generations. Another noteworthy observation is that the weak current breaks parity, primarily because ψ_R does not contribute to the current.

The CKM matrix, denoted by V_{CKM}, defines the relative coupling strengths for charged current interactions among different combinations of up- and down-type quarks. Its elements are arranged as follows

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2.4}$$

The strength of these matrix elements can be easily visualized using the Wolfenstein parametrization, which is an expansion in terms of a small parameter $\lambda = |V_{us}| \approx 0.22$. Up to $O(\lambda^3)$, the CKM matrix in this parametrization can be expressed as:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}. \tag{2.5}$$

The Wolfenstein parametrization illustrates that transitions within the same generation, such as $u \to d$, $c \to s$, and $t \to b$, experience no CKM suppression. Conversely, transitions between the first and second, second and third, and first and third generations are subject to CKM suppressions of order λ, λ^2, and λ^3, respectively. These suppressions arise from the mixing between different quark generations.

Exploiting the unitarity of the CKM matrix, we get six equations, out of which Eq. 2.6 is relevant for the $b \to s\ell^+\ell^-$ diagrams (Fig.2.1).

$$V_{ub}V_{us}^* + V_{cb}V_{cs}^* + V_{tb}V_{ts}^* = 0 \tag{2.6}$$

The electromagnetic current, mediated by the photon (A_μ), is responsible for electromagnetic interactions. Since the neutrino is electrically neutral, it does not contribute to the corresponding electromagnetic Lagrangian ($\mathcal{L}_{\text{em}}$). In electromagnetic interactions, both right- and left-handed spinors contribute, leading to the conservation of parity, in contrast to the weak interaction described earlier.

The electromagnetic Lagrangian can be expressed as follows:

$$-\mathcal{L}_{\text{em}} = \frac{gg'}{\sqrt{g^2+g'^2}} A_\mu Q_i \sum_{\psi\in\{U,D,E\}} (\overline{\psi_{Li}}\gamma^\mu\psi_{Li} + \overline{\psi_{Ri}}\gamma^\mu\psi_{Ri}) + \text{h.c.} \tag{2.7}$$

$$= eA_\mu J^{\text{EM}}_\mu, \tag{2.8}$$

where $e = Q\frac{gg'}{\sqrt{g^2+g'^2}}$ represents the elementary charge, and J^{EM}_μ is the electromagnetic current.

Finally, the neutral current mediated by the Z boson is given by

$$-\mathcal{L}_{\text{NC}} = \sqrt{g^2+g'^2} Z_\mu \left(\sum_{\psi\in U,D,E} \overline{\psi_{Li}}\gamma^\mu T^3 \psi_{Li} - \frac{g'^2}{g^2+g'^2} J^{\mu\text{EM}} \right) + \text{h.c.} \tag{2.9}$$

The neutral current in the SM does not connect to different generations, unlike the charged current, because of the non-diagonal nature of V_{CKM}. This is a very important feature of the SM, which is exploited to probe BSM physics.

2.1.3 Accidental Symmetries

Accidental symmetries in the SM are symmetries that arise as approximate or unintended outcomes of the model's structure, despite not being exact symmetries of the underlying theory. Unlike gauge invariance, which is the outcome of a fundamental symmetry of the theory, the accidental symmetries emerge as byproducts of gauge invariance and renormalizability constraints. These symmetries often manifest as global symmetries, particularly in situations where non-renormalizable terms are neglected, and they are aptly named 'accidental symmetries'. They play a role in simplifying the SM and making it easier to study.

The SM has an accidental symmetry:

$$U(1)_B \times U(1)_e \times U(1)_\mu \times U(1)_\tau,$$

where the $U(1)_B$ is the baryon number and other three $U(1)$ are lepton family numbers. To ensure proper normalization, a common practice is to set the proton's baryon number (B) to 1. Consequently, each quark within the proton is attributed one-third of a unit of baryon number. In the SM, it is essential to note that each lepton family carries its own distinct lepton number: L_e for electrons, L_μ for muons, and L_τ for taus.

One more approximate accidental global symmetry is lepton universality. The fundamental interaction for three charged leptons in the SM is the same except for the Yukawa interaction they have with the Higgs boson, which is responsible for the difference in their masses. Since leptons are colorless and possess unit electric charge, lepton universality boils down to the principle that weak interactions, mediated by the $W^{\pm}$ and Z^0 bosons, treat all three generations of charged leptons (electron, muon, and tau) in an identical manner.

Importantly, this symmetry does not derive from any fundamental axioms of the theory, such as gauge symmetries. Consequently, the prediction of lepton universality requires empirical verification, and it remains a subject of rigorous testing in the field of particle physics. Deviations from lepton universality in experimental measurements could provide valuable insights into potential extensions of the SM.

2.1.4 FCNCs

In the context of the SM, flavor-changing phenomena predominantly manifest in charged-current weak interactions mediated by $W^{\pm}$ bosons. These interactions enable transitions between different generations of elementary particles. However, in the case of neutral-current weak interactions, which involve the exchange of Z^0 bosons, as well as interactions mediated by gluons (carriers of strong force) and photons (representing electromagnetic force), the SM does not allow a change in flavor at tree level. Consequently, the occurrence of flavor-changing neutral current (FCNC) processes is not allowed at tree level within the SM.

The $b \to s$ quark-level FCNC transition is possible via quantum loops, and the Feynman diagram of the lowest order loop that allows this process is shown in Fig. 2.1. The loop corrections introduce an additional constraint, approximately the order of coupling$^2/(16\pi^2)$, compared to tree-level diagrams for the same process. While loop corrections typically have minimal impact on tree-level processes, decays solely arising from loop-level transitions can exhibit significant deviations from the SM expectation. In such cases, even minor deviations from the SM can result in noticeable effects due to the absence of tree-level contributions. Consequently, measurements of $b \to s\ell^+\ell^-$ transitions, despite being suppressed, have high sensitivity. This emphasizes the significance of FCNC processes in providing valuable insights into the potential existence of BSM physics.

In the Feynman diagram shown in Fig. 2.1, all three up-type quark propagators are permitted, and their dependence on the CKM matrix elements is expressed as follows:

$$\mathcal{M} \propto \sum_{i=u,c,t} V_{ib}V_{is}^{*}. \tag{2.10}$$

However, Eq. 2.6 dictates that any part of the diagram that relies solely on flavor through the term $\sum_{i=u,c,t} V_{ib}V_{is}^*$ will vanish. The only factor that prevents this vanishing is the mass of the quark itself, and the dependency is given by:

$$\mathcal{M} \propto \sum_{i=u,c,t} V_{ib}V_{is}^* f(m_i).$$

The function f is known as the Inami-Lim function [11], which can be expanded in a power series. For $m_i \ll m_W$, particularly when $i = u$ or c, we have $f(m_i \ll m_W) \propto \frac{m_i^2}{m_W^2}$. Conversely, for $i = t$, the assumption $m_i \ll m_W$ doesn't hold, and it remains constant. In the context of this specific diagram, the contribution from the top quark dominates.

Nevertheless, the FCNC contribution in the loop is very small, and the process mediated by such quark-level transition, like $B \to K^{(*)}\ell^+\ell^-$ where ℓ stands for an electron or a muon, are very rare. The decay branching fractions ($\mathcal{B}$) are of the order of 10^{-6}.

2.1.5 Lepton universality

As discussed earlier, lepton flavor universality (LFU) implies that the coupling of gauge bosons is the same for all lepton generations. A lepton universality ratio, defined as

$$R_K \equiv R_K(q^2_{\mathrm{min}}, q^2_{\mathrm{max}}) = \frac{\int_{q^2_{\mathrm{min}}}^{q^2_{\mathrm{max}}} \frac{d\Gamma(B^+\to K^+\mu^+\mu^-)}{dq^2} dq^2}{\int_{q^2_{\mathrm{min}}}^{q^2_{\mathrm{max}}} \frac{d\Gamma(B^+\to K^+e^+e^-)}{dq^2} dq^2}, \tag{2.11}$$

is introduced to test this unique but accidental symmetry of the SM in $B^+ \to K^+\ell^+\ell^-$ decays. Here Γ is the partial decay width for the channel within parentheses and q^2 is the dilepton invariant-mass squared. The expected value of R_K is unity, up to a small phase-space difference between the electron and muon channels. The effects of Yukawa couplings on rare B meson decays are predicted to be negligible, except near the kinematic threshold $4m_\ell^2$ [12]. The SM predicts $R_K = 1 \pm \mathcal{O}(10^{-3}) \pm \mathcal{O}(10^{-2})$ [13], where uncertainties account for phase-space effects and electroweak radiative corrections (soft-collinear photon emission [14] from the leptons), respectively. Deviations from unity in the LFU ratio would indicate BSM physics.

2.2 Challenges of the Standard Model

The SM has proven to be an incredibly successful framework for comprehending fundamental constituents of the universe and their interactions. Despite its remarkable predictive capabilities, the SM has several limitations and it cannot explain a number of phenomena, which provide motivation for the exploration of BSM physics.

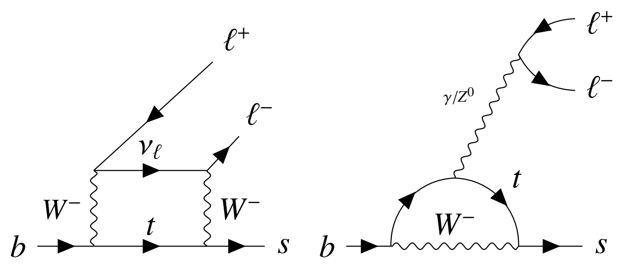


Figure 2.1: Feynman diagrams representing the $b \to s\ell^+\ell^-$ process, showcasing the box diagram on the left and the penguin diagram on the right.

- Dark Matter: One of the most profound mysteries in contemporary physics is the enigma of dark matter. Observations at various scales, from galactic rotation curves to the cosmic microwave background radiation, provide compelling evidence for the existence of an invisible, non-luminous form of matter that vastly surpasses ordinary matter in the universe. The SM does not provide a viable candidate for dark matter, leaving the identification of dark matter particles a central challenge. BSM theories, such as supersymmetry or axion models, offer potential solutions to this puzzle by introducing new particles that could be candidate for dark matter [15].

- Matter-Antimatter Asymmetry: The predominance of matter over antimatter in the universe $\left[O(10^{-9})\right]$ is a fundamental question in cosmology and particle physics. The SM predicts the presence of CP violation, a phenomenon necessary to explain the matter-antimatter asymmetry, but the observed level of CP violation $\left[O(10^{-20})\right]$ is insufficient to account for the matter we observe. BSM theories, such as those incorporating additional sources of CP violation or new physics in the early universe, offer insights into the origin of this asymmetry [16].

- Neutrino Oscillations: Neutrino oscillations, the phenomenon where neutrinos change between different flavor eigenstates as they propagate through the matter, are an established experimental fact. This phenomenon implies that neutrinos possess nonzero mass, which contradicts the assumption of their masslessness in the SM. The existence of neutrino masses suggests the need for BSM physics to explain their origin and the underlying neutrino mixing [17, 18].

- Hierarchy Problem: The hierarchy problem relates to the significant difference between the strength of gravitational and weak forces, as well as the apparent fine-tuning required for the Higgs boson mass. The SM predicts the Higgs boson's mass to be much larger than that observed in nature, necessitating precise adjustments to account for the observed Higgs mass. BSM theories, including supersymmetry and extra dimensions, offer potential resolutions to the hierarchy problem by introducing new symmetries and particles that stabilize the Higgs boson mass [19].

Though the SM has been remarkably successful in describing the behavior of subatomic particles and their interactions, its unresolved challenges and limitations, such as dark matter, matter-antimatter asymmetry, inclusion of gravity, neutrino oscillations, and hierarchy problem, drive the need for BSM physics. The pursuit of latter represents a compelling and essential avenue for expanding our understanding of fundamental constituents of the universe and underlying principles that govern their behavior.

2.3 The issue with FCNCs

In the SM, flavor-changing charged currents are mediated by the $W^{\pm}$ bosons, whether involving quarks or leptons. On the other hand, FCNC refers to processes in which a quark undergoes a flavor transition without altering its electric charge, as discussed in Section 2.1.4.

While FCNC processes are experimentally observed, they are inherently very rare, which is theoretically explained by the unitarity of the CKM matrix. A priori, there is no reason for the absence of FCNC processes at tree level. In fact, one intriguing possibility is that the CKM matrix could be a component of a larger matrix, incorporating quark-lepton interactions, thereby introducing additional sources of FCNCs. This extension could offer insights into the rarity of FCNC processes while potentially providing a framework to explain why the electric charges of protons and electrons are equal. Moreover, the observation of neutrino oscillations demonstrates that flavor is not conserved, suggesting a BSM flavor structure. It's worth noting that the values of the CKM matrix are input parameters in the SM, measured through experiments, rather than being derived from fundamental principles. These challenges motivate the search for flavor symmetries and deeper motivations for flavor conservation.

2.4 The Effective Hamiltonian approach

The rare B decay is governed by the exchange of a weak boson whose mass is significantly higher than that of the B meson. This property allows us to construct an effective field theory that distinguishes between low- and high-frequency modes. Since our primary interest lies in the low-energy regime ($\mu \sim m_b$), which is considerably smaller than the electroweak scale ($\mu \ll M_W, M_Z$), the high-energy modes become negligible, and our focus remains solely on the low-energy or long-distance operators.

The weak decay amplitude can be separated into two distinct components: the "short-distance" and "long-distance" contributions, with a threshold scale (μ) serving as the demarcation point. The "short-distance" effects account for perturbative contributions for energies above $\mu > m_B$ and are characterized by Wilson coefficients.

On the other hand, the "long-distance" effects involve non-perturbative contributions, often associated with low-energy interactions and hadronic physics. This approach allows us to systematically address both the short- and long-distance effects in rare B decays, facilitating a separation of calculable perturbative contributions from the more intricate non-perturbative aspects of these processes.

The effective Hamiltonian for the $b \to s\ell^+\ell^-$ process is given by [20]:

$$\mathcal{H}_{\text{eff}}^{\text{SM}}(b \to s\ell^+\ell^-) = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_{i=7,9,10} C_i(\mu)O_i(\mu). \tag{2.12}$$

Here, the OPE allows us to distinguish "long-distance" effects from "short-distance" contributions. The C_i's are scale-dependent couplings, namely Wilson coefficients (WCs), associated with the four-leg vertices described by local operators O_i's that govern the interactions. The complete basis consists of ten operators O_i with $i = 1, 2, \ldots, 10$. Among them, $O_{1,2}$ represent tree-level W operators, $O_{3-6,8}$ involve penguin diagrams mediated by gluons, and $O_{7,9,10}$ are the relevant operators for the $b \to s\ell^+\ell^-$ process. Specifically, the SM-relevant operators for $b \to s\ell^+\ell^-$ are as follows:

$$O_7 = \frac{e}{(4\pi)^2} m_b \left(\bar{s}\sigma_{\mu\nu} P_R b\right) F^{\mu\nu} \tag{2.13}$$

$$O_9 = \frac{e}{(4\pi)} m_b \left(\bar{s}\gamma_\mu P_L b\right)\left(\bar{\ell}\gamma^\mu \ell\right) \tag{2.14}$$

$$O_{10} = \frac{e}{(4\pi)} m_b \left(\bar{s}\gamma_\mu P_L b\right)\left(\bar{\ell}\gamma^\mu \gamma_5 \ell\right). \tag{2.15}$$

Here, $\sigma_{\mu\nu} = \frac{i}{2}\left[\gamma_\mu, \gamma_\nu\right]$, e represents the QED coupling constant, and $P_{L/R} = \frac{1\mp\gamma_5}{2}$.

To ensure an accurate description of SM processes, the effective theory is harmonized with the SM by requiring each term in the effective theory to match the corresponding term in the full theoretical calculation at a specific matching scale, typically set at the electroweak scale (μ_W). Subsequently, leveraging the scale independence of the effective Hamiltonian, a renormalization group equation for the WC is derived [21]. By considering only SM contributions and setting $\mu_W = m_b$, specific values for the WCs are obtained [22]:

$$C_7 = -0.304 \qquad C_9 = 4.211 \qquad C_{10} = -4.103.$$

The decay rate for $b \to s\ell^+\ell^-$ can be then expressed as:

$$\Gamma(b \to s\ell^+\ell^-) = \frac{G_F^2 M_W^2 m_B^3 f_B^2}{8\pi^5}|V_{tb}V_{ts}^*|^2 \frac{4m_\ell^2}{m_B^2}\sqrt{1-\frac{4m_\ell^2}{m_B^2}}|C_{9,10}|^2 + \text{higher-order corrections.} \quad (2.16)$$

In this equation, f_B represents the B decay constant. A typical spectrum of $b \to s\ell^+\ell^-$ decays is depicted in Fig. 2.2. The variable q^2 spans a range from $4m_\ell^2$ to $(m_B - m_K)^2$.

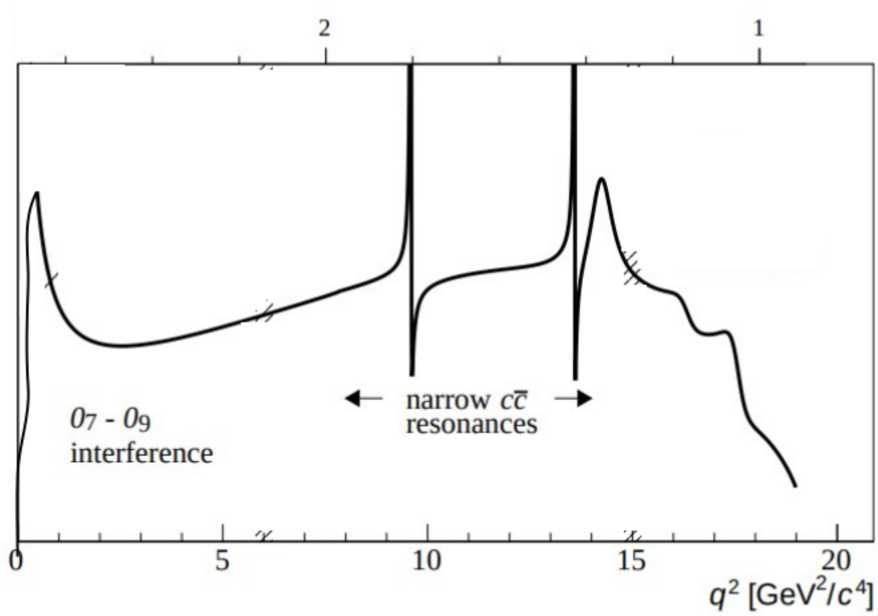


Figure 2.2: A typical q^2 spectrum of $b \to s\ell^+\ell^-$ process. The figure is taken from Ref. [1].

In the lowest q^2 region, namely $q^2 \lesssim 1\,\mathrm{GeV}^2/\mathrm{c}^4$, the process involves the emission of a slightly off-shell photon decaying into two charged leptons, with the primary contributing operator being the QED one, denoted as O_7.

Within the q^2 range of $[1, 6]\,\mathrm{GeV}^2/\mathrm{c}^4$, the dominant contribution arises from the interference between O_7 and O_9. Notably, this region is susceptible to the influence of new physics effects through the Wilson coefficients C_7 and C_9.

As we move into the q^2 range of $[7, 14]\,\mathrm{GeV}^2/\mathrm{c}^4$, the energy scale of hadrons involved in the process becomes comparable to the typical energy scale of quantum chromodynamics (Λ_{QCD}), which is approximately 1 GeV. The theoretical calculations for this region are carried out by approximating the bottom quark mass (m_b) as infinite and employing the Heavy Quark Effective Theory (HQET) [23], as a useful tool for simplifying the heavy quark dynamics.

2.4.1 New physics contributions

To address the earlier R_K anomaly and incorporate BSM physics, new operators are introduced alongside the SM ones. These new operators allow for minimal flavor violation, introducing chirality-flipped operators as follows:

$$O'_7 = \frac{e}{(4\pi)^2} m_b \left(\bar{s}\sigma_{\mu\nu} P_L b\right) F^{\mu\nu} \tag{2.17}$$

$$O'_9 = \frac{e}{(4\pi)} m_b \left(\bar{s}\gamma_\mu P_R b\right)\left(\bar{\ell}\gamma^\mu \ell\right) \tag{2.18}$$

$$O'_{10} = \frac{e}{(4\pi)} m_b \left(\bar{s}\gamma_\mu P_R b\right)\left(\bar{\ell}\gamma^\mu \gamma_5 \ell\right). \tag{2.19}$$

Additionally, scalar and pseudoscalar operators are introduced:

$$O^{(\prime)}_S = \frac{e}{(4\pi)} m_b \left(\bar{s}\gamma_\mu P_{R(L)} b\right)\left(\bar{\ell}\ell\right) \tag{2.20}$$

$$O^{(\prime)}_P = \frac{e}{(4\pi)} m_b \left(\bar{s}\gamma_\mu P_{R(L)} b\right)\left(\bar{\ell}\gamma_5 \ell\right). \tag{2.21}$$

The effective Hamiltonian incorporating both SM and BSM contributions is given by:

$$\mathcal{H}_{\mathrm{eff}}(b \to s\ell^+\ell^-) = -\frac{4G_F}{\sqrt{2}} V_{tb} V^*_{ts} \sum_{i=7,9,10,S,P} \left[C_i(\mu) O_i(\mu) + C'_i(\mu) O'_i(\mu)\right]. \tag{2.22}$$

In this context, the WCs are allowed to have imaginary components, unlike the SM case ($\mathcal{H}^{\mathrm{SM}}_{\mathrm{eff}}$) where WCs are real. Also, $C^{(\prime)}_i = C^{(\prime)}_i + C^{(\prime)\mathrm{NP}}_i$ to incorporate new physics contributions[24]. The $\mathcal{H}^{\mathrm{SM}}_{\mathrm{eff}}$ is a special case of $\mathcal{H}_{\mathrm{eff}}$ where only C_7, C_9, and C_{10} are nonzero.

Furthermore, in contrast to the SM, the LFU constraints on $C_{7,9,10}$ may not hold in BSM scenarios. Therefore, the corresponding $C'_{7,9,10}$ coefficients are labeled with the specific lepton flavor, denoted as $C'^{\mu}_{7,9,10}$ and $C'^{e}_{7,9,10}$.

The R_K observables in terms of these new parameters can be approximated as

$$R_K \approx \frac{\left|C_{10}^{\mathrm{SM}} + C_{10}^{\mu\mathrm{NP}} + C'^{\mu}_{10}\right|^2 + \left|C_9^{\mathrm{SM}} + C_9^{\mu\mathrm{NP}} + C'^{\mu}_9\right|^2}{\left|C_{10}^{\mathrm{SM}} + C_{10}^{e\mathrm{NP}} + C'^{e}_{10}\right|^2 + \left|C_9^{\mathrm{SM}} + C_9^{e\mathrm{NP}} + C'^{e}_9\right|^2} \text{ [25].}$$

Previous tensions [26, 27] of LHCb measurements with SM predictions have motivated the proposal of BSM scenarios, such as leptoquarks and Z' bosons (Feynman diagram shown in Fig. 2.3). The additional parameters, or the WCs that are zero in the SM, provide extra flexibility for phenomenological analyses of these new models and yield theoretical constraints on these WCs.

The leptoquark model postulates the existence of new particles, leptoquarks [28], which carry both lepton and quark quantum numbers, allowing them to mediate interactions between quarks and leptons. These models are discussed in Refs. [25, 29, 30]. The Z' model, an extension of the SM, introduces an additional neutral gauge boson, offering a theoretical avenue to explain BSM phenomena [31, 31, 32, 33, 34, 35, 36]. The latest result on R_K [37] agrees with the SM, still further cross-validation is an essential motivating factor for our measurements at Belle II.

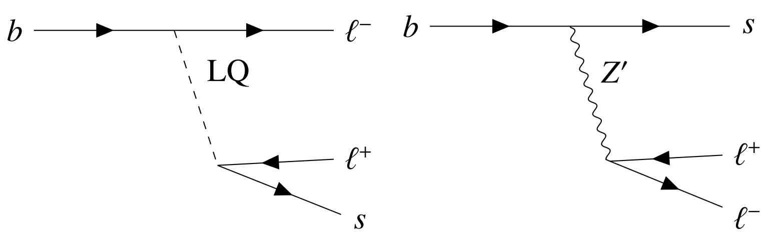


Figure 2.3: Examples of Feynman diagrams for the $b \to s\ell^+\ell^-$ process in the BSM domain. The left diagram shows the contribution from a leptoquark, while the right one represents the contribution from a Z' boson.

2.5 Experimental status of $B \to K\ell^+\ell^-$ observables

Observables are the measured quantities that play a crucial role in the verification and validation of theoretical models in particle physics. In this section, we present earlier measurements of the branching fraction and LFU ratio of $B \to K\ell\ell$ decay.

We also discuss observables that can be measured in the $B \to J/\psi[\ell\ell]K$ decay, which is a tree-level process with a branching fraction approximately two orders of magnitude larger than that of $B \to K\ell^+\ell^-$ and has the same final state as the latter. Therefore, we expect a highly precise result for the LFU ratio in $B \to J/\psi K$. Lepton universality hasn't been found to be violated in any reported tree-level processes [38], so we can check the LFU ratio in $B \to J/\psi[\ell\ell]K$ to make sure it is consistent with one. This would be a significant step toward our goal of measuring R_K.

2.5.1 Branching fraction

The branching fraction of a B meson decay channel is the fraction of the time it decays to that particular channel. Mathematically, the branching fraction can be expressed as

$$\mathcal{B} = \frac{N_{\rm sig}}{N_{B\overline{B}} \times 2 \times f^{\pm} \times \epsilon}, \tag{2.23}$$

where:

- $N_{\rm sig}$ is the number of reconstructed signal, generally estimated from a fit to the distribution of some variable.
- $N_{B\overline{B}}$: Number of $B\overline{B}$ events; for the pre-LS1 dataset we have $N_{B\bar{B}} = (387 \pm 6) \times 10^6$
- Signal efficiency (ϵ) refers to the fraction of generated signal events accurately reconstructed within a simulated sample, corrected for differences stemming from various sources, such as lepton and kaon identification, tracking, etc. These corrections account for imperfections in our simulation process.
- $f^{\pm} = \mathcal{B}(\Upsilon(4S) \to B^+B^-)/\mathcal{B}(\Upsilon(4S) \to \text{all}) = 0.519 \pm 0.029$ (described in Section 7.6)

The branching fraction of $B \to K\ell\ell$ has been reported by both Belle and BaBar experiments, with Belle's result obtained from a sample of $N_{B\overline{B}} = 772 \times 10^6$ events [39], and BaBar's result obtained from a sample of $N_{B\overline{B}} = 471 \times 10^6$ events, where the measurement is lepton-flavor inclusive [40]. The BaBar measurement is as follows:

- $\mathcal{B}[B^+ \to K^+\ell^+\ell^-] = [4.7 \pm 0.6\,(\text{stat.}) \pm 0.2\,(\text{syst.})] \times 10^{-7}$,

Belle measured lepton-flavor and charge exclusive branching fractions and these measurements have an uncertainty comparable with that of BaBar. The Belle measurement is as follows:

$$\mathcal{B}[B^+ \to K^+e^+e^-] = \left[5.75^{+0.64}_{-0.61}\ (\text{stat.}) \pm 0.15\ (\text{syst.})\right] \times 10^{-7}$$
$$\mathcal{B}[B^+ \to K^+\mu^+\mu^-] = \left[6.24^{+0.65}_{-0.61}\ (\text{stat.}) \pm 0.16\ (\text{syst.})\right] \times 10^{-7}.$$

The dominant source of uncertainty in the aforementioned results is statistical, with the associated systematic uncertainty being comparatively small. This analysis focuses on measuring the branching fraction over the full-q^2 bin, for $B^+ \to K^+e^+e^-$ and $B^+ \to K^+\mu^+\mu^-$ decays.

The $\mathcal{B}[B \to J/\psi K]$ value has been reported by BaBar ($N_{B\overline{B}} = 124 \times 10^6$) [41], Belle ($N_{B\overline{B}} = 772 \times 10^6$) [39] and Belle II ($N_{B\overline{B}} = 198 \times 10^6$) [42]. Among the three, the Belle measurement is the most precise one. The BaBar measurement is $\mathcal{B}[B^+ \to J/\psi K^+] = [1.061 \pm 0.015$ (stat.) ± 0.048 (syst.)$] \times 10^{-3}$. Belle reported a branching fraction of $\mathcal{B}[B^+ \to J/\psi K^+] = [1.032 \pm 0.007$ (stat.) ± 0.024 (syst.)$] \times 10^{-3}$. The uncertainties on the latter are dominated by systematics, with the most dominant sources being the number of $B\overline{B}$ events (1.4%) and $f^{\pm}$ (1.2%). Belle II has measured the branching fraction of $B^+ \to J/\psi(\ell^+\ell^-)K^+$ where the uncertainties are dominated by the number of $B\overline{B}$ events (1.5%) and $f^{\pm}$ (2.6%). The results are as follows,

$$\mathcal{B}[B^+ \to J/\psi(e^+e^-)K^+] = [6.00 \pm 0.10\ (\text{stat.}) \pm 0.19\ (\text{syst.})] \times 10^{-5},$$
$$\mathcal{B}[B^+ \to J/\psi(\mu^+\mu^-)K^+] = [6.06 \pm 0.09\ (\text{stat.}) \pm 0.19\ (\text{syst.})] \times 10^{-5}.$$

2.5.2 R_K

R_K refers to the ratio of the branching fractions of $B \to K^+\ell^+\ell^-$ decay channels, where $\ell =$ muon or electron. We use the following relation to calculate R_K in a specific $q^2 \in (q_0^2, q_1^2)$ bin:

$$R_K[q_0^2, q_1^2] = \frac{\int_{q_0^2}^{q_1^2} dq^2 \frac{d\Gamma(K^+\mu^+\mu^-)}{dq^2}}{\int_{q_0^2}^{q_1^2} dq^2 \frac{d\Gamma(K^+e^+e^-)}{dq^2}} = \frac{N^{K^+\mu^+\mu^-}}{\epsilon^{K^+\mu^+\mu^-}} / \frac{N^{K^+e^+e^-}}{\epsilon^{K^+e^+e^-}} = \frac{N^{K^+\mu^+\mu^-}}{N^{K^+e^+e^-}} \times \frac{\epsilon^{K^+e^+e^-}}{\epsilon^{K^+\mu^+\mu^-}}, \tag{2.24}$$

where N and ϵ represent the fitted signal yield and reconstruction efficiency for each channel. The latter is the product of the simulated signal efficiency and correction factors due to imperfect simulation, which are discussed in Section 2.5.1. The correction factors that are common between the muon and electron channels (such as kaon detection or tracking efficiency) cancel, making the observable more precise.

Previous measurements of this observable by the LHCb Collaboration [26] hinted at possible LFU violation, but the latest measurement [37] rules out any such possibilities in the decay rates. Belle and BaBar also carried out measurements of R_K in various q^2 bins, and their results, albeit with larger uncertainties, are consistent with LFU conservation. A summary of these results is presented in Fig. 2.4. The uncertainties in the Belle and LHCb measurements of R_K are dominated by the statistical uncertainty. In this analysis, we measure R_K for the lower ($q^2 \in [1.1, 6.0]\,\mathrm{GeV^2/c^4}$), higher ($q^2 > 14.18\,\mathrm{GeV^2/c^4}$), and full-$q^2$ bins.

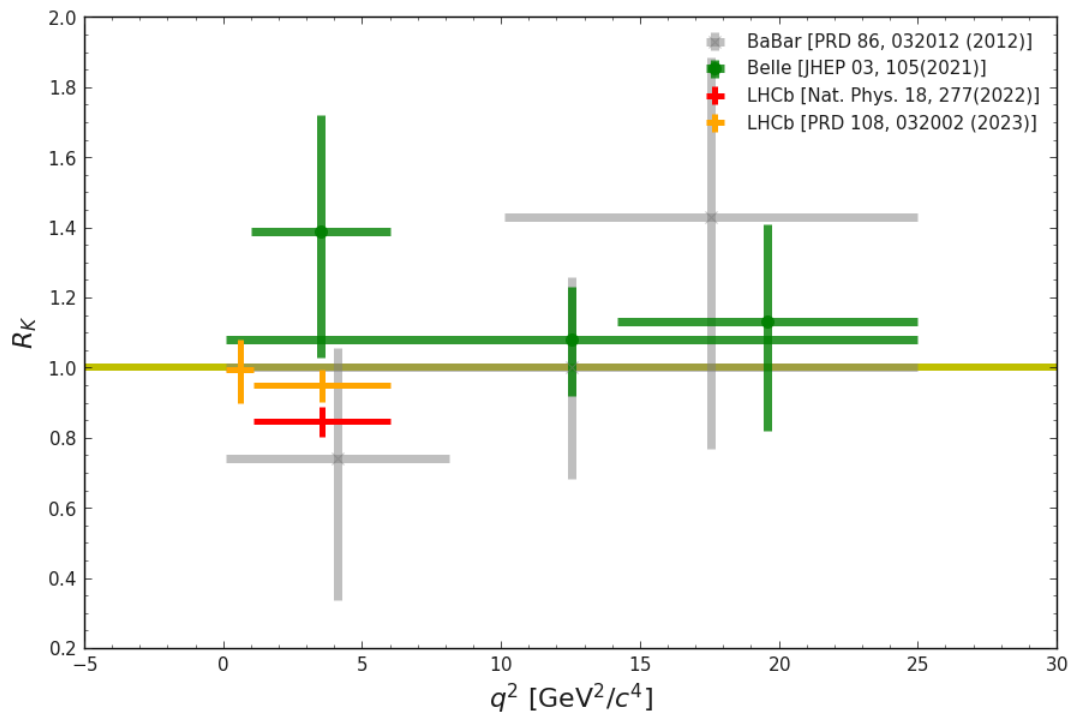


Figure 2.4: Summary of the R_K measurements for different q^2 bins.

2.5.3 $R_K(J/\psi)$

The relation we use is:

$$R_K(J/\psi) = \frac{\mathcal{B}[KJ/\psi(\mu^+\mu^-)]}{\mathcal{B}[KJ/\psi(e^+e^-)]} = \frac{N^{KJ/\psi(\mu^+\mu^-)}}{\epsilon^{KJ/\psi(\mu^+\mu^-)}} / \frac{N^{KJ/\psi(e^+e^-)}}{\epsilon^{KJ/\psi(e^+e^-)}} = \frac{N^{KJ/\psi(\mu^+\mu^-)}}{N^{KJ/\psi(e^+e^-)}} \times \frac{\epsilon^{KJ/\psi(e^+e^-)}}{\epsilon^{KJ/\psi(\mu^+\mu^-)}}. \tag{2.25}$$

where N and ϵ are the fitted signal yield and reconstruction efficiency for the respective channel.

Till date, Belle provides the most precise measurement on these observables, which is $R_{K^+}(J/\psi) = 0.994 \pm 0.011 \pm 0.010$. As we can see, statistical and systematic uncertainty are almost of the same order. The dominant systematic uncertainties are due to lepton identification; they are

0.79% and 0.57% for electron and muon, respectively. These systematics are obtained from a $J/\psi \to \ell^+\ell^-$ control sample, and the size of this sample decides the overall systematics. Hence, the lepton identification systematics is expected to decrease as Belle II collects more data.

Belle II measurement [42] of $R_K(J/\psi)$ is statistically limited ($N_{B\bar{B}} = 198 \times 10^6$, almost one fourth of the Belle dataset) with respect to Belle, but the systematic uncertainty for lepton identification (electron and muon are 0.6% and 0.4%, respectively) has improved. The measurement is $R_{K^+}(J/\psi) = 1.009 \pm 0.022 \pm 0.008$. LHCb also reported the measurement of $R_{K^+}(J/\psi) = 0.981 \pm 0.020$ (stat. $\oplus$ syst.) [26]. The dominant source of systematic uncertainty in the latter measurement comes from the statistical uncertainties of signal efficiencies.

Chapter 3

Experimental Setup

Accelerator-based particle physics experiments involve the acceleration (subsequent steering / focusing) of subatomic particles to relativistic speeds using electric (magnetic fields), followed by the collision of these accelerated particles at extreme energies. By studying the resulting final-state particles produced in these collisions, scientists gain insights into the fundamental constituents of matter, the forces that govern their interactions, and the fundamental laws of the universe, helping to answer questions about the nature of particles and the structure of the cosmos.

In this chapter, the experimental setup employed in the lepton universality study is discussed. The setup consists of two main components: the SuperKEKB collider, where electrons and positrons are accelerated to high energies and then collided, and the Belle II detector, which is used to detect the final-state particles resulting from these collisions.

3.1 SuperKEKB accelerator

SuperKEKB [43] is a double-ring asymmetric energy electron-positron accelerator located in Tsukuba, Japan. Its schematic view is presented in Fig. 3.1

The electron and positron ring energies are respectively $E_{e^-} = 7\,\mathrm{GeV}$ and $E_{e^+} = 4\,\mathrm{GeV}$, resulting in a center-of-mass (c.m.) energy of $E_{\mathrm{c.m.}} = \sqrt{(E_{e^-} + E_{e^+})^2 - (p_{e^-} + p_{e^+})^2} = 10.58\,\mathrm{GeV}$. This choice of c.m. energy, very close to the $\Upsilon(4S)$ resonance, predominantly produces two B_d^0 (or $B_u^\pm$) mesons in over 99% of the cases [9, 44].

The primary physics processes at SuperKEKB are $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ and $e^+e^- \to q\bar{q}, q \in$

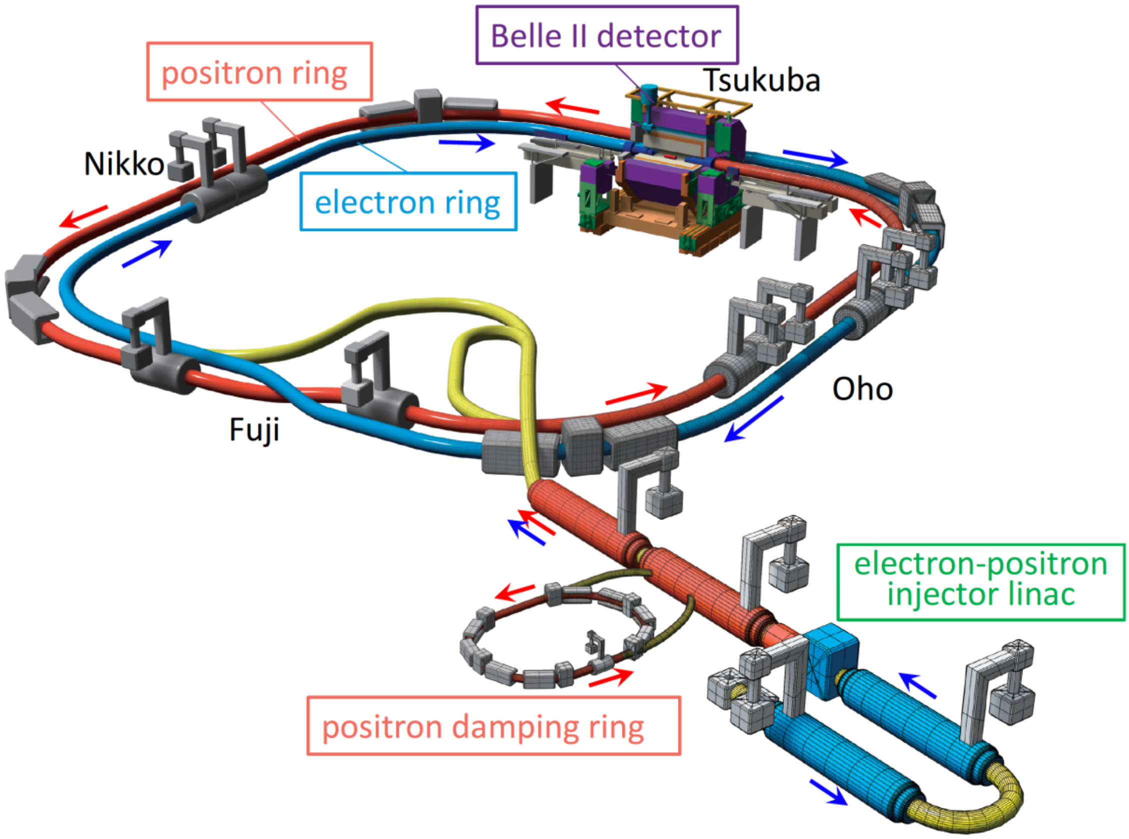


Figure 3.1: A schematic diagram of the SuperKEKB accelerator. Adapted from Ref.[2].

u, d, s, c. Leptonic processes such as $e^+e^- \rightarrow e^+e^-$, $e^+e^- \rightarrow e^+e^-e^+e^-$, $e^+e^- \rightarrow \mu^+\mu^-$ etc. characterized by fewer tracks, also occur. A summary of all major processes occurring at this energy is listed in Table 3.1.

Once a pair of $B_{d(u)}$ mesons is produced, their energy in the c.m. frame is approximately 5.29 GeV, which is just 11 MeV above their mass, rendering them almost at rest. However, due to the Lorentz boost factor $\beta\gamma = 0.284$, their average momenta are around $|\vec{p}_{e^-} + \vec{p}_{e^+}|/2 \sim 1.5$ GeV in the lab frame. This boost is designed to facilitate the study of CP violation, involving the measurement of the decay time difference between two B_d^0 mesons, determined by the distance (Δz). The typical mean lifetime difference between two decaying B mesons is on the order of their individual decay times, resulting in $\Delta z = \beta\gamma c\Delta t \sim 150\,\mu$m.

SuperKEKB aims to achieve an instantaneous luminosity 30 times greater than its predecessor KEKB, reaching $6 \times 10^{35}\text{cm}^{-2}\text{s}^{-1}$. This is achieved by doubling the beam current and reducing the cross-section at the collision point by a factor of 15, employing the novel nano-beam scheme described in Refs. [5, 43]. This scheme involves squeezing the longitudinal overlap size

Table 3.1: The production cross-section of various physics processes at the $\Upsilon(4S)$ resonance ($\sqrt{s} = 10.58\,\text{GeV}$)

Process	Cross-section (nb)
$e^+e^- \to \Upsilon(4S)$	1.11
$e^+e^- \to q\bar{q}$	3.69
$e^+e^- \to \tau^+\tau^-(\gamma)$	0.92
$e^+e^- \to e^+e^-(\gamma)$	300
$e^+e^- \to e^+e^-e^+e^-$	39.7
$e^+e^- \to e^+e^-\mu^+\mu^-$	18.9
$e^+e^- \to \gamma\gamma(\gamma)$	4.99
$e^+e^- \to \mu^+\mu^-(\gamma)$	1.15

between the two beams, thereby reducing the vertical overlap between them. This phenomena is called the "Hourglass effect" [45]. For a head-on or nearly head-on collision, the overlap region (σ_x) is the bunch length (d). But for a narrow, longer bunch with a large crossing angle, as shown in Fig.3.2, the overlap region is $d = \frac{\sigma_x}{\phi}$, where ϕ is the half-crossing angle. Compared to KEKB, σ_x is reduced by a factor of 15 ($150\,\mu$m $\to$ $10\,\mu$m), resulting in a 20-fold reduction in vertical overlap (σ_y) ($940\,\mu$m $\to$ $50\,\mu$m), as illustrated in Fig. 3.3.

3.2 Belle II detector

The Belle II detector is positioned precisely at the collision point of the SuperKEKB accelerator, where the magic of particle physics unfolds. Its location spans a remarkable range, extending from a mere 1.4 cm to a substantial 20 m around the heart of the experiment, known as the Interaction Point (IP). At this pivotal juncture, the beam pipe, with a minuscule radius of 1.4 cm, crafted from beryllium and boasting a 1 mm thickness, takes center stage. This choice of material, with its atomic number of four, ensures a low-Z characteristic, significantly diminishing the likelihood of multiple Coulomb scattering. A schematic diagram of Belle II is shown in Fig. 3.4.

Now, let's embark on an exploration of various subdetectors, each meticulously designed to serve its unique role in unravelling the mysteries of particle interactions, beginning from the innermost region and extending outward.

In the context of the nano-beam scheme at Belle II, the design of the beam pipe, with a mere

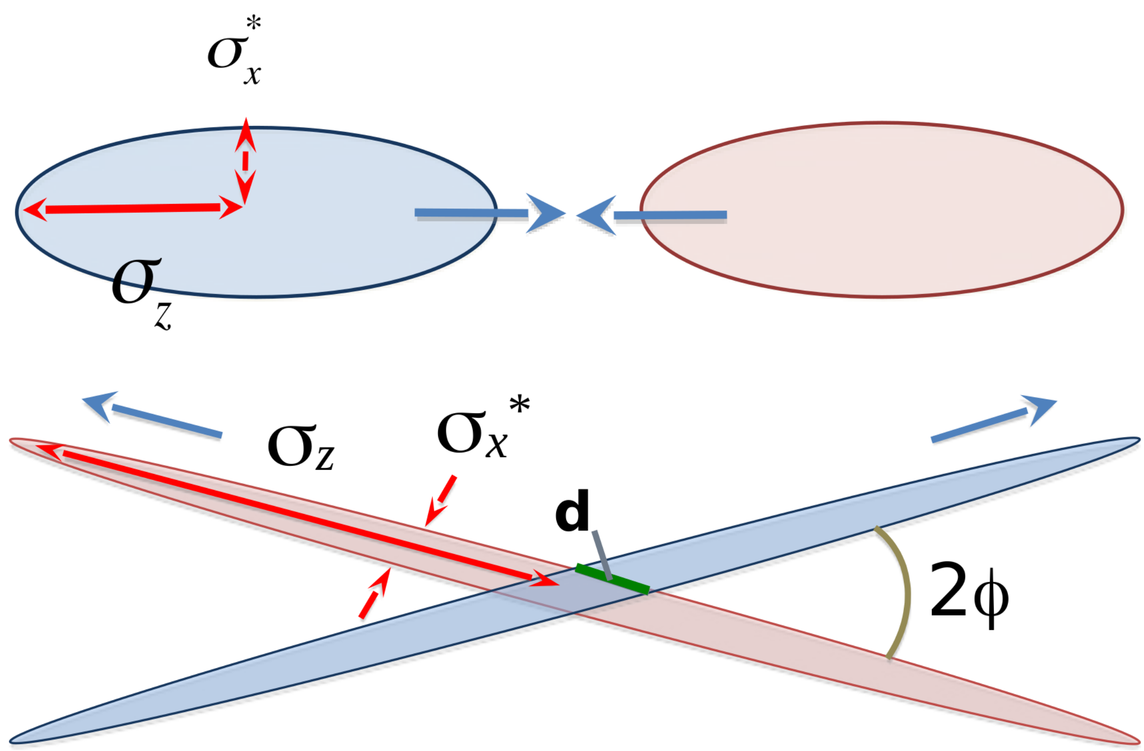


Figure 3.2: Pictorial presentation of the difference between head-on collision scheme (top) vs nano-beam scheme (bottom).

10 mm diameter, offers a remarkable opportunity to position the innermost detector in proximity to the IP. This placement is instrumental in achieving superior vertexing precision, a critical factor in our quest for accurate particle tracking and the reconstruction of primary and secondary vertices. However, this proximity also ushers in a new challenge in the form of intra-beam Coulomb scattering backgrounds, such as the troublesome Toucheck background. These backgrounds, inversely correlated with the beam pipe size, are anticipated to increase by a factor of 20 compared to the previous KEKB setup. The consequence of these enhanced beam backgrounds is a significant rise in occupancy, denoting the fraction of detector channels hit per triggered event. For the strip detectors, the intricate nature of these backgrounds poses a formidable challenge. When multiple particles traverse the strip detector, there are numerous possible hit positions, as illustrated in Fig. 3.5. To tackle this high occupancy issue arising from the intense beam backgrounds, a strategically designed solution has been implemented in the innermost region of the Belle II detector. Here, two pixelated detectors are placed, followed by four layers of strip detectors. The pixelated detectors offer a unique advantage in their ability to precisely reconstruct hit positions, a feat made possible by the discrete pixel arrange-

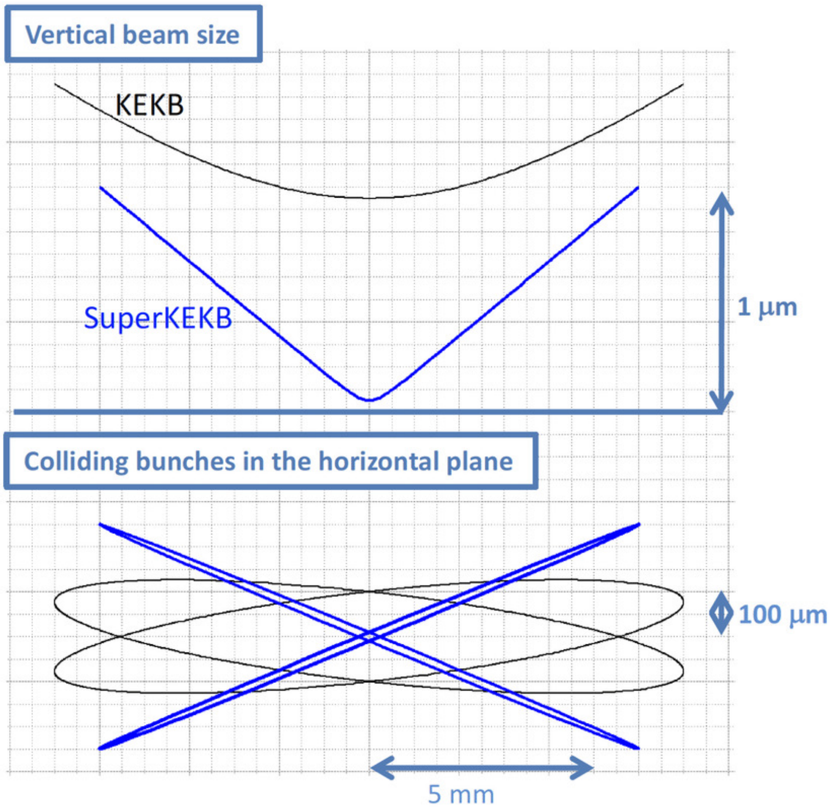


Figure 3.3: The beam profile of nano-beam scheme at SuperKEKB (blue) and KEKB (black) is portrayed. Adapted from Ref.[2].

ment. However, this innovation comes with its own set of challenges, particularly in terms of managing the substantial number of channels—equivalent to one channel for each pixel. This combination of pixelated and strip detectors strikes a harmonious balance between accurate hit reconstruction and high particle rates encountered in the nano-beam scheme.

3.2.1 PXD

The pixel detector (PXD) comprises two layers of pixelated sensors, employing the depleted p-channel field effect transistor (DEPFET) technology [46], positioned at distances of 14 mm and 22 mm from the IP. These sensors feature pixel areas of $50{\times}50\,\mu\mathrm{m}^2$ and $50{\times}75\,\mu\mathrm{m}^2$ for the inner and outer layers, respectively. Utilizing the principle of electron-hole pair generation within DEPFET sensors, it effectively detects the passage of charged particles. These sensors are organized into rectangular modules referred to as “ladders” collectively forming the dual tiers of the PXD. Specifically, the inner layer comprises 8 ladders, while the outer one is designed with 12 ladders in total. The dataset employed in this thesis predates the 2022 long shutdown, featuring a fully installed inner layer, while only two ladders of the outer layer had reached

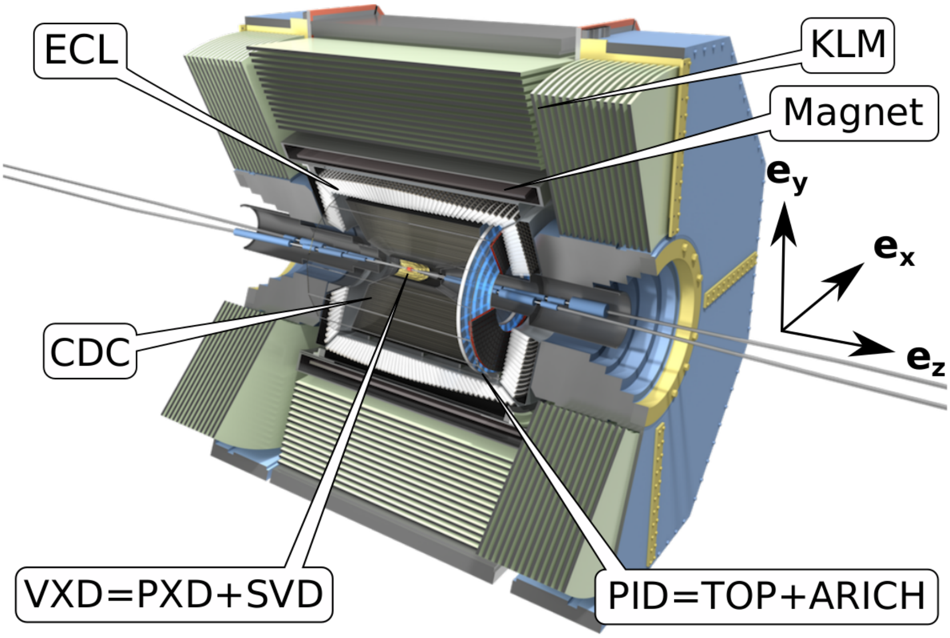


Figure 3.4: A schematic diagram of the Belle II detector. Adapted from Ref. [3].

completion.

3.2.2 SVD

The Belle II silicon vertex detector (SVD) comprises several cylindrically arranged DSSD (double-sided silicon strip detector) sensors placed around the IP. Each DSSD sensor is either rectangular or trapezoidal in shape and has a thickness of $300 - 320\,\mu$m. The sensor has an n-type bulk with one side doped by acceptors and the other side by donors. The longer strips on the acceptor-implanted side (p or U side) measure the r-ϕ coordinate. Similarly, the shorter strips on the donor-implanted side (n or V side) measure the z coordinate. In the SVD, the readout pitch of the layer-3 sensors is $50\,\mu$m in the r-ϕ plane and $160\,\mu$m along the z axis. On the other hand, the layer-4, -5, and -6 sensors have a readout pitch of $75\,\mu$m in the r-ϕ plane and $240\,\mu$m along the z axis.

In an SVD sensor, when a charged particle traverses the reverse-biased silicon sensor material, it imparts energy, generating electron-hole pairs. Due to the applied electric field, the electrons move towards the p-side of the sensor, while the positively charged holes move towards the n-side. This differential motion of charges creates a current signal, which is detected by the p and n side strips on the sensor.

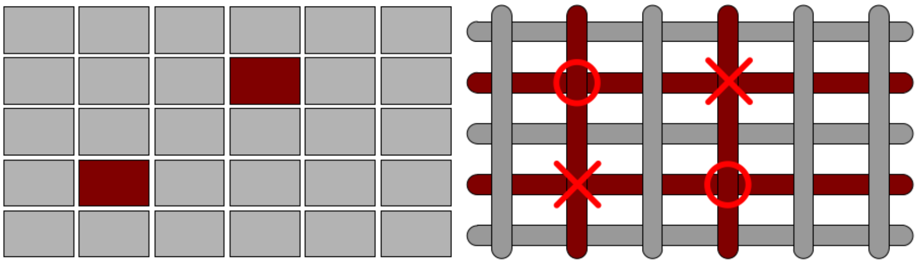


Figure 3.5: When impacted simultaneously by two particles, a pixel sensor (left) can unequivocally distinguish between the two hits. In a double-sided microstrip sensor (right), these dual impacts result in four potential combinations, with two representing genuine hits (crosses) and the remaining two referred to as 'ghost hits' (circles). Adapted from Ref. [4].

By analyzing the current signal from these orthogonal *p* and *n* side strips, the SVD can precisely determine the two-dimensional coordinates of the particle's trajectory on a specific layer of the detector. This information, when combined with data from multiple layers, allows the reconstruction of the particle's three-dimensional path. These reconstructed trajectories are essential for particle tracking and identification of decay vertices, contributing to our understanding of fundamental particle interactions in high-energy physics experiments. Figure 3.6 illustrates the working principle of the SVD. The working principle and performance of SVD are extensively discussed in Chapter 12.

3.2.3 CDC

The central drift chamber (CDC) forms an integral part of the Belle II tracking system, working in conjunction with the PXD and SVD sub-detectors. Beyond tracking, the CDC also aids in particle identification for particles with momenta less than 1 GeV/c, leveraging the ionization loss of charged particles as a distinguishing feature.

Constructed with a detection volume filled with a C_2H_6 + He gas mixture, the CDC comprises 56 layers of wire detectors. These layers are strategically positioned radially around the IP, spanning from 0.168 to 1.111 m. Within the CDC, two types of wires play pivotal roles. Axial wires run parallel to the beam axis, providing information in the $r-\phi$ plane, perpendicular to the beamline. Meanwhile, stereo wires are slightly tilted relative to the beam axis to gather information along the z axis. This dual-wire setup enables the CDC to capture precise three-dimensional positions as charged particles traverse the chamber. The layers of axial and stereo wires are grouped into superlayers. The first superlayer, primarily axial, encompasses eight

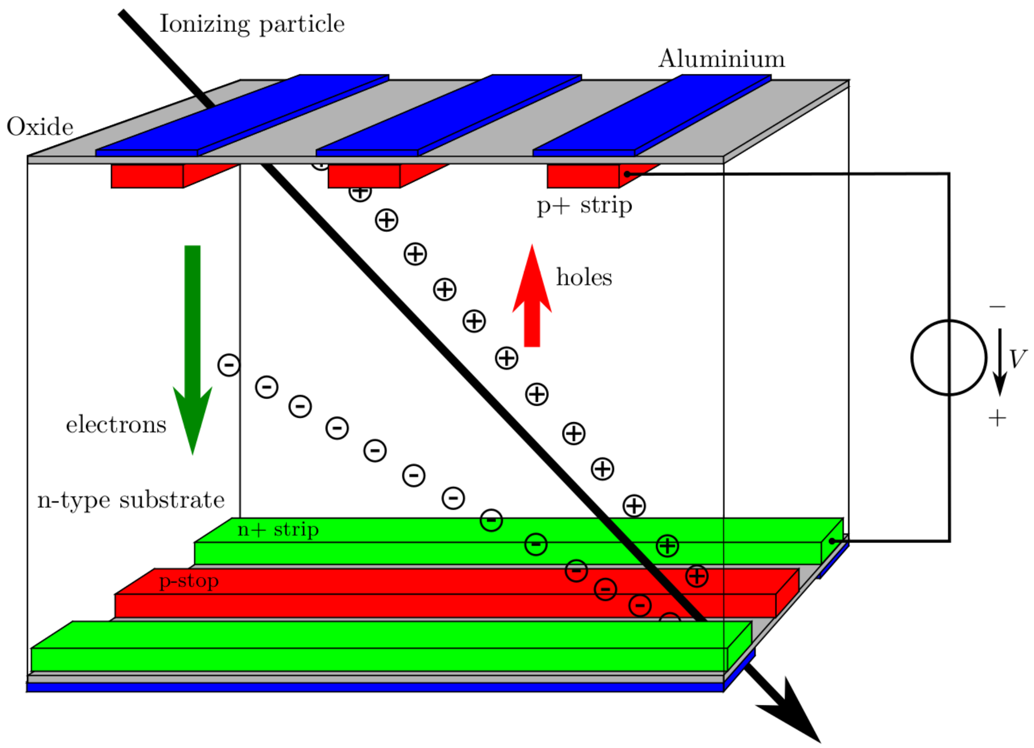


Figure 3.6: Working principle of the SVD: A reverse bias voltage is applied between the p+ and n+ strips. Conventionally, the reverse bias voltage is considered positive from the p- to n-side. When an ionizing particle deposits energy, electron-hole pairs are created, with electrons attracted towards the n+ strips and holes towards the p+ strips. Adapted from Ref. [4].

layers of wires, while the subsequent superlayers consist of six layers of wires each.

As charged particles traverse the gas medium, they ionize it, generating electron-ion pairs. Guided by the electric field within the CDC, the drifting electrons are directed towards the wire detectors. Here, their timing and position are meticulously recorded. This wealth of information allows for the reconstruction of particle trajectories and precise determination of their momenta in the context of high-energy collision events.

3.2.4 TOP

The time of propagation (TOP) subdetector, located in the barrel section, and the aerogel ring imaging cherenkov (ARICH) subdetector in the endcap region, play a crucial role in particle identification, particularly for distinguishing between kaons and pions. The particle identification is fundamental in determining the particle velocity, which, when combined with the track momentum measured by the tracking system, allows for a precise identification of particle mass.

The key principle underlying particle identification in these detectors is Cherenkov radiation. When charged particles travel through a clear medium like aerogel or quartz at a velocity greater than the speed of light in that medium, they emit Cherenkov radiation. The angle of emission (θ_C) is dependent on the particle's velocity (β) and the refractive index (n) of the material, as described by the equation:

$$\beta = \frac{1}{n \cos\theta_C}. \tag{3.1}$$

The TOP subdetector is primarily constructed using an array of fused silica (quartz) bars. Cherenkov light emitted within these bars undergoes total internal reflection and is retained within the material. In the forward end of the detector, spherical mirrors are used to focus and direct the emitted Cherenkov light. On the opposite end, photomultiplier tubes (PMTs) are mounted to measure the timing of the photons.

As charged particles pass through the quartz bars, they emit Cherenkov photons that propagate toward the array of PMTs as shown in Fig. 3.7. By measuring the arrival times and positions of these photons at the PMTs, the subdetector accurately determines the emission angle, providing critical information for particle identification in the Belle II experiment.

3.2.5 ARICH

The ARICH subdetector, located in the forward end-cap region outside the CDC, serves the essential purpose of identifying kaons and pions within a momentum range of 0.4 to 4 GeV$/c$, as well as muons and electrons below 1 GeV$/c$. This discrimination is achieved through a meticulously designed setup consisting of an aerogel radiator, where Cherenkov photons are generated by the passage of charged particles as shown in Fig. 3.7. These emitted photons subsequently travel through an array of imaging devices, precisely measuring both their positions and arrival times. By analyzing these data, the ARICH subdetector effectively calculates the Cherenkov angle and, in turn, the velocity of the particles using Eq. 3.1.

3.2.6 ECL

The electromagnetic calorimeter (ECL) serves a central role in our experiment, efficiently detecting photons, precisely measuring their energy and angular coordinates, identifying electrons, and contributing to event triggering and luminosity measurements, both in real-time and offline data analysis.

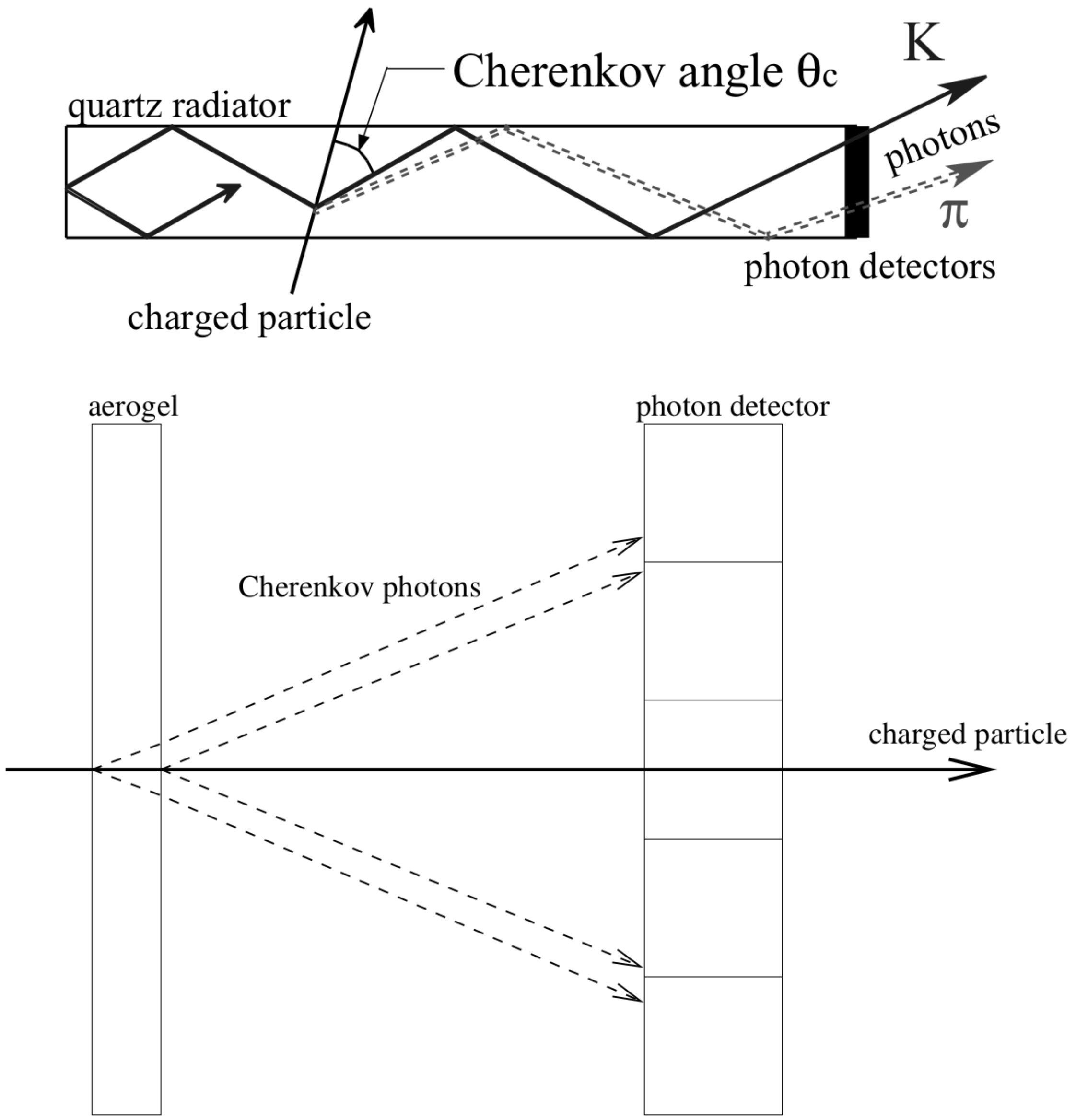


Figure 3.7: Working principles of TOP and ARICH subdetectors is depicted in the top and bottom plots respectively. Adapted from Ref. [5].

Constructed with CsI(Tl) crystals, cesium iodide doped with thallium, the ECL utilizes this exceptional scintillating material in both the barrel and forward sections, covering a polar angle range ($12.4^\circ < \theta < 155.1^\circ$) and providing extensive coverage, approximately 90%, across a wide solid angle. We add Tl impurity to pure CsI crystals to enhance their scintillation efficiency by modifying energy bands, thereby improving the emission of photons from excited electrons. Within the ECL, the barrel section houses 6624 CsI(Tl) crystals, each taking the form of a truncated pyramid with dimensions of approximately $6 \times 6\,\mathrm{cm}^2$ in cross-section and 30 cm in length. The endcaps add 2112 CsI(Tl) crystals with various shapes, resulting in a total crystal count of 8736. These crystals are connected to two photodiodes each, allowing for signal collection.

When high-energy photons or electrons traverse these crystals, they initiate electromagnetic showers. These showers are cascades of secondary particles, primarily electrons and positrons, generated as a high-energy photon or electron interacts with the material, emitting photons and creating new e^+e^- pairs in a rapidly amplifying chain reaction. These crystals respond to this interaction by producing scintillation light, which is then detected and quantified by photodetectors. The length of the crystals corresponds to 16.1 times radiation length, which ensures to absorb all the energy of high energy photons.

The energy resolution of the ECL [47] can be expressed in an approximate form by the following equation:

$$\frac{\sigma_E}{E} = \frac{0.066\%}{E} \bigoplus \frac{0.81\%}{E^{1/4}} \bigoplus 1.34\%. \tag{3.2}$$

In this expression, the initial term represents the contribution of electronic noise, while the second term and a portion of the third term arise from fluctuations in shower leakage. The remaining part of the third term is linked to systematic effects, including uncertainties in counter-to-counter calibration. The Belle II ECL subsetector yields a mass resolution of approximately $5\,\mathrm{MeV/c^2}$ for $\pi^0 \to \gamma\gamma$ and relative energy resolutions of about 4% at an energy of 100 MeV and 2% at an energy of 8 GeV [48].

3.2.7 Solenoid magnet

Encircling the ECL, there is a superconducting solenoid that generates a magnetic field strength of 1.5 T within a cylindrical volume measuring 3.4 m in diameter and 4.4 m in length. This solenoid relies predominantly on niobium-titanium (NbTi) for its superconducting coils and operates in conjunction with a cryogenic system utilizing liquid helium for cooling. In the presence of this magnetic field, charged particles follow curved trajectories, and by assessing the curvature through data collected from the tracking detectors, their momenta are determined using the formula:

$$p_\mathrm{T}[\mathrm{GeV}/c] = 0.3B[\mathrm{T}]r[\mathrm{m}]. \tag{3.3}$$

Here, B represents the magnetic field strength, and r denotes the radius of curvature.

3.2.8 KLM

The KLM, short form for the K Long and Muon subdetector, constitutes the outermost component of Belle II. Its primary role is to identify and track muons and neutral long-lived kaons (K_L^0), particularly those that exhibit partial showering in the ECL.

The KLM is divided into two sections: barrel and end-cap, collectively covering the polar angle range of $20° < \theta < 155°$. It is composed of an alternating arrangement of 4.7 cm thick iron plates and either glass-electrode resistive plate chambers (RPCs) in the barrel section or plastic scintillators in the end-caps.

In the barrel section, there are a total of 15 detector layers interspersed with 14 iron plates, while the end-caps consist of 14 detector layers combined with 14 iron plates. It is worth noting that the choice of plastic scintillators in the end-caps, as opposed to RPCs used in KEKB, was prompted by the higher beam background conditions expected at SuperKEKB.

The KLM employs RPCs to measure the ionization generated by charged particles, serving the purpose of identifying muons and K_L^0 mesons. Similarly, scintillators are utilized to detect the emitted light produced by these particles, which aids in their identification. The iron plates play a vital role as passive absorbers, effectively diminishing the particle flux that reaches the KLM subdetector. This passive shielding results in a more efficient and accurate muon identification process.

3.3 Triggers

The Belle II trigger system plays a vital role by selecting the rare, high-value events worthy of further analysis, while swiftly discarding the noise or known physics background. The trigger system is divided in two layers:

3.3.1 Level-1 or L1 trigger

The cross-section for $e^+e^- \to B\overline{B}$ is overshadowed by other physics processes, as listed in Table 3.1. To efficiently select B events, a field-programmable gate array (FPGA)-based configurable trigger system records events with a fixed latency of $5\,\mu$s. The most critical trigger information is derived from the CDC, which measures the number of tracks, as well as from the ECL, which assesses clusters and energy deposition. This information enables the categorization of events into high-multiplicity B/D decay, low-multiplicity, and potential dark sector events. High-multiplicity triggers, crucial for preserving $\Upsilon(4S)$ decay events (along with $e^+e^- \to q\bar{q}$) with high efficiency, typically require a minimum of three tracks or four clusters.

Even though low-multiplicity events are common in beam backgrounds, Belle II needs to retain events that are important for τ-physics, dark sector searches, and two-photon physics studies. We engineer L1 triggers to meet this challenge by effectively selecting such events while rejecting beam backgrounds.

Furthermore, approximately every 100th event of Bhabha and $e^+e^- \to \gamma\gamma$ processes is stored for luminosity measurements, identified by their unique characteristics of two tracks matched with two ECL clusters or simply two ECL clusters.

3.3.2 High level trigger or HLT

It is designed to rapidly analyze and filter the immense volume of data generated by the Belle II detector. By fully reconstructing the events, the HLT efficiently identifies events of interest that merit further investigation. This real-time decision-making capability is essential for reducing the data rate and allowing for selective storage of pertinent information.

Chapter 4

Event reconstruction

In this chapter, we embark on the journey of event reconstruction within the Belle II experiment, focusing on the critical task of reconstructing B mesons from their final-state particles. Event reconstruction is the key step that transforms raw detector-level data into a clear and meaningful picture of particle interactions. As we explore the methods and techniques used to identify and reconstruct B mesons, we also provide a succinct description of the dataset that forms the foundation of our analysis. Understanding the dataset is paramount as we unravel the complexities of lepton universality in $B \to K\ell\ell$ decays. Throughout the thesis, charge-conjugate processes are implicitly included.

The chapter is structured as follows: In Section 4.1, we delineate the overarching analysis approach. Section 4.3 provides an exposition of the dataset employed in this analysis. Sections 4.4 and 4.5 elucidate the process of reconstructing B mesons from the selected charged tracks.

4.1 Analysis strategy

In the context of e^+e^- collisions occurring at the c.m. energy of $\Upsilon(4S)$ within a unique experimental setup at SuperKEKB, pairs of $B_{d(u)}$ or B mesons are produced. A schematic diagram is shown in Fig. 4.1. Given the boost factor $\beta\gamma = 0.284$ at SuperKEKB, a B meson, on average, traverses a distance of approximately $150\,\mu$m before decaying into various final states.

The typical decay products from the B meson include five types of charged particles: electrons ($e^{\pm}$), muons ($\mu^{\pm}$), pions ($\pi^{\pm}$), kaons ($K^{\pm}$), and protons ($p^{\pm}$), as well as three kinds of neutral

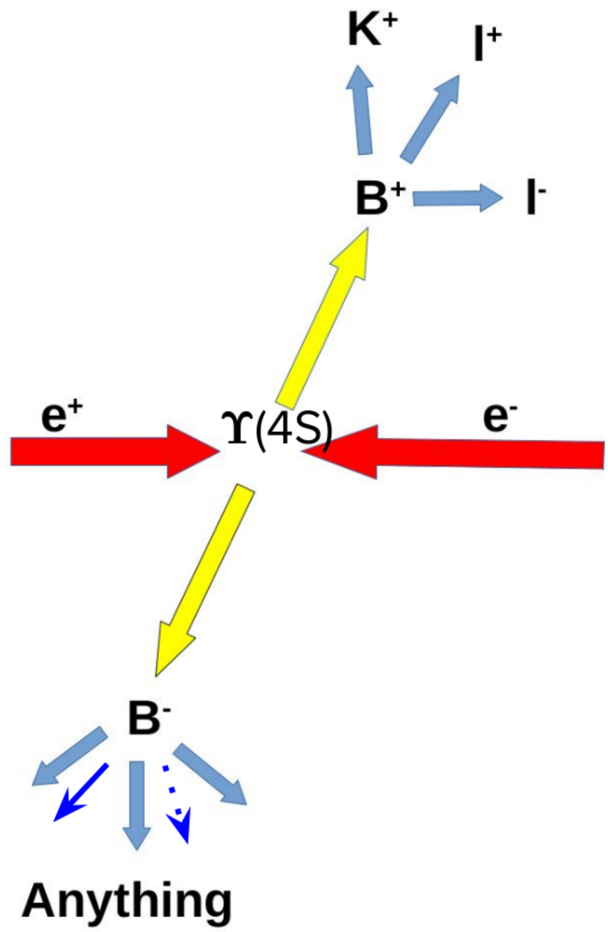


Figure 4.1: Schematic representation of e^+e^- collisions at $\Upsilon(4S)$, resulting in the production of a pair of B mesons. Subsequently, one of the B mesons undergoes a generic decay, while the other decays into the $K\ell\ell$ final-state, forming the signal event of our interest.

particles: photons (γ), long-lived neutral kaons (K_L^0), and neutrons (n). Importantly, these final-state particles either are stable or exhibit sufficiently long lifetimes, ensuring that they do not undergo decay within the detector material. Within Belle II, these particles leave distinct signatures, enabling us to construct likelihoods for various particle hypotheses in the observed data. On other hand, short-lived particles like K_S^0 and π^0 are reconstructed from their decay products.

In the exclusive analyses, where the explicit reconstruction of B mesons is undertaken, potential B meson candidates are formed by combining the four-momentum vectors of the final-state particles. Leveraging the inherent advantage of monoenergetic fixed-energy B mesons at the initial state, we are motivated to find out kinematic variables that exhibit high-resolution peaks, thereby enhancing the precision of our signal detection.

The background suppression forms a critical facet of our analysis methodology, achieved through the application of optimized selections on various kinematic variables. These selections are judiciously chosen to maximize a figure-of-merit (FOM), defined as:

$$\mathrm{FOM} = \frac{S}{\sqrt{S+B}}, \tag{4.1}$$

where S and B represent the number of signal and background events within the signal-enriched kinematic region.

Ultimately, the observables of interest are determined via a likelihood fit. An integral aspect of our strategy involves maintaining the blindness of the data within the signal-enriched region, a practice elaborated upon below.

4.1.1 Blind analysis

Blind analysis is a fundamental practice mostly implemented at B factories to safeguard the integrity and impartiality of data analysis. In the realm of precision measurements in particle physics, there is always a risk of unconscious bias stemming from prior expectations or hypotheses. To address this challenge, blind analysis entails the deliberate concealment of crucial data details or analysis procedures from researchers until all aspects of the analysis, including selection criteria and estimation of systematic uncertainties, are meticulously established. Only once these components are rigorously finalized do researchers remove the blinders, granting them access to the results without the influence of preconceived notions. This approach bolsters the credibility and robustness of our findings, instilling a heightened level of confidence in the exploration of new physics phenomena.

4.2 $B \to J/\psi K$ control sample

A control sample is a subset of the data used for various purposes, such as validation, calibration, and systematic uncertainty estimation. In our case, we have chosen decay processes that exhibit larger branching fractions and share similar decay kinematics with our primary signal, $B \to K\ell^+\ell^-$.

To validate the selection criteria applied to $B \to K\ell^+\ell^-$, we employ $B \to J/\psi K$ events. We use $B \to J/\psi K$ decays to calibrate the mean and width of $M_{\rm bc}$ and ΔE during the signal extraction phase of $B \to K\ell^+\ell^-$, which will be discussed in Chapter 7.4. These charmonium B

decays are also helpful in calibrating the MVA classifier output and correcting for any potential data-simulation disparity. Inclusive $J/\psi \to \ell^+\ell^-$, along with a few other samples, are used to calibrate the lepton identification selection as discussed in Section. 10.2.

4.3 Data and simulated samples

In the foundation of our analysis lie two indispensable components: the collision dataset, derived from real-world observations at Belle II, and the simulated sample, a critical tool for understanding the sources of background and validating our analysis techniques.

4.3.1 Simulated sample

To examine the properties of signal events, refine selection criteria, and calculate detection efficiency, simulated events for both signal and background events are produced. These events are often referred to as 'Monte Carlo' (MC) events because they utilize the Monte Carlo method, a statistical technique that relies on random sampling to simulate complex processes or systems in a controlled and statistically representative manner. B-meson decays are generated with the EvtGen [49] package, where the radiation effects from final-state charged particles are incorporated using Photos [50]. Geant4 [51] simulates the interaction of generated particles with the detector. The Belle II software package [52] is used to process both collision data and simulation events.

Signal MC sample

In our analysis, we draw upon an extensive dataset comprising 5 million events corresponding to each of the signal decays and the most significant backgrounds crucial to our investigations. These events are meticulously crafted to replicate $\Upsilon(4S)$ events, wherein one of the two produced B mesons decays as per the decay channel of interest, while the other B meson follows the observed branching fractions of B meson decays, as documented by the Particle Data Group [9]. The signal decay events are listed below:

- $B^+ \to K^+e^+e^-$
- $B^+ \to K^+\mu^+\mu^-$
- $B^+ \to K^+J/\psi(e^+e^-)$
- $B^+ \to K^+J/\psi(\mu^+\mu^-)$

- $B^+ \to \pi^+ J/\psi(e^+e^-)$
- $B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$.

It's important to note that the analyses were mostly conducted utilizing simulated samples characterized by run-independent conditions. Though the precise run conditions were not replicated, a stringent validation process was employed using run-dependent simulated samples. Nonetheless, in the absence of specific mentions, reference to the "simulated sample" inherently alludes to the run-independent variant.

The run-dependent simulated samples, on the other hand, are carefully designed to match the exact conditions of the detector. This includes things like masking or temporarily turning off detector part that are malfunctioning, as well as taking into account the effect of beam-induced backgrounds. The beam background is simulated in run-independent simulations, while the one in run-dependent simulations consists of randomly triggered events, making the latter more realistic.

Background MC sample

To study other background contributions, we use a simulated sample equivalent to 1 ab^{-1} of $\Upsilon(4S) \to B\overline{B}$ and $e^+e^- \to q\bar{q}$ continuum events. Here, q stands for a u, d, s, or c quark. Continuum events are simulated using the KKMC [53] generator interfaced with PYTHIA [54]. As we focus our analysis on B decays, it is important to note that low-multiplicity events are typically excluded from consideration. The minimum requirement of three tracks from HLT inherently vetoes these events. The production cross-sections for the $\Upsilon(4S) \to B\overline{B}$ and $e^+e^- \to q\bar{q}$ events are listed in Table 4.1.

Table 4.1: Production cross-sections of various physics processes at the $\Upsilon(4S)$ resonance ($\sqrt{s}$ = 10.58 GeV).

Process	Cross-section (nb)
$\Upsilon(4S) \to B\overline{B}$	1.11
$e^+e^- \to u\bar{u}$	1.61
$e^+e^- \to d\bar{d}$	0.40
$e^+e^- \to s\bar{s}$	0.38
$e^+e^- \to c\bar{c}$	1.30

4.3.2 Data sample

The data sample used in our study amounts to an integrated luminosity of 362 fb^{-1}, equivalent to $(387 \pm 6) \times 10^6$ $B\overline{B}$ events. The B counting technique used to calculate the number of $B\overline{B}$ events in the data is described in Appendix A.

4.4 Reconstruction of final-state particle

This section focuses on the reconstruction of B decay events with the detector objects obtained from track- and cluster-level information. The process of event reconstruction involves identifying and reconstructing the primary and secondary particles in the event, as well as applying preliminary selections to filter out unwanted background events.

4.4.1 Track selection

Using track- and cluster-level information obtained from the detector, we reconstruct the final-state charged particles, which include $K^{\pm}$, $\mu^{\pm}$, and $e^{\pm}$, as well as γ as primary particles. To ensure that the charged tracks originate from a region close to the e^+e^- interaction point, we impose requirements on their distance of closest approach in both the transverse x-y plane (d_0) and along the z axis (z_0). Specifically, we require $|d_0| < 0.5$ cm and $|z_0| < 2.0$ cm. We then identify tracks that meet these criteria as either $\mu^{\pm}$, $e^{\pm}$, or $K^{\pm}$ based on specific requirements. To suppress contamination from low-multiplicity $e^+e^- \to e^+e^-\ell^+\ell^-$, $e^+e^- \to e^+e^-\gamma$, and $e^+e^- \to \mu^+\mu^-\gamma$ backgrounds, we require that the event contains at least five charged particles (# of clean tracks), each with a transverse momentum greater than 100 MeV/c.

4.4.2 Particle identification

The magnetic field is responsible for curving the paths of charged tracks, a crucial element in determining the momentum of these charged particles. To identify a specific particle, it is imperative to ascertain its velocity, which, in conjunction with momentum, reveals its mass. The measurement of velocity relies on the dedicated particle identification subdetectors.

Based on data recorded by the CDC, TOP, and ARICH subdetectors, likelihood calculations serve as the main source of information for the identification of charged kaons. Furthermore, these subdetectors play an indispensable role in distinguishing charged leptons, encompassing electrons and muons (e, μ). The ensuing sections offer a succinct overview of the methodologies utilized to establish likelihood estimates for each of these detection systems.

Kaon identification

The SVD and CDC measures the ionization energy loss (dE/dx) of a traversing charged particle. The dE/dx value depends universally on β, or velocity. Hence, from dE/dx, we know $\beta\gamma$, and using the relation $\beta\gamma = p/m$, we can find the mass of a particle if its momentum is known. A likelihood function is formulated for each particle hypothesis m as follows:

$$L_m\left(\frac{dE}{dx}, p\right) = \prod_i P_m\left(\left(\frac{dE}{dx}\right)_i, p\right).$$

Here, i runs through all the dE/dx values associated with a given track. Below a momentum of 1 GeV/c, this helps in a very efficient manner.

The emission angle of Cherenkov light is uniquely determined by the particle velocity. The propagation time of photons in an internal-total-reflection light guide can be calculated as a function of the photon emission angle. This motivates us to form a likelihood function for each of the particle hypotheses as a function of the arrival time and two-dimensional information of a Cherenkov ring image from the TOP. The ARICH also measures the Cherenkov emission angle of a particle traversing through the end caps. The likelihood is formed as a function of the number of Cherenkov photons in each pixel of the ARICH subdetector.

Finally, combining all detector information by multiplying the likelihood of relevant subdetectors (CDC, SVD, TOP, ARICH for hadrons), we get the final likelihood for each particle hypothesis. Using the likelihood, we form a variable, namely binary kaonID, defined by

$$\mathcal{P}(K/\pi) = \frac{L_K}{L_K + L_\pi},$$

where L_K and L_π represent the likelihoods for a track to be a kaon or a pion, respectively. We select a track to be kaon satisfying $\mathcal{P}(K/\pi) > 0.6$. We estimate a kaon identification efficiency of 86% with a pion misidentification rate of 7%, for a momentum range of $[0.5, 4.5]$ GeV/c and a polar-angle acceptance of $[0.49, 2.62]$ rad. The kaon candidates with a hit count in the CDC subdetector greater than 20 are retained.

Muon identification

We identify the muons by their penetration depth and lateral scattering pattern in the KLM. The process begins by extrapolating the track from the outermost CDC layer to the KLM subdetector. A search is then conducted to find a potential match in the KLM's innermost layer.

The degree of matching is quantified by evaluating whether a hit falls within 3.5σ of the region defined by the extrapolated track, with σ representing the quadrature sum of the extrapolated track's uncertainty and the hit position's uncertainty.

Two distinct likelihoods are constructed during this process. The first a likelihood is established for the longitudinal penetration depth through the KLM. For this, we consider the hit pattern across all KLM layers that were crossed during the extrapolation and the pattern of matched hits in those crossed layers of the track. The second likelihood pertains to the matched track for each particle hypothesis, which is formulated as a function of the matched χ^2 value obtained from the Kalman filter [55] applied to the extrapolated track and the associated degrees of freedom. A factor of two in the degree-of-freedom calculation accounts for the presence of two independent coordinates (and therefore measurements) per KLM layer.

The overall likelihood for identifying muons is derived by multiplying the above two likelihoods, which respectively pertain to the longitudinal penetration depth and transverse scattering. The muon identification is then normalized using the following equation:

$$\text{muonID} = \mathcal{P}(\mu) = \frac{L_\mu}{L_\mu + L_e + L_K + L_\pi + L_p + L_d}. \tag{4.2}$$

In order to designate a track as a potential muon candidate, a selection criterion of muonID > 0.99 is applied. This criterion result in a muon identification efficiency of 89% with a pion misidentification rate of 3%, calculated for a momentum range $[0.4, 6.5]\,\mathrm{GeV}/c$ and polar-angle acceptance $[0.4, 2.6]\,\mathrm{rad}$. Furthermore, to ensure the track reaches the KLM subdetector, the muon candidate is required to have a momentum greater than $0.7\,\mathrm{GeV}/c$.

Electron identification

We employ a multivariate analysis (MVA) based identification to select electrons from the list of charged tracks. While the ECL is the primary contributor to this identification process, the most efficient input comes from the ratio E/p (ECL energy over momentum). In the low-momentum regime ($p < 600\,\mathrm{GeV}/c$), a number of other cluster shape variables provide additional information for electron identification.

The MVA classifier utilizes likelihoods from various subdetectors and incorporates additional ECL observables [56], such as the cluster lateral moment, the Zernike moment, and the projection of the distance between the track entry point in the ECL and the cluster centroid onto the extrapolated track direction. These observables are not part of the traditional likelihood-based electron identification variable, which is defined similarly to Equation 4.2. Instead, the additional ECL variables are combined with the likelihood-based electron identification to train a classifier. The output of that classifier, or electronID $\mathcal{P}(e)$, is then utilized to select electron candidates. To select electron candidates, a criterion of $\mathcal{P}(e) > 0.9$ is applied. The resulting electron identification efficiency is 95% and the pion misidentification rate is around 1%, estimated for a momentum range $[0.4, 7.0]$ GeV/c and polar-angle acceptance $[0.22, 2.71]$ rad. In addition, the momentum is required to be greater than $0.4\,\mathrm{GeV}/c$.

The bremsstrahlung process is particularly significant in the context of electron identification. Bremsstrahlung occurs when a charged particle undergoes radiation while being accelerated. Synchrotron radiation, in which the magnetic field accelerates charged particles, is an example of this process. Additionally, bremsstrahlung takes place when a charged particle traverses a material, influenced by the electric field of the material's nucleus. The likelihood of emitting a photon as a result of this process is approximately proportional to

$$P \sim \frac{q^2 Z^2 E}{m^2}, \tag{4.3}$$

where q, E, and m represent the charge, energy, and mass of the particle, respectively, and Z signifies the atomic number of the material.

Due to its $1/m^2$ dependency, electrons, being the lightest charged particles, have a notably higher probability of radiation compared to relatively heavier particles like muons, pions, protons, or kaons. It has been established that only electrons experience a deterioration in resolution due to bremsstrahlung. Before reaching the ECL, electrons must traverse a great deal of materials, including the beam pipe, layers of vertex detectors, and the CDC support structure. All of these materials are located within the first 20 cm of the detector. The radiated photons travel tangentially to the paths of electrons, and some of them possess sufficient energy to form clusters within the ECL. Since electrons curve due to the magnetic field, they form separate clusters from the bremsstrahlung-radiated photons. Consequently, the correct energy estimation from the ECL cluster associated with the track becomes challenging, leading to a worsened resolution. A schematic diagram illustrating this process is presented in Fig. 4.2.

The objective of bremsstrahlung recovery is to fix the energy of electron tracks by accounting for the energy loss caused by photons that are radiated by bremsstrahlung. For a given electron depositing energy in the ECL, the four-momenta of all photons detected within a cone of

50 mrad with respect to the initial direction of the electron and having an energy greater than 50 MeV are added to that of the electron to recover the energy loss due to bremsstrahlung. Subsequently, the momentum of the electron track becomes:

$$p_e = p_{\text{track}} + \sum p_\gamma.$$

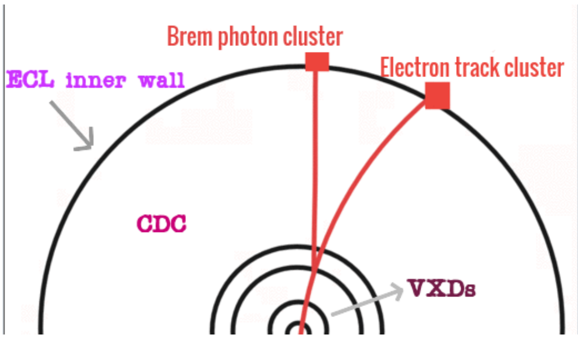


Figure 4.2: Schematic diagram of an electron track emitting a photon at one of the VXD layers.

This procedure aims to enhance the accuracy of electron energy determination by considering the influence of bremsstrahlung radiation. The selection criteria for final-state particles are summarised in Table 4.2.

Table 4.2: Summary of particle selection criteria.

	Selection	Cut value
Charged track	$\lvert d_0 \rvert$	< 0.5 cm
	$\lvert z_0 \rvert$	< 2.0 cm
K^+	$\mathcal{P}(K/\pi)$	> 0.6
	# of CDC hits	> 20
μ^+	p_μ	$> 0.7\,\text{GeV}/c$
	$\mathcal{P}(\mu)$	> 0.99
e^+	p_e	$> 0.4\,\text{GeV}/c$
	$\mathcal{P}(e)$	> 0.9

4.5 B candidate reconstruction

We reconstruct B mesons by combining a charged kaon with a pair of oppositely charged lepton candidates. We then perform a vertex fit on the B candidates and retain those for which the fit converges.

4.5.1 ΔE and M_{bc}

Two kinematic variables, the energy difference ΔE and the beam-energy constrained mass M_{bc} play a key role in suppressing the combinatorial background that arises due to random combinations of charged tracks. They are defined as:

$$\Delta E = E_B^* - E_{\mathrm{beam}}^*, \tag{4.4}$$

$$M_{\mathrm{bc}} = \sqrt{E_{\mathrm{beam}}^{*2} - p_B^{*2}}. \tag{4.5}$$

Here (E_B^*, p_B^*) are the energy and momentum of B the candidate, and E_{beam}^* is the beam energy; all three are calculated in the c.m. frame. For a correctly reconstructed B candidate, ΔE and M_{bc} are expected to peak around zero and known B mass [9], respectively.

The resolution of ΔE or $\sigma_{\Delta E}$ as seen in Eq. 4.6, gets contributions from both the beam energy and energy of final-state particles, but dictated by the latter, especially for the B candidate reconstruction involving electrons or photons where the energy measurement is solely dependent on the ECL.

$$\sigma_{\Delta E}^2 = \sigma_{E_B^*}^2 + \sigma_{E_{\mathrm{beam}}^*}^2. \tag{4.6}$$

On the other hand, the resolution of M_{bc} can be written as

$$\sigma_{M_{\mathrm{bc}}}^2 = \sigma_{E_{\mathrm{beam}}^*}^2 + \left(\frac{p_B^*}{m_B}\right)^2 \sigma_{p_B^*}^2. \tag{4.7}$$

Since the momentum of B meson in the c.m. frame is small ($p_B^*/m_B \sim 0.06$), the second term can be ignored. Hence, the resolution of M_{bc} is decided by the beam energy resolution which is around $3\,\mathrm{MeV/c^2}$.

Figure 4.3 shows the M_{bc} and ΔE distributions for correctly reconstructed $B^+ \to K^+\mu^+\mu^-$ candidates in the signal MC sample. We retain the B candidates with $|\Delta E| < 0.3\,\mathrm{GeV}$ and $M_{\mathrm{bc}} \in [5.20, 5.29]\,\mathrm{GeV}/c^2$ for further analysis. The event selection criteria are summarised in Table 4.3.

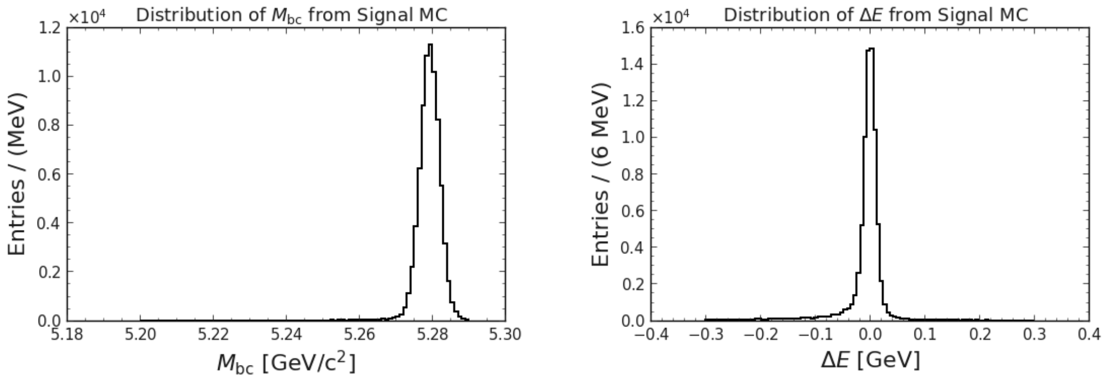


Figure 4.3: M_{bc} and ΔE distributions for correctly reconstructed $B^+ \to K^+\mu^+\mu^-$ candidates in the signal MC sample.

Table 4.3: Summary of event selection criteria.

Variable	Selection
Energy difference	$\lvert\Delta E\rvert < 0.3$ GeV
Beam-energy constrained mass	$5.20 < M_{bc} < 5.29$ GeV/c^2
χ^2 probability of the B vertex fit	$p(\chi^2_{vert}) > 0.001$
# of clean tracks in the event	> 4
($d_0 < 2$ cm, $\lvert z_0 \rvert < 4$ cm, and $p_T > 100$ MeV/c)	

Chapter 5

Backgrounds

A fundamental consideration in any experimental analysis is the meticulous examination of background sources. In this chapter, we delve into the diverse sources of background events that have the potential to mimic or interfere with the $B \to K\ell\ell$ signal. Understanding and effectively mitigating these backgrounds is of paramount importance in safeguarding the precision of our lepton universality measurement.

Before delving into the main content, it is essential to establish a common nomenclature for these backgrounds. We categorize backgrounds into two primary types: combinatorial and peaking. Combinatorial backgrounds show up as random combinations of final-state tracks, which makes ΔE and $M_{\rm bc}$ distributions without any clear peaks. In contrast, 'peaking backgrounds' arise from decays with relatively high branching fractions (B meson decay) that bear some resemblance to the $B \to K\ell\ell$ decay channel. Consequently, they can exhibit discernible peaks in the ΔE and/or $M_{\rm bc}$ distributions.

The chapter is organized as follows: In Section 5.1, we delve into the details of the backgrounds originating from $B \to \psi K$ decays. Section 5.2 talks about the π^0 Dalitz decay and the dilepton background coming from converted photons, as well as the corresponding suppression strategy. Sections 5.3 and 5.4 are dedicated to explaining the continuum and semileptonic B decay backgrounds, along with their characteristics. Their suppression strategies are explored in the subsequent chapter through multivariate analysis. Lastly, Section 5.5 discusses peaking backgrounds and outlines the strategy for their suppression.

5.1 $B \to \psi(nS)K$

We find a substantial contamination arising from $B^+ \to K^+\psi(nS)[\ell^+\ell^-]$ decays with $n = 1$ and 2, where the respective branching fractions are two and one order (Table 5.1) of magnitude larger than our charmless signal.

Table 5.1: Branching fractions of $B \to \psi(nS)K$ with $n \in 1, 2$ decays [9].

Main decay	Sub-decay	Branching fraction
$\mathcal{B}(B^+ \to J/\psi K^+) =$ $(1.020 \pm 0.019) \times 10^{-3}$	$\mathcal{B}(J/\psi \to e^+e^-) =$ $(5.971 \pm 0.032) \times 10^{-2}$	$\mathcal{B}(B^+ \to J/\psi[e^+e^-]K^+) =$ $(6.09 \pm 0.12) \times 10^{-5}$
	$\mathcal{B}(J/\psi \to \mu^+\mu^-) =$ $(5.961 \pm 0.032) \times 10^{-2}$	$\mathcal{B}(B^+ \to J/\psi[\mu^+\mu^-]K^+) =$ $(6.08 \pm 0.12) \times 10^{-5}$
$\mathcal{B}(B^+ \to \psi(2S)K^+) =$ $(6.24 \pm 0.20) \times 10^{-4}$	$\mathcal{B}(\psi(2S) \to e^+e^-) =$ $(7.93 \pm 0.17) \times 10^{-3}$	$\mathcal{B}(B^+ \to \psi(2S)[e^+e^-]K^+) =$ $(4.95 \pm 0.19) \times 10^{-6}$
	$\mathcal{B}(\psi(2S) \to \mu^+\mu^-) =$ $(8.0 \pm 0.6) \times 10^{-3}$	$\mathcal{B}(B^+ \to \psi(2S)[\mu^+\mu^-]K^+) =$ $(4.99 \pm 0.41) \times 10^{-6}$

Since the B decay products for the above charmonium modes are exactly the same as the signal, neither ΔE nor $M_{\rm bc}$ can help discriminate these backgrounds from the signal. As the charged leptons are byproducts of either a J/ψ or a $\psi(2S)$ decay, the respective dilepton invariant-mass distribution should peak around the J/ψ or $\psi(2S)$ mass. The two peaks in the dilepton mass squared ($q^2 = M^2_{\ell^+\ell^-}$) distribution of the generic simulated sample shown in Fig. 5.1 correspond to these two charmonia. The distributions of ΔE and $M_{\rm bc}$ in the respective sample are shown in Fig. 5.2.

In order to suppress the contribution from these modes, we apply a veto on $M_{\ell^+\ell^-}$ around the J/ψ and $\psi(2S)$ mass region. The vetoes are

- $M_{\mu^+\mu^-}(M_{e^+e^-}) \notin [2.96(2.91), 3.19]$ GeV/c^2
- $M_{\mu^+\mu^-}(M_{e^+e^-}) \notin [3.60(3.57), 3.74]$ GeV/c^2.

For the electron mode, we apply the charmonium veto on four kinds of dilepton invariant mass $M_{e^{\rm rec}e^{\rm rec}}$, $M_{e^{\rm unrec}e^{\rm unrec}}$, $M_{e^{\rm unrec}e^{\rm rec}}$, and $M_{e^{\rm rec}e^{\rm unrec}}$, where $e^{\rm rec}$ and $e^{\rm unrec}$ are the brem-recovered and brem-unrecovered electrons, respectively. This helps in reducing background from the events where incorrectly brem-recovered electrons drive the invariant mass to escape the charmonium veto.

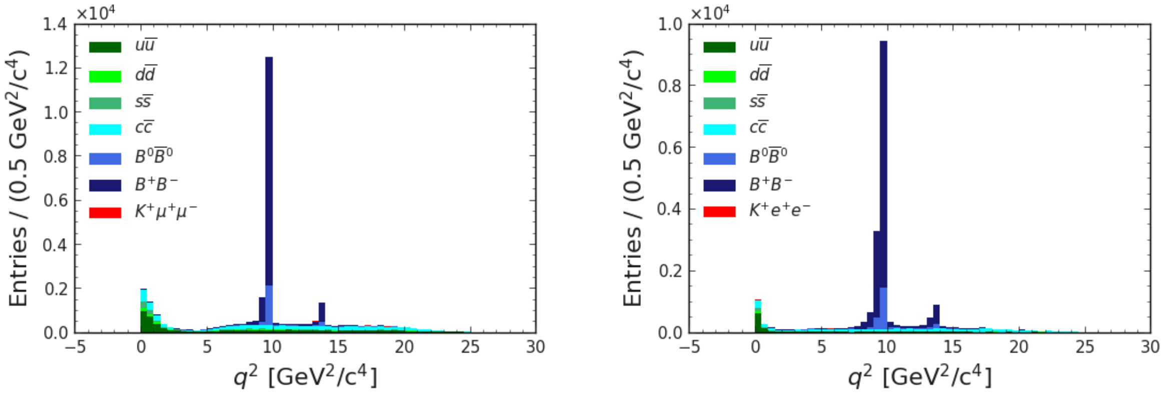


Figure 5.1: q^2 distributions for $B^+ \to K^+e^+e^-$ and $B^+ \to K^+\mu^+\mu^-$ in the generic simulated sample.

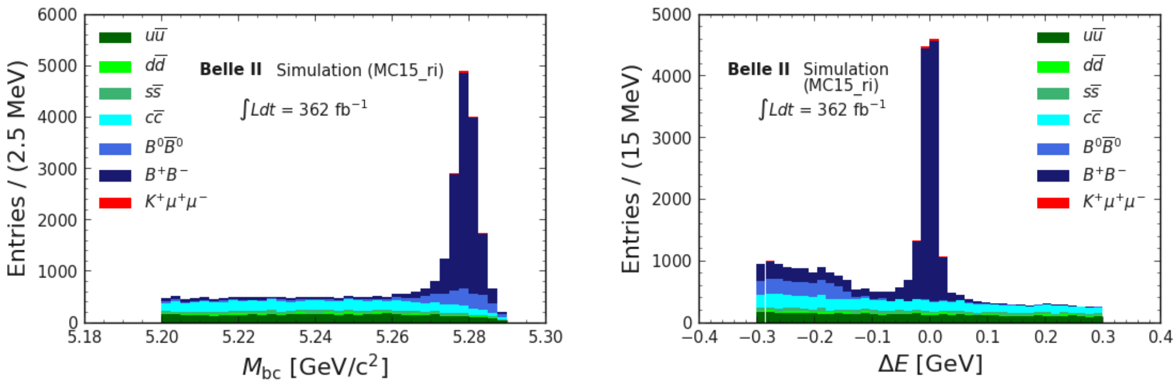


Figure 5.2: M_{bc} and ΔE distributions obtained from the generic MC sample of the $B^+ \to K^+\mu^+\mu^-$ decay, with different colours representing various types of background. The signal (in red) is hardly visible.

The impact of the veto on the distributions of ΔE and M_{bc} for both the electron and muon channels can be seen in Figs. 5.3 and 5.4.

To reconstruct the control sample $B \to J/\psi K$, we apply all selection requirements except for the q^2 veto (and MVA classifier in Sect. 6) focusing on the narrow resonance region $M_{\ell\ell} \in [2.91(2.96), 3.19]$ GeV$/c^2$ for $\ell = e(\mu)$.

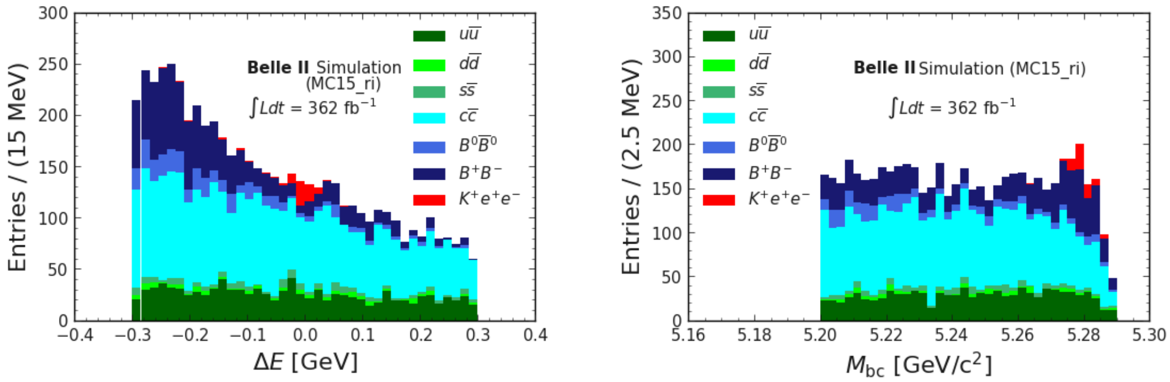


Figure 5.3: ΔE and $M_{\rm bc}$ distributions obtained from the generic MC sample of the $B^+ \to K^+e^+e^-$ channel after applying the veto to remove $B \to \psi(nS)K$ $[n \in 1,2]$, with different colours representing various types of background.

5.2 Converted photon and π^0 Dalitz decay

The electron channel encounters additional sources of background from the π^0 Dalitz decay $\pi^0 \to e^+e^-\gamma$ and photon conversion $\gamma^* \to e^+e^-$ in the detector material. Although the branching fraction of $\pi^0 \to e^+e^-\gamma$ is quite small (1.174%), since the pions are the most dominant (around 70%) particles at Belle II, the Dalitz decay can be a major background for $B^+ \to K^+e^+e^-$. A kaon track may be erroneously combined with the lepton pair from either of the above two processes, leading to a false signal. As shown in Fig. 5.5, such events exhibit a peak in the low dielectron invariant mass, $M_{e^+e^-}$. To mitigate this background, a criterion of $M_{e^+e^-} > 0.15\,\mathrm{GeV}/c^2$, just above the pion mass, is applied.

5.3 Continuum background

The cross-section for $e^+e^- \to \mu^+\mu^-$ can be expressed as:

$$\sigma(e^+e^- \to \mu^+\mu^-) = \frac{4\pi\alpha^2}{3s}(\hbar c)^2. \tag{5.1}$$

Here, α is the fine-structure constant and s is the c.m. energy square. Similarly, the cross section for hadronic events, produced in e^+e^- collisions can be obtained using the relation:

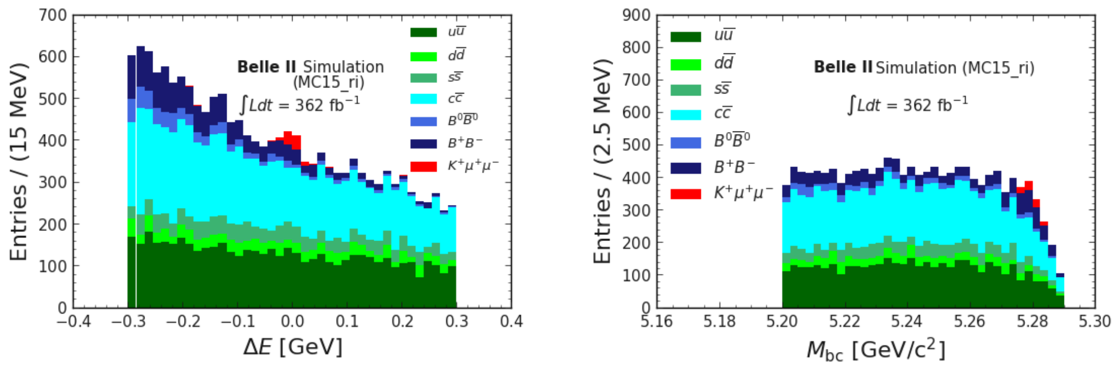


Figure 5.4: ΔE and $M_{\rm bc}$ distributions obtained from the generic MC sample of the $B^+ \to K^+\mu^+\mu^-$ channel after applying the veto to remove $B \to \psi(nS)K$ $[n \in 1, 2]$, with different colours representing various types of background.

$$\sigma(e^+e^- \to \text{hadrons}) = 3 \sum_f z_f^2 \; \sigma(e^+e^- \to \mu^+\mu^-). \tag{5.2}$$

Here, z_f represents the fractional charge of the quark pairs produced in e^+e^- collisions, and the summation over f encompasses the different quark flavors that can be produced given the available $\sqrt{s}$. Additionally, resonance states are produced depending on $\sqrt{s}$, which lead to peaks in the cross-section. A graphical representation of the cross-section of hadronic events in e^+e^- collisions as a function of the c.m. energy is shown in Fig. 5.6.

The term 'continuum' refers to background events that exhibit a continuous distribution of energy or momentum spectra spanning a wide range, e.g., as shown in Fig.5.6. At $\sqrt{s}$ = 10.58 GeV, $e^+e^- \to \Upsilon(4S) \to B\overline{B}$ events are considered part of the Υ resonance, while the rest of the events are classified as continuum. Since Bhabha, dimuon, and other leptonic events are filtered out by our trigger, for this analysis we consider only hadronic continuum events. This continuity of continuum events is utilized for B-meson counting (Appendix A).

The dominant source of background in our analysis is hadronic continuum events, originating from e^+e^- collisions that subsequently undergo a hard scattering process, resulting in pairs of light quarks $q \in (u, d, s, c)$ with much lower mass compared to B mesons. The decay products from these $q\overline{q}$ events can mimic signal, as a random combination of charged tracks from light-quark jets may resemble the final state of our signal decay. The high rate of events originating from this background underscores the importance of efficiently suppressing it.

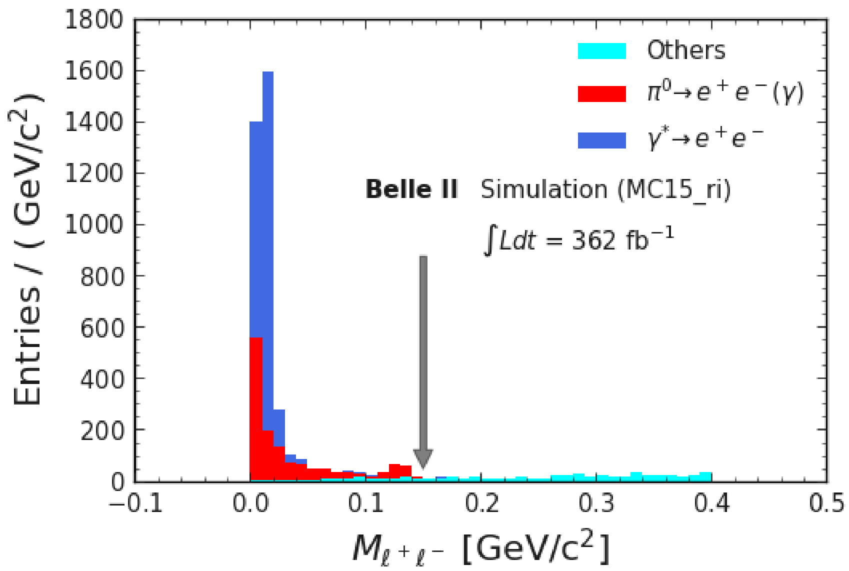


Figure 5.5: $M_{e^+e^-}$ distribution obtained from the simulated sample. The blue and red filled histograms represent the backgrounds coming from a π^0 Dalitz decay and a converted photon, respectively. The cyan filled histogram represents the other backgrounds, primarily originating from η, η', and ω decays. The downward arrow shows the selection applied on the dilepton invariant mass.

The light quarks in $e^+e^- \to q\overline{q}$ continuum events are produced with high momenta in the c.m. frame, resulting in a back-to-back fragmentation of two jets of light hadrons, as shown in Fig. 5.8. On the other hand, B mesons are produced almost at rest in the c.m. frame, resulting in isotropically distributed decay products. Therefore, continuum events can be distinguished from B decays by their event topology. The variables related to event topology are harnessed in a multivariate analysis to distinguish signal from background events, as discussed in the next chapter.

5.4 Semileptonic B decay background

The kinematic constraints of signal candidates can be satisfied by the decay products of $B\overline{B}$ events, resulting in a generic background that mimics our signal. Semileptonic decays of both B mesons in $B\overline{B}$ events can produce leptons that can mimic signal candidates when combined with a kaon, making such random combinations the dominant background due to their higher branching fraction. The branching fraction of semileptonic B decays, such as $B \to D^{(*)}\ell\nu$, is much larger (Table 5.2) than that of the signal $B \to K\ell^+\ell^-$. In addition, a less significant back-

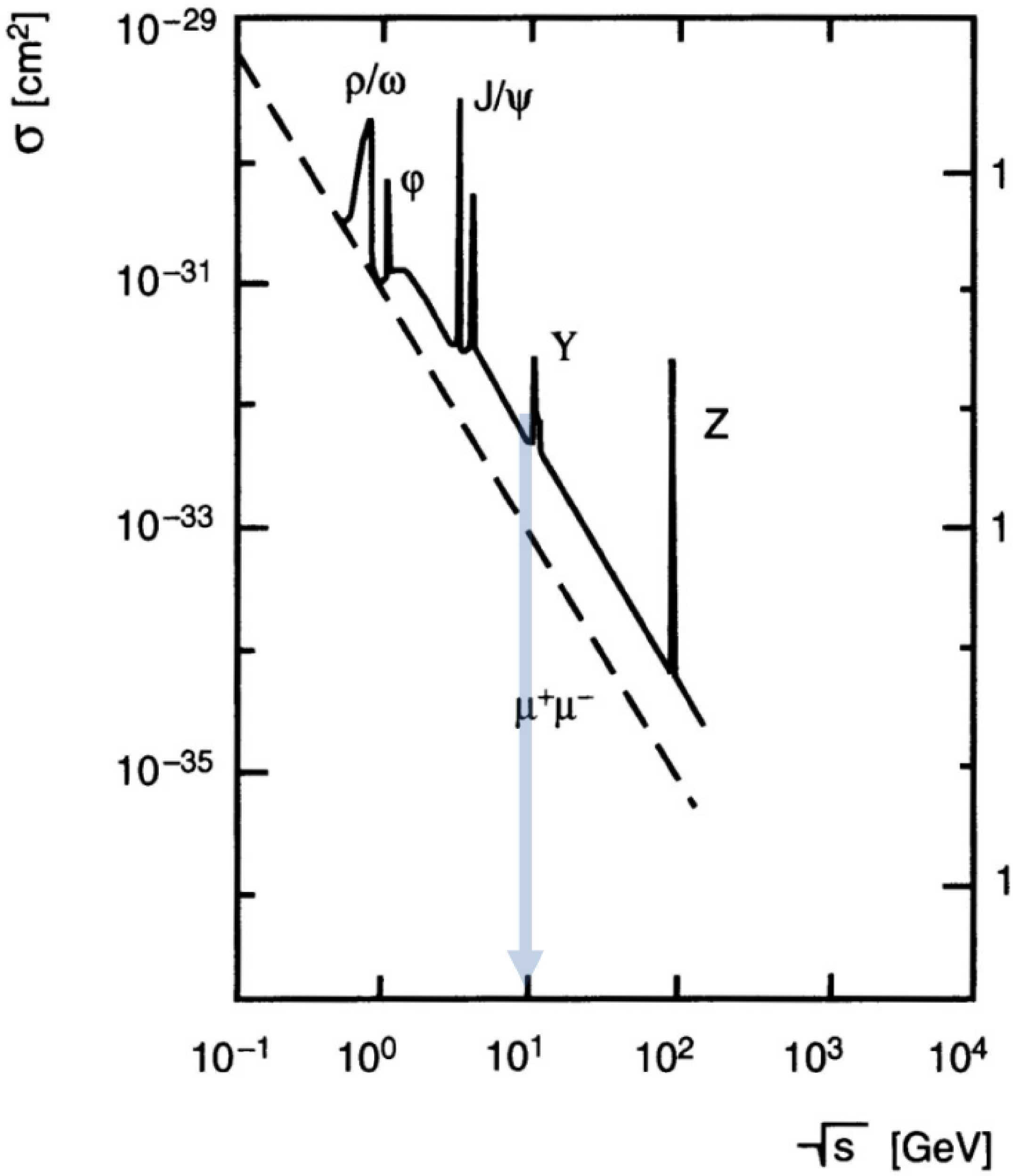


Figure 5.6: Cross-section of e^+e^- collisions resulting in hadronic events as a function of the c.m. energy ($\sqrt{s}$). The dashed line in the graph represents the $e^+e^- \to \mu^+\mu^-$ process, while the downward arrow points to the c.m. energy corresponding to SuperKEKB. At this energy value, the cross-section exhibits a peak due to the presence of Υ resonances. The figure is from Ref. [6]

ground can arise from one-sided cascade semileptonic B decay, where two leptons are produced in the decay of B and D mesons, and they can also appear as signal candidates when combined with a kaon. A schematic diagram of the semileptonic B decay background is presented in

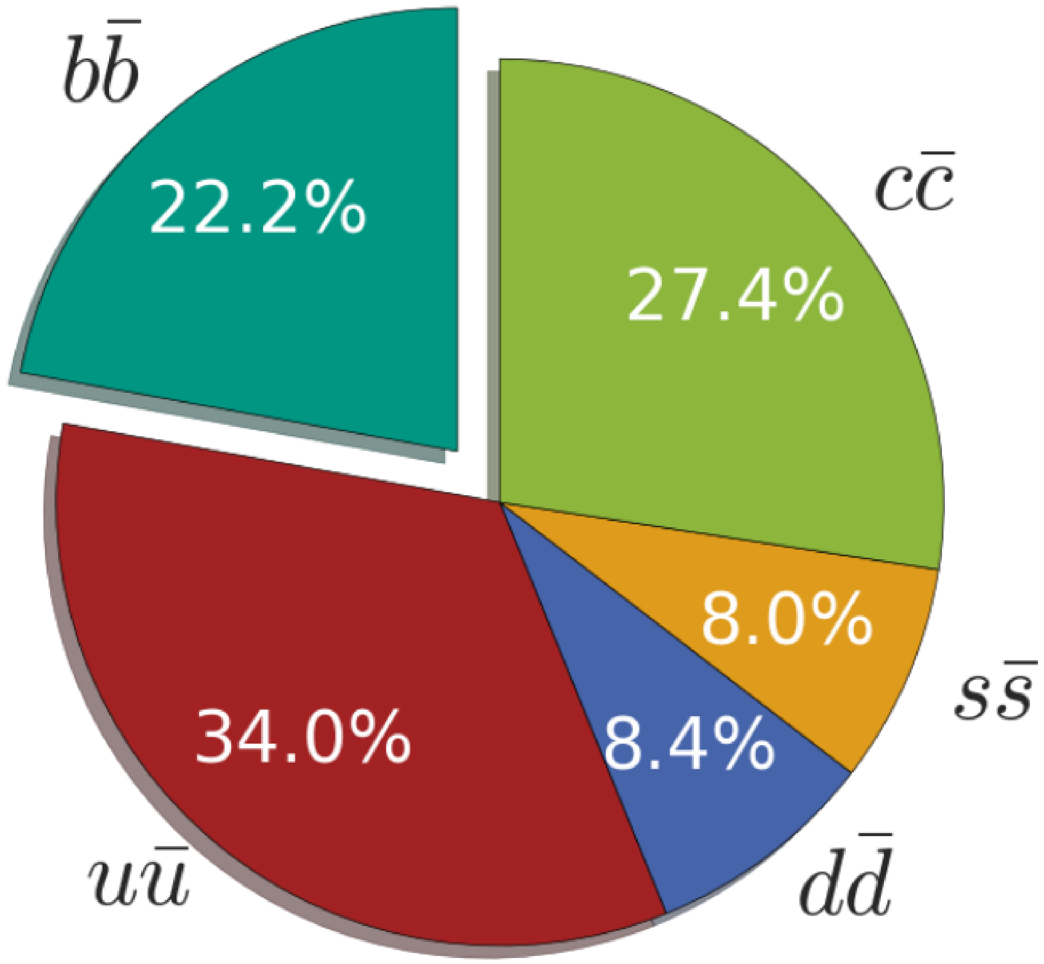


Figure 5.7: Pie chart illustrating the cross-section distributions of various continuum events and $B\overline{B}$ pairs originating from the $\Upsilon(4S)$ resonance(shown by $b\bar{b}$) in e^+e^- collisions at $\sqrt{s}$ = 10.58 GeV. Adapted from Ref. [7].

Fig 5.9.

The generic B decay background mostly originates from events where both B mesons decay semileptonically. Signal B candidate for this type of background is formed from the decay products of both B mesons, which include at least one lepton from each decay. However, not all charged tracks can meet the requirement that all three of them originate from a single point, in contrast to the signal decay. This is due to the B and D mesons travelling some distance before decaying. To leverage this observation, vertex-related variables are employed in a multivariate classifier, aiding in the suppression of these backgrounds, as discussed in the next chapter.

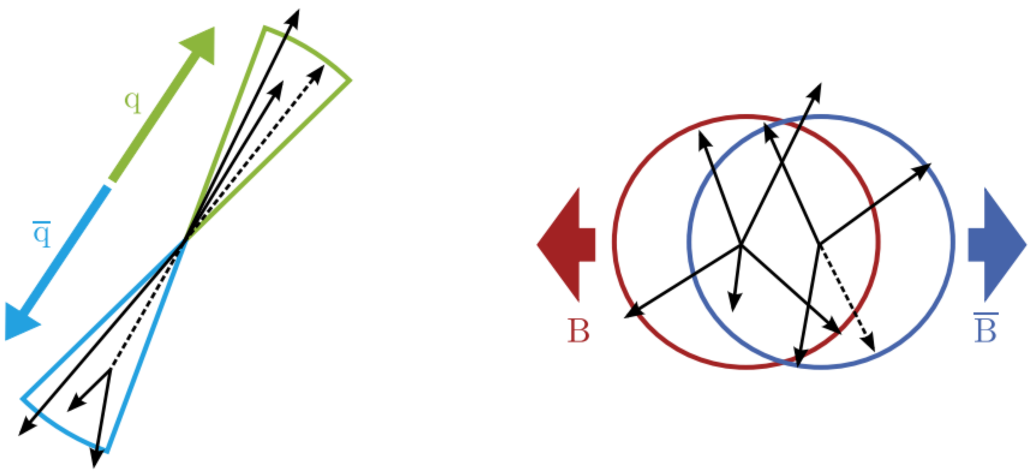


Figure 5.8: Event topology of (left) jet-like $e^{+}e^{-} \to q\bar{q}$ and (right) spherical $e^{+}e^{-} \to \Upsilon(4S) \to B\overline{B}$ event.

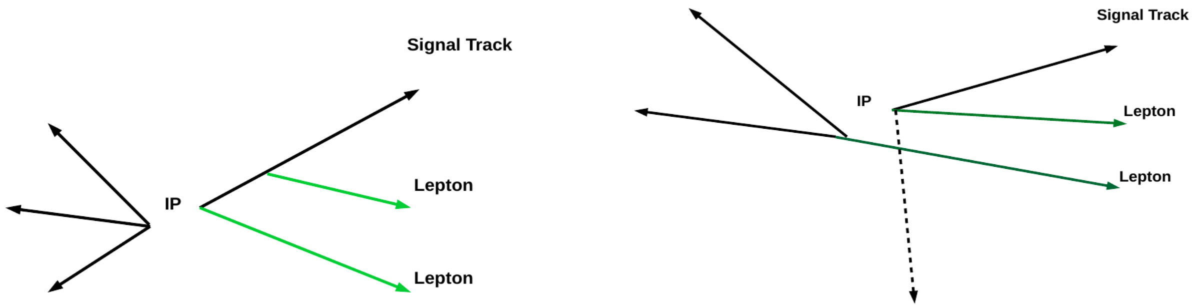


Figure 5.9: Schematic diagram of (left) single-sided semileptonic $B\overline{B}$ decay and (right) both-sided semileptonic $B\overline{B}$ decay events, that have potential to fake signal events.

5.5 Other peaking backgrounds

As discussed earlier, peaking backgrounds are a type of background that closely resemble the signal, rather than being a random combination of tracks in the event. These backgrounds arise from B-meson decays into certain final states that can be erroneously identified as the final state of the signal. Two types of peaking backgrounds are discussed below.

5.5.1 Reducible peaking background

The types of peaking backgrounds that can be reduced by applying a veto are known as reducible peaking backgrounds. Below we list the possible reducible peaking backgrounds for different channels:

Table 5.2: Branching fractions of some dominant B decays.

Decay	Branching fraction
$B^+ \to \overline{D}^{*0}\mu^+\nu_\mu$	$(5.58 \pm 0.22)\%$
$B^+ \to \overline{D}^{0}\mu^+\nu_\mu$	$(2.30 \pm 0.09)\%$
$B^0 \to D^{*-}\mu^+\nu_\mu$	$(4.97 \pm 0.12)\%$
$B^0 \to D^{-}\mu^+\nu_\mu$	$(2.24 \pm 0.09)\%$

$B^+ \to K^+\mu^+\mu^-$:

- $B^+ \to \overline{D}^0(\to K^+\pi^-)\pi^+$ with both pions being misidentified as muons. In Fig. 5.10 (left) we can see the distribution of invariant mass ($M_{K^+\mu^-}$) of the $K^+\mu^-$ system. We use the pion mass hypothesis for the muon candidate while calculating $M_{K^+\mu^-}$. A peak in the $M_{K^+\mu^-}$ distribution near the D^0 mass [9] can be seen, implying the contribution from $B^+ \to \overline{D}^0(\to K^+\pi^-)\pi^+$. To suppress it, we apply a veto on $M_{K^+\mu^-}$ corresponding to a $\pm 3\sigma$ window around the known D^0 mass: $M_{K^+\mu^-} \notin [1.849, 1.881]\,\mathrm{GeV}/c^2$.

- $B^+ \to J/\psi(\to \mu^+\mu^-)K^+$ with the kaon being misidentified as a muon and vice versa. This background can be removed by applying a veto on the invariant mass ($M_{K^+\mu^-}$) of the $K^+\mu^-$ system around the J/ψ mass: $3.06 < M_{K^+\mu^-} < 3.13\ \mathrm{GeV}/c^2$. We change the mass hypotheses of kaon candidate to muon mass while calculating $M_{K^+\mu^-}$. Figure 5.10 (right) shows the distribution of $M_{K^+\mu^-}$ for such double misidentification case.

$B^+ \to K^+e^+e^-$:

- $B^+ \to \overline{D}^0(\to K^+\pi^-)\pi^+$ with both pions being misidentified as electrons. We use the pion mass hypothesis for the electron candidate while calculating $M_{K^+e^-}$. A peak in the $M_{K^+e^-}$ distribution near the D^0 mass can be seen, implying the contribution from $B^+ \to \overline{D}^0(\to K^+\pi^-)\pi^+$. To suppress it, we apply a veto on $M_{K^+e^-}$ corresponding to a $\pm 3\sigma$ window around the known D^0 mass: $M_{K^+e^-} \in [1.849, 1.881]\,\mathrm{GeV}/c^2$.

5.5.2 Irreducible peaking backgrounds

These backgrounds arise from various sources, typically involving wide resonances (Fig 5.11), where implementing a veto is not practical. Instead, we count them in respective MC samples and estimate their yields in data by applying a correction factor. The $B^+ \to K^+\pi^+\pi^-$ decays contribute as these backgrounds due to misidentifications of $\pi \to e$ and $\pi \to \mu$. To estimate their contributions, we generate a signal sample of 10 million events for various channels, as listed in Table 5.3, and reconstruct them to count the number of peaking backgrounds surviving all selection criteria. The ΔE and $M_{\rm bc}$ distributions are shown in Fig. 5.12. As the mass difference between pions and muons is not significant, $B^+ \to K^+\pi^+\pi^-$ can contribute as a peaking background. The $\pi \to \mu$ fake-rate is larger than that of $\pi \to e$, resulting in 5.9 ± 0.2 (0.02 ± 0.01) such backgrounds in the muon (electron) channel in the overall q^2 bin. Table 5.4 summarizes the number of $B^+ \to K^+\pi^+\pi^-$ backgrounds in different q^2 bins for the muon channel. For the electron channel, the number of such peaking backgrounds is consistent with zero. To estimate the peaking background in data, we multiply the MC yield by a correction factor, which is the square of the $\pi \to \mu$ fake-rate calculated from inclusive $K_S^0 \to \pi^+\pi^-$ and $e^+e^- \to \tau^\pm(1P)\tau^\mp(3P)$ events in two-dimensional bins of momentum and polar angle.

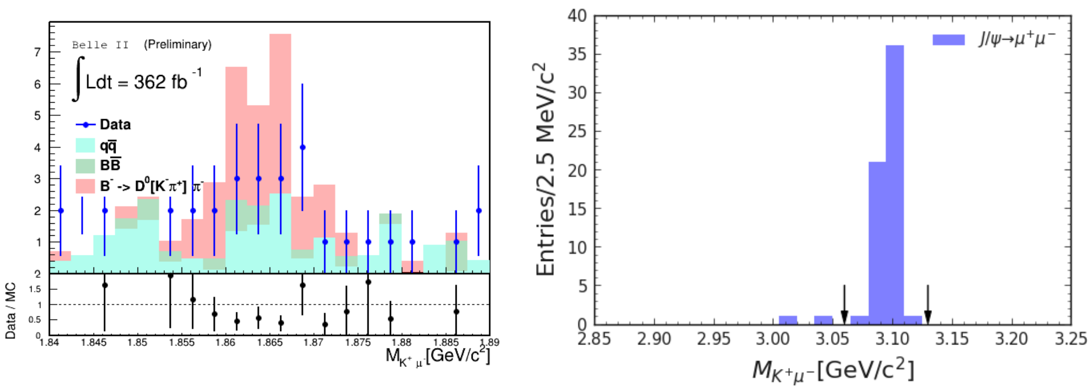


Figure 5.10: Distributions of $M_{K^+\mu^-}$ with (left) the mass hypothesis of the muon candidate replaced by the pion mass in the data and MC sample and (right) the mass hypothesis of the kaon candidate replaced by the muon mass in MC sample.

Table 5.3: List of various decay channels contributing to $B^+ \to K^+\pi^+\pi^-$. Their branching fractions and uncertainties are taken from PDG [9].

channel	Branching fraction ($\times10^{-6}$)
$B^+ \to K_0^*(1430)^0\pi^+$	39.05 ± 8
$B^+ \to K^+\pi^+\pi^-$	16.30 ± 8
$B^+ \to K_2^*(1430)^0\pi^+$	5.56 ± 0.1
$B^+ \to K^*(892)^0\pi^+$	10.06 ± 0.8
$B^+ \to K^+ f_0(980)$	9.2 ± 1.6
$B^+ \to K^+\rho(770)^0$	3.7 ± 0.5

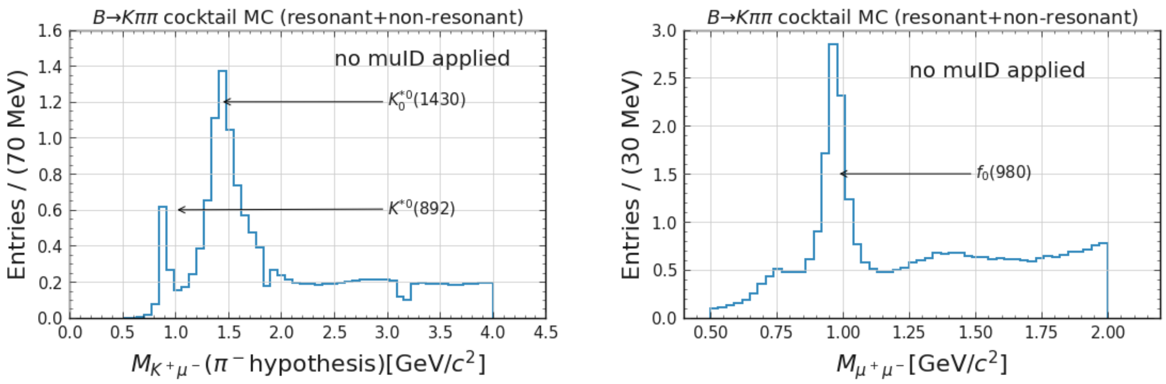


Figure 5.11: Invariant-mass distributions of the $K^+\mu^-$ and $\mu^+\mu^-$ systems without any muonID selection. The histograms are scaled to unit area and wide resonance peaks are shown by arrows.

Table 5.4: Irreducible $B \to K\pi\pi$ peaking background in the $B^+ \to K^+\mu^+\mu^-$ channel.

q^2 bin (GeV^2/c^4)	$n_{\text{peak}}^{\text{MC}}$	$f^{\text{data/MC}}(\pi^+ \to \mu^+) \times f^{\text{data/MC}}(\pi^- \to \mu^-)$	$n_{\text{peak}}^{\text{data}}$
$\in [0.1, 8.12]$	1.90 ± 0.09		1.09 ± 0.07
> 14.18	1.40 ± 0.08	$(0.77 \pm 0.02) \times (0.75 \pm 0.02)$	0.81 ± 0.05
full-q^2	3.30 ± 0.12		1.91 ± 0.15

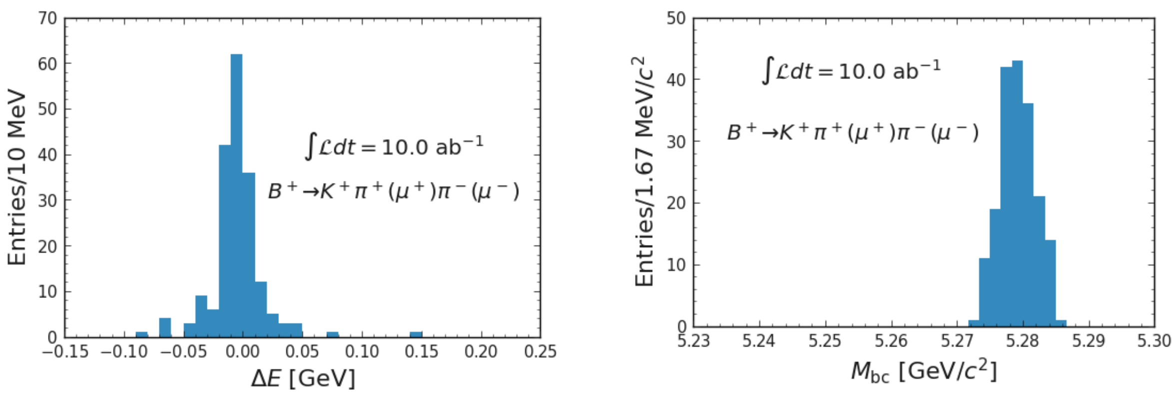


Figure 5.12: (Left) ΔE and (right) M_{bc} distributions for the $K^+\pi^+\pi^-$ peaking background reconstructed as $B^+ \to K^+\mu^+\mu^-$.

Chapter 6

Multivariate analysis

In this chapter, we unravel the intriguing world of multivariate analysis (MVA), a robust tool in our toolkit at Belle II for distinguishing the signal we seek from backgrounds. Think of it as a sophisticated filter that helps us extract meaningful information from the data.

We'll delve into the basics of implementing MVA and its critical role in improving the precision of our measurements. As we explore the implementation, fine-tuning, and outcomes of MVA, we'll gain insights into how it enhances our data analysis at Belle II.

The chapter is structured as follows: Sections 6.1 and 6.2 provide concise introductions to MVA and boosted decision trees. Sections 6.3 and 6.4 delve into feature selection and the training of the dataset. Moving forward, Sections 6.5, 6.6, and 6.7 comprehensively explore the concepts of bias-variance trade-off and tools employed to check for overtraining. Section 6.8 is dedicated to the optimization of hyperparameters, while Sections 6.9 and 6.10 address the performance of MVA and the techniques applied for background suppression.

6.1 Introducing MVA

MVA is a powerful technique in the field of data analysis that involves examining the discrimination power of multiple variables simultaneously. Unlike traditional univariate methods, which consider only one variable at a time, MVA allows us to analyse complex relationships and patterns within datasets. In the realm of particle physics, where intricate interactions occur, MVA plays a crucial role in distinguishing signals from background events.

MVAs are particularly valuable when dealing with high-dimensional data. By exploring correlations among variables, they provide a holistic understanding of the datasets, enabling researchers to make more accurate predictions and classifications. Boosted Decision Trees, an example of MVA used in our analysis, harness the collective power of multiple variables to enhance discrimination between the signal and background events.

6.2 Introducing Boosted Decision Tree (BDT)

For our analysis, the "Stochastic gradient-boosted decision tree" is used. The implementation of this classifier in the Belle II software is explained in Ref. [57]. This section will briefly explain this multivariate classifier.

- Decision Tree: A decision tree is a fundamental machine learning algorithm used for both classification and regression tasks. It is a graphical representation of a decision-making process that resembles an inverted tree, where each internal node represents a decision based on a specific feature and each leaf node represents a prediction or outcome. Decision trees are known for their simplicity and interpretability, making them valuable for understanding the logic behind a classification or regression model. The tree is constructed by recursively partitioning the dataset into subsets based on the best features, which ultimately results in a set of rules for making predictions. A schematic diagram of the decision tree is shown in Fig. 6.1.

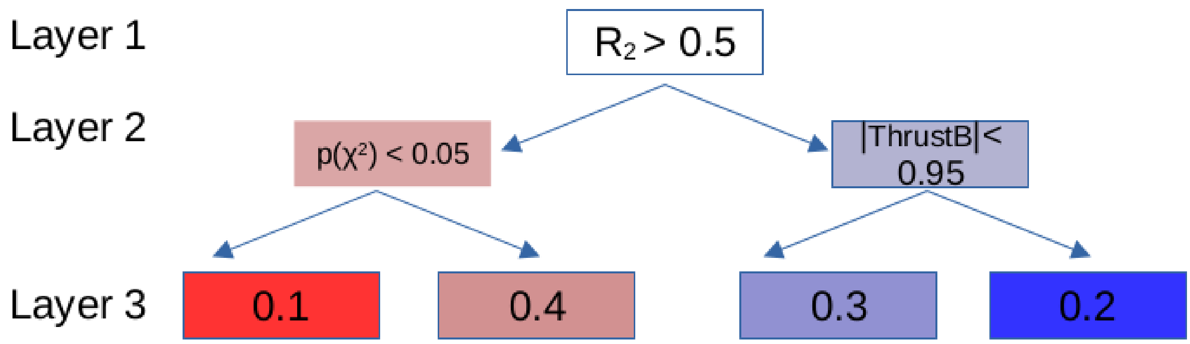


Figure 6.1: Three-Layer Decision Tree: as test data points, the label of which is unknown, move through the tree, the tree progresses from the top to the bottom. At each node within the tree, a binary decision is made until the traversal reaches a terminal node. The probability that the test data points belong to the signal category is determined by the signal fraction, which is the ratio of the number of training data points (with known labels) that ultimately end up in the same terminal node and the total number of training data points.

- Boosted Decision Trees: Boosted decision trees are an ensemble learning method that combines multiple decision trees to improve predictive accuracy. This technique iteratively builds a series of decision trees, with each subsequent tree focusing on correcting the errors made by the previous ones. Boosted decision trees are highly effective for handling complex datasets and are known for their robustness and ability to capture intricate patterns in the data. They are widely used in various fields, including machine learning competitions and real-world applications.

- Gradient Decision Trees: Gradient decision trees are a variant of decision trees that leverage gradient-based optimization techniques to enhance their performance. Traditional decision tree algorithms, like CART (Classification and Regression Trees), use a greedy approach to select the best feature to split the data at each node. Unlike traditional decision tree algorithms, gradient decision trees incorporate gradients derived from a loss function with respect to predictions made by the current ensemble of decision trees. This allows for more informed decisions during tree construction, leading to faster convergence and more accurate models, which is especially beneficial in handling large datasets.

- Stochastic Gradient Decision Trees: Stochastic gradient decision trees combine the principles of stochastic gradient descent (SGD) with decision trees. The SGD is an optimization technique commonly used in training machine learning models. In the context of decision trees, stochastic gradient methods introduce randomness into the tree-building process. Instead of considering the entire dataset at each split, a random subset (mini-batch) is used, which can lead to a faster training and improved generalization. The SGD trees are particularly useful when dealing with massive datasets that do not fit into memory.

6.3 Selection of features

The MVA classifier is meant to effectively reduce the large amounts of background coming from continuum and semileptonic B decay processes. This action follows the initial selection, which has been discussed in Chapter 5 and can be visually assessed in Figs. 5.3 and 5.4.

The detailed characteristics of these backgrounds have been thoroughly explored in Chapter 5. Utilizing insights gained from the analysis of these properties, specific features have been meticulously chosen to train the classifier. The distributions of the three variables per channel are shown in Figs. 6.2 and 6.3, and additional top-ranked features are detailed in Appendix B.

6.3.1 Features to suppress continuum background

As elaborated upon in Section 5.3, distinguishing continuum events from B decays hinges on their distinct topologies. In order to tackle the continuum background, we incorporate a set of following 22 event-shape variables into the MVA framework as features

- KSFW moments: We use Kakuno's Super Fox-Wolfram [58] (KSFW) moments H^{so}_{xl} and H^{oo}_{l}. All the particles in an event are divided into two categories: reconstructed signal B decay products denoted as 's' and those from the rest-of-event (ROE) denoted as 'o'. They are further decomposed into three categories: charged ($x = 0$), neutral ($x = 1$), and missing ($x = 2$), where the missing momentum is treated as a particle. The KSFW moments are then defined using the relation:

$$H^{so}_{xl} = \sum_{i}\sum_{j,x} |\vec{p}_{jx}| P_l(\cos\theta_{i,jx}),$$

  where i runs over the B decay products, j and x run over the ROE in the category x, $\vec{p}_{jx}$ is the momentum of the particle jx, and $P_l(\cos\theta_{i,jx})$ is the l^{th} order Legendre polynomial of the cosine of the angle between particles i and jx. Similarly, we define

$$H^{oo}_{l} = \begin{cases} \sum_{i}\sum_{j,x} |\vec{p}_{jx}| P_l(\cos\theta_{i,jx}), & l = \text{even} \\ \sum_{j}\sum_{k} Q_j Q_k |\vec{p}_j||\vec{p}_k| P_l(\cos\theta_{j,k}), & l = \text{odd} \end{cases}$$

  where j and k run over the ROE, and Q_j and Q_k are the charges of j^{th} and k^{th} particles, respectively. We used 10 out of 18 KSFW moments. The remaining eight KSFW moments are not used because of their high correlation with higher ranking features. The KSFW moments used are listed below.

  - H^{so}_{14}
  - H^{oo}_{1}
  - H^{so}_{20}
  - H^{so}_{04}

- H^{so}_{24}
- H^{so}_{00}
- H^{so}_{01}
- H^{so}_{22}
- H^{so}_{12}
- H^{so}_{02}

- CLEO cones: These variables [59], introduced by the CLEO collaboration, are based on the sum of the magnitudes of momenta of all particles within angular regions around the thrust axis in intervals of 10°, resulting in nine concentric cones. We skip the first two CleoCones because of their high correlation with the kinematic variable $M_{\rm bc}$.

- Thrust: The magnitude of thrust for the reconstructed B candidate, where the thrust is defined as:
$$T = \max_{\hat{n}} \frac{\sum_i \vec{p}_i.\hat{n}}{\sum_i |\vec{p}_i|}.$$
Here $\vec{p}_i$ is the momentum of the $i^{\rm th}$ decay product of the B candidate and $\hat{n}$ is a unit vector along the direction that maximizes the total longitudinal projection of momenta.

- $\cos\theta_{T_B,T_{\rm ROE}}$: The cosine of the angle between the thrust axis of signal B and that of ROE.

- $\cos\theta^*_B$: The cosine of the polar angle of B momentum, calculated in the c.m. frame. As the $\Upsilon(4S)$, a spin-1 particle, decays into $B\overline{B}$, two spin-0 particles, $\cos\theta^*_B$ should be distributed as $\sin^2\theta^*_B = 1 - \cos^2\theta^*_B$. Whereas for continuum events it should be distributed flat because of the random combination of final-state particles.

- R_2: The normalized second Fox-Wolfram moment defined as:
$$R_2 = \frac{H_2}{H_0}, \;\text{ with }\; H_k = \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_k(\cos\theta_{ij}),$$
where $\vec{p}_i$ ($\vec{p}_j$) is the momentum of the $i^{\rm th}$ ($j^{\rm th}$) decay product of the reconstructed B candidate, θ_{ij} is the angle between $\vec{p}_i$ and $\vec{p}_j$, and P_k is the $k^{\rm th}$-order Legendre polynomial.

- ΔZ_{BB}: The separation between the signal-side B vertex and that of the tag-side B along the z axis. Though this is not a topological variable; rather it exploits the difference in lifetime between prompt continuum events and $B\overline{B}$ decays containing B mesons.

6.3.2 Features to suppress semileptonic B backgrounds

As elucidated in Section 5.4, the discrimination of semileptonic B decay events can be achieved through the utilization of vertex-related variables. In the MVA classifier designed for this purpose, we incorporate a set of following seven vertex-related features.

- $p(\chi^2_{\text{vert}})$: The probability of the vertex-fit χ^2 is a good variable to discriminate fake events from semileptonic B decays. It is also useful in separating correctly reconstructed signal from misreconstructed background.
- $\Delta Z_{\ell\ell}$: The separation between two leptons along the z axis. This is useful in suppressing background events from semileptonic B decays, where both leptons do not necessarily come from the reconstructed B candidate.
- dr and dz: The transverse and longitudinal distance between the B vertex and the IP.
- The following variables are used to exploit the presence of kaon candidates originating from D^0 decays, specifically $D^0 \to K^-\pi^+$, which predominantly arises from semileptonic B decays. These variables also help distinguish genuine kaon candidates from other particles that may mimic signal kaon candidates.

  We define the D^0 candidates by combining signal kaon with pion candidates from the remaining tracks (other than signal or ROE tracks). Subsequently, a vertex fit is executed on the reconstructed D^0 candidates. These candidates are then ranked based on the quality of their vertex fit, which provides the valuable information for identifying and selecting the most reliable candidate. The feature list includes:

  - $p(\chi^2)[D^0(ROE)]$: the probability of the best D^0 vertex-fit χ^2 ;
  - Median-$p(\chi^2)[D^0(ROE)]$: the median χ^2 probability of the vertices within a specific event,
  - $M[D^0]$: the reconstructed mass of $D^0 \to KX$ for the optimal D^0 vertex, where, the symbol X denotes the mass calculated based on the most probable particle hypothesis. It is important to note that the reconstruction of the D^0 vertex involves combining the signal kaon candidate with the ROE tracks assuming the most likely particle hypothesis, rather than the default pion hypothesis.

During the feature selection process, we carefully examine the correlations among different variables. If this correlation exceeds 70%, we exclude the less important feature to prevent the redundancy in our analysis.

Furthermore, we conduct a thorough comparison of the shape between the data and simulation within the background dominated region, specifically in the $q^2 \in J/\psi$ range. Any feature that displays a poor agreement between the two is omitted from further consideration. We correct the signal efficiency of nonresonant $B^+ \to K^+\ell^+\ell^-$ events for potential mismodeling of the feature variables, using $B^+ \to K^+J/\psi(\ell^+\ell^-)$ events. This process is discussed in detail in Section 10.9.

Additionally, we pay close attention to the correlation of each feature with $M_{\rm bc}$ and ΔE. Features exhibiting a correlation exceeding 10% with these two kinematic variables are not included in the analysis, as it can potentially introduce unwanted bias.

6.4 Training dataset

For the purpose of suppressing continuum and $B\overline{B}$ backgrounds, we employ a single MVA for each channel. The training and testing samples consist of 600 and 400 fb^{-1} generic run-independent MC events, respectively. We utilize MC truth information to eliminate the contribution from possible peaking background while training the MVA. The primary modes that are removed are residual charmonium events ($B \to J/\psi X$), $B^+ \to \overline{D}^0[K^+\pi^-]\pi^+$ events, and $B^+ \to K^+\pi^+\pi^-$ events as listed in Table 5.3. The motivation for removing these peaking backgrounds, which have similar characteristics to signal, is to reduce the probability of the MVA erroneously classifying signal as background events.

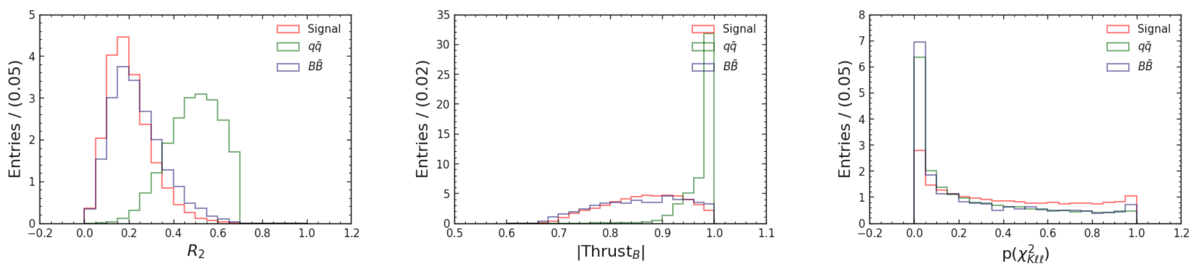


Figure 6.2: MC distributions of some feature variables for $B^+ \to K^+e^+e^-$. Red, green, and blue histograms correspond to signal, continuum, and generic B background, respectively. The histogram area are normalized to unity.

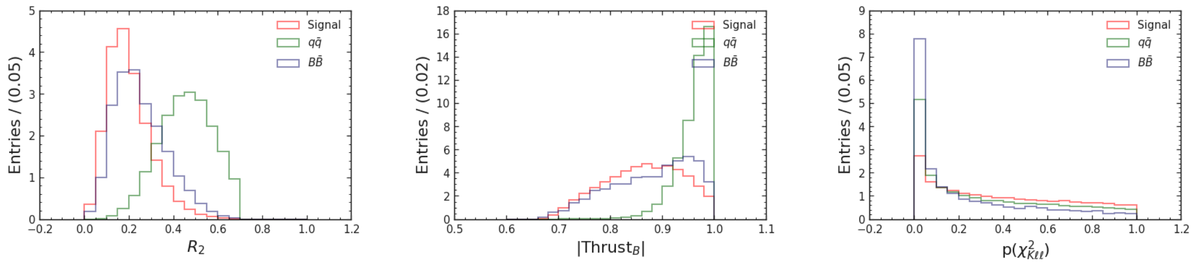


Figure 6.3: MC distributions of some feature variables for $B^+ \to K^+\mu^+\mu^-$. Red, green, and blue histograms correspond to signal, continuum, and generic B background, respectively. The histogram area are normalized to unity.

6.5 Bias variance trade off

This is a fundamental phenomenon in machine learning. For a given dataset, a specific model complexity will yield optimal performance and robust predictions for the unseen data. Let's illustrate this concept with a schematic diagram as shown in Fig. 6.4. In this plot, the blue and red dots represent true signal and background points in the two-dimensional x-y feature space.

In the top plot, the decision boundary is suboptimal, leading to a significant number of misclassified points due to a high bias. Even increasing the training sample size won't improve performance because of this bias. This situation is known as undertraining, and it can be addressed by introducing more model parameters to better capture the underlying data patterns.

In the bottom-left plot of the same Fig. 6.4, the black decision boundary does a reasonable job of separating signal and background data points. It slightly increases the number of model parameters to create a smoother boundary, though some misclassifications can be attributed to dataset fluctuations. With more training data, the performance may further improve.

Finally, in the bottom-right plot, the decision boundary is overly complex, accurately classifying each data point but with many parameters. However, this complexity makes the model less robust when making predictions on the unseen data. It tends to overfit by overly considering statistical fluctuations, resulting in high variance. This condition is referred to as overtraining.

The relationship between the model complexity and the bias-variance trade-off is illustrated in Fig. 6.5. A simpler model, characterized by fewer parameters, tends to exhibit a higher bias but a lower variance. In this scenario, both the training error (the disparity between the model's predictions and the actual training dataset values) and the testing error (evaluating the model's ability to generalize to unseen data) are relatively large. As we incrementally increase the model complexity, the training error diminishes, potentially reaching zero with an appropriate decision boundary that correctly classifies all training data points. Conversely, increasing complexity reduces bias, making underfitting optimal. However, there comes a point where testing error reaches a minimum and then begins to rise due to overfitting, even though bias decreases. This phenomenon highlights the intricate interplay between model complexity, training and testing errors.

6.6 ROC-AUC and optimized classifier

The Receiver Operating Characteristic Area Under the Curve (ROC AUC) is a performance metric used to evaluate the quality of a binary classification model. It quantifies the model's ability to distinguish between the positive and negative classes.

The ROC curve is a graphical representation of the true positive rate (sensitivity) against the false positive rate (1−specificity) at various threshold settings for the model's predictions. The ROC AUC score is a single numerical value that captures the overall performance of the model. A perfect model would have an ROC AUC score of 1, indicating that it can perfectly distinguish between the two classes, while an ineffective model would have an ROC AUC score of 0.5, equivalent to random guessing. Generally, higher ROC AUC scores signify better classification performances.

The ROC-AUC score is inversely related to the classification error. A higher ROC-AUC score signifies a better discrimination between classes, leading to a lower classification error rate. Conversely, a lower ROC-AUC score suggests a weaker class separation and, consequently, a higher classification error rate. The ROC-AUC measures a model's ability to distinguish between positive and negative instances, with higher scores indicating reduced classification errors.

As discussed in Section 6.5, when the test error is minimized, it is equivalent to maximizing the test ROC-AUC score, indicating the optimal point for the classifier. Beyond this optimal point, the training ROC-AUC score begins to increase while the test ROC-AUC score decreases, causing a gap to form between the scores of the two datasets. This understanding serves as the basis for optimizing the classifier by tuning hyperparameters, which control the model's complexity.

6.7 KS test

The Kolmogorov-Smirnov (KS) test is a commonly used statistical method to identify overfitting. It is a hypothesis test that quantifies the maximum vertical distance between the empirical cumulative distribution function (ECDF) of two distributions. If the calculated test statistic, known as the KS statistic, exceeds a threshold value based on the chosen significance level, it suggests that the dataset significantly deviates from the expected distribution, leading to the rejection of the null hypothesis. Conversely, if the statistic falls below the threshold value, it implies a good fit between the dataset and the expected distribution, resulting in the acceptance of the null hypothesis. To see how well an MVA classifier can predict unknown datasets, we check to see if the KS p-values for the training and testing classifier output distributions exceed 10%. This helps detect potential overtraining.

6.8 Optimization

To improve the performance of the classifier, we optimize its hyperparameters. The hyperparameters are parameters of a machine learning model that are not learned from the training data but are set prior to the training process. They govern various aspects of the model's behaviour and influence how the latter learns from the data. The hyperparameters are distinct from model parameters, which are the internal variables learned during training to make predictions. In our study, the following hyperparameters are optimized:

- nTree (Number of Trees): This hyperparameter defines the number of decision trees that are sequentially trained and added to the ensemble during the boosting process. A larger value of 'nTree' can improve the model's ability to fit the training data but may increase the risk of overfitting.
- nLevel (Maximum Depth of Trees): 'nLevel' represents the maximum depth or levels allowed for each individual decision tree in the ensemble. Limiting the depth helps prevent overfitting and can lead to a more robust model. However, setting it too low may result in underfitting.
- Learning Rate (or Shrinkage Rate): The learning rate is a crucial hyperparameter that controls the contribution of each individual tree to the final prediction. A smaller learning

rate means that each tree's contribution is diminished, and the training process is slower but more stable. A larger learning rate leads to faster convergence but can make the model overly sensitive to noise and overfitting.

We conduct a grid search to optimize the model's hyperparameters, as depicted graphically in Figure 6.6 for the $B^+ \to K^+\mu^+\mu^-$ channel classifier. The grid search showcases ROC-AUC scores and KS p-values, offering insights into the classifier's performance on training (in red) and testing (in blue) data, as well as signal (in green) and background (in yellow) samples across various hyperparameter combinations.

When determining the final hyperparameters, we consider three key factors. First, the goal is to maximize the test ROC-AUC score to minimize bias or underfitting. Second we aim at minimizing the disparity between ROC-AUC scores for training and testing datasets to prevent overfitting. Finally, we strive for a classifier that maintains similarity between the $B^+ \to K^+e^+e^-$ and $B^+ \to K^+\mu^+\mu^-$ channels. The optimized hyperparameters employed during training are listed in Table 6.1.

Table 6.1: The hyperparameters used for the two classifiers.

Hyperparameter	$B^+ \to K^+e^+e^-$ classifier	$B^+ \to K^+\mu^+\mu^-$ classifier
nTree	40	50
nLevel	3	3
Learning rate	0.08	0.10

The relationship between sample size, model flexibility, and the transition from high variance to high bias regions influences the selection of different hyperparameters for the electron and muon channels. With a fixed model flexibility, increasing sample size shifts scores from high variance to high bias. Within the high variance region, the ROC-AUC score for the training sample becomes notably higher compared to the test sample, indicating a risk of overtraining. In this high-bias region, the train and test scores agree, but the ROC-AUC score can be further enhanced by introducing more model complexity or increasing flexibility. Conversely, with a constant sample size, increasing model flexibility leads to high variance. For the $K^+\mu^+\mu^-$ channel, the higher muon fake rate results in nearly triple the available background events for training compared to the electron channel. Consequently, to leverage the larger training dataset, we optimize hyperparameters for increased complexity or flexibility, using higher nTree and learning rate values.

6.9 Performance of MVA

Figure 6.7 shows the distribution of the MVA classifier output for the test dataset. To avoid overtraining of classifiers, necessary checks are carried out. We ensure that the respective ROC AUC (Figure 6.8) for the training and testing samples are the same by adjusting the hyperparameters of the classifier. Furthermore, we perform the KS test, as shown in Figure 6.9, for the signal and background distributions between the training and testing samples to verify consistency.

6.10 Background suppression

Figure 6.10 shows the signal retention and background rejection achieved for different selections on the classifier output. To maintain a good balance between preserving signal efficiency and attaining substantial background rejection, we apply a selection on the classifier output to be greater than 0.5. The corresponding distributions of M_{bc} and ΔE can be seen in Figs. 6.11 and 6.12. This particular selection on the classifier output is chosen judiciously, as it offers an approximately 90% background rejection while retaining around 90% of signal efficiency. We do not optimize on the MVA output, rather apply a loose criterion and use the residual distribution to extract signal that increases the sensitivity on R_K. Loosening the criterion further, for instance, to 0.4, leads to a smaller background rejection without significantly impacting the expected precision on R_K. Conversely, tightening the cut, e.g., to 0.6, marginally reduces the precision on R_K.

A significant amount of background is observed in the region $\Delta E < -0.1\,\mathrm{GeV}$, primarily originating from residual $B^+ \to J/\psi K^+$ and semileptonic B decays, as elaborated in Chapter 5. Figure 6.13 illustrates the contribution of residual backgrounds arising from the semileptonic $B \to D^{(*)}\ell\nu$ decay process and the charmonium veto evading $B^+ \to J/\psi K^+$ background in the ΔE distribution. To mitigate this background contribution, a selection criterion of $\Delta E > -0.1\,\mathrm{GeV}$ has been imposed.

6.11 Best Candidate Selection

After applying all the selections reported till now, the average candidate multiplicity per event is found to be 1.01. Only 1.3% of events have a candidate multiplicity greater than one. Figure 6.14 shows the candidate multiplicity distribution on signal MC events before BCS for the $B^+ \to K^+\mu^+\mu^-$ channel. We use the probability of the B-vertex fit χ^2 or $p(\chi^2_{\mathrm{vtx}})$ to suppress the self-crossfeed (SCF) background. In the case of multiple B candidates, we retain the one with the highest value of $p(\chi^2_{\rm vtx})$. The best candidate selection (BCS) efficiency is defined as,

$$\text{BCS efficiency} = \frac{\text{No. of correctly reconstructed candidates selected by BCS}}{\text{No. of events with multiple } B \text{ candidates}},$$

and the SCF fraction is defined as,

$$\text{SCF fraction} = \frac{\text{No. of misreconstructed candidates selected by BCS}}{\text{No. of total events after BCS}}.$$

The BCS efficiencies and SCF fractions, as calculated in signal MC events, are given in Table 6.2.

Table 6.2: Results for best candidate selection.

Channel	BCS efficiency (%)	SCF fraction (%)
$B^+ \to K^+e^+e^-$	67.0 ± 1.7	1.36 ± 0.02
$B^+ \to K^+\mu^+\mu^-$	68.1 ± 1.7	1.40 ± 0.02

After applying all the aforemention selections, the distributions of $M_{\rm bc}$ and ΔE are shown in Figs. 6.15 and 6.16.

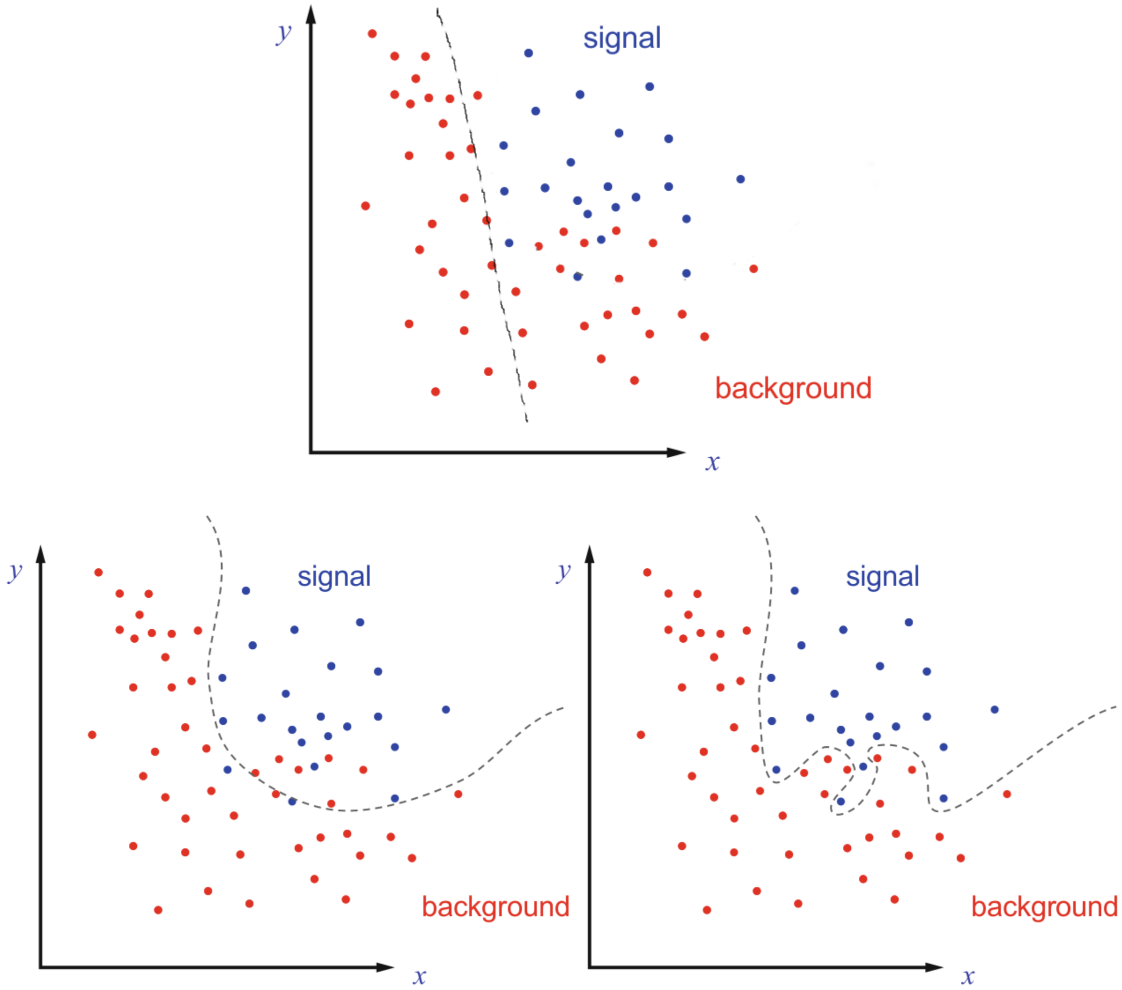


Figure 6.4: (top) A schematic diagram of the bias-variance trade-off. Blue dots represent true signal points, while red dots denote background points in a two-dimensional feature space. dashed line exhibits a suboptimal decision boundary with high bias, exemplifying undertraining. In the bottom-left, a decision boundary (black dashed curve) slightly increases model complexity, improving the signal-background separation. The bottom-right plot, however, displays an overly complex decision boundary, achieving high accuracy but suffering from the overfitting and excessive variance. These plots underscore that a delicate balance between the bias and variance is a must, emphasizing the need for the right level of complexity to achieve optimal performance and robust predictions. The figure is from Ref. [8]

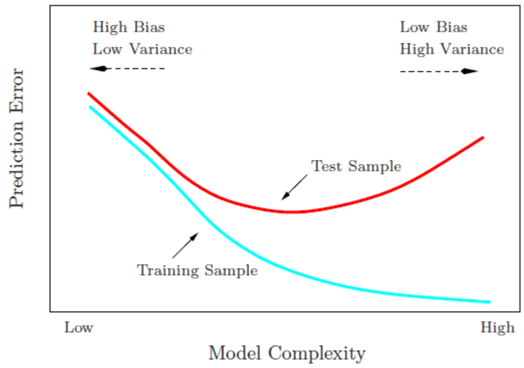


Figure 6.5: Bias variance.

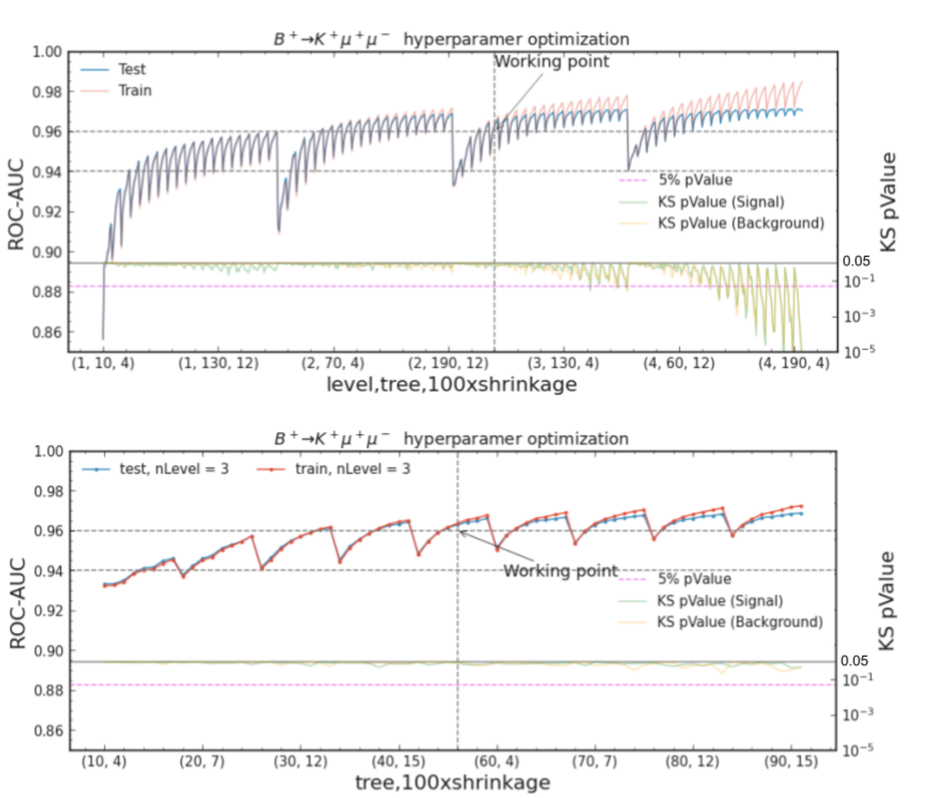


Figure 6.6: Training (red) and testing (blue) comparison between ROC-AUC scores and KS p-Value for signal (green) and background (yellow) for different sets of hyperparameters for the $B^+ \to K^+\mu^+\mu^-$ channel. The top represents the all three hyperparameters varied and the bottom focuses on a particular hyperparameter nLevel = 3.

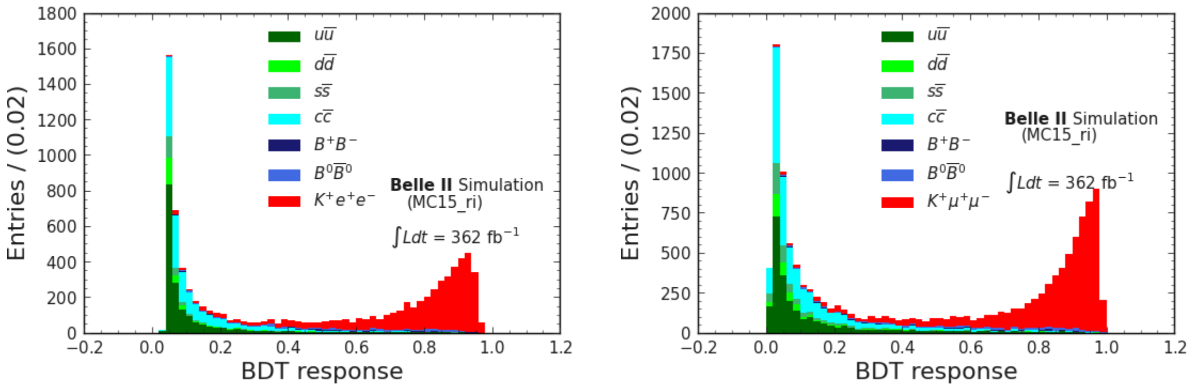
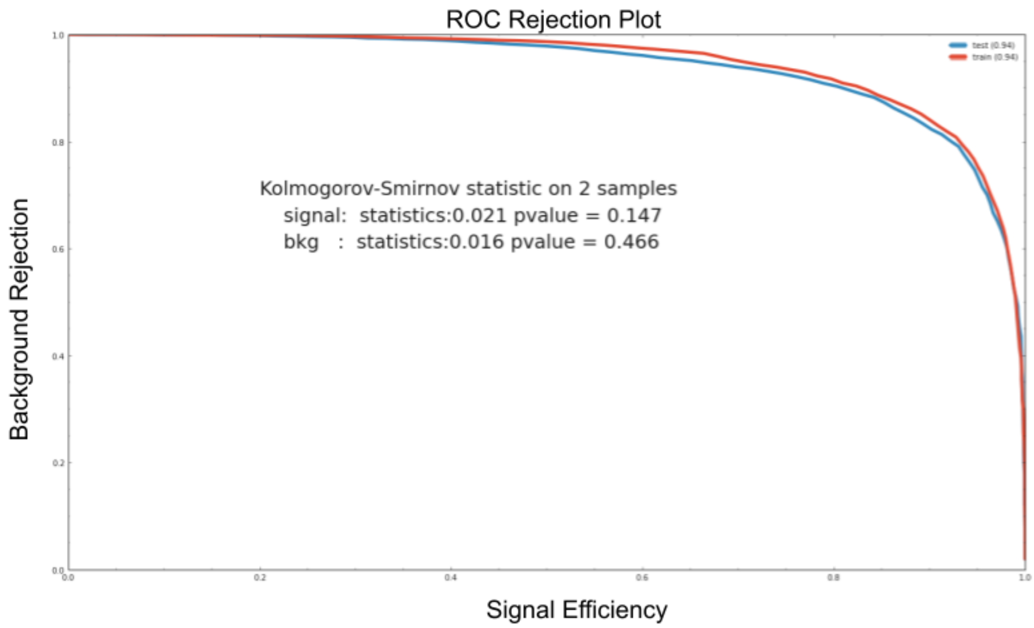


Figure 6.7: Distributions of the MVA classifier output for (left) $B^+ \to K^+e^+e^-$ and (right) $B^+ \to K^+\mu^+\mu^-$ channels. The background samples correspond to an integrated luminosity of 362 fb^{-1}, and an equal number of signal events are taken from the signal MC sample. For $B^+ \to K^+e^+e^-$ channel, both the training and testing ROC-AUC score is 0.94 and for $B^+ \to K^+\mu^+\mu^-$ the same 0.96.



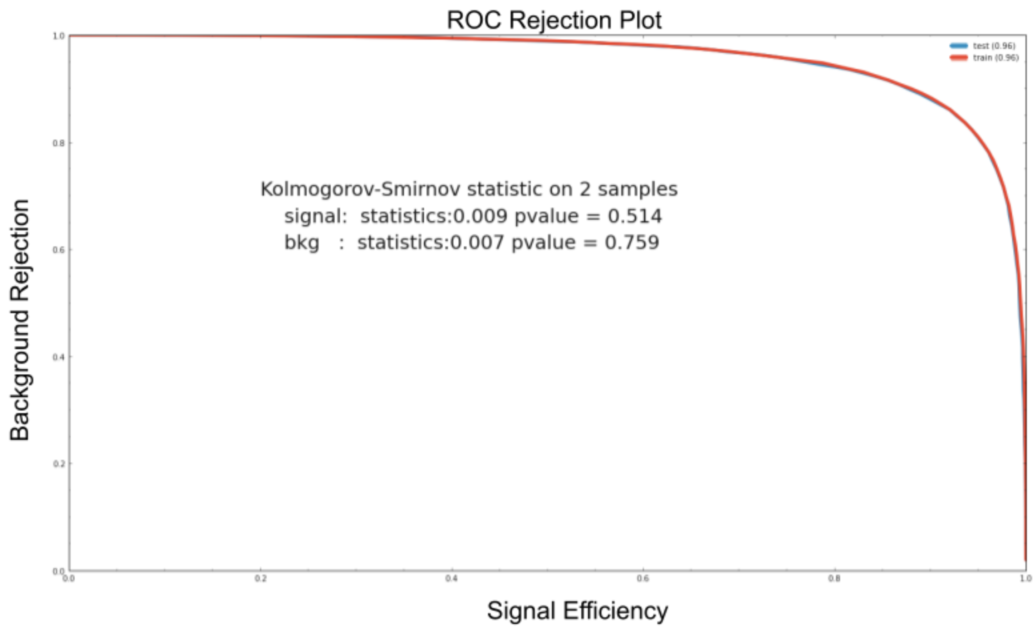


Figure 6.8: Comparison of the ROC-AUC between the training and testing samples for two channels: (top) $B^+ \to K^+e^+e^-$ and (bottom) $B^+ \to K^+\mu^+\mu^-$. The KS test p-Values for the $B^+ \to K^+e^+e^-$ channel are 0.14 (signal) and 0.46 (background), and for the $B^+ \to K^+\mu^+\mu^-$ channel those numbers are 0.51 (signal) and 0.76 (background).

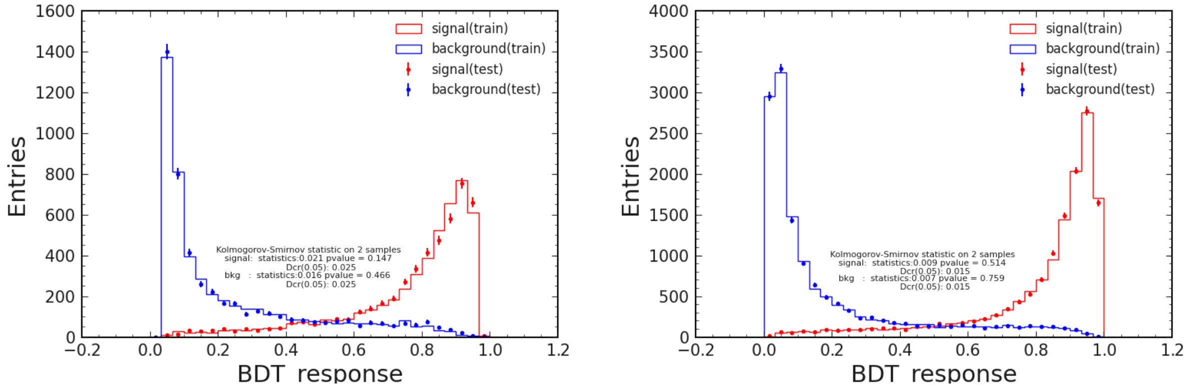


Figure 6.9: KS test results for the (left) $B^+ \to K^+e^+e^-$ and (right) $B^+ \to K^+\mu^+\mu^-$ channels. The 'pValue' is greater than 5%, indicating a good agreement of the shapes between the training and testing samples.

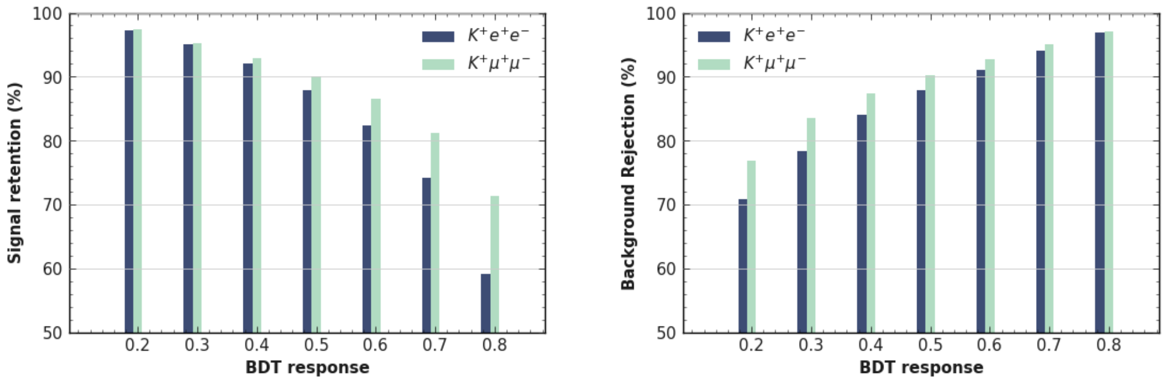


Figure 6.10: (left) Signal retention and (right) background rejection for both $K^+e^+e^-$ and $K^+\mu^+\mu^-$ channels.

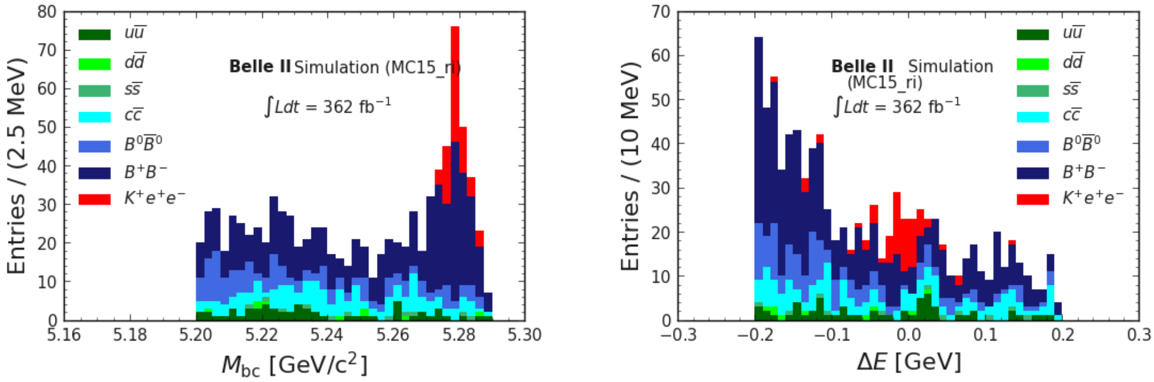


Figure 6.11: $M_{\rm bc}$ and ΔE distributions for the $B^+ \to K^+e^+e^-$ channel after applying selection on the MVA classifier output.

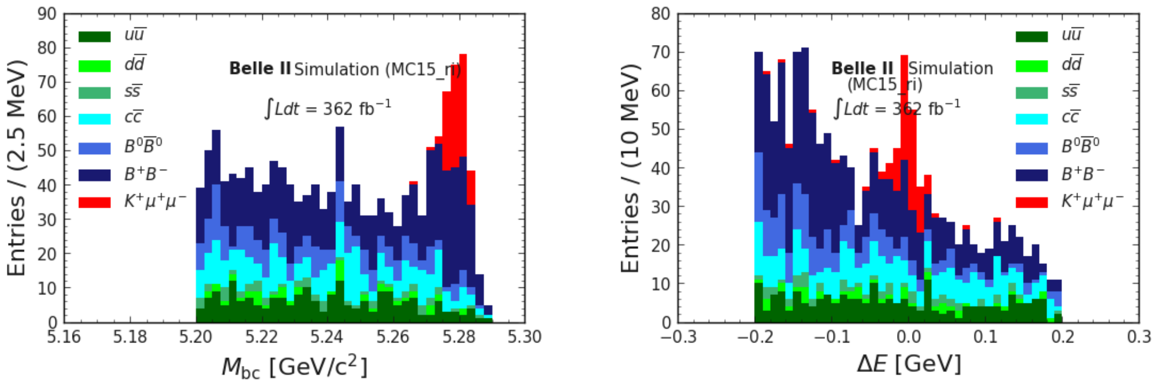


Figure 6.12: $M_{\rm bc}$ and ΔE distributions for the $B^+ \to K^+\mu^+\mu^-$ channel after applying selection on the MVA classifier output.

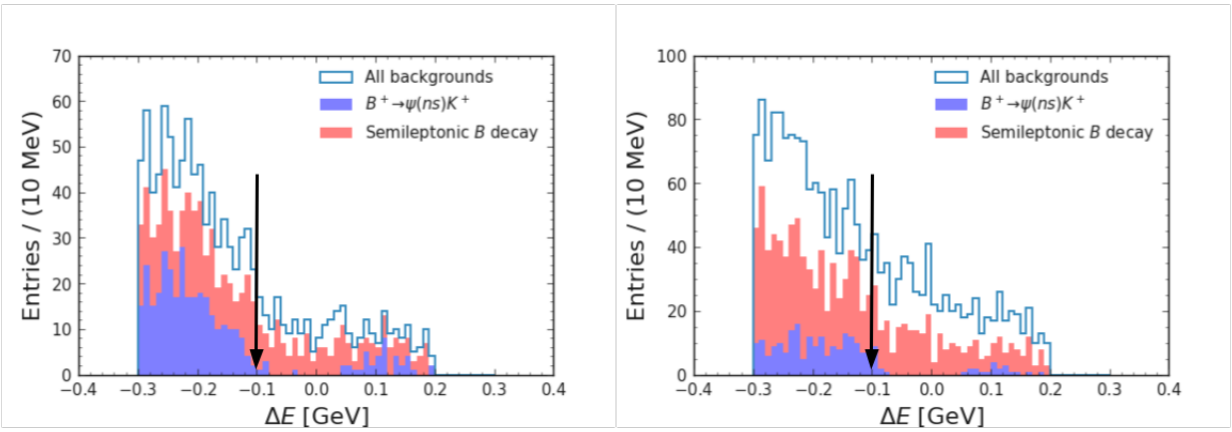


Figure 6.13: ΔE distributions of various backgrounds for the (left) $B^+ \to K^+e^+e^-$ and (right) $B^+ \to K^+\mu^+\mu^-$ channel. The red filled histograms correspond to the residual backgrounds coming from the semileptonic $B \to D^{(*)}\ell\nu$ decay process and the blue filled histograms correspond to the $B^+ \to J/\psi K^+$ background that evade the charmonium veto.

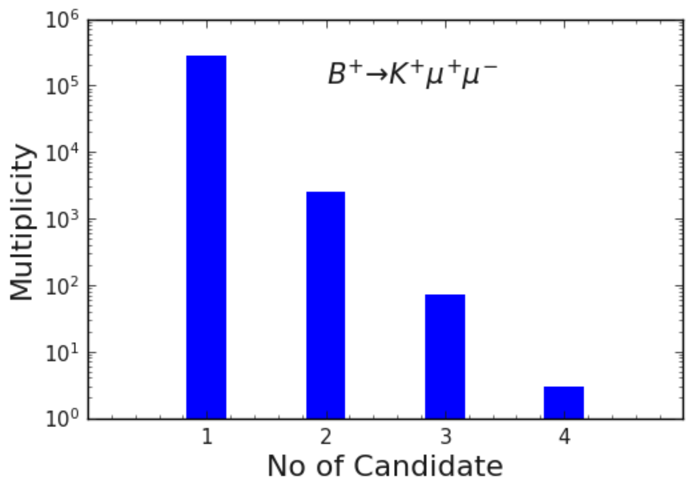


Figure 6.14: Candidate multiplicity distribution before BCS for the $B^+ \to K^+\mu^+\mu^-$ channel.

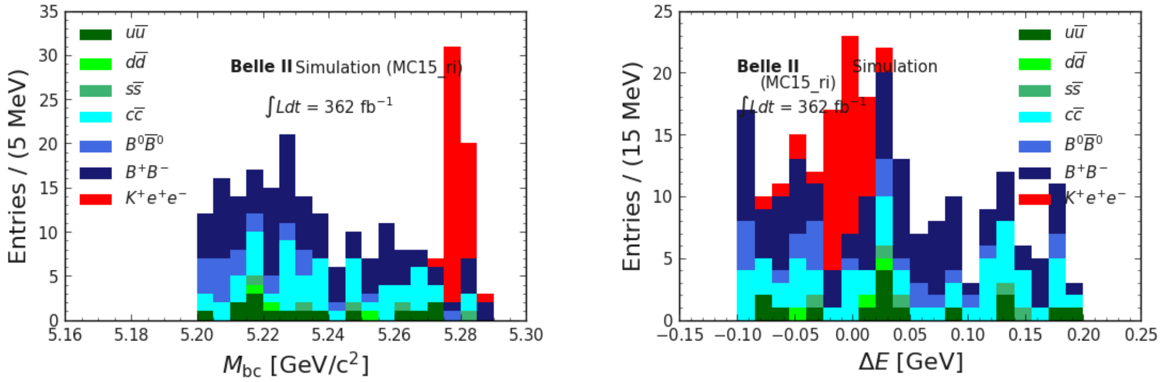


Figure 6.15: M_{bc} and ΔE distribution of the $B^{+} \to K^{+}e^{+}e^{-}$ channel after applying all the selection, including that on the MVA classifier output.

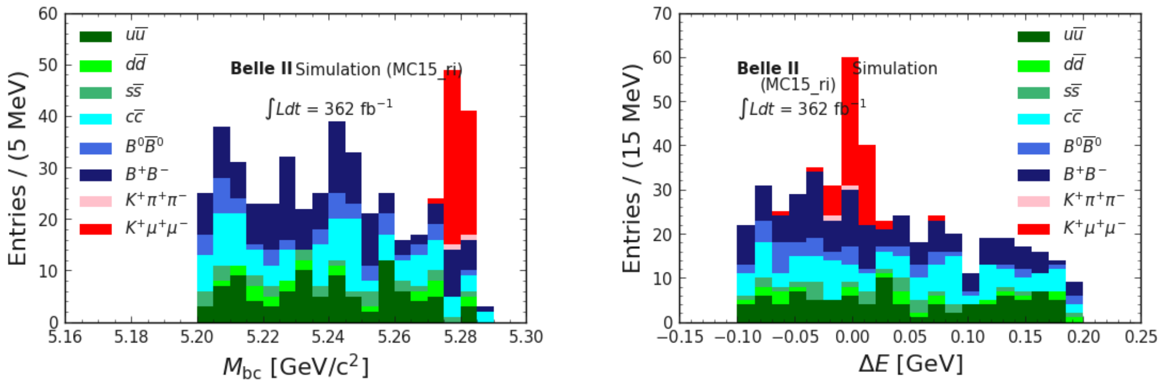


Figure 6.16: M_{bc} and ΔE distributions of the $B^{+} \to K^{+}\mu^{+}\mu^{-}$ channel after applying all the selection, including that on the MVA classifier output.

Chapter 7

Signal extraction and measurement of observables

After putting significant effort into background suppression, the next step in our analysis is to extract essential observables from the collected data. This chapter serves as a comprehensive guide through the intricate process of signal extraction, where precision and accuracy hold paramount importance. At its core lies the powerful maximum-likelihood fit method, a tool that enables us to distil valuable insights from the intricate data landscape. Throughout this chapter, we navigate with meticulous organization, paving the way for a refined and precise measurement of observables that are central to our analysis.

The sections are organized as follows: Section 7.1 provides an introduction to the maximum-likelihood fit, while Section 7.2 discusses the transformation of the MVA classifier output to a new fit variable. Sections 7.3 and 7.4 elaborate on the fit methodology, and Section 7.5 focuses on the toy study. Lastly, Section 7.6 addresses the measurement of observables in the simulated sample.

7.1 Maximum likelihood, a brief introduction

The likelihood function is a combined probability distribution function (PDF) of all measurements in the data sample. Observables are treated as parameters of the likelihood function, which is obtained by maximizing the latter.

To delve deeper into the concept, let's consider a dataset comprising N similar measurements, denoted as $\mathbf{x} = \{x_1, x_2, \dots, x_N\}$, where each x_i represents a statistically independent measurement and follows the PDF $f(x_i; \boldsymbol{\theta})$. Here, $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_m\}$ is a set of m unknown parameters to be estimated. The joint PDF for each of the observed values of x is the likelihood function:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\theta}) = \prod_{i=1}^{N} f(x_i, \boldsymbol{\theta}). \tag{7.1}$$

The maximum-likelihood estimate of the set of parameters $\boldsymbol{\theta}$ are the values $\hat{\theta}$ for which the likelihood function $\mathcal{L}(\mathbf{x}, \boldsymbol{\theta})$ is maximized.

7.2 Translation of MVA output

The analytical modeling of the shape of BDT output (C) shown in Fig. 6.7 is challenging, thus we use a transformation to simplify the modeling process. This transformation is known as the probability integral or μ transformation [60], where the cumulative distribution of C is denoted as F_C. The transformed BDT output, denoted as $C' = F_C(C)$, can be more easily modeled. The signal shape of C', shown in Fig. 7.1, is uniform, which is guaranteed by the 'Probability Integral Transform'.

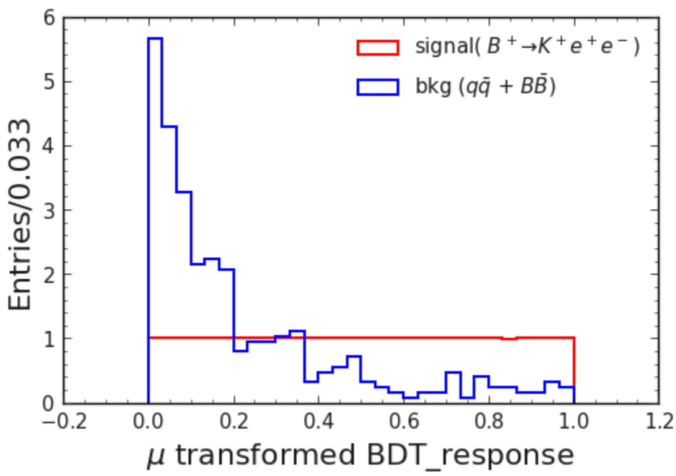


Figure 7.1: Normalized distribution of transformed BDT output for the $B^+ \to K^+e^+e^-$ channel.

The μ transformation is performed with correctly reconstructed signal MC events, and a separate transformation is done for each q^2 bin. The resulting signal C' distributions for both channels are presented in Fig. 7.2.

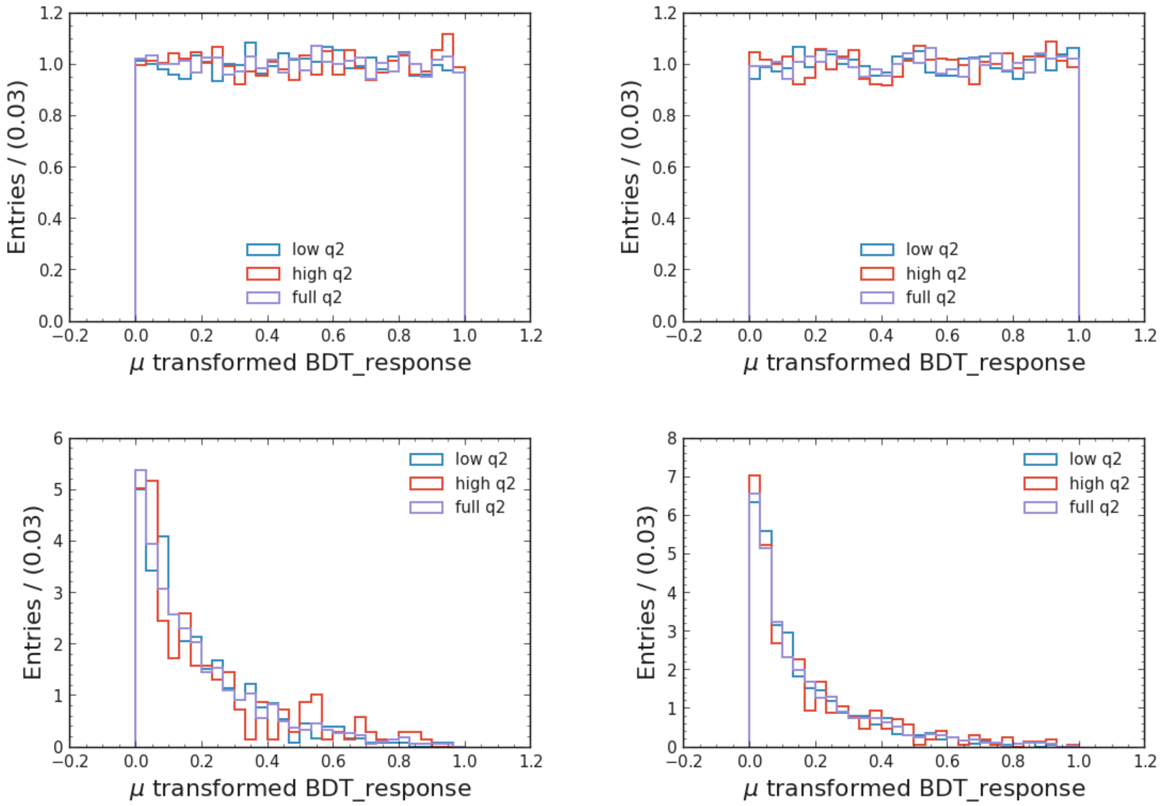


Figure 7.2: Distributions of C' for signal (top) and background (bottom) in case of the electron (left) and muon (right) channels.

7.3 Fit methodology

To determine the signal yield, we utilize a three-dimensional unbinned extended maximum-likelihood fit. The variables used in the fit are $M_{\rm bc}$, ΔE, and C'. The linear correlation coefficients among these variables are less than 10% for both signal and background.

For a dataset of N candidates, the likelihood function can be written as:

$$\mathcal{L}(\vec{\alpha},\vec{\beta},\vec{\gamma},\vec{n}) = \frac{e^{-\sum n_j}}{N!}\prod_{i=1}^{N}\sum_{j} n_j \times \mathcal{P}_j(\Delta E^i : \vec{\alpha}) \times \mathcal{Q}_j(M^i_{\rm bc} : \vec{\beta}) \times \mathcal{R}_j(C'^i : \vec{\gamma})$$

where $\mathcal{P}_j$, $\mathcal{Q}_j$ and $\mathcal{R}_j$, and n_j are the PDF of ΔE, $M_{\rm bc}$ and C' distribution, and the number of events corresponding to the j^{th} component, respectively. The argument ΔE^i, $M^i_{\rm bc}$ and C'^i denote the value of ΔE, $M_{\rm bc}$ and C' for the i^{th} candidate; $\vec{\alpha}, \vec{\beta}$ and $\vec{\gamma}$ are the set of shape parameters for $\mathcal{P}$'s, $\mathcal{Q}$'s and $\mathcal{R}$'s. The likelihood function is maximized with respect to $\vec{\alpha},\vec{\beta},\vec{\gamma},\vec{n}$ to calculate the shape parameters and number of events corresponding to different components.

The fitting procedure for $B^+ \to K^+e^+e^-$ involves two components, denoted by $j = 1, 2$: signal and combinatorial background. In the case of $B^+ \to K^+\mu^+\mu^-$, an additional component ($j = 3$) is included to account for the $B^+ \to K^+\pi^+\pi^-$ background.

7.4 PDF Modeling

7.4.1 Signal modeling

The distributions of fit variables are shown in Figs. 7.3 and 7.4. The signal ΔE PDF is a sum of a Gaussian and a Cruijff function [61]. The Cruijff function is a function having a Gaussian core with radiative tails on both sides, given by

$$\mathcal{P}^{\text{Cruijff}}(\Delta E : \mu, \alpha_L, \alpha_R, \sigma_L, \sigma_R) = \begin{cases} \exp[-\frac{(\Delta E-\mu)^2}{\alpha_L(\Delta E-\mu)^2+2\sigma_L^2}], & \text{if } \Delta E < \mu \\ \exp[-\frac{(\Delta E-\mu)^2}{\alpha_R(\Delta E-\mu)^2+2\sigma_R^2}], & \text{if } \Delta E \geq \mu. \end{cases}$$

The signal M_{bc} is parametrized by an empirical function introduced by the Crystal Ball Collaboration. The so-called Crystal Ball function accounts for tails in the distribution, defined as

$$\mathcal{P}^{\text{CB}}(M_{\text{bc}} : m_0, \sigma, \alpha, n) = \begin{cases} e^{-\frac{(M_{\text{bc}}-m_0)^2}{2\sigma^2}}, & \text{if } M_{\text{bc}} > m_0 - \alpha\sigma \\ (\frac{n}{\alpha})^n e^{-\frac{\alpha^2}{2}} (\frac{m_0-M_{\text{bc}}}{\sigma} + \frac{n}{\alpha} - \alpha)^{-n}, & \text{if } M_{\text{bc}} \leq m_0 - \alpha\sigma \end{cases}$$

where m_0 and σ are the mean and width of the distribution. The Crystal Ball function can be interpreted as a Gaussian shape that transits to a power-law tail, below a certain threshold, which is defined by the parameters α and n.

The signal C' is parametrized by a first-order Chebychev polynomial. Since we have used the signal BDT cumulative distribution for the μ transformation, we expect the signal C' to be uniformly distributed. Other than the mean and width, we use MC shape parameters to fit in data. The mean and width are calculated from a fit to the data in the $B \to J/\psi K$ dominated q^2 region.

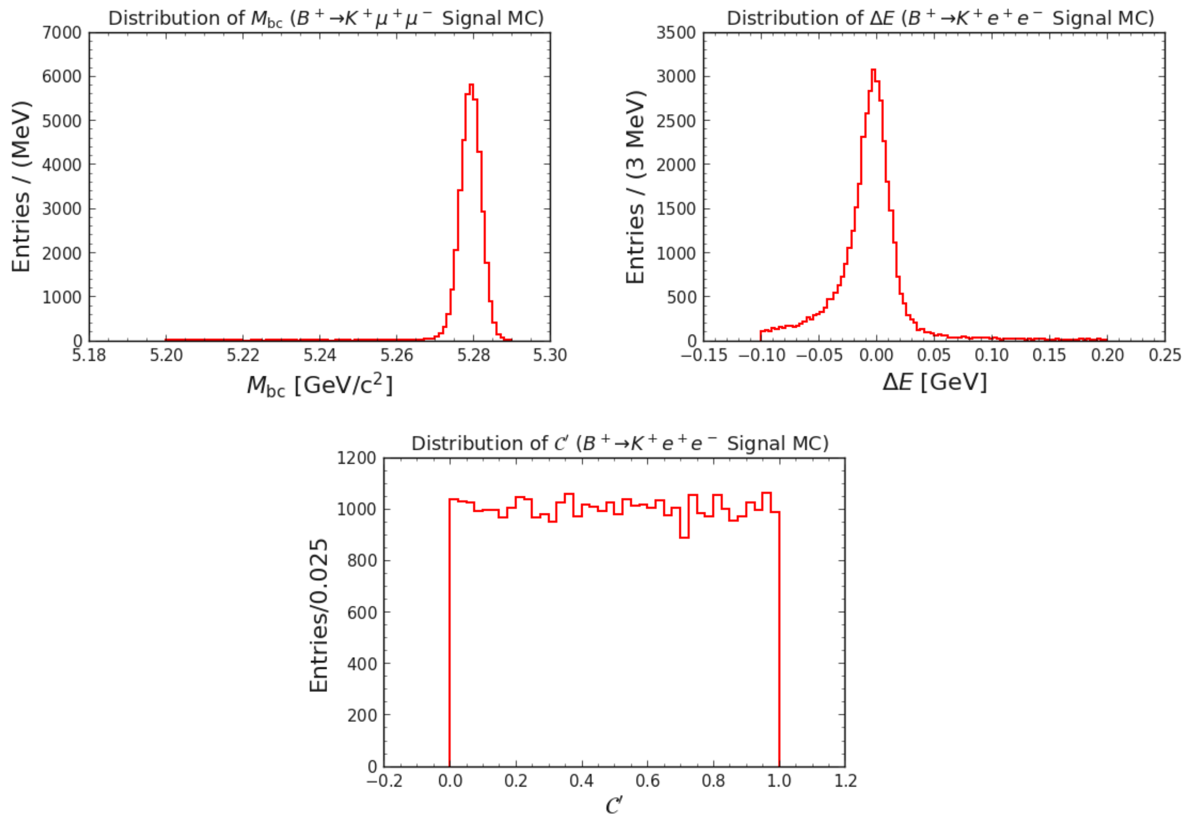


Figure 7.3: (top left) $M_{\rm bc}$, (top right) ΔE, and (bottom) C' distributions of $B^+ \to K^+e^+e^-$ signal after applying all the selection.

7.4.2 Combinatorial background modeling

The ΔE and C' distributions of combinatorial background are modeled by an exponential function. The $M_{\rm bc}$ distributions is parametrized by an empirical function introduced by the ARGUS Collaboration[62]. This so-called ARGUS function is defined as

$$\mathcal{P}^{\rm ARGUS}(M_{\rm bc} : m_0, \chi) = M_{\rm bc}\sqrt{1-\left(\frac{M_{\rm bc}}{m_0}\right)^2}\, e^{-\chi\left(1-\left(\frac{M_{\rm bc}}{m_0}\right)^2\right)},$$

where χ describes the slope and m_0 is the cut-off value of the distribution. Due to kinematic constraints, m_0 corresponds to half the c.m. energy ($m_0 = E_{\rm c.m.}/2$) in the reconstruction of B mesons from $\Upsilon(4S)$ decays at Belle II. Hence, we have fixed it to 5.29 GeV/c^2.

7.4.3 $K\pi\pi$ background

The $B^+ \to K^+\pi^+\pi^-$ yield is determined by multiplying the known branching fraction [9] with the misidentification rate of π as μ, obtained from auxiliary data. The resulting estimated yields for the low-, high-, and full-q^2 bins are 1.09, 0.81, and 1.91, respectively.

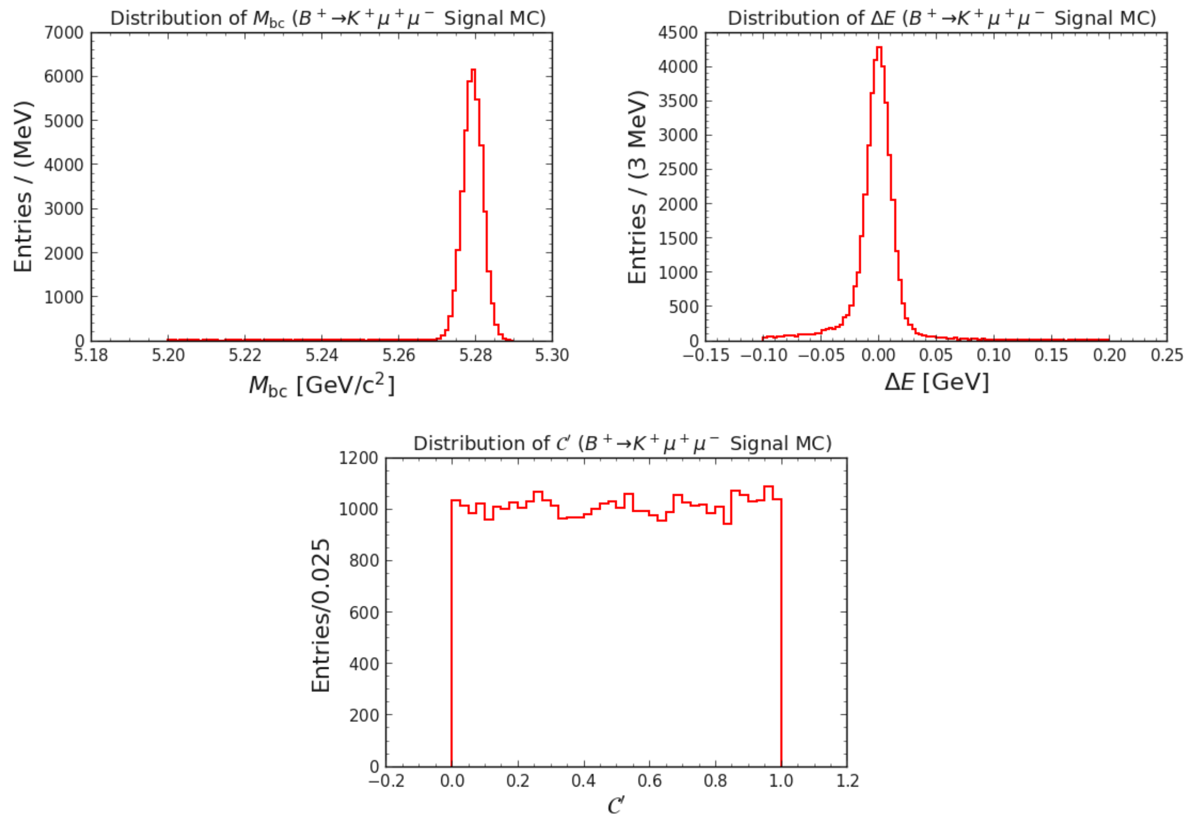


Figure 7.4: (top left) $M_{\rm bc}$, (top right) ΔE, and (bottom) C' distributions of $B^+ \to K^+\mu^+\mu^-$ signal after applying all the selection.

The overall fit strategy is summarised in Table 7.1. The distributions of $M_{\rm bc}$, ΔE, and C' in the data-equivalent simulated sample for low-q^2 bin are shown for $B^+ \to K^+\ell^+\ell^-$ events in Figs. 7.5, 7.6, and 7.7, respectively; the fit results are superimposed.

7.5 Fit validation

We prepare one thousand samples, each of which corresponds to an integrated luminosity of 362 fb^{-1}, using simulated events by employing the bootstrap method. This involves repeatedly resampling smaller samples with replacement from a larger sample [63]. The number of signal and background events in each sample are determined by a Poisson random number generator, with the means taken from MC simulation (S_0 and B_0 denote the mean signal and background yields, respectively). Subsequently, each distribution is fitted, and the pull is calculated for each fit. If the signal yield obtained from the fit is $N_s \pm \delta N_s$, the pull is given by $\frac{N_s - S_0}{\delta N_s}$. The distribution of pull is then fitted with a Gaussian function. The obtained mean and width of the Gaussian function are consistent with zero and one, respectively, as expected for an unbiased

Table 7.1: Summary of overall fit strategy. The parameters written in **red** are floated in the fit to data, those written in blue are fixed from MC events and those written in orange are fixed from the control sample $B^+ \to J/\psi(\ell^+\ell^-)K^+$.

Components/ Yield	Signal	Combinatorial background	$B \to K\pi\pi$ background (for $B^+ \to K^+\mu^+\mu^-$)
Yield	$N_{\mathbf{sig}}$	$N_{\mathbf{bkg}}$	$N_{K\pi\pi}$
ΔE	$f \times \mathcal{P}^{\text{Cruijff}}(\mu_{\Delta E}, \alpha_L, \alpha_R, f_L \times \sigma_{\Delta E}, f_R \times \sigma_{\Delta E})$ + $(1- f)\times\mathcal{P}^{\text{Gauss}}\ (\mu_{\Delta E},\sigma_{\Delta E})$	$\mathcal{P}^{\text{Exp}}(\tau_{\Delta E})$	$\mathcal{P}^{\text{Gauss}}\ (\mu_{\Delta E},\sigma_{\Delta E})$
M_{bc}	$\mathcal{P}^{CB}\ (\mu_{M_{\text{bc}}},\sigma_{M_{\text{bc}}}, \alpha,n)$	$\mathcal{P}^{\text{ARGUS}}\ (m_0,\chi)$	$\mathcal{P}^{\text{Gauss}}\ (\mu_{M_{\text{bc}}},\sigma_{M_{\text{bc}}})$
C'	$\mathcal{P}^{\text{Cheby}}(a_1)$	$\mathcal{P}^{\text{Exp}}(\tau_{C'})$	$\mathcal{P}^{\text{Cheby}}(a_1)$

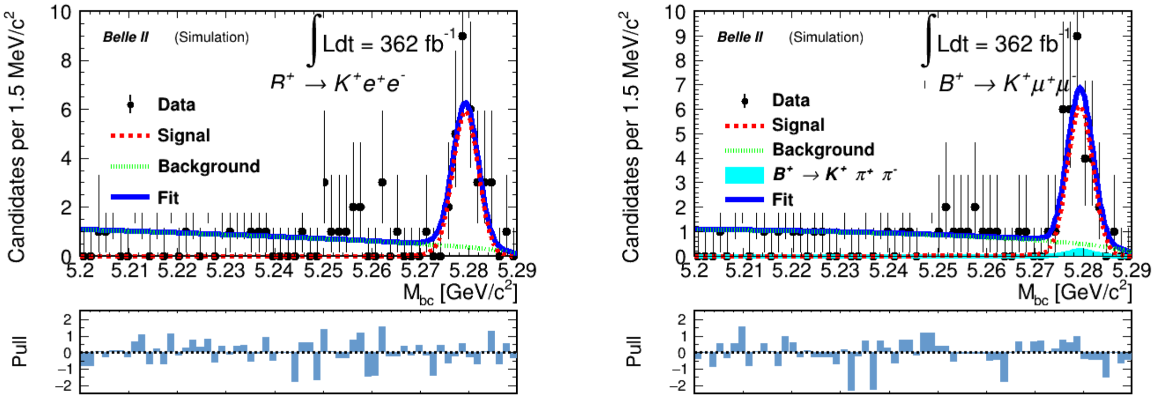


Figure 7.5: M_{bc} distributions for both $B^+ \to K^+\ell^+\ell^-$ channels in $q^2 \in [0.10, 8.12]\,\text{GeV}^2/c^4$ bin with the fit result superimposed (top) and pull distribution with respect to the fit result (bottom). The projection is obtained in a window $-0.05 < \Delta E < 0.05$ GeV. Black dots with error bars denote the data, blue solid curves denote the total fit, dashed red curves are the signal component, dotted green curves are the background component, and the filled cyan region in the muon channel represents the $B^+ \to K^+\pi^+\pi^-$ component.

estimator. This check is performed for all the fitters: low-, high-, and full-q^2 region as well as two control samples. Figure 7.8 shows the pull distribution for the $B^+ \to K^+e^+e^-$ channel. The mean and width extracted from the Gaussian fit are consistent with zero and one.

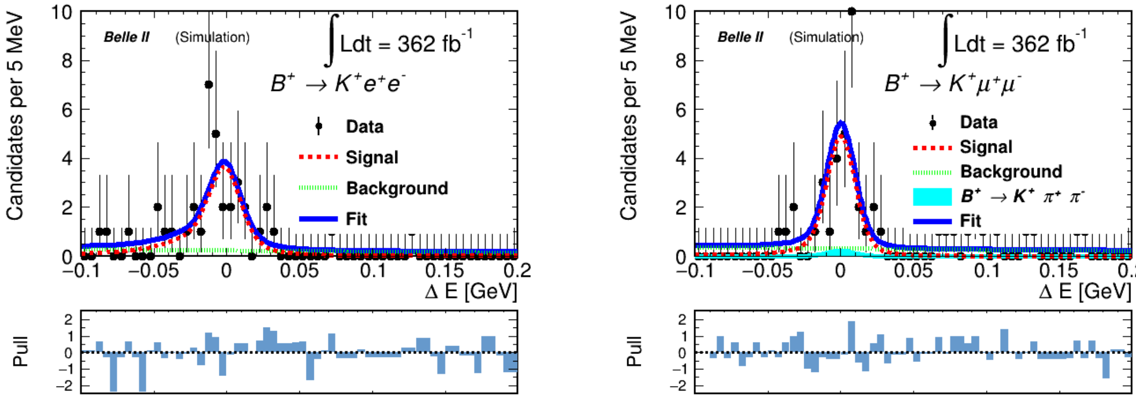


Figure 7.6: ΔE distributions for both $B^+ \to K^+\ell^+\ell^-$ channels in $q^2 \in [0.10, 8.12]\,\mathrm{GeV}^2/c^4$ bin with the fit result superimposed (top) and pull distribution with respect to the fit result (bottom). The projection is obtained in a window $5.27 < M_{\rm bc} < 5.29\,\mathrm{GeV}/c^2$. Black dots with error bars denote the data, blue solid curves denote the total fit, dashed red curves are the signal component, dotted green curves are the background component, and the filled cyan region in the muon channel represents the $B^+ \to K^+\pi^+\pi^-$ component.

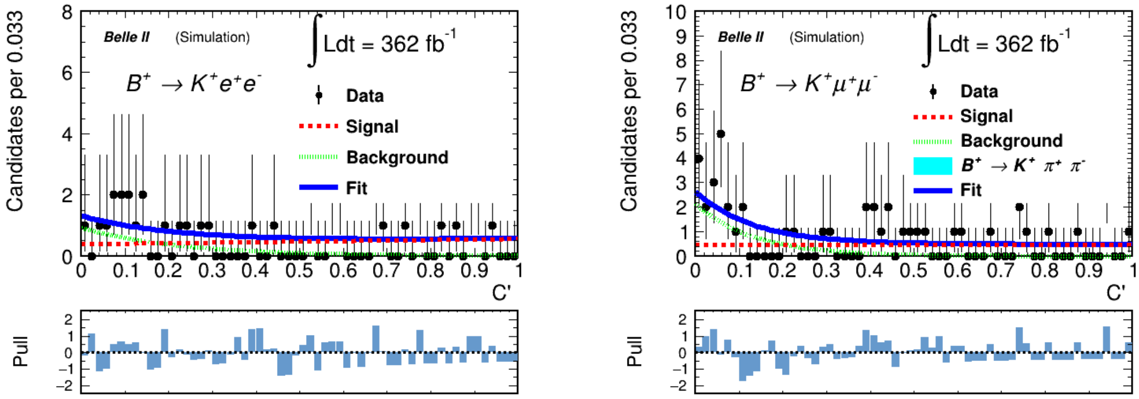


Figure 7.7: C' distributions for both $B^+ \to K^+\ell^+\ell^-$ channels in $q^2 \in [0.10, 8.12]\,\mathrm{GeV}^2/c^4$ bin with the fit result superimposed (top) and pull distribution with respect to the fit result (bottom). The projection is obtained in a window $5.27 < M_{\rm bc} < 5.29\,\mathrm{GeV}/c^2$. Black dots with error bars denote the data, blue solid curves denote the total fit, dashed red curves are the signal component, dotted green curves are the background component, and the filled cyan region in the muon channel represents the $B^+ \to K^+\pi^+\pi^-$ component.

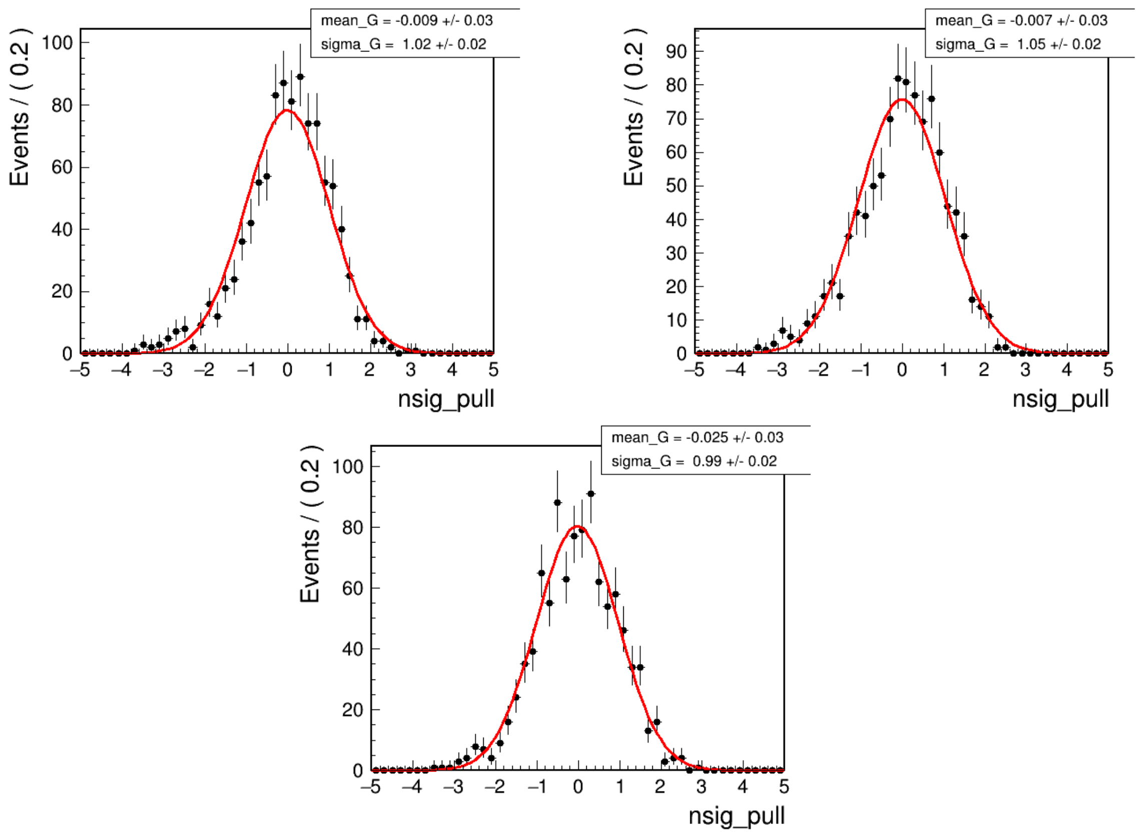


Figure 7.8: Fit to pull distribution for the bootstrapped MC sample of (a) $B^+ \to K^+e^+e^-$ (low-q^2), (b) $B^+ \to K^+e^+e^-$ (high-q^2), and (c) $B^+ \to K^+e^+e^-$ (full-q^2)

7.6 Measurement of observables

The branching fractions are determined with the relation

$$\mathcal{B} = \frac{n_{\rm sig}}{2\, N_{B\overline{B}}\, \epsilon\, f^{\pm}}, \tag{7.2}$$

where $n_{\rm sig}$ is the signal yield determined by the fit, $N_{B\overline{B}}$ is the number of $B\overline{B}$ events, ϵ is the signal detection efficiency, and $f^{\pm}$ is the branching fraction of the $\Upsilon(4S)$ decay to B^+B^- pairs assuming $\mathcal{B}\left[\Upsilon(4S) \to B\overline{B}\right] = 1$. We use the value $f^{\pm} = 0.519 \pm 0.029$ obtained from two independent measurements [41, 64].

According to Ref. [44], the world average of $f^{\pm}$ is mostly based on measurements that assume isospin symmetry in $B \to J/\psi K$ decays. However, this assumption fails to account for potential theoretical uncertainties. To address this limitation, we employ a recent measurement of $f^{\pm}/f^{00}$ [64] that explicitly considers theoretical uncertainties associated with the isospin assumption. Multiplying this $f^{\pm}/f^{00}$ measurement with that of f^{00} [41], we obtain the value $f^{\pm} = 0.519 \pm$

0.029.

We calculate R_K using the relation

$$R_K = \frac{\mathcal{B}(K^+\mu^+\mu^-)}{\mathcal{B}(K^+e^+e^-)} = \frac{n_{\text{sig}}^{K^+\mu^+\mu^-}}{n_{\text{sig}}^{K^+e^+e^-}} \times \frac{\epsilon^{K^+e^+e^-}}{\epsilon^{K^+\mu^+\mu^-}}, \tag{7.3}$$

The MC results are listed in Table 7.2.

Table 7.2: Reconstruction efficiency, signal yield and branching fraction for both channels and R_K for different q^2 bins. Where two uncertainties are given, the first is statistical and the second is systematic, and the singly quoted uncertainty corresponds to the statistical component. The generated values for branching fractions are indicated in parentheses.

q^2 bin $\left[\mathrm{GeV}^2/c^4\right]$	Channel	ϵ (%)	n_{sig}	$\mathcal{B}(10^{-7})$	R_K
$(0.10, 8.12)$	$K^+\mu^+\mu^-$	29.27	34.2 ± 6.4	$3.0 \pm 0.6 \pm 0.2$ (2.72)	$1.25 \pm 0.35 \pm 0.02$
	$K^+e^+e^-$	30.79	28.8 ± 5.5	$2.4 \pm 0.5 \pm 0.2$ (2.78)	
> 14.18	$K^+\mu^+\mu^-$	39.70	25.8 ± 5.6	$1.7 \pm 0.4 \pm 0.1$ (1.37)	$1.41 \pm 0.50 \pm 0.02$
	$K^+e^+e^-$	35.25	16.2 ± 4.3	$1.2 \pm 0.3 \pm 0.1$ (1.31)	
full-q^2 bin	$K^+\mu^+\mu^-$	22.34	59.8 ± 8.6	$5.2 \pm 0.8 \pm 0.4$ (6.00)	$1.31 \pm 0.29 \pm 0.02$
	$K^+e^+e^-$	21.95	44.7 ± 7.0	$6.8 \pm 1.0 \pm 0.4$ (6.00)	

Chapter 8

Measurements of branching fraction, isospin asymmetry, and LFU ratio in $B \to J/\psi K$ decays

In Chapter 5.1, we delved into the $B \to J/\psi(\ell\ell)K$ backgrounds, which exhibit similar characteristics to the signal events of $B \to K\ell\ell$ due to their identical decay final states. However, these background events originate from a narrow dilepton invariant-mass window, rendering them easily distinguishable and eliminable. In this chapter, we will leverage the resemblance between these background and our signal events to measure observables from the data. Before delving into the intriguing data region, this exercise serves as a validation of our methodology. By applying the same selection criteria to this SM candle, we establish the robustness of our analysis approach.

The sections are organized as follows: Sections 8.1, 8.2, and 8.3 delve into the details of event reconstruction, background suppression, and signal extraction procedures, respectively. Section 8.4 discusses the measurement of observables in a relatively smaller dataset with an integrated luminosity of $189\,\mathrm{fb}^{-1}$. Section 8.5 focuses on the measurement of observables in a simulated sample equivalent to $362\,\mathrm{fb}^{-1}$ and includes a comparison of the MVA classifier output distribution between data and simulated samples.

8.1 Reconstruction of $B \to J/\psi K$

We reconstruct J/ψ candidates by combining two oppositely charged leptons of the same flavor. The invariant mass of the J/ψ candidate must lie in the range $(2.91, 3.19)\,\mathrm{GeV}/c^2$ and $(2.96, 3.19)\,\mathrm{GeV}/c^2$ for the electron and muon channels, respectively. Imperfect recovery of the energy lost due to bremsstrahlung calls for a wider asymmetric interval around the known J/ψ mass in the former case.

The K_S^0 candidates are reconstructed from two oppositely charged particles, assumed to be pions. A kinematic fit is performed on the K_S^0 vertex to ensure that the pion tracks originate from a common vertex. We retain candidates with an invariant mass in the range $[0.487, 0.508]\,\mathrm{GeV}/c^2$, which is $\pm 3\sigma$ Gaussian resolution around the known K_S^0 mass [9]. In addition, momentum-dependent selections are applied on the K_S^0 flight length in the transverse plane, the azimuthal angle between the K_S^0 momentum vector and the vector connecting the IP with the decay vertex of the K_S^0 candidate, and the difference between the distances of closest approach of the two pion tracks along the z axis. Figure 8.1 shows the $M_{K_S^0}$ distribution of simulated sample of the $B^0 \to K_S^0 J/\psi(\mu^+\mu^-)$ channel. Finally, we combine J/ψ candidates with kaons, which are either K^+ or K_S^0 candidates, to reconstruct B mesons.

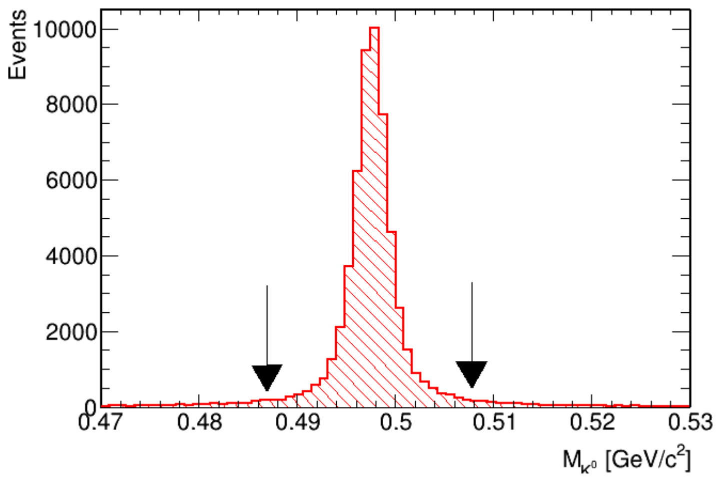


Figure 8.1: $M_{K_S^0}$ distribution for the $B^0 \to K_S^0 J/\psi(\mu^+\mu^-)$ channel in the simulated sample.

8.2 Backgrounds of $B \to J/\psi K$ events

Because the resonance region is so narrow, backgrounds from both continuum and B decay events are very small compared to charmonium signal events. The dominant background contribution arises from the $B \to K^*(K\pi)J/\psi$ process, where the B candidate is reconstructed without the inclusion of the pion from the K^*. Consequently, the ΔE values are shifted towards the lower side. Figure 8.2 illustrates the presence of $B\overline{B}$ background events, which populate the region of $\Delta E < -0.1$ GeV. To mitigate this background contamination, we require ΔE to be greater than -0.1 GeV.

After implementing these selection criteria, the resulting sample demonstrates a purity of approximately 94% (96%) for the muon (electron) mode, as shown in Fig. 8.3. A minor background is observed beneath the J/ψ invariant mass peak, originating from decays of B mesons to an inclusive J/ψ final state.

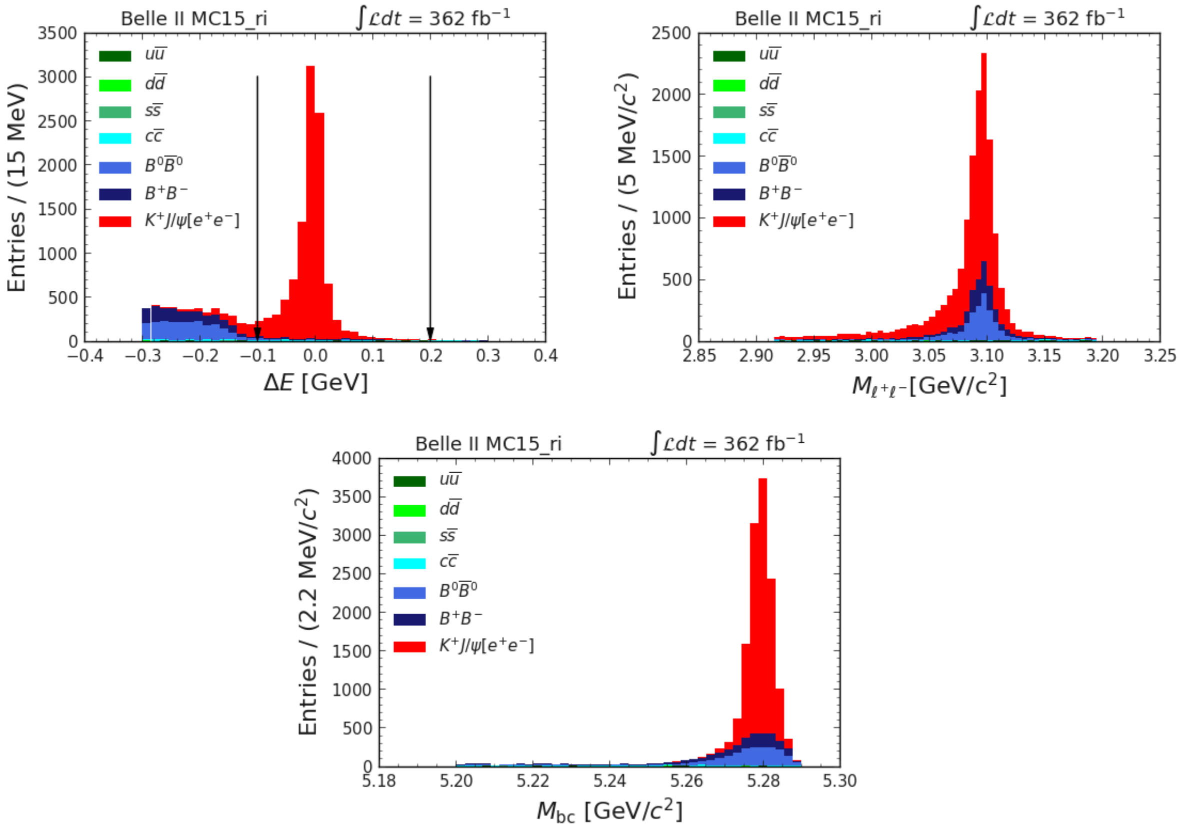


Figure 8.2: ΔE (top left), $M_{e^+e^-}$ (top-right) and $M_{\rm bc}$ (bottom) distributions for $B^+ \to K^+J/\psi(e^+e^-)$. The arrows in the top-left plot indicate the cut position.

Figures 8.4 and 8.5 show the distribution of ΔE and $M_{\rm bc}$ for two channels. The backgrounds, shown in Fig. 8.6, arise from random combinations of B decay and continuum events, as well as from nonresonant $B^+ \to K^+\ell^+\ell^-$, $B^+ \to \pi^+J/\psi$ with a misidentified pion as a kaon, and $B^0 \to K^*(\to K\pi)J/\psi$ where the pion from the K^{*0} decay is not reconstructed in the B candidate. The backgrounds from random combinations exhibit an almost flat distribution in ΔE and $M_{\rm bc}$. The ΔE distribution of π^+J/ψ events is slightly shifted to the positive side due to an incorrect

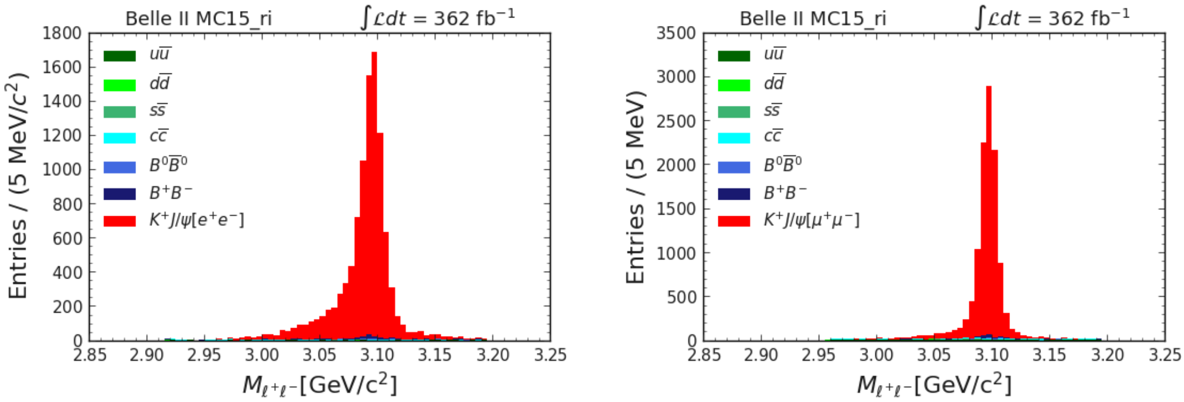


Figure 8.3: $M_{\ell^+\ell^-}$ distributions for the $B^+ \to K^+J/\psi(\ell^+\ell^-)$ channel (with $\ell = e$ in the left and $\ell = \mu$ in the right) after applying $\Delta E > -0.1$ GeV.

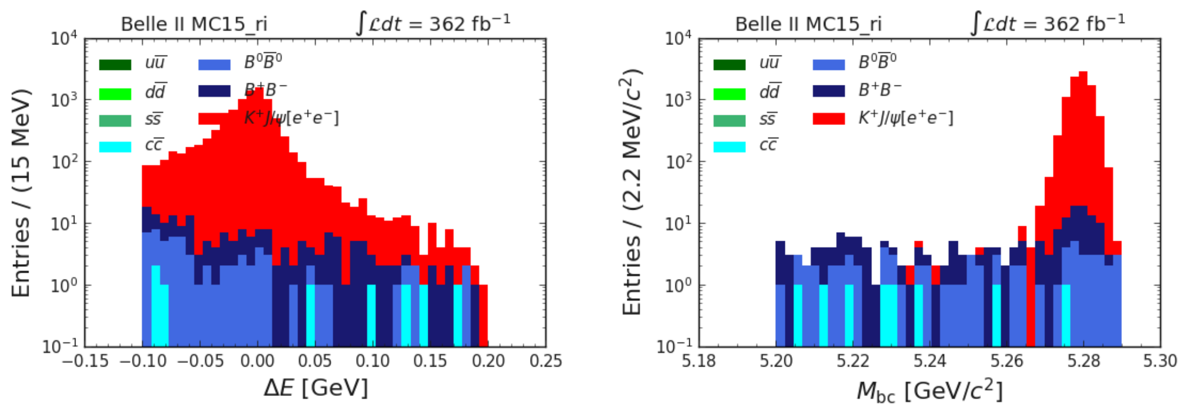


Figure 8.4: (left) ΔE, and (right) $M_{\rm bc}$ distributions of $B^+ \to J/\psi(ee)K^+$ after applying all the selection.

($m_K > m_\pi$) mass hypothesis. Similarly, the ΔE distribution for $K^*(K\pi)J/\psi$ is shifted to the negative side because the charged pion is not accounted for when calculating ΔE.

The goal of estimating the number of backgrounds from $B^+ \to J/\psi\pi^+$ in data is to facilitate the signal extraction. The number of backgrounds in the MC sample is determined by counting the events from 10 million $B^+ \to J/\psi(\ell\ell)\pi^+$ signal MC events and multiplying it by the corresponding branching fraction. To translate this number to the background yield in data, the effective fake-rate of pion to kaon, denoted as $f(\pi \to K)$ at $\mathcal{L}_{K/\pi} > 0.6$, is multiplied by the MC background yield. The fake-rate is obtained from the performance group, and is calculated by weighting the normalized bin entries of pions in $B^+ \to J/\psi(\ell\ell)\pi^+$ MC events in the $p - \theta$ bins. The effective fake-rate and background yields in data and simulation can be found in Table 8.1.

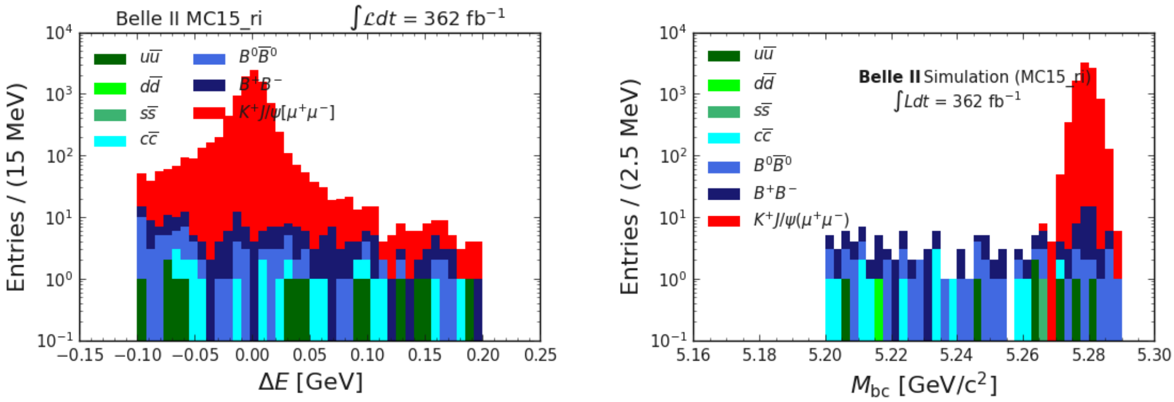


Figure 8.5: (left) ΔE, and (right) M_{bc} distributions of $B^+ \to J/\psi(\mu\mu)K^+$ after applying all the selection.

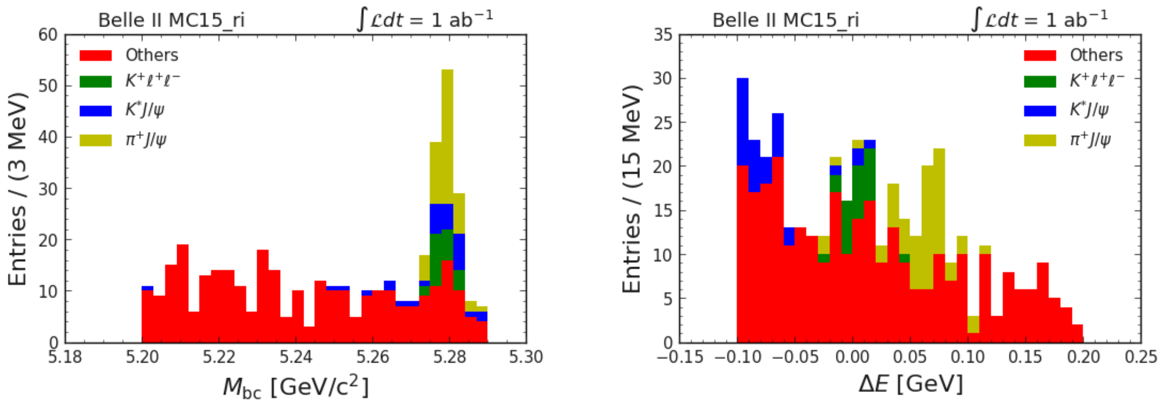


Figure 8.6: (left) ΔE, and (right) M_{bc} distributions for remaining backgrounds of $B^+ \to J/\psi K^+$ after applying all the selection.

8.3 Signal extraction

The signal yield is obtained through an extended maximum-likelihood fit to the unbinned distributions of M_{bc} and ΔE. For the neutral B channels, the fit consists of two components: one for the correctly reconstructed signal and another for misreconstructed background events. In case of the charged channels, there is an additional component in the fit to model $B^+ \to J/\psi\pi^+$ background events.

Details regarding the signal and background PDFs can be found in Section 7.4. During the fit to the data, all parameters of the signal PDF, except for the mean and width, are floated. Additionally, the shapes of all background PDFs are floated in the fit. The distributions of M_{bc} and ΔE for each B channel are shown in Figs. 8.7 and 8.8, respectively, with the fit results

Table 8.1: Estimation of the $B^+ \to J/\psi\pi^+$ background in data.

Channel	$f_{\mathrm{data/MC}}$	N_{MC} (362 fb^{-1})	N_{data}(362 fb^{-1})
$B^+ \to \pi^+ J/\psi(e^+e^-)$	1.109 ± 0.008	22.1 ± 0.1	24.5 ± 0.2
$B^+ \to \pi^+ J/\psi(\mu^+\mu^-)$	1.109 ± 0.008	23.0 ± 0.1	25.6 ± 0.2

superimposed.

8.4 Observable measurements and results in 189 fb^{-1}

We perform a standalone analysis in a smaller dataset and measure the branching fraction, lepton universality, and isospin asymmetry from the $B \to J/\psi K$ sample. In this analysis, we have not applied any MVA selection. The branching fractions are determined using the relation

$$\mathcal{B} = \frac{n_{\text{sig}}}{2\, N_{B\overline{B}}\, f^i\, \epsilon}, \tag{8.1}$$

where n_{sig} is the signal yield determined by the fit, $N_{B\overline{B}}$ is the number of $B\overline{B}$ events, ϵ is the signal selection efficiency, f^i is $f^{\pm}$ for the charged channels and f^{00} for the neutral channels; $f^{\pm}(f^{00})$ is the branching fraction of $\Upsilon(4S)$ to charged (neutral) $B\overline{B}$ assuming $\mathcal{B}\left(\Upsilon(4S) \to B\overline{B}\right) = 1$. We use the values $f^{00} = 0.487 \pm 0.013$ and $f^{\pm} = 1 - f^{00} = 0.513 \pm 0.013$ reported in Ref. [41] instead of the world average [44], as the latter is dominated by measurements that assume isospin symmetry in $B \to J/\psi K$ decays. The measurement $f^{\pm}/f^{00}$ [64] was not published by the time this dataset was analyzed, hence we used a different strategy as compared to Section 7.6 to calculate $f^{\pm}$.

We calculate $R_K(J/\psi)$ using the relation

$$R_K(J/\psi) = \frac{\mathcal{B}\,(J/\psi(\mu^+\mu^-)K)}{\mathcal{B}\,(J/\psi(e^+e^-)K)} = \frac{n_{\text{sig}}^{J/\psi(\mu^+\mu^-)K}}{n_{\text{sig}}^{J/\psi(e^+e^-)K}} \times \frac{\epsilon^{J/\psi(e^+e^-)K}}{\epsilon^{J/\psi(\mu^+\mu^-)K}}, \tag{8.2}$$

The branching fraction predictions suffer from uncertainties related to form factors. In contrast, observables like R_K or A_I are theoretically clean due to cancellation of these factors in the ratio. The isospin asymmetry can be defined as,

$$A_I = \frac{\Gamma[B^0 \to K^0 J/\psi(\ell^+\ell^-)] - \Gamma[B^+ \to K^+ J/\psi(\ell^+\ell^-)]}{\Gamma[B^0 \to K^0 J/\psi(\ell^+\ell^-)] + \Gamma[B^+ \to K^+ J/\psi(\ell^+\ell^-)]}, \tag{8.3}$$

where Γ is the rate of the particular decay. From the above expression, we can easily show,

$$A_I = \frac{2\,(\tau_{B^+}/\tau_{B^0})(f^{\pm}/f^{00})(n_{\text{sig}}/\epsilon)|_{J/\psi(\ell^+\ell^-)K_S^0} - (n_{\text{sig}}/\epsilon)|_{J/\psi(\ell^+\ell^-)K^+}}{2\,(\tau_{B^+}/\tau_{B^0})(f^{\pm}/f^{00})(n_{\text{sig}}/\epsilon)|_{J/\psi(\ell^+\ell^-)K_S^0} + (n_{\text{sig}}/\epsilon)|_{J/\psi(\ell^+\ell^-)K^+}}, \tag{8.4}$$

where $(\tau_{B^+}/\tau_{B^0}) = 1.076 \pm 0.004$ [9] is the ratio of the lifetimes of the charged and neutral B meson, and $(f^{\pm}/f^{00}) = 1.053 \pm 0.052$ [41]. The factor of 2 arises because a K^0 forms a K_S^0 meson half the time.

The signal efficiencies are initially measured with simulated samples and are subsequently corrected for potential data-simulation differences. In the simulated sample, the signal efficiency is determined as the number of correctly reconstructed signal events divided by total the number of generated events. To account for discrepancies between data and simulation arising from hadron and lepton identification selections, a correction factor is applied. This factor is calculated with a dedicated control sample, which is discussed in detail in Chapter 10. Results are listed in Tables 8.2 and 8.3.

Table 8.2: Reconstruction efficiency, signal yield, branching fraction, and world average [9] of the latter for each channel. Where two uncertainties are given, the first is statistical, and the second is systematic, and the single uncertainty corresponds to the statistical component.

Channel	ϵ (%)	n_{sig}	$\mathcal{B}$ (10^{-5})	World average (10^{-5})
$B^+ \to J/\psi(e^+e^-)K^+$	30.4	3706 ± 62	$6.00 \pm 0.10 \pm 0.19$	6.09 ± 0.12
$B^+ \to J/\psi(\mu^+\mu^-)K^+$	37.2	4578 ± 62	$6.06 \pm 0.09 \pm 0.19$	6.08 ± 0.12
$B^0 \to J/\psi(e^+e^-)K_S^0$	20.4	1052 ± 33	$2.67 \pm 0.08 \pm 0.12$	2.66 ± 0.06
$B^0 \to J/\psi(\mu^+\mu^-)K_S^0$	25.0	1343 ± 37	$2.78 \pm 0.08 \pm 0.12$	2.66 ± 0.06

Table 8.3: Measured A_I and $R_K\,(J/\psi)$, where the first quoted uncertainty is statistical and the second is systematic. Their world averages are also listed.

Observable	Measured value	World average
$A_I\,(J/\psi(ee)K)$	$-0.022 \pm 0.016 \pm 0.030$	-0.002 ± 0.025
$A_I\,(J/\psi(\mu\mu)K)$	$-0.006 \pm 0.015 \pm 0.030$	-0.002 ± 0.025
$R_{K^+}\,(J/\psi)$	$1.009 \pm 0.022 \pm 0.008$	0.990 ± 0.011
$R_{K_S^0}\,(J/\psi)$	$1.042 \pm 0.042 \pm 0.008$	1.001 ± 0.015

Measurements of the $B \to J/\psi K$ decay were performed with a dataset of $189\,\mathrm{fb}^{-1}$ in 2022. At that time, we had planned to explore lepton universality in both charged and neutral FCNC channels using larger data samples. Since $B^0 \to J/\psi K_S^0$ serves as a validation channel for the neutral FCNC decay $B^0 \to K_S^0\ell\ell$, this channel was considered in 2022. However, considering the level of integrated luminosity ($362\,\mathrm{fb}^{-1}$), which was not significantly large, we decided to drop $B^0 \to K_S^0\ell\ell$. This decision was motivated by the five-times lower branching fraction [9] expected for the neutral FCNC decay compared to its charged counterpart.

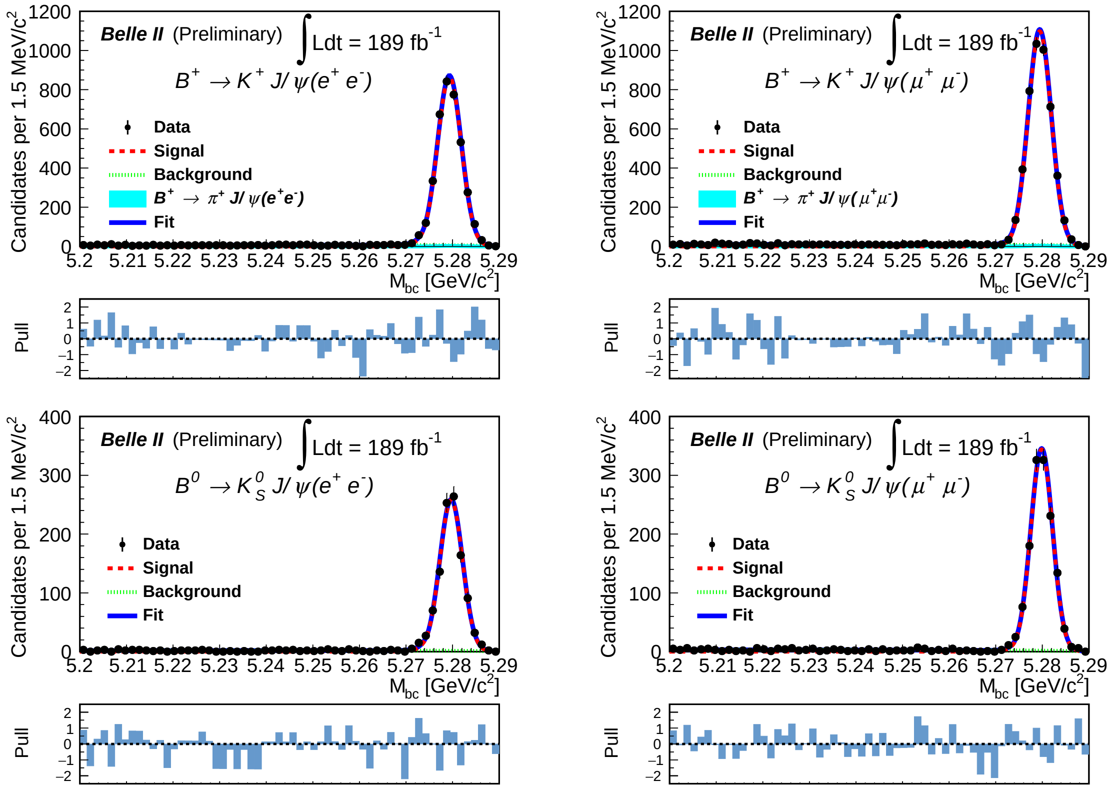


Figure 8.7: M_{bc} distributions for each $B \to J/\psi(\ell^+\ell^-)K$ channel with the fit result superimposed (top) and pull distribution with respect to the fit result (bottom). Black dots with error bars denote the data, solid blue curves denote the total fit, dashed red curves are the signal component, dotted green curves are the background component, and filled cyan regions in the charged channel are the $B^+ \to J/\psi\pi^+$ component.

8.5 Observable measurements in 362 fb^{-1}

The branching fraction of $B \to J/\psi K$ holds a special importance in the context of hadron collider experiments, serving as a crucial input for various other measurements. To ensure the integrity of our analysis and maintain blindness in the process, we intend to measure the branching fraction without applying any selection to MVA classifier output, as the classifier does not really help to suppress backgrounds to $B \to J/\psi K$. We use $f^{\pm} = 0.519 \pm 0.029$ as described in Section 7.6 for this measurement. On the other hand, to validate the $B \to K\ell\ell$ analysis, we perform a data-MC comparison of the MVA output distribution. The result from the simulated sample equivalent to data luminosity is shown in Fig. 8.4.

Figure 8.9 shows the MVA classifier output distributions in both the simulated and data samples. We find the simulation to exhibit a stronger peak compared to the data. Given that we

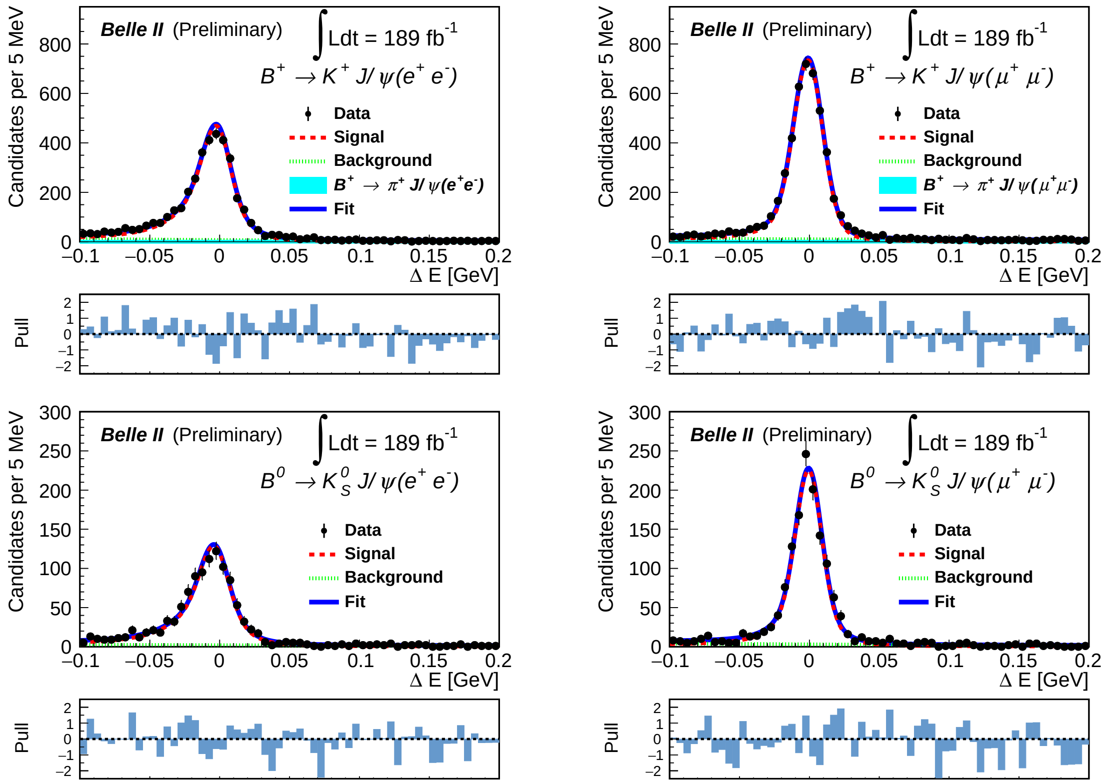


Figure 8.8: ΔE distributions for each $B \to J/\psi(\ell^+\ell^-)K$ channel with the fit result superimposed (top) and pull distribution with respect to the fit result (bottom). Black dots with error bars denote the data, solid blue curves denote the total fit, dashed red curves are the signal component, dotted green curves are the background component, and filled cyan regions in the charged channel are the $B^+ \to J/\psi\pi^+$ component.

Table 8.4: Reconstruction efficiency (simulation w/o data-MC correction), signal yield, and branching fraction for each channel. The uncertainty corresponds to the statistical component.

Channel	ϵ (%)	n_{sig}	$\mathcal{B}$ (10^{-5})	$R_K(J/\psi)$
$B^+ \to J/\psi(e^+e^-)K^+$	36.9	8567 ± 93	5.93 ± 0.06	
				1.000 ± 0.015
$B^+ \to J/\psi(\mu^+\mu^-)K^+$	39.2	9095 ± 96	5.93 ± 0.06	

are not applying any BDT selection to measure the observables of $B \to J/\psi K$, it is very unlikely that this discrepancy will directly impact those observables. We attribute this difference to minor discrepancies found in the variable R_2 between the data and the simulation. We intentionally allowed for these differences since R_2 is a key feature of the MVA classifier. The rationale behind this decision was to address any inconsistencies introduced in the analysis of

$B \to K\ell\ell$ by applying the correction factors discussed in Section. 10.9.

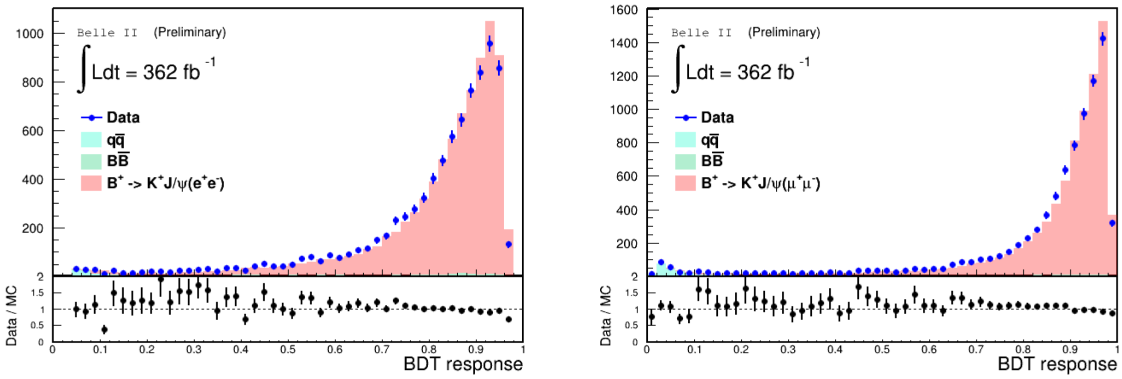


Figure 8.9: Distributions of the MVA classifier output in the $q^2 \in J/\psi$ region. The blue dots represent the data corresponding to an integrated luminosity of $362\,\mathrm{fb}^{-1}$. The stacked red, cyan, and green histograms represent MC events, also equivalent to $362\,\mathrm{fb}^{-1}$, after applying data-simulation correction factors that account for discrepancies arising from lepton and hadron identification.

Chapter 9

Overseeing signal with a smaller dataset

While the data results are yet to be reported, the analysis has undergone rigorous review processes, including the completion of the working group review phase. Currently, a committee comprising three members has been actively reviewing the analysis, and upon their permission, we will examine the data. The progress made in this section highlights the careful and meticulous approach taken in preparing for the final measurement of R_K on the data.

Though we have not shown the result of $B^+ \to K^+\ell^+\ell^-$ measurement in data as it is under the internal review, we have searched for its signal in the early dataset of Belle II. Using 63 fb^{-1} of data recorded by the experiment, a preliminary measurement of the $B^+ \to K^+\ell^+\ell^-$ signal yield is performed. The previously outlined analysis strategy includes an optimized selection based on an FOM defined in Eq. 4.1 of the MVA classifier output, resulting in a suppression of approximately 98% of background events. The $B^\pm \to K^\pm\ell^+\ell^-$ signal yield is extracted using an extended maximum-likelihood fit to the unbinned $M_{\rm bc}$ and ΔE distributions. In Fig 9.1, the signal enhanced projections for a fit to the distributions of $M_{\rm bc}$ and ΔE of $B^+ \to K^+\ell^+\ell^-$ is shown. The fit model for $M_{\rm bc}$ contains the following three components: a Crystal Ball function for the signal (red dot-dashed curve), another Crystal Ball function for the peaking background from $B^+ \to K^+\pi^+\pi^-$ decays (magenta filled area), and an ARGUS function for the combinatorial background from $q\overline{q}$ continuum and other B decays (green dashed curve). Similarly, the fit model for ΔE contains the following three components: a double-sided Crystal Ball summed with a Gaussian function for the signal (red dot-dashed curve), another double-sided Crystal Ball summed with a Gaussian function for the peaking background from $B^+ \to K^+\pi^+\pi^-$ decays (magenta filled area), and an exponential function for the combinatorial background from $q\overline{q}$

continuum and other B decays (green dashed curve). Black markers with error bars are data. The signal shape parameters obtained from correctly reconstructed simulated events are kept fixed in the fit. The ARGUS endpoint is fixed to the kinematic threshold of 5.29 GeV/c^2 and the second parameter is determined from the fit to data. We also determine the parameter of the exponential from the fit to data. The number of peaking background events is fixed to a value of 0.7, as estimated from simulation detailed in Section 5.5.2. The signal component has an overall significance of 2.7 standard deviations, with a yield of $8.6^{+4.3}_{-3.9} \pm 0.4$, where the quoted uncertainties are statistical and systematic, respectively. We obtain a signal yield of $8.6^{+4.3}_{-3.9}$ (stat.) $\pm$ 0.4 (syst.) events.

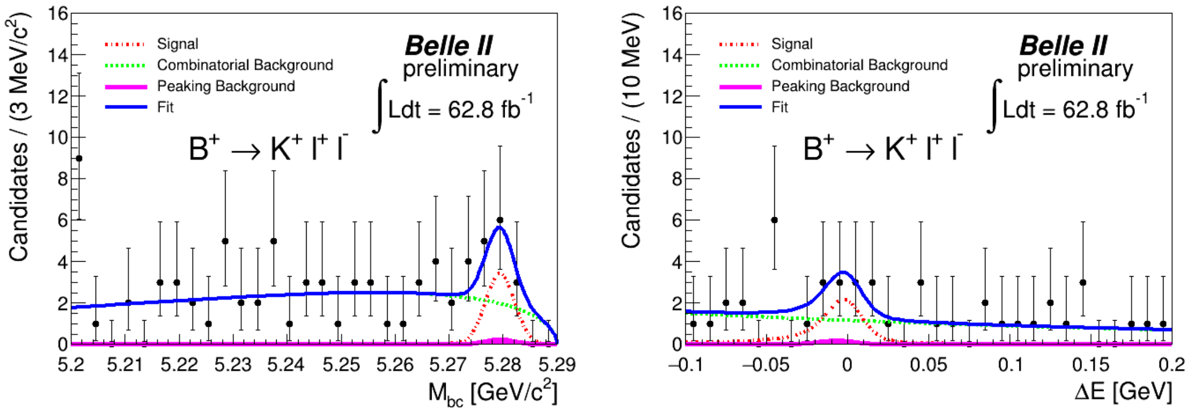


Figure 9.1: Signal enhanced projections for a fit to the distributions of M_{bc} and ΔE of $B^+ \to K^+\ell^+\ell^-$ candidates, where ℓ is an electron or a muon. The M_{bc} projection is obtained in a window $-0.06 < \Delta E < 0.04$ GeV whereas the ΔE projection is obtained for $M_{\text{bc}} > 5.27$ GeV/c^2.

Chapter 10

Systematic uncertainty

“Systematic uncertainties are all uncertainties that are not directly due to the statistics of the data.” [65]

This section presents a comprehensive study of the various sources of systematic uncertainties that affect the measurement of the observables of interest. Such sources include detector effects, theoretical uncertainties, and experimental limitations. Systematic uncertainties are evaluated to understand their impact on the final results and to ensure that the conclusions drawn are robust and reliable. The study of these uncertainties is an essential aspect of particle physics analyses, and it is crucial to establish confidence in the obtained results. Below we describe the potential sources of uncertainty that could affect the measurement of branching fraction, lepton universality, and isospin asymmetry. Before delving into the detailed analysis of each source, let’s first understand the uncertainty associated with data-simulation differences. Uncertainty can be attributed to each parameter in the final observable.

10.1 Types of systematic uncertainty

There are various types of systematic uncertainty. They are described below.

10.1.1 Imperfect simulation

The signal efficiency, calculated from simulation, is corrected for potential data-simulation differences. The uncertainty associated with these correction factors contributes to systematic uncertainty. Due to imperfect simulation, the data and MC efficiencies are not exactly the same. This introduces a systematic uncertainty, known as the multiplicative systematic uncertainty.

Let's assume the following for the efficiencies:

$$\epsilon_{\rm MC} = \epsilon_0 \times \epsilon_{\rm pid}^{\rm mc} \times \epsilon_{\rm track}^{\rm mc} \times \ldots \times \epsilon_n^{\rm mc},$$

where ϵ_0 is the efficiency with no selection applied, $\epsilon_{\rm pid}^{\rm mc}$ is the efficiency with PID selection applied at the MC level, and so on. Similarly, for data, we have:

$$\epsilon_{\rm data} = \epsilon_0 \times \epsilon_{\rm pid}^{\rm data} \times \epsilon_{\rm track}^{\rm data} \times \ldots \times \epsilon_n^{\rm data},$$

so:

$$\epsilon_{\rm data} = \epsilon_{\rm MC} \times r_{\rm pid} \times r_{\rm track} \times \ldots \times r_{\rm n},$$

where $r_n = \epsilon_n^{\rm data}/\epsilon_n^{\rm mc}$.

Assuming that various sources of systematic uncertainty causing data-MC differences are uncorrelated, we can calculate the uncertainty as:

$$\frac{\sigma_{\epsilon_{\rm data}}}{\epsilon_{\rm data}} = \sqrt{\left(\frac{\sigma_{\epsilon_{\rm MC}}}{\epsilon_{\rm MC}}\right)^2 + \left(\frac{\sigma_{r_{\rm pid}}}{r_{\rm pid}}\right)^2 + \left(\frac{\sigma_{r_{\rm track}}}{r_{\rm track}}\right)^2 + \ldots + \left(\frac{\sigma_{r_{\rm n}}}{r_{\rm n}}\right)^2}.$$

The following sources of systematic uncertainty fall into this category:

- Particle identification
  - Lepton identification
  - Kaon identification
- Track reconstruction
- Signal MC sample size
- MVA classifier output
- K_S^0 identification

10.1.2 Fixed shape parameter

The uncertainty for fixed shape parameters while extracting signal from the data using any maximum-likelihood fit is also a common source of uncertainty.

- Charmless $B \to K\pi\pi$ background
- PDF shape parameters

10.1.3 Theoretical modeling and external input

Nearly every data analysis relies on input from theoretical models. It is quite common to have multiple theoretical descriptions, each with its own limitations when representing the real world. Typically, one of these descriptions is chosen as the default, while the others are varied to observe their impact on the final observable. The variation in the observable is then treated as a source of systematic uncertainty. The "MC decay model" is an example of a theoretical description falling into this category.

There are some parameters involved in the analysis which do not directly come from the theoretical modeling but from other experimental result. Number of $B\overline{B}$ events and $f^{\pm}$ fall into this category that come with their respective uncertainty.

10.2 Lepton identification

The ratio of efficiency (r_k) for lepton ID selection are calculated in a three-dimensional (momentum, polar angle and charge) bin k from $J/\psi \to \ell^+\ell^-$, $e^+e^- \to e^+e^-(\gamma)$, $e^+e^- \to \mu^+\mu^-\gamma$, and $e^+e^- \to (e^+e^-)\ell^+\ell^-$ control samples. Along with the ratio of efficiencies, their statistical (δr^{stat}) and systematic (δr^{syst}) uncertainties are provided for each bin.

We use Eq. 10.1 to calculate correction factors corresponding to each lepton. Each bin is statistically independent with no correlated systematic uncertainties among the bins. We use Eq. 10.2 to calculate the uncertainty for each correction factor.

$$R = \frac{\sum_l n_l R_l}{\sum_l n_l} \tag{10.1}$$

$$\delta r = \frac{\sqrt{\sum_l (n_k \delta r_k^{\text{stat+sys}})^2}}{\sum_k n_k} \tag{10.2}$$

These correction factors are listed in Table 10.1. Finally, the overall correction factor for two leptons is a multiplication of individual correction factor and overall uncertainty is the propagation of uncertainties of individual lepton correction factor (Table 10.2).

10.3 Kaon identification

The $D^{*+} \to D^0[K^-\pi^+]\pi^+$ sample is used to estimate the difference in kaon identification efficiency between data and simulation. The correction factors are calculated as functions of

Table 10.1: Correction factor for each lepton, and its relative uncertainties (stat+sys) (in %) for two channels.

Lepton	Channel	Correction factor			
		$q^2 \in [0.10, 8.12]$	$q^2 > 14.18$	full-q^2	$q^2 \in [8.50(8.75), 10.20]$ for $KJ/\psi(\ell\ell)$
e^+	$B^+ \to K^+e^+e^-$ or $B^+ \to K^+J/\psi(e^+e^-)$	$0.9975^{+0.27}_{-0.04}$	$1.0020^{+0.11}_{-0.07}$	$0.9991^{+0.17}_{-0.14}$	$0.9980^{+0.35}_{-0.04}$
e^-		$0.9937^{+0.38}_{-0.10}$	$0.9991^{+0.20}_{-0.05}$	$0.9956^{+0.29}_{-0.10}$	$0.9941^{+0.46}_{-0.13}$
μ^+	$B^+ \to K^+\mu^+\mu^-$ or $B^+ \to K^+J/\psi(\mu^+\mu^-)$	$0.9418^{+0.16}_{-0.03}$	$0.9639^{+0.20}_{-0.19}$	$0.9508^{+0.17}_{-0.09}$	$0.9538^{+0.16}_{-0.05}$
μ^-		$0.9443^{+0.04}_{-0.04}$	$0.9697^{+0.90}_{-0.90}$	$0.9547^{+0.39}_{-0.39}$	$0.9573^{+0.31}_{-0.17}$

particle momentum and polar angle.

Suppose $R_l = \epsilon_l^{\text{data}}/\epsilon_l^{\text{MC}}$ is the ratio of data to MC efficiency for a particular 2D bin of the above-mentioned table and n_l is the number of true charged kaons in that bin, then Eq. 10.1 refers to the overall efficiency correction factor associated to the hadronID.

Now suppose the uncertainty associated to R_ℓ is given by $\pm\delta R_\ell^{\text{stat}} \pm \delta R_\ell^{\text{syst}}$. Since the statistical component of the uncertainty in each bin is independent, they will be added in quadrature in calculating the uncertainty of R. On the other hand, the systematic components of each bin are correlated to each other, hence they are simply added. So the overall systematic uncertainty is given as

$$\delta R = \frac{\sqrt{\sum_l (n_l \delta R_l^{\text{stat}})^2}}{\sum_l n_l} \bigoplus \frac{\sum_l n_l \delta R_l^{\text{syst}}}{\sum_l n_l}. \tag{10.3}$$

The correction factor and associated systematic uncertainty calculated for three q^2 bins of non-resonant $B^+ \to K^+\ell^+\ell^-$ channel and for $B^+ \to K^+J/\psi(\ell^+\ell^-)$ are shown in Table 10.3.

Table 10.2: Overall systematics due to Lepton ID.

Channel	q^2 bins $\left[\mathrm{GeV}^2/c^4\right]$	Overall correction factor (R)	Rel. uncert. (δR) in %
$B^+ \to K^+e^+e^-$	$q^2 \in [0.10, 8.12]$	0.9912	$^{+0.47}_{-0.11}$
$B^+ \to K^+e^+e^-$	$q^2 > 14.18$	1.0011	$^{+0.20}_{-0.16}$
$B^+ \to K^+e^+e^-$	full-q^2	0.9947	$^{+0.35}_{-0.11}$
$B^+ \to K^+J/\psi(e^+e^-)$	$q^2 \in [8.50, 10.20]$	0.9921	$^{+0.58}_{-0.14}$
$B^+ \to K^+\mu^+\mu^-$	$q^2 \in [0.10, 8.12]$	0.8893	$^{+0.17}_{-0.05}$
$B^+ \to K^+\mu^+\mu^-$	$q^2 > 14.18$	0.9347	$^{+0.98}_{-0.95}$
$B^+ \to K^+\mu^+\mu^-$	full-q^2	0.9077	$^{+0.45}_{-0.42}$
$B^+ \to K^+J/\psi(\mu^+\mu^-)$	$q^2 \in [8.75, 10.20]$	0.9131	$^{+0.37}_{-0.19}$

10.4 K_S^0 identification

From studies performed with a sample of $D^{*+} \to D^0\left(\to K_S^0\pi^+\pi^-\right)\pi^+$ decays, we observe that the data–simulation ratio of K_S^0 reconstruction efficiency changes linearly as a function of the flight distance, which corresponds to an uncertainty of 0.4% per cm of average flight length. The average flight length of K_S^0 for the signal channel is estimated to be 7.6 cm with simulated events. Hence, we assign a 3.0% systematic uncertainty for branching fraction and isospin asymmetry measurements involving a K_S^0 meson.

10.5 Systematic from number of $B\overline{B}$ events

The number of $B\overline{B}$ events in the LS1 is $N_{B\overline{B}} = (387 \pm 6) \times 10^6$. An overall uncertainty of 1.5% is thus assigned as a systematic uncertainty on the measured branching fractions. The B meson counting is briefly discussed in Appendix A.

Table 10.3: Overall systematics due to Hadron ID.

Channel	q^2 bins $\left[\mathrm{GeV}^2/c^4\right]$	Overall correction (R) factor	Rel. uncert. (δR) in %
$B^+ \to K^+e^+e^-$	$q^2 \in [0.10, 8.12]$	0.9566	0.09
$B^+ \to K^+e^+e^-$	$q^2 > 14.18$	0.9770	0.13
$B^+ \to K^+e^+e^-$	full-q^2	0.9633	0.07
$B^+ \to K^+J/\psi(e^+e^-)$	$q^2 \in [8.50, 10.20]$	0.9525	0.10
$B^+ \to K^+\mu^+\mu^-$	$q^2 \in [0.10, 8.12]$	0.9564	0.09
$B^+ \to K^+\mu^+\mu^-$	$q^2 > 14.18$	0.9764	0.13
$B^+ \to K^+\mu^+\mu^-$	full-q^2	0.9543	0.07
$B^+ \to K^+J/\psi(\mu^+\mu^-)$	$q^2 \in [8.75, 10.20]$	0.9527	0.10

10.6 $f^{\pm}$

Since $N_{B^+B^-} = f^{\pm}N_{B\bar{B}}$ (Section 7.6), we need to assign an additional 5.6% relative uncertainty for the branching fraction measurements.

10.7 Tracking efficiency

The track finding efficiency is determined with $e^+e^- \to \tau^+\tau^-$ events, where one tau lepton decays leptonically and the other decays hadronically, denoted as 1-prong and 3-prong τ-decays, employing the tag and probe method. Tracking efficiency is assessed using three high-quality tracks with total charge ±1, alongside an extra probe track inferred from charge conservation. A systematic uncertainty of 0.24% per charged track (signal side B decay) is assigned due to the difference in track selection efficiency between data and simulation, resulting in a total multiplicative systematic uncertainty of 0.72% for the branching fraction measurements. For R_K measurements, systematic uncertainty due to tracking cancels in the ratio, hence no systematic uncertainty is assigned for this source.

10.8 Limited MC statistics

The uncertainty on signal efficiency due to limited MC size is considered as a source of systematic. This is calculated using the formula:

$$\sqrt{\frac{\epsilon(1-\epsilon)}{N}},$$

where ϵ is the efficiency and N is the total number of events generated (50 million for nonresonant and 10 million for resonant channel). The systematic uncertainties are summarized in Table 10.4.

Table 10.4: Systematic due to limited signal MC statistics.

Channel	ϵ (%)	Systematic due to ϵ (%)
$B^+ \to K^+e^+e^-$	23.18	0.03
$B^+ \to K^+\mu^+\mu^-$	24.03	0.03
$B^+ \to K^+J/\psi(e^+e^-)$	35.23	0.04
$B^+ \to K^+J/\psi(\mu^+\mu^-)$	37.68	0.04

10.9 MVA classifier output

We estimate the difference in efficiency due to BDT classifier selection between data and simulation using a control region. By applying a BDT selection threshold of 0.5, the data are divided into two sub-samples. The yields in these sub-samples, denoted as $N_{\text{sig}}^{\text{low}}$ and $N_{\text{sig}}^{\text{high}}$, are related to the overall signal yield $N_{\text{sig}}^{\text{tot}}$ through the efficiency ϵ as below

$$\begin{aligned} N_{\text{sig}}^{\text{low}} &= (1-\epsilon)N_{\text{sig}}^{\text{tot}} \\ N_{\text{sig}}^{\text{high}} &= \epsilon N_{\text{sig}}^{\text{tot}}. \end{aligned} \tag{10.4}$$

A simultaneous fit is performed to extract ϵ_{MC} and ϵ_{data}. The fit projections for the muon channel is shown in Fig 10.1. To correct for the MC efficiency, the ratio between the data and MC efficiency, denoted as $r = \epsilon_{\text{data}}/\epsilon_{\text{MC}}$, is calculated. The uncertainty on r is assigned as a systematic uncertainty. In the case of branching fraction measurement of non-resonant $B^+ \to K^+\ell^+\ell^-$ modes, systematics uncertainties of 0.38% and 0.30% are assigned for electron and muon modes, respectively. For the measurement of R_K a systematic uncertainty of 0.48% is assigned due to the selection of MVA classifier output. The MVA classifier systematic uncertainty is summarised in Table 10.5.

10.10 PDF shape

During the nominal 3D fit, various PDF shape parameters and the charmless $B^+ \to K^+\pi^+\pi^-$ yield, are fixed in the fit. To determine the systematics uncertainty due to these fixed parameters, they are subsequently varied by $\pm 1\sigma$ around their mean values. To consider the correlation

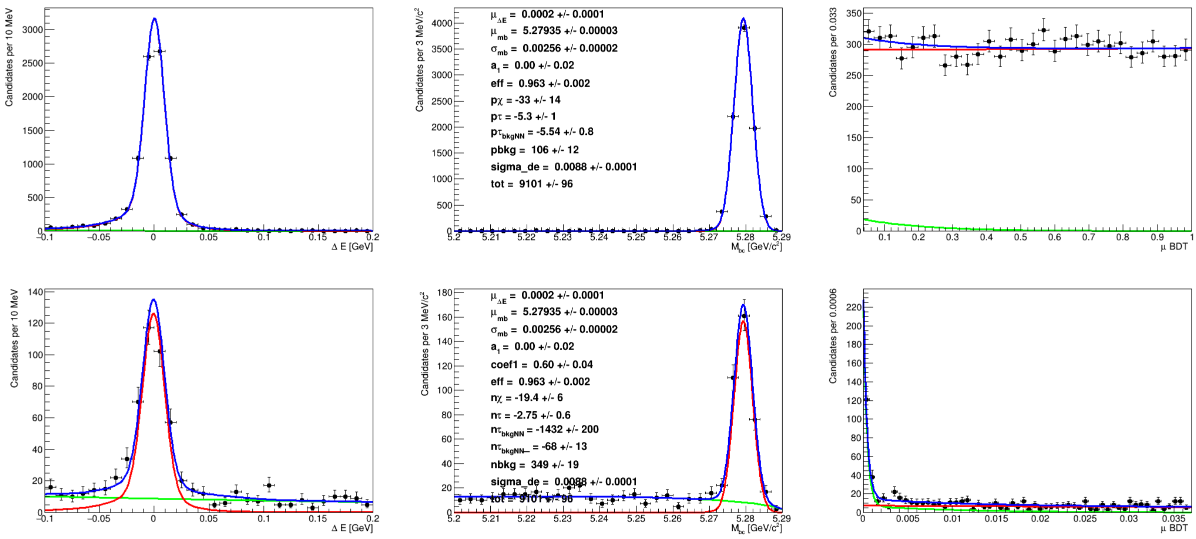


Figure 10.1: Simultaneous fit projection to MC distributions of ΔE (left), $M_{\rm bc}$ (center), and C' (right) of the $B^+ \to K^+ J/\psi(\mu^+\mu^-)$ channel. The top row corresponds to the subsample with BDT > 0.5, while the bottom row corresponds to the subsample with BDT < 0.5.

Table 10.5: Systematic due to MVA classifier output.

Channel	$\epsilon_{\rm MC}$	$\epsilon_{\rm data}$	$r = \epsilon_{\rm data}/\epsilon_{\rm MC}$	$\Delta r/r$ (%)
$B^+ \to K^+ J/\psi(e^+e^-)$	0.960 ± 0.002	0.942 ± 0.003	0.981 ± 0.004	0.38
$B^+ \to K^+ J/\psi(\mu^+\mu^-)$	0.963 ± 0.002	0.952 ± 0.002	0.988 ± 0.003	0.30

among the shape parameters, they are varied simultaneously using a Gaussian constrained random variable, with the mean and width values obtained from an MC fit. The resulting set of parameters is then used to fit an MC sample, and the fractional change in yield with respect to the nominal one is calculated. After repeating the process 1000 times, a Gaussian distribution is fitted to the relative yield changes, and the width of this distribution is used as the systematic uncertainty. The PDF shape systematic uncertainties are summarised in Table 10.6.

10.11 MC decay model

We utilize the BTOSLLBALL [66] model from the EVTGEN [49] package to simulate $B \to K\ell\ell$ events. To assess the systematic uncertainties associated with the BTOSLLBALL decay model, we compute weight factors in 2D bins of q^2 and $\cos\theta$, where θ denotes the angle between the direction of motion of the $\overline{B}$ meson and the negatively charged lepton in the dilepton center-of-mass frame.

Table 10.6: Systematic due to PDF shape parameters.

Channel	Systematic (%)
$B^+ \to K^+ J/\psi(e^+e^-)$	0.04
$B^+ \to K^+ J/\psi(\mu^+\mu^-)$	0.07
$B^+ \to K^+ e^+ e^-$	0.56
$B^+ \to K^+ \mu^+ \mu^-$	1.23

These weight factors represent the ratio of the partial decay width in the SM to that in the BTOSLLBALL model. We utilize Eq. 4.1 from Ref. [67] to derive the distribution of q^2 vs. $\cos\theta$ and Ref. [68] to get the latest form factor, from the SM.

For different q^2 bins, we compute the weight factors using the expression:

$$w_i = \frac{\left[\Delta\Gamma/\Delta q_i^2\right]\text{Theory}}{\left[\Delta\Gamma/\Delta q_i^2\right]\text{BTOSLLBALL}}. \tag{10.5}$$

Subsequently, after applying all the selection criteria, we adjust the $q^2 - \cos\theta$ distribution by rescaling it with these weight factors. The systematic uncertainty was quantified as the fractional change in the integrated area of the 2D distribution. For both the electron and muon channels of the non-resonant $B \to K\ell\ell$ branching fraction, we assign a 0.5% systematic uncertainty.

10.12 Propagation of uncertainty

Up to this point, we have covered several sources of systematic uncertainty, though some of them cancel and do not contribute to the final uncertainty. Table 10.7 provides a summary of each source of uncertainty and its impact on the measurement of branching fraction and R_K. We apply error propagation formulas to calculate the total uncertainty on the final observable. If we have a function of several variables, say $y = y(x_1, x_2, ..x_n)$, then the error propagation formula says the standard deviation of the y or σ_y to be

$$\sigma_y^2 = \sum_{i,j}^{n} \left[\frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}\right]_{\vec{X}=\vec{\mu}} V_{ij}, \tag{10.6}$$

where $\vec{\mu}$ is the mean of the variables $\vec{X} = (x_1, x_2, ..x_n)$ and V_{ij} is the covariance matrix of the set of variables $(x_1, x_2, ..x_n)$.

Table 10.7: Systematic uncertainties on $\mathcal{B}$, R_K and A_I

Source	BF	R_K or $R_K(J/\psi)$	A_I
Electron identification	✓	✓	X
Muon identification	✓	✓	X
Kaon identification	✓	X	✓
K_S^0 reconstruction	✓	X	✓
Number of $B\overline{B}$ events	✓	X	X
$f^{\pm}/f^{00}$	✓	X	✓
Track reconstruction	✓	X	✓
Limited MC statistics	✓	✓	✓
$B \to K\pi\pi$ or $B \to \pi J/\psi$ background	✓	✓	✓
PDF shape parameters	✓	✓	✓
MC decay model	✓(X for $J/\psi K$)	X	X
τ_{B^+}/τ_{B^0}	X	X	✓

Branching fraction measurement

The working formula of $\mathcal{B}$ is Eq. 8.1. The equation can be rearranged as

$$\mathcal{B} = \frac{N^s}{N_{B^{\pm}(B^0)} \times \epsilon_{\text{MC}} \times (f^{\text{tracking}} f^{\text{kaonID}} f^{\text{leptonID}})_{\text{data/MC}}},$$

where N^s is fitted yield, and f's are efficiency correction factors due to imperfect simulation. From Eq. 10.6, we can straight away derive, that if the individual terms in the functional form are only in multiplicative or divisive form then the relative error of different sources will be added in quadrature. Hence,

$$\Delta\mathcal{B}/\mathcal{B} = \sqrt{(\Delta N_s/N_s)^2 + (\Delta N_{B^{\pm}}/N_{B^{\pm}}))^2 + (\Delta\epsilon_{\text{MC}}/\epsilon_{\text{MC}})^2 + (\Delta f^{\text{tracking}}/f^{\text{tracking}})^2 \ldots} \; .$$

R_K measurement

The formula used to calculate the observable is given by Eq. 8.2. Since the efficiency of the two modes is proportional, the data-MC correction factors obtained from kaonID, tracking, and so on, cancel because they are the same for both channels. The total systematic uncertainty is obtained by taking the square root of the quadrature sum of relative uncertainties from various sources of systematic.

The systematic uncertainties for individual modes are listed in Table 10.8.

Table 10.8: Relative systematic uncertainties (%) on $\mathcal{B}(B^+ \to K^+\ell^+\ell^-)$, $\mathcal{B}(B^+ \to K^+J/\psi[\ell^+\ell^-])$, R_K and $R_K(J/\psi)$.

Source	$\mathcal{B}(B^+ \to K^+\ell^+\ell^-)$		R_K	$\mathcal{B}(B^+ \to K^+J/\psi[\ell^+\ell^-])$		$R_K(J/\psi)$
	e^+e^-	$\mu^+\mu^-$		e^+e^-	$\mu^+\mu^-$	
Number of $B\overline{B}$ events	1.50	1.50	–	1.50	1.50	–
PDF shape	0.56	1.23	1.35	0.04	0.07	0.08
Electron identification	$^{+0.47}_{-0.16}$	–	$^{+0.47}_{-0.16}$	$^{+0.58}_{-0.11}$	–	$^{+0.58}_{-0.14}$
Muon identification	–	$^{+0.45}_{-0.42}$	$^{+0.45}_{-0.42}$	–	$^{+0.37}_{-0.19}$	$^{+0.37}_{-0.19}$
Kaon identification	0.13	0.13	–	0.10	0.10	–
Tracking efficiency	0.72	0.72	–	0.72	0.72	–
BDT classifier	0.38	0.30	0.48	–	–	–
$\Upsilon(4S)$ branching fraction	5.60	5.60	–	5.60	5.60	–
Limited MC statistics	0.03	0.03	0.04	0.04	0.04	0.06
MC decay model	0.5	0.5	–	–	–	–
Total	6.0	7.1	1.5	5.9	5.9	0.8

A_I measurement

The working formulae for A_I is Eq. 8.4. From Eq. 10.6, we can derive

$$\Delta A_I = \alpha \sqrt{\sum_i (\Delta s_i/s_i)^2},$$

with,

$$\alpha = \frac{4(\tau_{B^+}/\tau_{B^0})(f^{\pm}/f^{00})(N_{\text{sig}}/\epsilon)\,|_{K^0_S\ell\ell}\,(N_{\text{sig}}/\epsilon)\,|_{K^+\ell\ell}}{[(2(\tau_{B^+}/\tau_{B^0})(f^{\pm}/f^{00})(N_{\text{sig}}/\epsilon)\,|_{K^0_S\ell\ell} + (N_{\text{sig}}/\epsilon)\,|_{K^+\ell\ell}]^2}.$$

and s_i stands for ϵ_{MC}, (τ_{B^+}/τ_{B^0}) and $(f^{\pm}/f^{00})$, and correction factors for kaonID, tracking efficiency (for charged kaon or K^0_S daughter tracks), K^0_S reconstruction efficiency, for different i. The systematic uncertainties on branching fraction for $B \to J/\psi K$, $R_K(J/\psi)$ and A_I as discussed in Section 8.4, are listed in Table 10.9.

Table 10.9: Relative systematic uncertainties (%) on $\mathcal{B}(B \to J/\psi K)$, $R_K\,(J/\psi)$, and absolute uncertainty on $A_I\,(B \to J/\psi K)$.

Source	$\mathcal{B}(B \to KJ/\psi)$				R_K		A_I	
	K^+	K^+	K^0_S	K^0_S	K^+	K^0		
	e^+e^-	$\mu^+\mu^-$	e^+e^-	$\mu^+\mu^-$			e^+e^-	$\mu^+\mu^-$
Number of $B\overline{B}$ events	1.5	1.5	1.5	1.5	–	–	–	–
PDF shape	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.1
Electron identification	0.6	–	0.6	–	0.6	0.6	–	–
Muon identification	–	0.4	–	0.4	0.4	0.4	–	–
Kaon identification	0.2	0.2	–	–	–	–	0.1	0.1
K^0_S reconstruction	–	–	3.0	3.0	–	–	1.5	1.5
Tracking efficiency	0.9	0.9	1.2	1.2	–	–	0.4	0.4
Limited MC statistics	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$\Upsilon(4S)$ branching fraction	2.6	2.6	2.6	2.6	–	–	2.6	2.6
τ_{B^+}/τ_{B^0}	–	–	–	–	–	–	0.2	0.2
Total	3.2	3.2	4.4	4.4	0.8	0.8	3.0	3.0

Chapter 11

Measurement of the branching fraction for $B \to K^*(892)\ell^+\ell^-$

11.1 Introduction

The decays $B \to K^*\ell^+\ell^-$ are another example of flavor-changing-neutral-current transitions of the $b \to s\ell^+\ell^-$ type. The capability of these decays to explore BSM physics has already been discussed in Chapter 2. The measurement of R_{K^*} [69] garnered a significant attention due to its deviation from the SM prediction. Although its most recent measurement [70] agrees with the SM, the significance of searching for LFU violation remains intact. Detecting these decays represents the first step towards measuring R_{K^*} at Belle II.

11.2 Event reconstruction and dataset

This study is done with a Belle II dataset of an integrated luminosity of $189\,\mathrm{fb}^{-1}$, equivalent to 198×10^6 $B\overline{B}$ events. The analysis strategy is similar to the $B^+ \to K^+\ell^+\ell^-$ analysis except some points, which are discussed in this section.

The π^0 candidates are reconstructed by pairing two photons, each having a minimum energy of 80, 30, or 60 MeV, depending on whether they are detected in the forward, barrel, or backward region of the ECL, respectively. The different energy thresholds are used to suppress beam background, which is higher in the endcap compared to the barrel section. The $K^*(892)$ candidates are formed by combining three final states, namely $K^+\pi^-$, $K_S^0\pi^+$, and $K^+\pi^0$. To select K^* candidates, we retain those falling within an invariant-mass window of $[0.796, 0.996]\,\mathrm{GeV}/c^2$, which is approximately four times the natural width of the K^* meson. The invariant-mass distribution in the simulated sample is shown in Fig. 11.1.

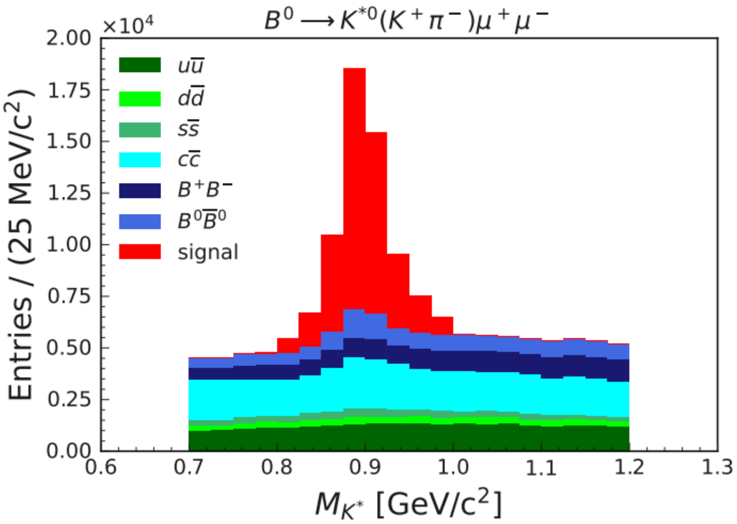


Figure 11.1: M_{K^*} distribution for the $B^0 \to K^*\mu^+\mu^-$ channel in the simulated sample.

For π^0 candidates, the allowed range of invariant mass is $[0.122-0.142]$ GeV$/c^2$, encompassing $\pm3\sigma$ around the known π^0 mass [9]. The $B \to K^*\ell\ell$ candidates are then reconstructed by combining K^* mesons with two oppositely charged leptons of same flavor.

To suppress continuum and B decay backgrounds, a BDT classifier similar to $B^+ \to K^+\ell^+\ell^-$ trained. For this preliminary study, we have applied a tight figure-of-merit based optimized selection on the BDT classifier output. Such a criterion rejects approximately 98% of the background, with a relative signal loss ranging between 30% and 35% depending on the decay channel. An example of the classifier output and its optimization are shown in Fig. 11.2.

For the dimuon decay modes, misreconstructed events for which $M_{\rm bc}$ and ΔE lie in the signal region are observed. These backgrounds are due to hadrons misidentified as muons and mistakenly associated with the signal decay. They are suppressed using vetoes. The veto windows are decided by changing particle mass hypotheses depending on the decay mode. For the $B^0 \to K^{*0}(K^+\pi^-)\mu^+\mu^-$ channel, the following event veto selections are applied to suppress peaking background: (a) a veto on the pion-muon invariant mass $M(\pi^-\mu^+) \notin [3.06, 3.11]$ GeV$/c^2$ is applied to suppress $B^0 \to J/\psi(\mu^+\mu^-)K^{*0}(K^+\pi^-)$ decays where the π^- is misidentified as a μ^-, and the μ^- as a π^-; (b) a veto on the kaon-pion-muon invariant mass $M(K^+\pi^-\mu^-) \notin [1.86, 1.885]$ GeV$/c^2$ is applied to suppress $B^0 \to D^-(K^+\pi^-\pi^-)\pi^+$ decays where a π^- from D^-

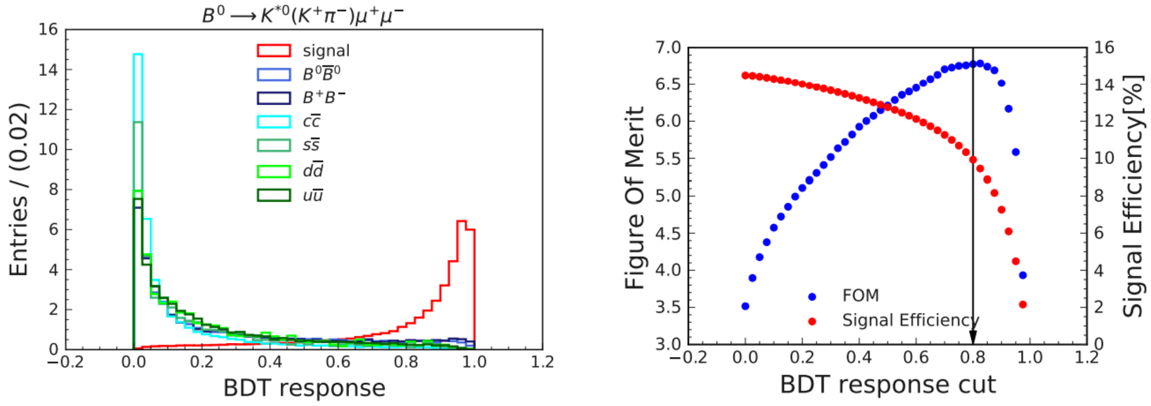


Figure 11.2: MVA classifier output distribution (left) and optimization of the classifier output for the $B^0 \to K^*\mu^+\mu^-$ channel in the simulated sample (right).

decay and the π^+ from B decay are misidentified as muons; (c) a veto on the kaon-muon invariant mass $M(K^+\mu^-) \notin [1.82, 1.9]$ GeV$/c^2$ is applied where the π^- from $\overline{D^0} \to K^+\pi^-$ decays is misidentified as a μ^-. For the $B^+ \to K^{*+}(K^0_S\pi^+)\mu^+\mu^-$ channel, two vetoes are applied: (a) a veto on the pion-muon invariant mass $M(\pi^+\mu^-) \notin [3.085, 3.105]$ GeV$/c^2$ is applied to suppress $B^+ \to J/\psi(\mu^+\mu^-)K^{*+}(K^0_S\pi^+)$ decays where the π^+ candidate is misidentified as a μ^+ and vice versa; (b) a veto in the kaon-pion-muon invariant mass, $M(K^0_S\pi^+\mu^-) \notin [1.857, 1.87]$ GeV$/c^2$ is applied to suppress $B^+ \to \overline{D^0}(K^0_S\pi^+\pi^-)\pi^+$ decays where a π^- from $\overline{D^0}$ decay and the π^+ from B decay are misidentified as muons. For the $B^+ \to K^{*+}(K^+\pi^0)\mu^+\mu^-$ channel, a veto on the kaon-pion-muon invariant mass $M(K^+\pi^0\mu^-) \notin [1.855, 1.87]$ GeV$/c^2$ is applied to suppress $B^+ \to \overline{D^0}(K^+\pi^0\pi^-)\pi^+$ decays where a π^- from $\overline{D^0}$ decay and a π^+ from B decay are misidentified as muons. These vetoes suppress the peaking backgrounds with a loss of 0.5% – 6.0% in signal efficiency, depending on the decay mode.

After applying all the selection criteria, the average candidate multiplicity ranges between 1.04 and 1.24. It is higher for the $B^+ \to K^{*+}(K^+\pi^0)\ell^+\ell^-$ channel due to larger probability of misreconstructed π^0's. In case of multiple candidates, we retain the one having ΔE closest to zero.

The signal yield is extracted from an extended maximum-likelihood fit to the unbinned $M_{\rm bc}$ and ΔE distributions, combining both charged and neutral B samples. The fit procedure is validated on simplified simulated experiments as well as with events reconstructed in the $B \to J/\psi(\ell^+\ell^-)K^*$ control channel. Signal-enhanced projection plots for $B \to K^*\mu^+\mu^-$, $B \to K^*e^+e^-$, and $B \to K^*\ell^+\ell^-$ are shown in Fig. 11.3.

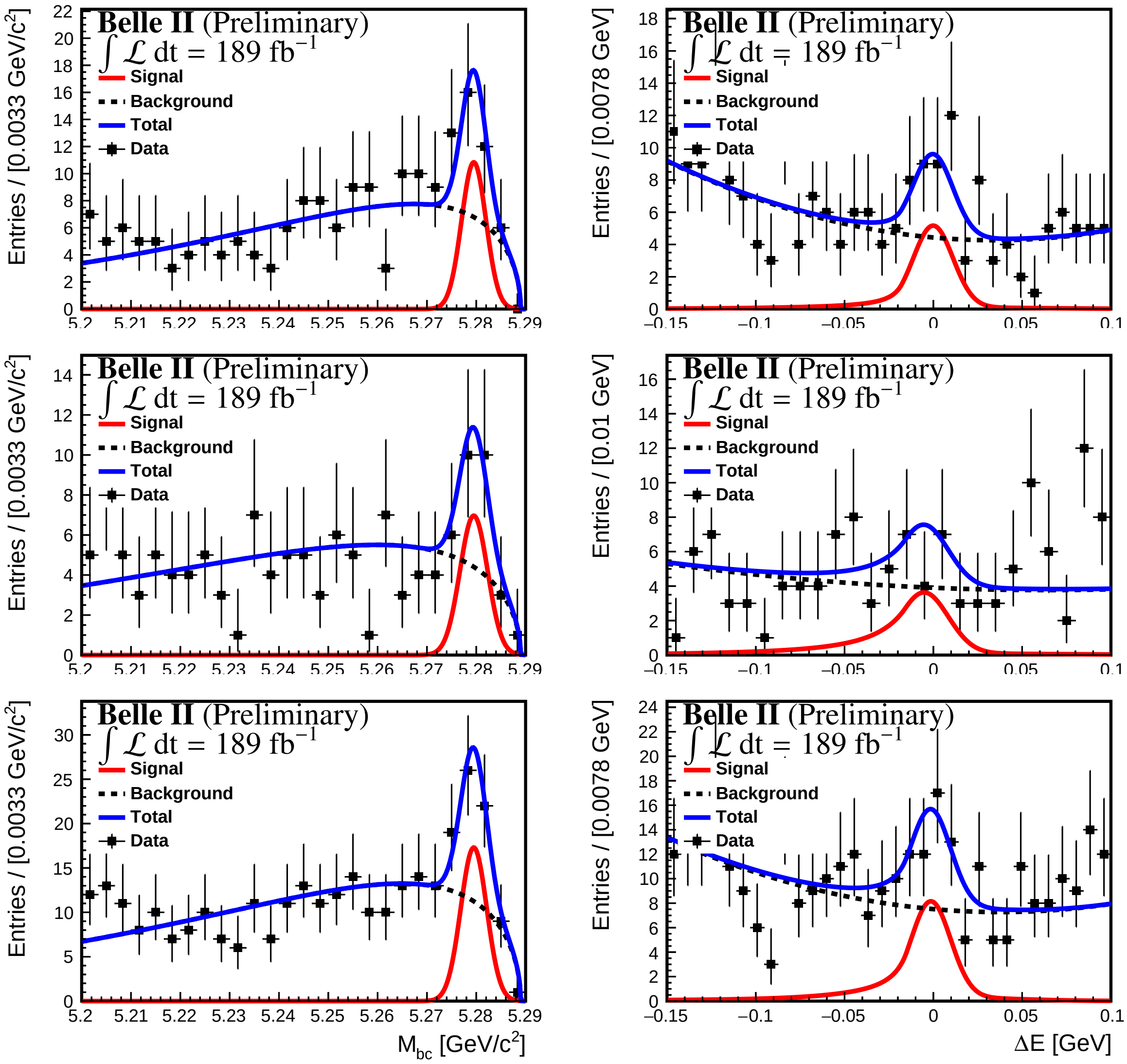


Figure 11.3: Distributions of $M_{\rm bc}$ (left) and ΔE (right) for $B \to K^*\mu^+\mu^-$ (top), $B \to K^*e^+e^-$ (middle), and $B \to K^*\ell^+\ell^-$ (bottom). Data points with error bars are superimposed on the blue solid curves, which show the total fit function, and red solid and black dotted curves represent the signal and background components, respectively. Candidates shown in the ΔE distributions are restricted to $M_{\rm bc} \in [5.27, 5.29]$ GeV$/c^2$ range and those in the $M_{\rm bc}$ distributions are restricted to $\Delta E \in [-0.05, 0.05]$ GeV.

11.3 Result and Summary

We reconstruct 22 ± 6, 18 ± 6, and 38 ± 9 signal events for $B \to K^*\mu^+\mu^-$, $B \to K^*e^+e^-$, and $B \to K^*\ell^+\ell^-$, corresponding to 4.8σ, 3.6σ, and 5.9σ, respectively. The significance σ is given as $\sqrt{-2\ln(\mathcal{L}_0/\mathcal{L})}$, where $\mathcal{L}_0$ is the likelihood with the signal yield fixed to zero and $\mathcal{L}$ is the maximum likelihood. The quoted uncertainties are statistical. The branching fractions for the entire q^2 region, excluding the charmonium resonances [J/ψ and $\psi(2S)$] and low-q^2 region to remove the $B \to K^*\gamma(\to e^+e^-)$ background, are

$$\begin{aligned}
\mathcal{B}(B \to K^*\mu^+\mu^-) &= (1.19 \pm 0.31^{+0.08}_{-0.07}) \times 10^{-6}, \\
\mathcal{B}(B \to K^*e^+e^-) &= (1.42 \pm 0.48 \pm 0.09) \times 10^{-6}, \\
\mathcal{B}(B \to K^*\ell^+\ell^-) &= (1.25 \pm 0.30^{+0.08}_{-0.07}) \times 10^{-6}.
\end{aligned}$$

Here, the first and second uncertainties are statistical and systematic, respectively. The precision of the results is limited by the sample size, and the results are compatible with world-average values [9].

Chapter 12

SVD performance in the commissioning phase

From April to July 2018, the Belle II collaboration undertook a commissioning phase marking a pivotal stage in the deployment of its vertex detector. During this critical period, before the full detector installation, a comprehensive data collection effort was undertaken. This chapter delves into a thorough investigation of the performance of the SVD throughout this commissioning phase. The aim is to assess and analyze how the SVD functioned under these early operational conditions, shedding light on its capabilities and potential enhancements, thereby contributing valuable insights to the broader objectives of Belle II.

12.1 Belle II commissioning

At Belle II, the decision to position the SVD even closer to the IP than its predecessor, Belle, is driven by the unique requirements of the experiment. This increased proximity does expose the SVD to higher levels of radiation damage, but it is a strategic choice made with careful consideration. The rationale behind this decision is the relatively lower boost of particles in the lab frame compared to the previous KEKB setup. Such lower boost results in shorter decay flight-lengths for B mesons and other short-lived particles. By placing the SVD closer to the IP, Belle II aims to compensate for this lower boost and enhance the precision of time-dependent CP violation analyses, where an accurate reconstruction of decay vertices is crucial. This strategic positioning allows Belle II to achieve the high precision required to explore rare physics phenomena and advance our understanding of fundamental principles in physics.

In the context of the Belle II commissioning, one of the primary challenges is the substantial increase in beam background due to the novel nano-beam scheme, where the beam compression is enhanced 20-fold compared to its predecessor, Belle. This increased beam background, attributed to Touschek and beam-gas scatterings at SuperKEKB, surged by factors of 20 and 100, respectively, in contrast to KEKB [71, 72].

The enhanced beam background, particularly concentrated at the VXD near the IP, posed a significant concern as it induces radiation damage to the detector, adversely impacting its operational performance and longevity. This background also results in increased hit occupancy, i.e., the number of hits per sensor per event, thereby degrading the tracking performance.

To comprehensively assess the impact of the beam background before full installation, a commissioning run, also known as Phase 2 (with physics data taking designated as Phase 3), is conducted. During this data-collection phase, all detectors except the VXD are fully installed, with one ladder per layer being deployed for the VXD. This practical exploration unveiled the real-world effects of beam background levels within the unique Belle II environment, providing crucial insights that complement our simulation studies. These findings play a pivotal role in our understanding of how high beam background influences detector performance during the Belle II commissioning phase. In this collected dataset, the performance study conducted by us examines the hit efficiency and spatial resolution of the SVD.

12.2 SVD subdetector

As discussed in Chapter 3 the Belle II SVD comprises several cylindrically arranged DSSD sensors placed around the IP. Each sensor is either rectangular or trapezoidal and has a thickness of $300-320\,\mu$m. The sensor has an n-type bulk with one side doped by acceptors and the other side by donors. The longer strips on the acceptor-implanted side (p or U side) measure the r-ϕ coordinate. Similarly, the shorter strips on the donor-implanted side (n or V side) measure the z coordinate.

In the SVD, the readout pitch of the layer-3 sensors is $50\,\mu$m in the r-ϕ plane and $160\,\mu$m along the z axis. On the other hand, the layer-4, -5, and -6 sensors have a readout pitch of $75\,\mu$m in the r-ϕ plane and $240\,\mu$m along the z axis. In order to improve the spatial resolution, a floating strip is introduced between two readout strips on both the P- and N-side. This floating strip effectively redistributes the induced charge, sharing it between the adjacent strips. Consequently, this arrangement reduces the effective strip pitch to half of the readout pitch, contributing to an improved spatial resolution. A detailed description of the SVD can be found in Ref. [73]. Figure 12.1 shows a photo of the Phase-2 VXD (PXD+SVD) layout taken before installation inside the Belle II spectrometer.

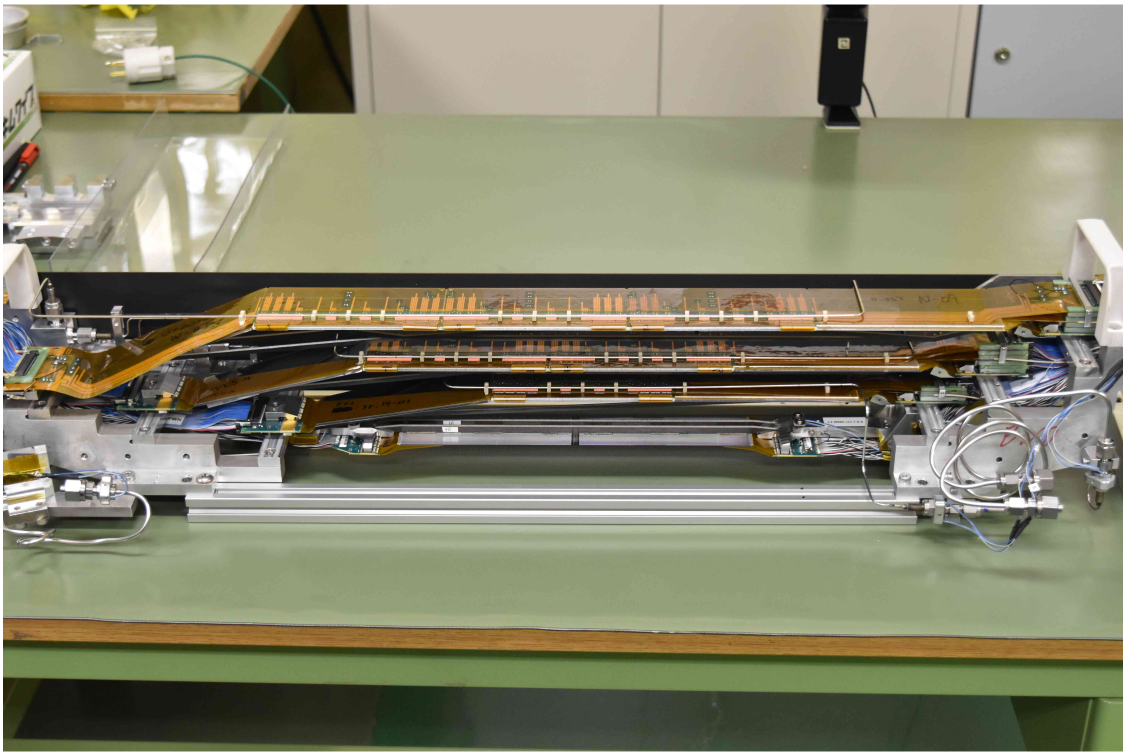


Figure 12.1: Phase-2 VXD subdetector, featuring a single ladder for each layer, placed on a table in preparation for installation. The innermost layers, comprising PXD layers 1 and 2, as well as SVD layer 3, are faintly discernible.

12.3 SVD hit reconstruction and position determination

The SVD readout strips are connected to APV25 chips [74], which are responsible for amplifying, shaping, and digitizing the electronic analogue signals. Signal quality is ensured, by retaining only strips with an signal-to-noise ratio (SNR) value greater than three for analysis. Phase 2 introduces a comprehensive signal characterization, collecting six samples from the digitized signals. The charge calculation follows, with each strip's charge accurately determined in electron units through the gain information obtained from calibration runs.

The cluster identification becomes crucial because when a charged particle traverses the sensor material, it activates one or more strips on either side of the SVD. Clusters are defined as serial strips, all having an SNR value greater than three, and at least one of them surpassing an SNR value of five. In cases where the signal (charge) of a single strip surpasses the threshold, but none of its adjacent strips exhibit signals above the threshold, the available information about the particle's position is limited to the strip's own position. Within this so called single-strip cluster scenario, the charge of the strip does not yield precise information regarding the particle's location; it is essentially equivalent to a binary readout, indicating whether the signal from the strip exceeds the threshold or not. The measured position of the cluster can be expressed as

$$\tilde{x}_{CL1} = x_i = i \cdot p \tag{12.1}$$

where x_i is the position of the i^{th} strip and p is the pitch of the sensor side. In Fig.

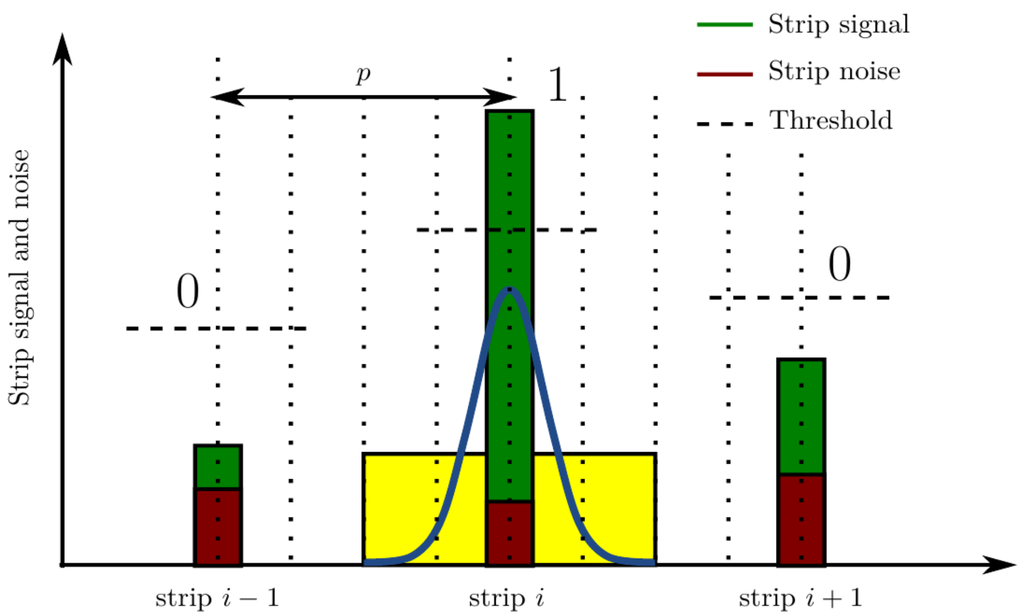


Figure 12.2: A cluster featuring a single strip above the threshold. The actual positions of the particles that generated this signal configuration exhibit a uniform distribution (depicted in yellow). Adapted from Ref. [4].

In this context, the true position of the particle could potentially exist anywhere within the range from half a pitch to the left of the strip to half a pitch to the right, relative to the strip's position. In other words, the real position follows a uniform distribution spanning from half a pitch to the left to half a pitch to the right concerning the strip. The relationship between the measured position ($\tilde{x}$) and the actual position (x) can be expressed as a probability distribution function $f(x)$:

$$f(x) = \begin{cases} 1/p \text{ for } \tilde{x} - p/2 \leq x \leq \tilde{x} + p/2, \\ 0 \text{ othewise.} \end{cases} \tag{12.2}$$

The resolution of a binary one-strip cluster in such cases, where the measured position follows a uniform distribution, is one standard deviation of the distribution of x or $p/\sqrt{12}$, often referred to as the digital resolution.

$$\sigma_1 = \sqrt{\int x^2 f(x)dx - \left(\int xf(x)dx\right)^2} = \sqrt{\frac{1}{p}\int_{\tilde{x}-p/2}^{\tilde{x}+p/2} x^2 dx} = \frac{p}{\sqrt{12}}. \tag{12.3}$$

It is important to note that in real-world scenarios, the position resolution of one-strip clusters is influenced by factors such as capacitive coupling between strips and the SNR of the strips, which eventually improve the resolution of a single-strip cluster with respect to the digital resolution. This deviates from the ideal single-strip cluster resolution and can be expressed as:

$$\Delta x_{CL1} = \frac{p}{1 + S_1/3N_1}. \tag{12.4}$$

Here, S_1 and N_1 represent the signal and noise of the strip, with the factor of 3 signifying the SNR threshold. In the case of a two-strip cluster scenario, the cluster's position is determined utilizing the center-of-gravity (COG) method, as expressed by the equation:

$$\tilde{x}_{CL2} = \frac{x_1 S_1 + x_2 S_2}{S_1 + S_2}. \tag{12.5}$$

Here, S_i and x_i's represent the charge and position of the individual strips that constitute the cluster. Since we utilize the charge of strips to measure the position of the traversed particle, the resolution of two-strip clusters is better than that of the single-strip cluster. The resolution for the two-strip cluster can be expressed as

$$\Delta x_{CL2} = \frac{p}{S/3N_1}. \tag{12.6}$$

where S and N_1 are the total charge of the cluster and noise of the first strip, respectively. In Fig. 12.3, a schematic diagram of the measured particle position for a two-strip cluster is shown.

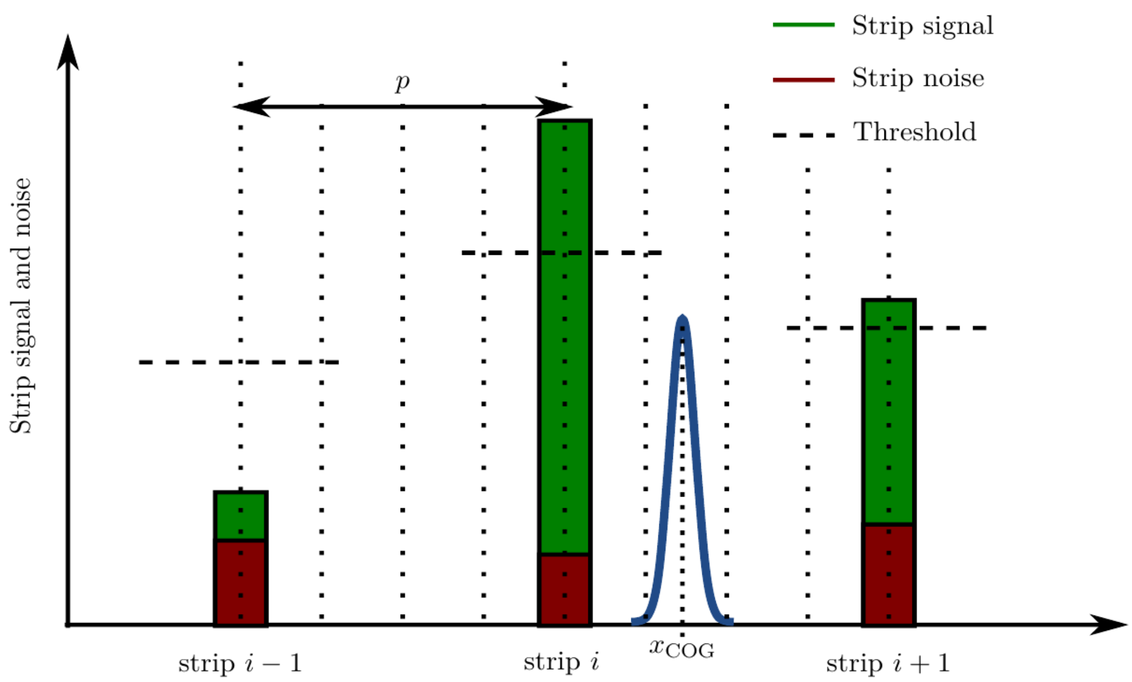


Figure 12.3: A cluster featuring two strips above the threshold. The signal height provides supplementary data that can be utilized to enhance the precision of the determined position within the region between the readout strips. Adapted from Ref. [4].

For clusters more than two-strips, the head-tail algorithm is used. The position and resolution of such clusters is given by

$$\tilde{x}_{CL3} = \frac{1}{2}\left[x_{\text{head}} + x_{\text{tail}} + p\frac{S_{\text{head}} - S_{\text{tail}}}{S_{\text{center}}}\right], \tag{12.7}$$

$$\Delta X_{CL3} = \frac{p}{2}\sqrt{\left(\frac{3N_{CL}}{S_{\text{center}}}\right)^2 + \frac{1}{2}\left(\frac{S_{\text{tail}}}{S_{\text{center}}}\right)^2 + \frac{1}{2}\left(\frac{S_{\text{head}}}{S_{\text{center}}}\right)^2}, \tag{12.8}$$

Here, the subscripts 'head' and 'tail' correspond to the two edge strips within the cluster, and S_{center} represents the average charge per strip within the cluster, excluding the 'head' and 'tail' strips. A comprehensive analysis of the position resolution in various scenarios can be found in Ref. [75]. The 3D hit point reconstruction combines clusters from opposite sides, delivering a holistic representation of SVD hits of particle trajectories.

12.4 Event selection

The Phase-2 running of Belle II, which took place between February and June 2018, collected approximately 500 pb^{-1} of e^+e^- collision data. This sample corresponds to around 6×10^5 $e^+e^- \to \mu^+\mu^-$ and 22×10^6 $e^+e^- \to e^+e^-$ events given the detector acceptance $15° < \theta < 170°$. The reconstruction of charged tracks relies on the hit on PXD, SVD, and CDC subdetectors. To maintain the robustness of our analysis, we specifically concentrate on dimuon events, although for cross-check we have used the Bhabha sample. Muons are favored over electrons in studies of detector performance for the following reason. Unlike electrons, which are more prone to processes like bremsstrahlung that can complicate track measurement and reconstruction, muons maintain more predictable paths through the detector. Their reduced susceptibility to such interactions allows for cleaner signatures and more reliable assessments of tracking and momentum resolution in particle detectors. Dimuon events are chosen for their distinct and unambiguous signatures, as well as their propensity for reduced background contamination in comparison to alternative event types. The track selection criteria mandate that the impact parameters of muon tracks satisfy $d_0 < 2$ cm and $|z_0| < 4$ cm, thereby confirming their association with the IP and excluding the possibility of random cosmic ray events. We apply a criterion that limits the angular separation, measured in the c.m. frame, between the momentum vector of one charged track and the direction opposite to the momentum vector of other track to be less than 10°. To classify a track as a muon, the energy deposit in the ECL must be lower than 0.7 GeV. To effectively mitigate contamination from Bhabha events, we implement a strict criterion on the polar angles of the tracks, restricting them to the barrel region, defined as $35° < \theta < 125°$. This angular confinement is vital because Bhabha events exhibit a higher probability of being found near the beam axis, making this selection crucial in maintaining the purity of the sample. For the measurement of SVD hit efficiency, Bhabha events are also utilized along with dimuon events as a crosscheck. The selection criteria for these events closely resemble those of dimuon events, with slight modifications. Specifically, polar angle selections on the tracks are no more applied, and there are no criteria related to energy deposits in the ECL. Instead, we require the tracks to have a momentum greater than 3 GeV/c for inclusion in this analysis. In order to minimize potential bias, hits from the SVD layer, for which the spatial resolution or hit efficiency is being determined, is excluded from track reconstruction.

In the final selection of valid tracks for analysis, two prerequisites are imposed: each track must comprise at least one PXD and one SVD hit. This sequence of processes, from signal conditioning to track selection, forms the foundation of our SVD hit reconstruction methodology.

12.5 Sensor hit resolution

12.5.1 Definition and measurement strategy

The spatial resolution of an SVD sensor refers to its ability to accurately measure the position of a charged particle. It quantifies how well the sensor can distinguish the location where the particle passed through it. In particle physics experiments, high spatial resolution is crucial to accurately reconstruct particle trajectories and vertices, which in turn helps in identifying the particles.

The spatial resolution of sensor is defined by the uncertainty linked to its measured cluster hit position, which can be quantified through an analysis of the residual distribution. The residual (x_{res}) corresponds to the difference between the measured hit position (x_{hit}) and the position of the track extrapolated to the sensor plane or the intercept position (x_{track}).

$$x_{\text{res}} = x_{\text{hit}} - x_{\text{track}} \tag{12.9}$$

A schematic diagram of intercept and hit position is shown in Fig. 12.4.

For the sake of simplicity and to a first-order approximation, we can assume that the hit and intercept positions are uncorrelated. Consequently, employing the propagation of uncertainty, we find that the uncertainty on the residual, denoted as σ_{res}, is obtained by taking the quadrature sum of uncertainties associated with the hit position (σ_{hit}) and the extrapolated track position (σ_{track}).

$$\begin{aligned} \sigma_{\text{res}}^2 &= \sigma_{\text{hit}}^2 + \sigma_{\text{track}}^2 \\ \sigma_{\text{hit}} &= \sqrt{\sigma_{\text{res}}^2 - \sigma_{\text{track}}^2} \end{aligned} \tag{12.10}$$

So, we have Eq. 12.10 where the spatial resolution is expressed in terms of measurable quantities. The residual width (σ_{res}) is obtained by fitting a sum of two Gaussian functions with a common mean:

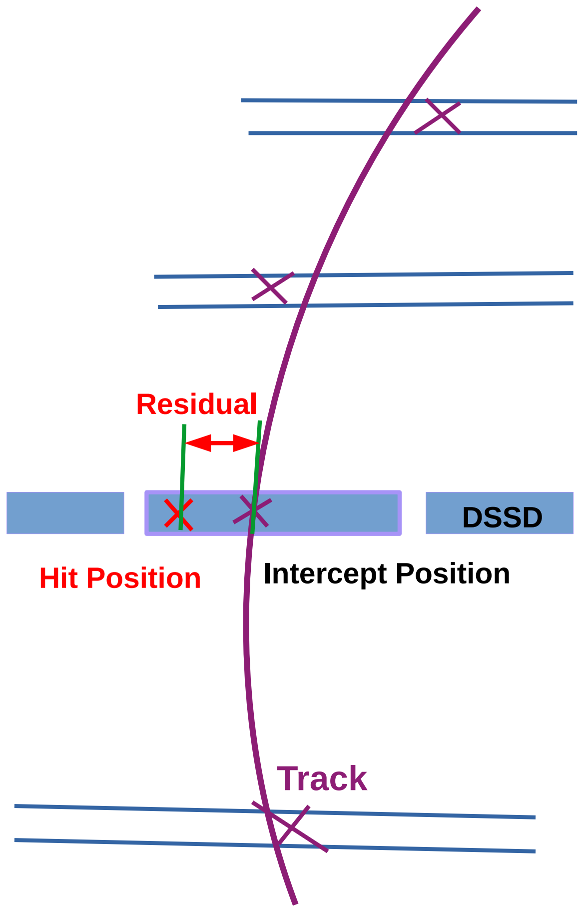


Figure 12.4: Schematic diagram of SVD layers, fitted track, and intercept position or track extrapolated to the sensor plane.

$$f(x) = \frac{s}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu)^2}{2\sigma_1^2}} + \frac{1-s}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu)^2}{2\sigma_2^2}}. \tag{12.11}$$

The effective width is determined as $\sigma_{\text{res}}^2 = s\sigma_1^2 + (1-s)\sigma_2^2$, where s is the fraction of first Gaussian. The uncertainty in track extrapolation (σ_{track}) is calculated using the true position of the intercept. In Fig. 12.5, we show an example of the residual distribution for the central sensor of the 6^{th} layer, for clusters composed of a single strip.

During our examination of the simulated dataset, it has come to our attention that the σ_{track} value obtained through extrapolation is underestimated. Consequently, we've chosen an alter-

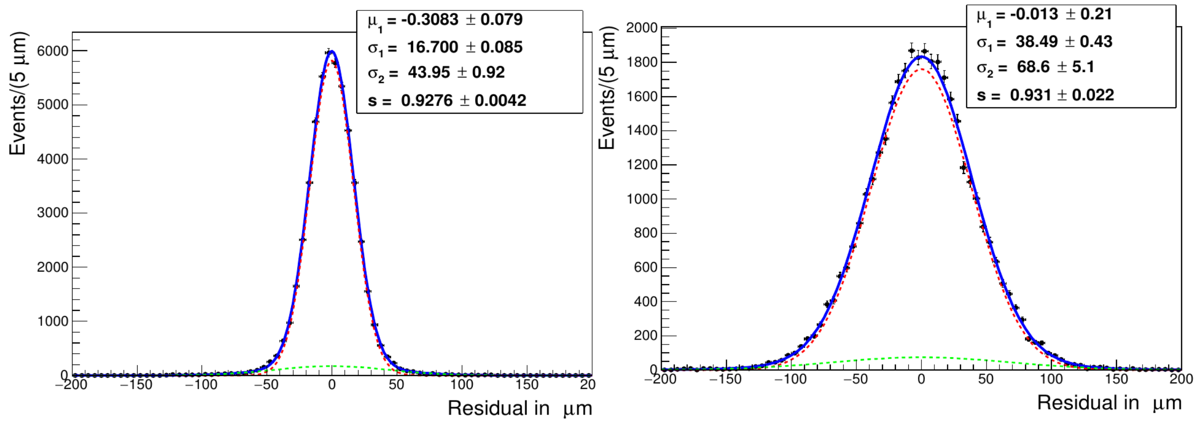


Figure 12.5: Residual distributions observed for the 6th layer, representing the U-side (left) and V-side (right) for a cluster size of one.

native approach by calculating σ_{track} based on the actual position where the track intersects with the sensor. In essence, we employ a fitting procedure on the distribution of true residuals, representing the difference between the intercept and true-hit location. This fitting involves modeling the distribution using a combination of two Gaussian functions with a common mean, and from this distribution, we extract the effective width as our estimate for σ_{track}, a method previously mentioned. The statistical uncertainty on σ_{hit} is

$$\sqrt{\left(\frac{\sigma_{\text{res}}}{\sigma_{\text{hit}}}\delta(\sigma_{\text{res}})\right)^2 + \left(\frac{\sigma_{\text{track}}}{\sigma_{\text{hit}}}\delta(\sigma_{\text{track}})\right)^2}.$$

with $\delta(\sigma_{\text{res}})$ and $\delta(\sigma_{\text{track}})$ being obtained from the fitting. In Fig. 12.6, we show an example of the true residual distribution for the central sensor of the 6th layer, for clusters composed of a single strip.

12.5.2 Result

Table 12.1 lists the estimated spatial resolution for each SVD layer. The digital resolution expression accounts for the presence of one floating strip between two readout strips, resulting in a factor of 2. Single-strip clusters outperform the digital resolution because they utilize adjacent strips with an SNR value below the threshold discussed in Eq. 12.4. Multistrip clusters have a better resolution than single-strip clusters due to the additional charge information. Smaller cluster position residuals on the p side reflect a smaller strip pitch and a better track parameter estimation compared to the n side.

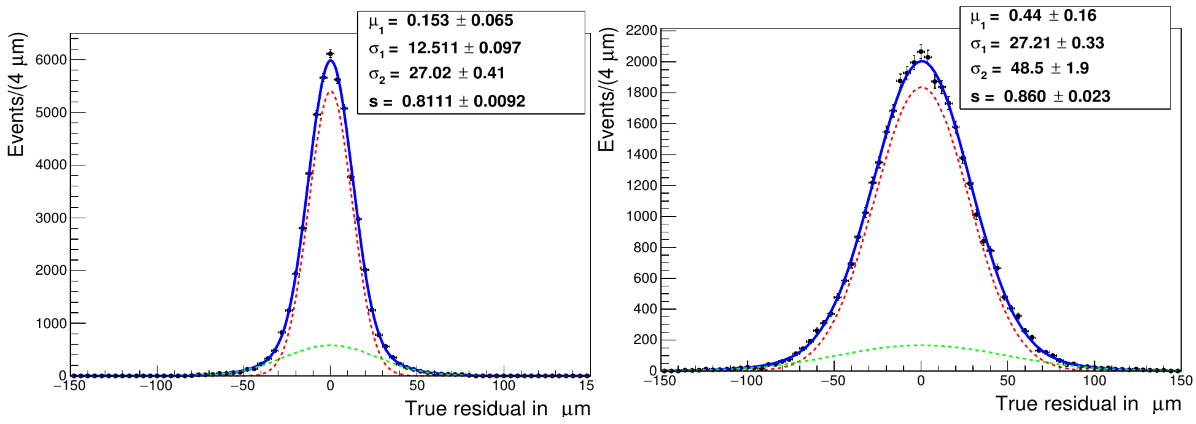


Figure 12.6: True residual distributions observed for the 6$^{\text{th}}$ layer, representing the U-side (left) and V-side (right) for a cluster size of one.

12.6 Sensor hit efficiency

12.6.1 Definition and measurement startegy

The hit efficiency refers to the effectiveness of an SVD sensor in successfully detecting and recording hits or signals from charged particles passing through it. This efficiency quantifies the likelihood that a charged particle, which should leave a measurable signal as it traverses the SVD, is indeed detected and recorded by the sensor. Mathematically, the SVD hit efficiency is expressed as a ratio or percentage, and is calculated as follows:

$$\epsilon = \frac{\text{Number of successfully detected hits}}{\text{Total number of expected hits}} \times 100\%. \tag{12.12}$$

In this formula, “Number of successfully detected hits” represents the actual hits recorded by the SVD, while “Total number of expected hits” is the number of hits that should have been ideally recorded based on the expected particle trajectories.

To measure the hit efficiency, we first exclude hits within the layer of interest from the track fitting process. Subsequently, the track undergoes a refitting procedure. Following this, the track is extrapolated to the sensor plane, and a search is conducted to identify a hit within a specified window spanning $\pm 500\,\mu$m centred around the extrapolated point. The efficiency is calculated by dividing the total number of hits (N_{hit}) within the window by the number of charged tracks (N_{trk}).

Table 12.1: Spatial resolutions for different sensors of layer 3–6 for both p and n sides.

Layer	Side	Pitch (in μm)	Digital resolution $\frac{p}{2\sqrt{12}}$(in μm)	σ_{hit} (one)	σ_{hit} (two)
3	U	50	7.21	5.78 ± 0.12	5.02 ± 0.12
3	V	160	23.09	17.86 ± 0.08	14.66 ± 0.19
4	U	75	10.82	10.0 ± 0.09	6.58 ± 0.12
4	V	240	34.64	26.83 ± 0.17	19.55 ± 0.28
5	U	75	10.82	10.01 ± 0.34	6.53 ± 0.22
5	V	240	34.64	26.84 ± 0.26	18.77 ± 0.43
6	U	75	10.82	11.56 ± 0.51	8.52 ± 0.47
6	V	240	34.64	27.16 ± 1.77	20.44 ± 0.88

Owing to background clusters, the number of hits within the ±500 μm window can be affected by noise. To take care of it, we need to estimate the number of noisy hits (N_{bkg}) within the signal region. This is done by counting N_{bkg} from the sideband [±0.5 mm, ±1.0 mm] and extrapolating it to the signal region, assuming that the number of noisy hits is uniform.

The efficiency is calculated by

$$\epsilon = \frac{N_{\text{hit}} - N_{\text{bkg}}}{N_{\text{trk}}}. \tag{12.13}$$

The efficiency for the second sensor on the U-side in the 4$^{\text{th}}$ and 5$^{\text{th}}$ layers, as well as the second and third sensors in the 6$^{\text{th}}$ layer, is recalculated by excluding tracks that fall within the stripID range of 108 < stripID < 276. These regions correspond to APV chips that were masked for majority runs in Phase-2. The hitmap of fourth layer U-side second sensor is shown in Fig. 12.7.

According to the definition of efficiency, the number of hits for a given number of tracks (n) follows the binomial distribution, where the probability (p) of having one hit for a given track is efficiency. So the uncertainty in the number of hits is $\sqrt{np(1-p)}$ that gives rise to the uncertainty in efficiency as:

$$\Delta\epsilon = \sqrt{\frac{p(1-p)}{n}}. \tag{12.14}$$

12.6.2 Result

Table 12.2 shows hit efficiency results for sensors in the four SVD layers in Phase-2. In certain cases, certain sensors lack sufficient hits for calculation, and as a result, they remain empty. The first sensor in each layer, with the exception of layer-3, has a trapezoidal shape, which results in a non-constant pitch on the U-side of these sensors. Such non-uniform pitch adds complexity to the efficiency calculation. Therefore, during the initial commissioning phase, we choose not to address this particular aspect. The efficiency uncertainty determined from dimuon events exhibits a slight elevation due to limited statistics. Consequently, we make an effort to reevaluate the efficiency using Bhabha events, and observe that the results obtained from both methods largely agree. Most sensors have an efficiency greater than 98%.

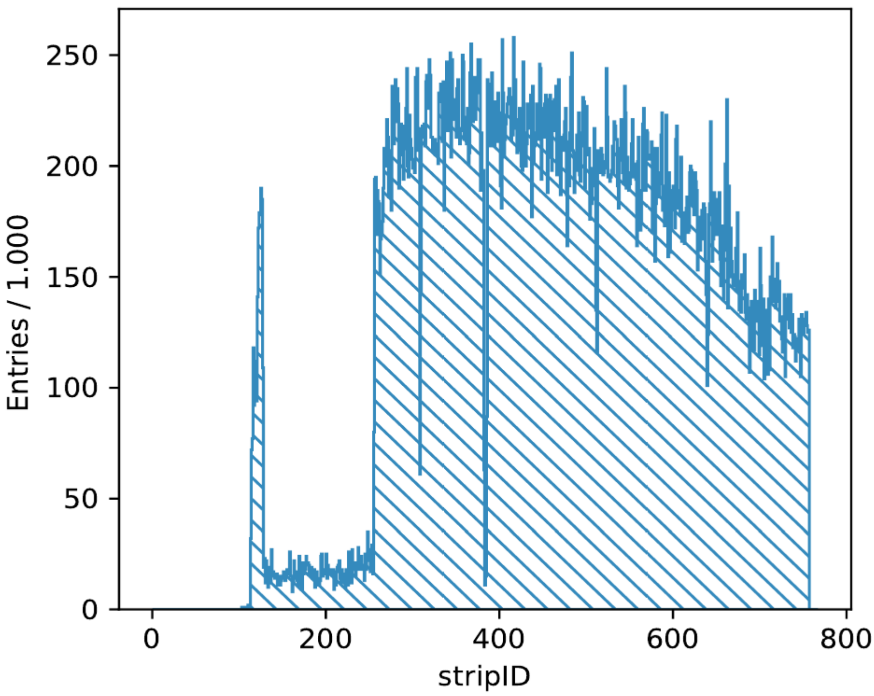


Figure 12.7: Distribution of hitmap for the fourth layer, second U-side sensor. The dip around 200 corresponds to masked APV chips.

12.7 Comparison of residual between data and simulation

This section delves into the comparison of residual shapes between data and simulation for certain sensors while also exploring potential causes for discrepancies. Figure 12.8 compares the shape of residuals for selected sensors between data and simulation. Notably, misalignment or the absence of optimized alignment parameters, particularly in the early stages of the

Table 12.2: Summary of hit efficiency (in %) of the Bhabha and dimuon samples for Phase-2 data.

		Bhabha events		Dimuon events	
Layer No	**Sensor No**	U side	V side	U side	V side
3	1	99.08 ± 0.02	99.44 ± 0.02	98.58 ± 0.29	99.59 ± 0.16
3	2	99.09 ± 0.02	99.09 ± 0.02	98.81 ± 0.13	98.97 ± 0.12
4	1	—	99.31 ± 0.02	—	99.84 ± 0.16
4	2	98.47 ± 0.04	97.32 ± 0.05	98.40 ± 0.17	97.30 ± 0.19
4	3	99.53 ± 0.02	99.51 ± 0.02	99.31 ± 0.69	100.00 ± 0.00
5	1	—	97.81 ± 0.06	—	—
5	2	99.11 ± 0.03	97.90 ± 0.05	98.82 ± 0.22	99.15 ± 0.17
5	3	98.48 ± 0.04	95.69 ± 0.06	98.25 ± 0.20	94.27 ± 0.35
5	4	99.82 ± 0.02	99.50 ± 0.03	100.00 ± 0.00	100.00 ± 0.00
6	1	—	99.42 ± 0.04	—	—
6	2	99.46 ± 0.03	97.37 ± 0.06	99.39 ± 0.25	98.24 ± 0.38
6	3	98.70 ± 0.06	99.23 ± 0.04	98.77 ± 0.21	99.25 ± 0.15
6	4	98.12 ± 0.04	98.71 ± 0.04	98.15 ± 0.32	98.90 ± 0.25
6	5	99.70 ± 0.04	99.17 ± 0.06	—	—

experiment, causes noticeable difference between the data and the simulation.

Tracking detectors play a pivotal role in measuring the positions of charged particles relative to detector elements, with intrinsic resolutions typically in the tens of microns. These measurements are amalgamated into a fit based on precise knowledge of relative detector positions. However, uncertainties arising from construction and assembly procedures can exceed intrinsic resolutions, impeding track-fit precision. Discrepancies between assumed and real detector geometries may introduce biases in measured positions, thereby affecting track parameter extraction. Misalignments within the tracker can manifest as deviations from expected particle trajectories, especially if they occur in planes perpendicular to particle motion, a phenomenon known as planar deformation misalignment. Such misalignments can induce distortions in recorded trajectories, leading to inaccuracies in reconstructing particle properties. This, in turn, results in a bias in residual positions depending on various positions within the same sensor plane.

Getting the detector alignment parameters right is crucial to achieving the tracking performance required for physics goals. This makes sure that local measurements are accurate within a global reference frame. In Fig. 12.9, we observe that the residual mean in simulation remains around zero across the sensors, whereas in data, the same varies depending on the cluster position, indicating possible misalignment. Furthermore, for certain sensors, the extent of misalignment is notably pronounced, often attributable to planar deformation, as illustrated in Fig. 12.10.

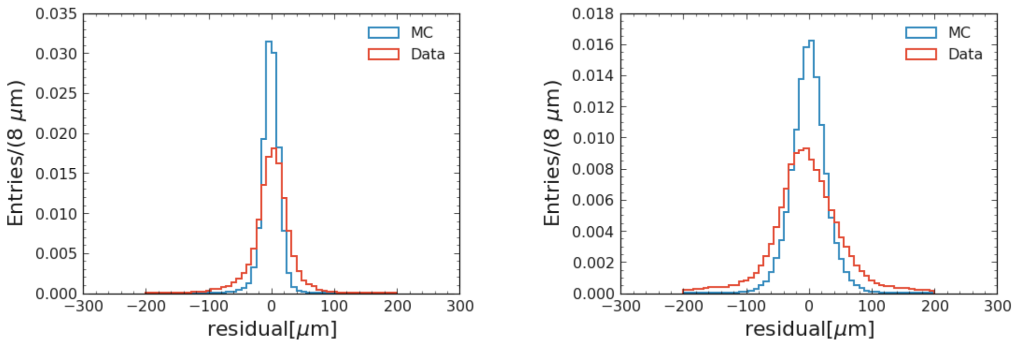


Figure 12.8: Comparison of residual distributions between data and simulation for (left) the fifth layer, third sensor, U-side, and (right) the fifth layer, fifth sensor, V-side.

12.8 Discussion

In summary, the analysis of the hit efficiency and spatial resolution of the Belle II SVD has yielded very good results. The spatial resolution, determined from the residual distribution, provides valuable insights into the precision of hit position measurements. The hit efficiency is found to be high (> 90%), indicating a high probability of detecting hits in the SVD sensors when charged particles traverse through them. We compare the residual distributions between data and simulation for specific sensors, highlighting differences likely due to misalignment. These findings demonstrate the effectiveness and reliability of the Belle II SVD in accurately reconstructing tracks and obtaining precise spatial information.

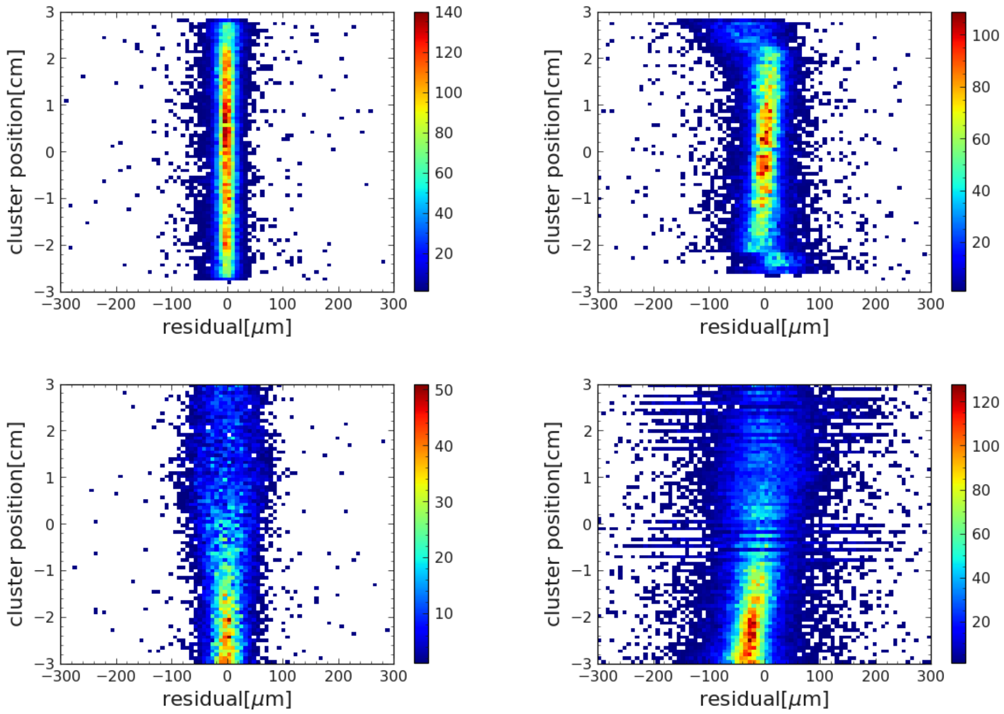


Figure 12.9: Two-dimensional histogram of residual position versus cluster position for (left) simulation and (right) data in the fifth layer, third sensor, showing (top) the U-side and (bottom) the V-side.

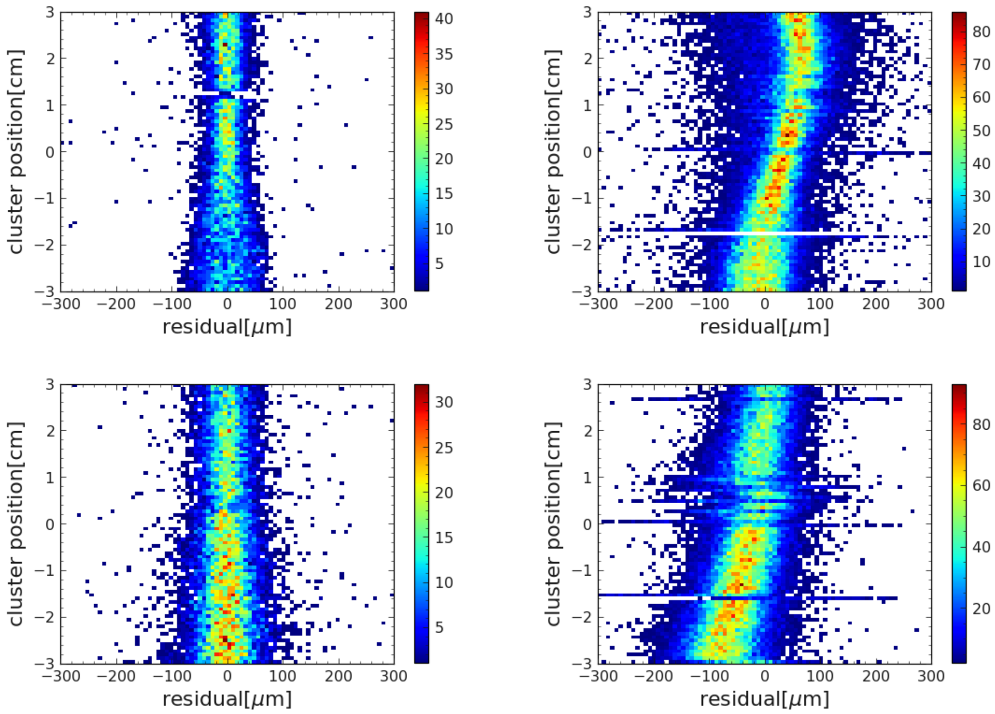


Figure 12.10: Two-dimensional histogram of residual position versus cluster position for (left) simulation and (right) data in the (top) fourth layer, second sensor, and (bottom) sixth layer, fourth sensor, V-side.

Chapter 13

Conclusion

This thesis chronicles a captivating journey into the realm of studying rare decays. I embarked on this exploration in 2019, when Belle II first began its collision data recording. Originally setting our sights on achieving a luminosity of $40\,\mathrm{ab}^{-1}$ by 2024 [72], our trajectory faced unforeseen challenges, with the global COVID-19 pandemic and other technical difficulties to achieve high peak luminosity hindering our progress toward this ambitious goal. Despite the setbacks, the dedication and perseverance of the research team prevailed, and the data collected, while falling short of the initial target, serves as a testament to the resilience required in the pursuit of scientific exploration.

The essence of venturing into the realm of BSM physics has been expounded in Chapter 2 of the thesis, laying the foundation for our scientific pursuit. Reflecting this spirit, the intricacies of the analysis methodology have been meticulously detailed across Chapters 4 to 7. Chapter 8 illuminates our investigation into the branching fraction and LFU ratio $R_K(J/\psi)$, using the SM candle decay $B \to J/\psi K$ with $189\,\mathrm{fb}^{-1}$ of data. The consistency with SM expectations not only validates the analysis methodology but also underscores the quality and calibration of the data. Moving forward to Chapter 9, a preliminary measurement of $B^+ \to K^+\ell^+\ell^-$ is presented, leveraging 63 fb^{-1} of recorded data. Extending beyond $B \to K\ell^+\ell^-$, Chapter 11 delves into a signal search for the $b \to s\ell^+\ell^-$ decay $B \to K^*\ell^+\ell^-$—key channel for LFU measurement. Despite the absence of journal a publication on LFU in $b \to s\ell^+\ell^-$ with the current Belle II dataset, this thesis encapsulates the journey toward that goal. As the great philosopher Lao Tzu once wisely remarked, 'The journey of a thousand miles begins with a single step.' In traversing the uncharted territories of particle physics, this thesis stands as a testament to the pursuit of understanding, acknowledging that the road to LFU measurement is paved with the intricate beauty of scientific inquiry.

The essence of this study lies not just in the expected final measurement of LFU but also in the rigorous analysis undertaken, a process that has undergone meticulous review, including the completion of the stringent working-group phase. Currently, a committee comprising three members is actively engaged in the careful examination of our analysis. Upon their approval, we will examine the signal region of the 362 fb^{-1} data, unlocking the potential for groundbreaking insights. The progress made in this section underscores a careful and meticulous approach taken in preparing for the final measurement of R_K, shedding light on the challenges faced and the unwavering commitment to advancing our comprehension of the fundamental forces that govern the universe.

In a broader picture, as we navigate the frontiers of particle physics and venture into the realm beyond the SM, we evoke the spirit of pioneers like Albert Einstein, who once remarked, 'The most beautiful experience we can have is the mysterious.' In measuring LFU, we stand at the crossroads of discovery, reminiscent of the times when luminaries such as Niels Bohr and Werner Heisenberg challenged the conventional boundaries of understanding. Our pursuit echoes their sentiment that 'Every valuable human being must be a radical and a rebel, for what he must aim at is to make things better than they are.' In this journey to measure physics beyond the established norms, we honor the tradition of those who pushed against the constraints of their era, recognizing that, as in the past, the most profound revelations often lie beyond the familiar equations and expected symmetries. Through the lens of LFU, we extend our reach into the unknown, acknowledging that the quest to measure the unmeasured is the essence of advancing our comprehension of the fundamental forces that shape the universe.

Appendix A

B counting

At the $\Upsilon(4S)$ resonance ($\sqrt{s} = 10.58\,\text{GeV}$), the hadronic events come from both the $B\overline{B}$ and continuum events. Thus, the determination of hadronic events stemming from $B\overline{B}$ events can be expressed as follows:

$$N_{B\overline{B}} = N_{H}^{\text{on-res}} - N_{H(q\overline{q})}^{\text{on-res}} \tag{A.1}$$

Here, $N_{B\overline{B}}$ represents the number of $B\overline{B}$ events, $N_{H}^{\text{on-res}}$ denotes the overall number of hadronic events, and $N_{H(q\overline{q})}^{\text{on-res}}$ signifies the number of hadronic events originating from $q\overline{q}$ processes in the on-resonance dataset.

To determine $N_{H(q\overline{q})}^{\text{on-res}}$, we utilize an off-resonance data collected at a c.m. energy of $\sqrt{s} = 10.52\,\text{GeV}$, which is only 60 MeV lower than the energy at which the on-resonance data are acquired. Concerning the continuum events, encompassing both leptonic and hadronic processes around the $\Upsilon(4S)$ resonance, the cross-sections exhibit continuity, as evident from Fig. 5.6. The leptonic events, such as $e^+e^- \to \mu^+\mu^-$, display a continuous spectrum, while the hadronic peak primarily arises from $B\overline{B}$ events. Excluding the latter, we refer to the remaining events as 'light quark $q\overline{q}$ events', which also exhibit a continuous distribution.

Due to the continuous nature, we can confidently assume that the ratio of hadronic on-resonance to off-resonance events originating from continuum is equal to the ratio of leptonic on-resonance to off-resonance events. This relationship can be expressed as follows:

$$\frac{N_{H(q\bar{q})}^{\text{on-res}}}{N_{H(q\bar{q})}^{\text{off-res}}} = \frac{N_{\ell^+\ell^-}^{\text{on-res}}}{N_{\ell^+\ell^-}^{\text{off-res}}} \tag{A.2}$$

$$\Rightarrow N_{H(q\bar{q})}^{\text{on-res}} = N_{H(q\bar{q})}^{\text{off-res}} \frac{N_{\ell^+\ell^-}^{\text{on-res}}}{N_{\ell^+\ell^-}^{\text{off-res}}} \tag{A.3}$$

$$\tag{A.4}$$

The quantity $\frac{N_{\ell^+\ell^-}^{\text{on-res}}}{N_{\ell^+\ell^-}^{\text{off-res}}}$ is determined based on the integrated luminosity ratio $\frac{\mathcal{L}_{\text{on-res}}}{\mathcal{L}_{\text{off-res}}}$ between highly abundant events, specifically $e^+e^- \to \gamma\gamma$ (4.99 nb), and $e^+e^- \to e^+e^-$ (300 nb).

So finally Eq. A.1 becomes

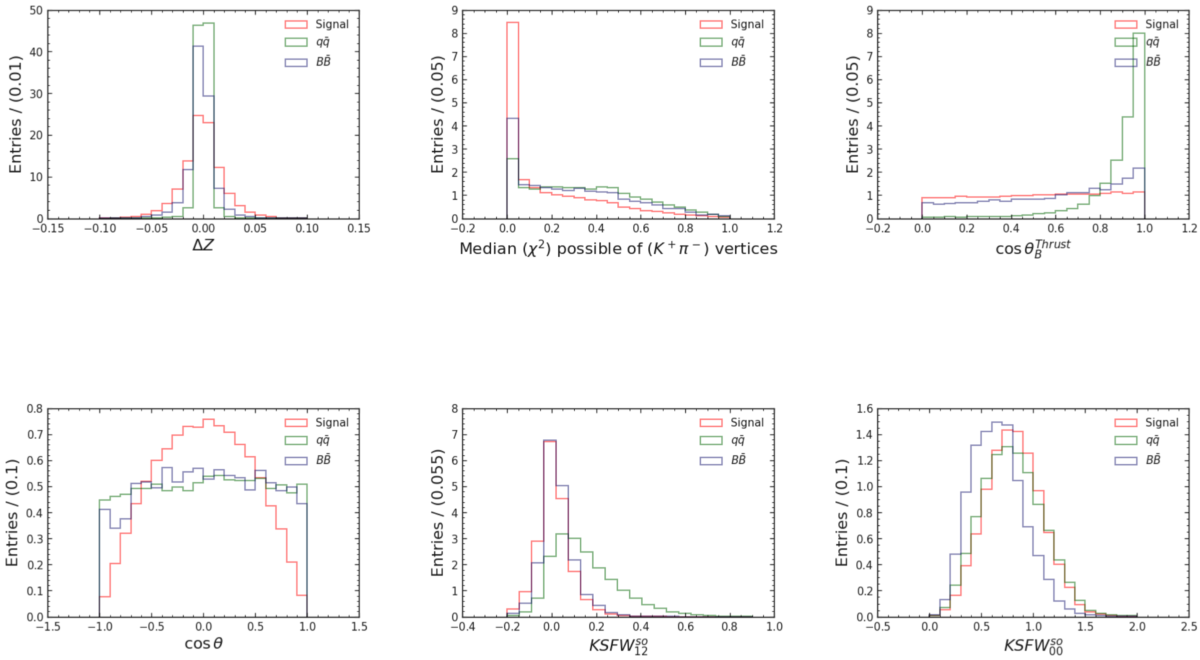
$$N_{B\bar{B}} = \frac{N_{H}^{\text{on-res}} - \mathcal{K} \cdot N_{H(q\bar{q})}^{\text{off-res}} \cdot \frac{\mathcal{L}_{\text{on-res}}}{\mathcal{L}_{\text{off-res}}}}{\epsilon_{B\bar{B}}} \tag{A.5}$$

$\epsilon_{B\bar{B}}$ represents the selection efficiency for the hadronic selection applied to $B\overline{B}$ events, while $\mathcal{K}$ factors in the variations in non-$B\overline{B}$ efficiencies and cross-sections as a function of beam energy. For more details, refer to Ref. [76].

Appendix B

Additional information of MVA classifiers

The distributions of feature in the simulation sample are shown in this appendix.



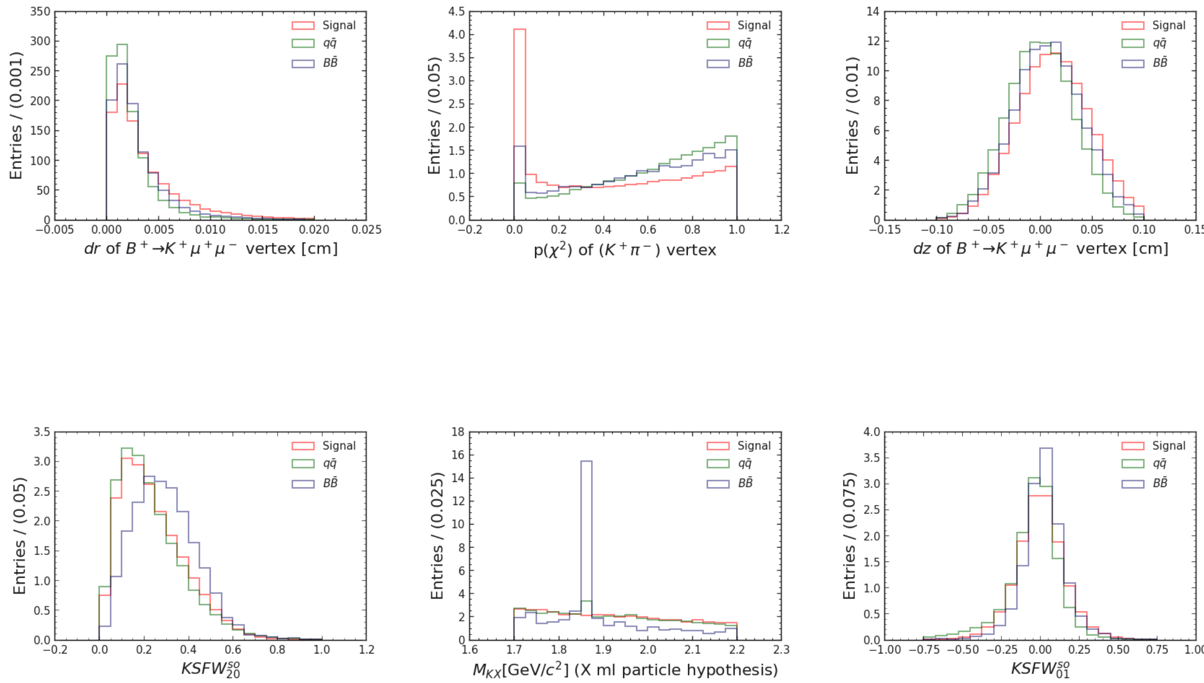
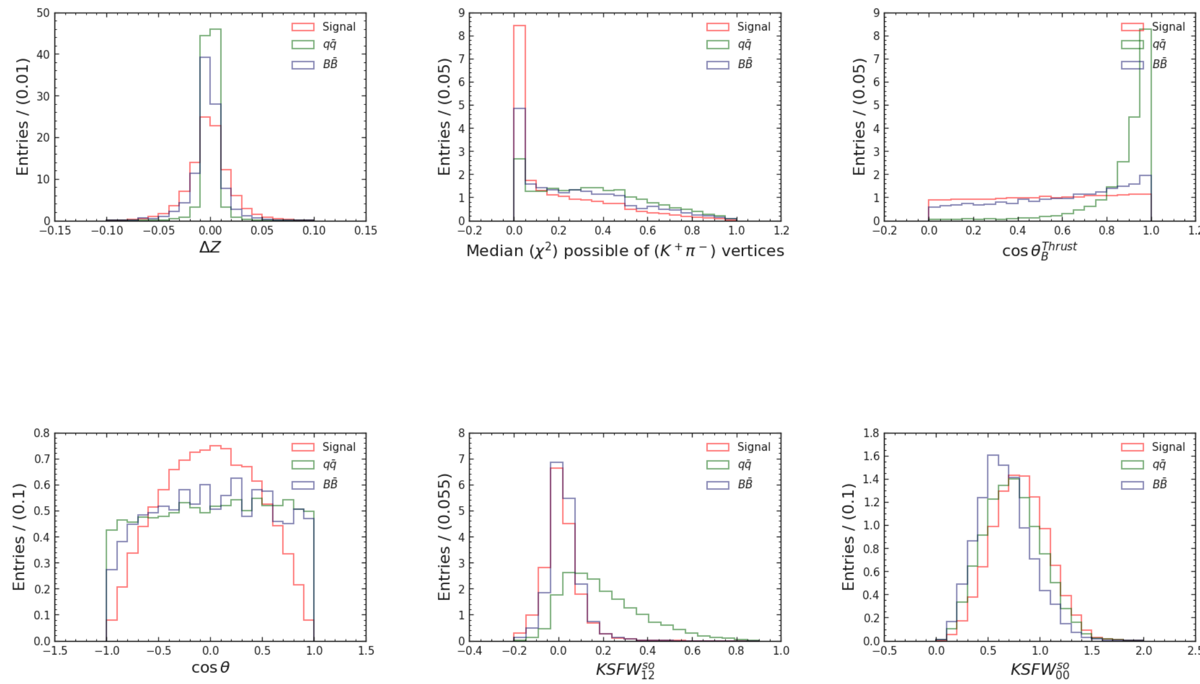


Figure B.1: MC distributions of some highly ranked feature variables for $B^+ \to K^+\mu^+\mu^-$. Red, green, and blue histograms correspond to signal, continuum, and generic B background, respectively.



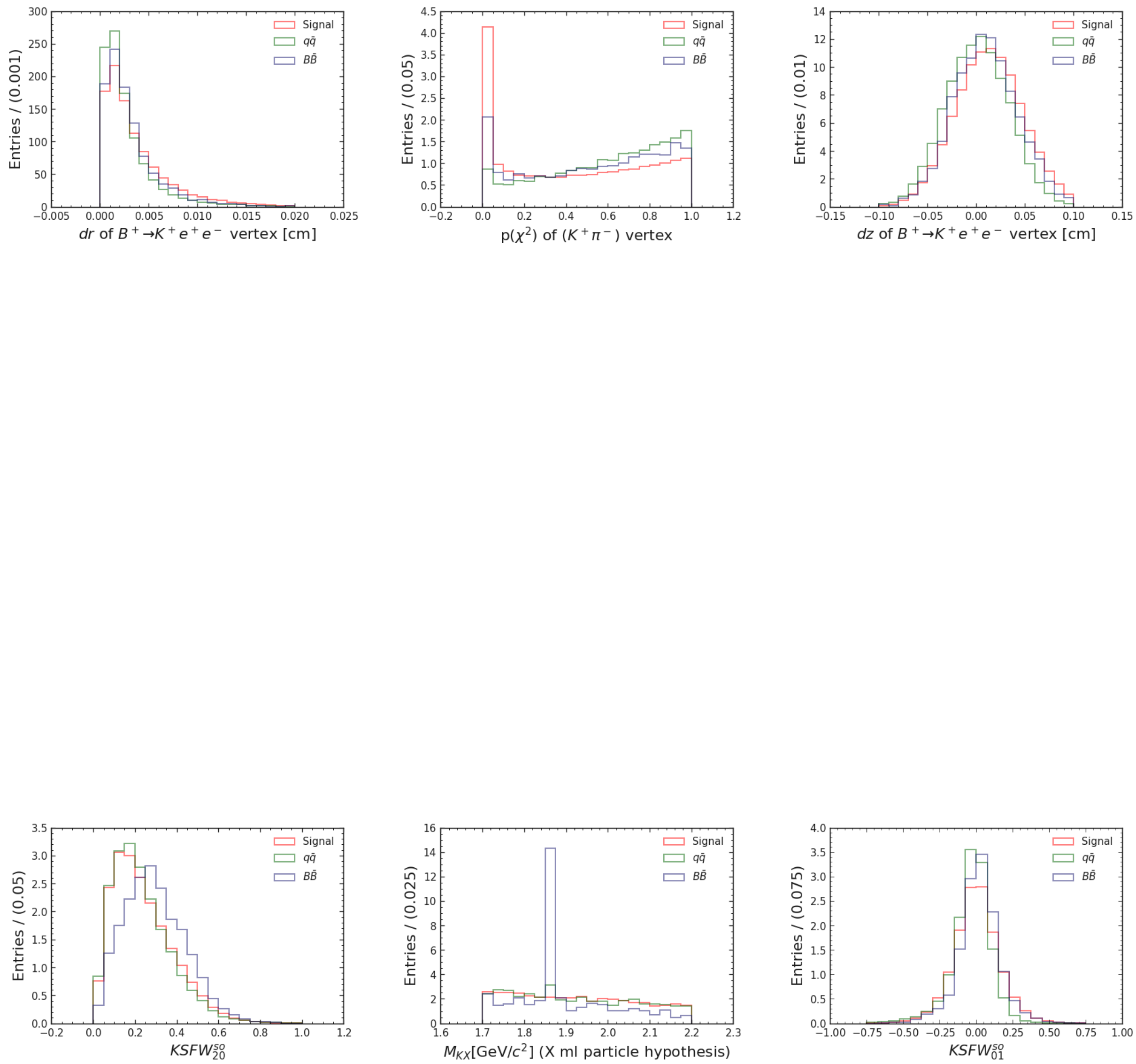


Figure B.2: MC distributions of some highly ranked feature variables for $B^+ \to K^+e^+e^-$. Red, green, and blue histograms correspond to signal, continuum, and generic B background, respectively.

Bibliography

[1] T. Blake, T. Gershon and G. Hiller, *Rare b hadron decays at the LHC*, *Ann. Rev. Nucl. Part. Sci.* **65** (2015) 113 [1501.03309].

[2] SuperKEKB collaboration, *SuperKEKB Collider*, *Nucl. Instrum. Meth. A* **907** (2018) 188 [1809.01958].

[3] C. Praz, *Search for $B \to K\nu\bar{\nu}$ decays with a machine learning method at the Belle II experiment*, Ph.D. thesis, Universität Hamburg, Hamburg, Germany, 2022.

[4] M. Valentan, *The Silicon Vertex Detector for b-tagging at Belle II*, Ph.D. thesis, Vienna, Tech. U., Atominst., 2013.

[5] Belle-II collaboration, *Belle II Technical Design Report*, 1011.0352.

[6] W. Demtröder, *Physics of elementary particles*, in *Nuclear and Particle Physics*, (Cham), pp. 173–212, Springer International Publishing (2022), DOI.

[7] D. Weyland, *Continuum Suppression with Deep Learning techniques for the Belle II Experiment*, 2017, https://api.semanticscholar.org/CorpusID:126015991.

[8] L. Lista, *Statistical Methods for Data Analysis in Particle Physics*, Spinger, second edition ed. (2017, Volume 941).

[9] P.A. Zyla et al., *Review of Particle Physics (Particle Data Group)*, *PTEP* **2020** (2020) 083C01.

[10] V. Brdar, A.J. Helmboldt, S. Iwamoto and K. Schmitz, *Type i seesaw mechanism as the common origin of neutrino mass, baryon asymmetry, and the electroweak scale*, *Phys. Rev. D* **100** (2019) 075029.

[11] T. Inami and C.S. Lim, *Effects of Superheavy Quarks and Leptons in Low-Energy Weak

Processes $K_L \to \mu\bar{\mu}$, $K^+ \to \pi^+\nu\bar{\nu}$ and $K^0 \leftrightarrow \overline{K^0}$, *Progress of Theoretical Physics* **65** (1981) 297 [https://academic.oup.com/ptp/article-pdf/65/1/297/5252099/65-1-297.pdf].

[12] W. Altmannshofer, P. Stangl and D.M. Straub, *Interpreting hints for lepton flavor universality violation*, *Phys. Rev. D* **96** (2017) 055008.

[13] M. Bordone, G. Isidori and A. Pattori, *On the Standard Model predictions for R_K and R_{K^*}*, *Eur. Phys. J. C* **76** (2016) 440 [1605.07633].

[14] T. Becher, A. Broggio and A. Ferroglia, *Introduction to Soft-Collinear Effective Theory*, vol. 896, Springer (2015), 10.1007/978-3-319-14848-9, [1410.1892].

[15] F. Zwicky, *Spectral displacement of extra-galactic nebulae*, *Helv. Phys. Acta* **6** (1933) 110.

[16] M.B. Gavela, P. Hernandez, J. Orloff and O. Pene, *Standard model CP violation and baryon asymmetry*, *Mod. Phys. Lett. A* **9** (1994) 795 [hep-ph/9312215].

[17] Super-Kamiokande collaboration, *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562 [hep-ex/9807003].

[18] KamLAND collaboration, *First results from KamLAND: Evidence for reactor anti-neutrino disappearance*, *Phys. Rev. Lett.* **90** (2003) 021802 [hep-ex/0212021].

[19] J.L. Feng, *Naturalness and the Status of Supersymmetry*, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 351 [1302.6587].

[20] D. Bečirević, O. Sumensari and R. Zukanovich Funchal, *Lepton flavor violation in exclusive $b \to s$ decays*, *Eur. Phys. J. C* **76** (2016) 134 [1602.00881].

[21] A.J. Buras, *Weak Hamiltonian, CP violation and rare decays*, in *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions*, pp. 281–539, 6, 1998 [hep-ph/9806471].

[22] C. Bobeth, M. Misiak and J. Urban, *Photonic penguins at two loops and m_t dependence of $BR[B \to X_s l^+ l^-]$*, *Nucl. Phys. B* **574** (2000) 291 [hep-ph/9910220].

[23] M. Della Morte, J. Heitger, H. Simma and R. Sommer, *Non-perturbative heavy quark*

effective theory: An application to semi-leptonic b-decays, Nuclear and Particle Physics Proceedings **261-262** (2015) 368.

[24] F. Beaujean, C. Bobeth and S. Jahn, *Constraints on tensor and scalar couplings from $B \to K\bar{\mu}\mu$ and $B_s \to \bar{\mu}\mu$*, *Eur. Phys. J. C* **75** (2015) 456 [1508.01526].

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

[26] LHCb collaboration, *Test of lepton universality in beauty-quark decays*, *Nature Phys.* **18** (2022) 277 [2103.11769].

[27] LHCb collaboration, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, *JHEP* **08** (2017) 055 [1705.05802].

[28] I. Doršner, S. Fajfer, A. Greljo, J.F. Kamenik and N. Košnik, *Physics of leptoquarks in precision experiments and at particle colliders*, *Phys. Rept.* **641** (2016) 1 [1603.04993].

[29] G. Hiller and M. Schmaltz, *R_K and future $b \to s\ell\ell$ physics beyond the standard model opportunities*, *Phys. Rev. D* **90** (2014) 054014 [1408.1627].

[30] L. Di Luzio, A. Greljo and M. Nardecchia, *Gauge leptoquark as the origin of B-physics anomalies*, *Phys. Rev. D* **96** (2017) 115011 [1708.08450].

[31] B. Bhattacharya, A. Datta, J.-P. Guévin, D. London and R. Watanabe, *Simultaneous Explanation of the R_K and $R_{D^{(*)}}$ Puzzles: a Model Analysis*, *JHEP* **01** (2017) 015 [1609.09078].

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

[33] 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 [1501.00993].

[34] A. Celis, J. Fuentes-Martin, M. Jung and H. Serodio, *Family nonuniversal Z' models with protected flavor-changing interactions*, *Phys. Rev. D* **92** (2015) 015007 [1505.03079].

[35] A. Falkowski, M. Nardecchia and R. Ziegler, *Lepton Flavor Non-Universality in B-meson Decays from a U(2) Flavor Model*, *JHEP* **11** (2015) 173 [1509.01249].

[36] S.F. King, *Flavourful Z models for $R_{K^{(*)}}$*, *JHEP* **08** (2017) 019 [1706.06100].

[37] LHCb collaboration, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **131** (2023) 051803 [2212.09152].

[38] BESIII collaboration, *Precision measurements of $B[\psi(3686) \to \pi^+\pi^- J/\psi]$ and $B[J/\psi \to l^+l^-]$*, *Phys. Rev. D* **88** (2013) 032007.

[39] BELLE collaboration, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, *JHEP* **03** (2021) 105 [1908.01848].

[40] BABAR collaboration, *Measurement of branching fractions and rate asymmetries in the rare decays $B \to K^{(*)}\ell^+\ell^-$*, *Phys. Rev. D* **86** (2012) 032012.

[41] BABAR collaboration, *Measurement of the branching fraction of $\Upsilon(4S) \to B^0\overline{B}^0$*, *Phys. Rev. Lett.* **95** (2005) 042001.

[42] Belle-II collaboration, *Measurements of the branching fraction, isospin asymmetry, and lepton-universality ratio in $B \to J/\psi K$ decays at Belle II*, 2207.11275.

[43] K. Akai, K. Furukawa and H. Koiso, *Superkekb collider*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **907** (2018) 188.

[44] Heavy Flavor Averaging Group collaboration, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2021*, *Phys. Rev. D* **107** (2023) 052008.

[45] M.A. Furman, *Hourglass effects for asymmetric colliders*, *Conf. Proc. C* **910506** (1991) 422.

[46] J. Kemmer and G. Lutz, *New detector concepts*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **253** (1987) 365.

[47] Belle collaboration, *A detailed test of the CsI(Tl) calorimeter for BELLE with photon beams of energy between 20-MeV and 5.4-GeV*, *Nucl. Instrum. Meth. A* **441** (2000) 401.

[48] K. Miyabayashi, *Belle II electromagnetic calorimeter and its performance during early SuperKEKB operation*, *JINST* **15** (2020) C10016.

[49] D.J. Lange, *The evtgen particle decay simulation package*, *Nuclear Instruments and*

Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152.

[50] E. Barberio, B. van Eijk and Z. Was, *Photos — a universal monte carlo for qed radiative corrections in decays*, *Computer Physics Communications* **66** (1991) 115.

[51] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce et al., *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250.

[52] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth and N. Braun, *The belle ii core software*, *Computing and Software for Big Science* **3** (2018) 1.

[53] S.A. Yost, S. Jadach, B. Ward and Z. Was, *KKMC-hh: a precision event generator for electroweak radiative corrections in hadron scattering*, *PoS* **RADCOR2017** (2018) 023.

[54] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to pythia 8.1*, *Computer Physics Communications* **178** (2008) 852.

[55] R. Frühwirth, *Application of kalman filtering to track and vertex fitting*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **262** (1987) 444.

[56] Marco Milesi, Justin Tan and Phillip Urquijo, *Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers*, *EPJ Web Conf.* **245** (2020) 06023.

[57] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, *Comput. Softw. Big Sci.* **1** (2017) 2.

[58] G.C. Fox and S. Wolfram, *Observables for the analysis of event shapes in e^+e^- annihilation and other processes*, *Phys. Rev. Lett.* **41** (1978) 1581.

[59] W.G.D. Dharmaratna and G.R. Goldstein, *Single quark polarization in quantum chromodynamics subprocesses*, *Phys. Rev. D* **53** (1996) 1073.

[60] J.E. Angus, *The probability integral transform and related results*, *SIAM Review* **36** (1994) 652 [`https://doi.org/10.1137/1036146`].

[61] BABAR collaboration, *Study of $B \to X\gamma$ decays and determination of $|V_{td}/V_{ts}|$*, *Phys. Rev. D* **82** (2010) 051101.

[62] H. Albrecht, T. Hamacher, R. Hofmann, T. Kirchhoff, R. Mankel, A. Nau et al., *Measurement of the polarization in the decay $B \to J/\psi K^*$*, *Physics Letters B* **340** (1994) 217.

[63] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, *The Annals of Statistics* **7** (1979) 1 .

[64] Belle collaboration, *Measurement of the B^+/B^0 production ratio in e^+e^- collisions at the $\Upsilon(4S)$ resonance using $B \to J/\psi(\ell\ell)K$ decays at belle*, *Phys. Rev. D* **107** (2023) L031102.

[65] R. Wanke, *How to deal with systematic uncertainties*, in *Data Analysis in High Energy Physics*, pp. 263–296, John Wiley Sons, Ltd (2013), DOI [`https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527653416.ch8`].

[66] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited*, *Phys. Rev. D* **71** (2005) 014029 [`hep-ph/0412079`].

[67] C. Bobeth, G. Hiller and G. Piranishvili, *Angular distributions of $\bar{B} \to \bar{K}\ell^+\ell^-$ decays*, *JHEP* **12** (2007) 040 [`0709.4174`].

[68] HPQCD collaboration, *$b \to k$ and $d \to k$ form factors from fully relativistic lattice qcd*, *Phys. Rev. D* **107** (2023) 014510.

[69] LHCb collaboration, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, *JHEP* **08** (2017) 055 [`1705.05802`].

[70] LHCb collaboration, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **131** (2023) 051803 [`2212.09152`].

[71] Z. Liptak, A. Paladino, L. Santelj, J. Schueler, S. Stefkova, H. Tanigawa et al., *Measurements of beam backgrounds in superkekb phase 2*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1040** (2022) 167168.

[72] BELLE-II collaboration, *The Belle II Physics Book*, *PTEP* **2019** (2019) 123C01 [1808.10567].

[73] BELLE-II SVD collaboration, *The design, construction, operation and performance of the Belle II silicon vertex detector*, *Journal of Instrumentation* **17** (2022) P11042.

[74] M. French, L. Jones, Q. Morrissey, A. Neviani, R. Turchetta, J. Fulcher et al., *Design and results from the APV25, a deep sub-micron cmos front-end chip for the CMS tracker*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **466** (2001) 359.

[75] R. Turchetta, *Spatial resolution of silicon microstrip detectors*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **335** (1993) 44.

[76] F. Abudinén, I. Adachi, P. Ahlburg, H. Aihara, N. Akopov, A. Aloisio et al., *Measurement of the integrated luminosity of the phase 2 data of the belle ii experiment **, *Chinese Physics C* **44** (2020) 021001.