Search for Lepton Flavour Violating $B^+ \to K^+\tau^\pm\ell^\mp$ Decays at Belle and Belle II

A thesis submitted for the degree of

Doctor of Philosophy

by

Junaid Ur Rehman

under the supervision of

Prof. Andrzej Bożek

and

Dr. Jarosław Wiechczyński

Division of Particle Physics and Astrophysics,
The Henryk Niewodniczański Institute of Nuclear Physics
Polish Academy of Sciences,
31-342 Kraków, Poland

Dedicated to

my late father, who nurtured in me curiosity,
a bibliophilic spirit, and a nonconformist mind.

Declaration

I declare that the thesis is original except for the source materials explicitly acknowledged and that, as a whole or in part, of has not been previously submitted for the degree.

Acknowledgement

My heartfelt gratitude goes to Prof. Andrzej Bożek for accepting me as a Ph.D. candidate and for his continuous support during this long journey. I really appreciate the working environment that he provided during the studies and his all-time availability for the discussions and feedback. His consistent and kind mentorship played a crucial role in completing this work. I would also like to express my sincere gratitude to Emeritus Prof. Maria Różańska for her expert opinions on my work and all the feedback during these four years. I have learned a lot from her, and she played a phenomenal role throughout the analysis.

A special thanks to Dr. Jarosław Wiechczyński for introducing me to the technicalities of the Basf2 and other important aspects of experimental high-energy physics. His support during this study is remarkable. I would like to thank Dr. Karol Adamczyk for his guidance and for introducing me to the ML techniques used in this study. I also extend my thanks to Dr. Devender Kumar for his vital support and comments during the later stages of this study. Many thanks to Mateusz and Oliwia for the time spent together and for the interesting conversations.

For my work in the Belle II SVD group, I would like to say thanks to the INFN Pisa, Belle II group, particularly to Dr. Luigi Corona, Prof. Giuliana Rizzo, and Prof. Giulia Casarosa. They provided tremendous support in accomplishing the tasks assigned to me in the Belle II SVD group.

I also would like to acknowledge the crucial administrative and financial support provided by KISD during these years. Moreover, I would like to say thanks to Prof. Shohei Nishida for hosting me at KEK Tsukuba, Japan, for two months in summer 2024 and to Prof. Emi Kou for hosting me at IJCLab Paris, France, for one month in 2025. I am grateful to NAWA for funding these trips under the NAWA-STER program.

Many thanks to the friends in Krakow (Dr. Yasir, Sultan, Waseem, Hafeez, Fasih, Surajit, Ahsan, Bilal, Sara, Hassan, and many others) for making this journey enjoyable and memorable. I am extremely grateful for all the moments and memories that I have shared with you. A huge thanks goes to Zeeshan, Dr. Kinga, Dr. Rehman, and Laraib for their constant moral and emotional support during this tedious journey.

Last but not least, I would like to express my sincere gratitude and thanks to my family for providing me with this time to concentrate on my studies and supporting me during the difficult times. I am forever grateful for all this unconditional support.

To all those whose names I cannot mention here, I sincerely apologize; please know that every person who supported me throughout this journey holds a special place in my heart. I would like to again express my sincere thanks to all the individuals involved in this long journey for helping me make it possible.

List of Conference Talks

The analysis reported in this thesis is presented at the following main conferences/meetings.

- XXXII Cracow Epiphany Conference on the recent results from Heavy Ion Physics, $12 - 16$ January 2026 Cracow.

  Talk title: "Towards the Search for Lepton Flavor Violating $B \to K\tau\ell$ Decays at Belle and Belle II"

- XXXI Cracow Epiphany Conference on the recent LHC Results, $13 - 17$ January 2025 Cracow.

  Talk title: "Search for Lepton Flavor Violating $B \to K\tau\ell$ Decays at Belle"

- 50th B2GM (Belle II General Meeting), Feb $26-$Mar $1, 2025$, KEK Japan.

  Talk title: "Search for $B^+ \to K^+\tau^\pm\ell^\mp$ Decays Using Inclusive Semileptonic Tagging at Belle/Belle II"

- XXX Cracow Epiphany Conference on Precision Physics at High Energy Colliders, $8 - 12$ January 2024 Cracow.

  Talk title: "Monte Carlo Studies of Lepton Flavor Violating $B \to K\tau\ell$ Decays at Belle/Belle II"

List of Publications

- Three conference publications (two published, one in preparation).
- Author in 40+ publication (as a result of active contribution in the Belle II Silicon Vertex Detector group) from Belle/Belle II Collaborations.
- The complete list of publications can be accessed through the following link iNSPIRE-HEP.

Abstract

Lepton-flavor-violating (LFV) decays of B mesons provide unambiguous evidence of physics beyond the Standard Model (SM), as they are essentially forbidden within the SM framework. The decays $B \to K\tau\ell$ ($\ell = e, \mu$) are particularly sensitive to new physics scenarios involving leptoquarks, extended Higgs sectors, or other sources of charged lepton flavor violation.

This doctoral work presents the development of novel reconstruction techniques dedicated to the challenging topology of the $B \to K\tau\ell$ decay at the Belle II experiment operating at the SuperKEKB e^+e^- collider.

A central element of the analysis is the use of semileptonic tagging of the companion B meson. While semileptonic tagging introduces missing energy on the tag side, it provides significantly higher efficiency compared to fully hadronic tagging approaches. This choice becomes particularly powerful through the dedicated exploitation of the hadronic signal decay mode $\tau \to (\pi/\rho)\nu$.

By selecting $\tau \to (\pi/\rho)\nu$, the signal side contains only a single neutrino, resulting in a constrained kinematic configuration. Combining the well-defined initial state of the e^+e^- collisions with the reconstructed tag-side kinematics enable the signal decay to be reconstructed up to a two-fold kinematic ambiguity arising from quadratic constraints. This strategy allows the recovery of otherwise inaccessible observables and provides strong discrimination against the background processes.

The thesis presents the full analysis framework, including signal modeling, background suppression strategies, multivariate classification techniques, systematic uncertainty evaluation, and projected sensitivity with increasing integrated luminosity. The developed methodology establishes a scalable and experimentally robust approach to LFV searches in the $B \to K\tau\ell$ channel.

This work demonstrates that the combination of semileptonic tagging with kinematically constrained τ decays opens a new avenue for charged lepton flavor violation searches at Belle II, significantly enhancing the experiment's discovery potential in precision flavor physics.

Streszczenie

Rozpady mezonów B naruszające zapach leptonowy stanowią jednoznaczny sygnał fizyki poza Modelem Standardowym (MS), gdyż w jego ramach są praktycznie zabronione. Rozpad $B \to K\tau\ell$ ($\ell = e, \mu$) jest szczególnie czułym testem scenariuszy nowej fizyki, takich jak modele z lepto-kwarkami czy rozszerzonym sektorem Higgsa.

Niniejsza rozprawa doktorska przedstawia rozwój nowych technik rekonstrukcji dedykowanych wymagającej topologii rozpadu $B \to K\tau\ell$ w eksperymencie Belle II działającym przy zderzaczu SuperKEKB.

Kluczowym elementem analizy jest zastosowanie pół-leptonowego znakowania mezonu towarzyszącego. Choć metoda ta wprowadza brakującą energię po stronie znakującej, zapewnia istotnie wyższą efektywność w porównaniu z pełnym znakowaniem hadronowym. Spójność i skuteczność tej strategii osiągnięto poprzez wykorzystanie hadronowego rozpadu sygnałowego $\tau \to (\pi/\rho)\nu$.

Wybór kanału $\tau \to (\pi/\rho)\nu$ powoduje, że po stronie sygnałowej występuje tylko jedno neutrino, co prowadzi do silnie ograniczonej konfiguracji kinematycznej. W połączeniu z dobrze zdefiniowanym stanem początkowym zderzeń e^+e^- oraz rekonstrukcją mezonu znakującego możliwe jest odtworzenie kinematyki rozpadu sygnałowego z dwuwartościową niejednoznacznością rozwiązań wynikającą z równań kwadratowych. Strategia ta pozwala odzyskać informacje niedostępne w standardowych metodach oraz znacząco poprawia rozdzielczość sygnału względem procesów tła.

W pracy przedstawiono kompletny schemat analizy obejmujący modelowanie sygnału, strategie tłumienia tła, zastosowanie metod wielowymiarowej klasyfikacji, ocenę niepewności systematycznych oraz projekcje czułości przy rosnącej zakumulowanej świetlności. Opracowana metodologia stanowi skalowalne i eksperymentalnie solidne podejście do poszukiwań LFV w kanale $B \to K\tau\ell$.

Wykazano, że połączenie pół-leptonowego znakowania z kinematycznie ograniczonymi rozpadami τ otwiera nową perspektywę w badaniach naruszeń zapachu leptonu wzmacniając potencjał odkrywczy eksperymentu Belle II w dziedzinie precyzyjnej fizyki zapachu.

Contents

Chapter 1

Introduction

"The important thing is not to stop questioning.
Curiosity has its own reason for existing."
Albert Einstein

Particle physics aims to identify the fundamental building blocks of matter and the forces that govern them. This effort led to the formulation of the Standard Model (SM) in the 1970s [1, 2, 3, 4, 5]. The SM describes electromagnetic, weak, and strong interactions within a single quantum field theory framework. Gravity is not included.

Experimentally, the SM works extremely well. It has passed every precision test at colliders and low-energy experiments. However, it is not a complete theory. It does not explain neutrino masses, the baryon asymmetry of the Universe, dark matter, or dark energy. It also gives no insight into the origin of flavour structure or fermion mass hierarchies. These gaps strongly suggest that the SM is only an effective theory valid at accessible energies.

There are two main strategies to search for physics beyond the Standard Model (BSM). The first is the direct production of new particles at high energies. The second is indirect searches through rare or forbidden processes. The latter approach probes energy scales well beyond the direct reach of colliders. This thesis follows the indirect path.

Lepton flavour violation (LFV) is a particularly clean probe of new physics. In the SM with massless neutrinos, individual lepton flavours are conserved [6]. Even after including neutrino masses, charged LFV processes remain extremely suppressed in the SM. Their expected branching fractions are far below the experimental sensitivity [7]. Therefore, any observable signal would be a clear indication of new physics.

Neutrino oscillations already demonstrate LFV in the neutral lepton sector [8]. Many BSM models also predict measurable LFV effects in processes involving charged leptons. In this work, we search for the decay

$$B^+ \to K^+ \tau^\pm (\to \pi^\pm/\rho^\pm \nu_\tau) \ell^\mp, \quad \ell = e, \mu.$$

This decay violates the charged lepton flavour. It is forbidden at tree level in the SM and highly suppressed in loops. From the outset, this analysis was designed with the expectation of an integrated luminosity of approximately 5 ab^{-1}, which would allow for a statistically meaningful sensitivity to the targeted branching fraction.

Some BSM scenarios, such as models with vector leptoquarks [9], predict branching fractions of order $\mathcal{O}(10^{-7})$. This is within reach of B-factory experiments. The most stringent current limits, $(0.59 - 2.45) \times 10^{-5}$ at 90% confidence level, were obtained by the Belle Collaboration using hadronic tagging and mainly leptonic τ decays [10, 11].

In this thesis, we improve the sensitivity using semileptonic tagging [12], which provides higher efficiency, and by reconstructing hadronic τ decays ($\tau \to \pi/\rho\,\nu_\tau$). This reduces the number of neutrinos in the final state and improves kinematic constraints. Any evidence of this decay would be a direct sign of physics beyond the SM.

1.1 Standard Model and Flavour Structure

The SM is a gauge theory based on

$$SU(3)_C \times SU(2)_L \times U(1)_Y,$$

which describes strong and electroweak interactions. Fermions form the matter content, while gauge bosons mediate the forces. The Higgs mechanism generates particle masses through spontaneous symmetry breaking [13].

A key feature of the SM is its flavour structure. In the quark sector, weak interaction eigenstates differ from mass eigenstates. Their relation is encoded in the Cabibbo–Kobayashi–Maskawa (CKM) matrix [14, 15]. This unitary 3×3 matrix contains three mixing angles and one complex phase. The complex phase is the only source of CP violation in the SM quark sector.

Flavour-changing neutral currents (FCNC) are absent at the tree level. They occur only via loops and are, therefore, suppressed. This suppression makes rare $b \to s$ transitions highly sensitive to new physics effects. Small contributions from heavy virtual particles can produce measurable deviations.

1.2 Beyond the Standard Model

In the SM, individual lepton flavour numbers arise as accidental symmetries. They are not protected by a fundamental principle. Many BSM scenarios predict that these symmetries are broken.

Lepton flavour violation can appear in radiative, purely leptonic, or semileptonic processes [16]. The decay studied in this thesis belongs to the semileptonic category. It probes operators that couple quarks and leptons directly.

Among the proposed explanations of flavour anomalies, leptoquark (LQ) models are particularly attractive [17]. A leptoquark is a bosonic particle that carries both baryon and lepton quantum numbers. It couples directly to a quark–lepton pair. Such a coupling naturally induces semileptonic transitions and, in general, lepton flavour violation.

Recent measurements of $R(D^{(*)})$ and $R(K^{(*)})$ show deviations from SM predictions ([18], [19],[20],[21]). A single weak-singlet vector leptoquark with predominantly left-handed couplings can accommodate both sets of anomalies in a minimal framework. In this scenario, the same couplings that modify flavour-conserving observables also generate LFV processes such as

$$B \to K\tau\mu.$$

For perturbative Yukawa couplings and leptoquark masses in the TeV range, the predicted branching fraction for $B \to K\tau\mu$ is of the order $\mathcal{O}(10^{-7})$. This is several orders of magnitude above the SM expectation, yet still below the current experimental limits.

An improvement of experimental sensitivity by about two orders of magnitude would, therefore, decisively test this class of models. The decay $B \to K\tau\ell$ is not only a clean LFV probe but is also directly connected to the broader flavour anomaly program.

1.3 Why B Physics?

The B mesons contain a heavy bottom quark. Their weak decays probe flavour structure, CP violation, and rare transitions. Because of the heavy mass scale, b decays are sensitive to virtual contributions from new particles.

Precision measurements of $b \to c$ and $b \to u$ transitions determine V_{cb} and V_{ub}. Measurements of CP-violating observables constrain the CKM phase. At the same time, rare decays such as $b \to s$ transitions provide clean tests of the SM.

Anomalies observed in ratios like $R(D^{(*)})$ and $R(K^{(*)})$ have intensified interest in semileptonic B decays. These deviations may hint at lepton flavour non-universality.

1.4 Motivation

The decay $B \to K\tau\ell$ is highly sensitive to new physics. It involves both a heavy b quark and a τ lepton, which enhances the potential contributions of heavy mediators.

In the SM, this decay is forbidden at tree level due to LFV. It can only occur through higher-order effects induced by neutrino mixing, leading to an extremely small branching fraction [22]. The expected rate is far below the experimental sensitivity.

Many extensions of the SM predict sizable enhancements. In particular, leptoquark scenarios can raise the branching fraction to $\mathcal{O}(10^{-7})$, within reach of current or near-future experiments [17]. This makes $B \to K\tau\ell$ a clean and well-motivated probe of charged lepton flavour violation.

A schematic representation of the decay is shown in Fig. 1.1.

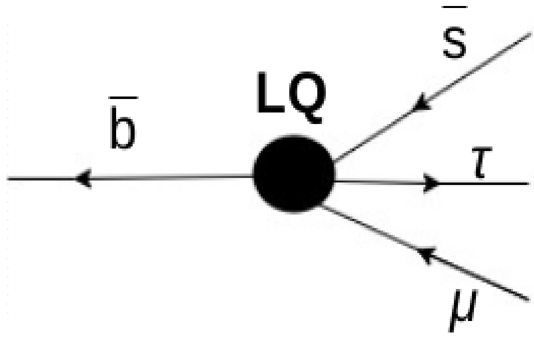


Figure 1.1: The visual description of $B \to K\tau\mu$ decays. The region shown in dark can contain some new physics (e.g., leptoquark) signatures.

1.5 Previous Searches

The first dedicated search for $B \to K\tau\ell$ was performed by the BABAR Collaboration [23] using 472×10^6 $B\overline{B}$ pairs. No signal was observed, and upper limits at the level of a few times 10^{-5} were set at 90% confidence level.

The most stringent limits to date, $(0.59 - 2.45) \times 10^{-5}$ at 90% confidence level, were obtained by the Belle Collaboration [10]. The analysis used the complete Belle data set of 772×10^6 $B\overline{B}$ pairs and hadronic tagging based on the full event interpretation (FEI) algorithm [24]. Mainly leptonic τ decay modes were considered, which involve multiple neutrinos and therefore reduce reconstruction efficiency.

The signal yield was extracted using the recoil mass variable M_{recoil}[1], which peaks at the τ mass due to the known initial kinematics and full reconstruction of the accompanying B meson.

The dominant background originated from semileptonic B^+ decays of the type

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

which were suppressed using kinematic vetoes. Multivariate techniques, including boosted decision trees (BDT), were employed to enhance signal sensitivity. Upper limits were determined using a frequentist approach.

[1] $M^2_{recoil} = m^2_B + m^2_{K\ell} - 2(E_B E_{K\ell}/c^4 + p_{Btag}p_{K\ell}cos\theta/c^2)$, where θ is the angle between p_{Btag} and $p_{K\ell}$.

Chapter 2

Belle and Belle II Experiments

"An experiment is a question which science poses to Nature, and a measurement is the recording of Nature's answer."
Max Planck

This chapter describes the experimental setups (generally known as the Belle and Belle II experiments) in detail. The Belle experiment was located in Tsukuba, Japan, and has collected data from 1999 - 2010 as a B factory experiment [25]. The Belle II experiment is an upgrade of the Belle experiment with the ambitious goal of collecting 50 times more data in its lifetime compared to the Belle dataset [26]. Like other particle physics experiments, they also had a source (KEKB accelerator for Belle and SuperKEKB for Belle II) to produce the particles (mainly B mesons) and a detector (Belle detector and Belle II detector) to detect them. Belle collected approximately 1 ab^{-1} of the data on the different $\Upsilon(nS)$ resonances[1], while Belle II started collecting data from 2019 and during its first run period (2019 - 2022) it accumulated 424 fb^{-1} of data. The second run period of Belle II started in early 2024, and so far (to February 2026), it has collected more than 200 fb^{-1} of data [27]. The main goal of the Belle experiment was to observe the CP violation in the B mesons, which was achieved and cited in the 2008 Nobel Prize in physics. The primary objective of the Belle II experiment is to search for BSM physics in various ways, e.g., improving the precision in the CKM sector, searching for missing mass in the decays involving dark matter, etc. After reviewing the basic information about the Belle/Belle II experiments, in the subsequent section, the KEKB and SuperKEKB accelerators are explained.

2.1 KEKB and SuperKEKB Accelerators

The KEKB accelerator is a two-ring asymmetric energy electron-positron collider with a circumference of approximately 3.0 km and located 11.0 m below ground

[1] where $n = 1, 2, 3, 4, 5$.

level. Its schematic layout is shown in Fig. 2.1. The main function of the KEKB collider was to produce the B mesons in the processess $e^-e^+ \to \Upsilon(4S) \to B\overline{B}$. Its design luminosity was 1×10^{34} cm^{-2}s^{-1} which it exceeded and achieved the maximum luminosity of 2.11×10^{34} cm^{-2}s^{-1} in 2009. In-depth technical details on its history, construction, performance, and other functionalities can be found in [28].

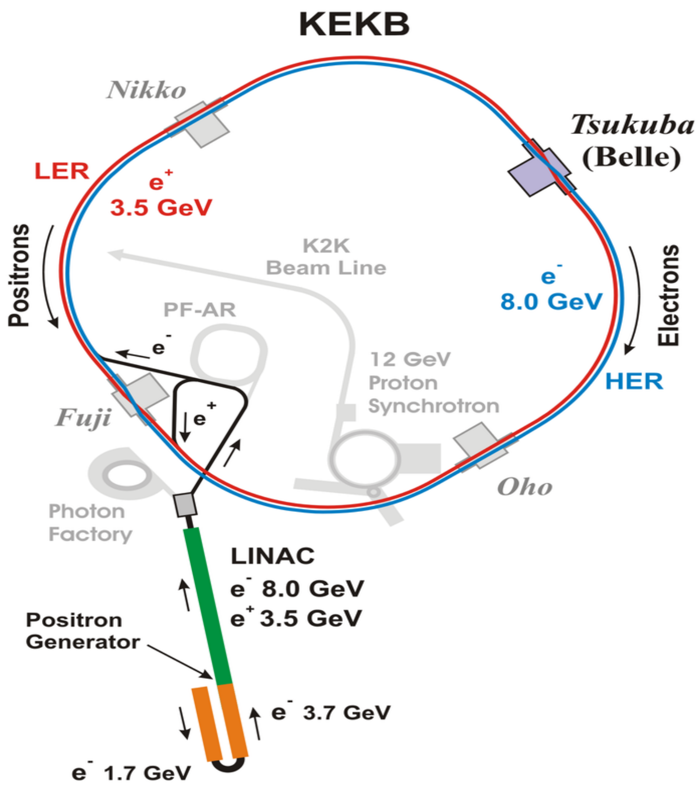


Figure 2.1: KEKB accelerator schematic layout. Credit [29]

It consists of two 3.016 km long rings, one is an 8 GeV electron ring (high energy ring - HER) and the other is a 3.5 GeV positron ring (low energy ring - LER). Electrons and positrons are directly injected into the rings through LINAC (linear accelerator) at full energies. The ideal conditions for obtaining higher luminosity are high beam currents and a small beam size. The KEKB radio frequency (RF) accelerator system enabled to store beam currents of 1.1 A in HER and 2.6 A in LER [30]. The beam is distributed among 5000 bunches with 59 cm of bunch spacing. The two rings have equal circumference and cross each other at the interaction point (IP), where the electron beam and the positron beam collide at a finite angle of ±11 mrad, allowing the bunches to separate quickly after the collision. At the IP, the beams are squeezed by using a pair of superconducting focus quads that are immersed in a 1.5 T magnetic field. This results in the size of the beams in IP 90 µm in the horizontal direction and 1.9 µm in the vertical direction.

The centre of mass (CM) energy at the IP is estimated by

$$\sqrt{s} \approx 2\sqrt{E_{HER}E_{LER}} \ ,$$

where E_{HER} and E_{LER} are the energies in the HER and LER, respectively, in the lab frame. This CM energy is scanned on different $\Upsilon(nS)$ resonances that have mass ranges from 9 to 11 GeV/c^2. The hadronic cross section for the different $\Upsilon(nS)$ states is shown in Fig. 2.2. One of the resonance states is $\Upsilon(4S)$ (with CM energy of 10.579 GeV), which further decays to a pair of B mesons, i.e. $B\overline{B}$. Due to the difference of the energies in the HER and LER beams, CM gets a Lorentz boost along the forward direction (electron beam), whose value is given by

$$\beta\gamma = \frac{E_{HER} - E_{LER}}{E_{CM}} = 0.425$$

This Lorentz boost increases the separation between the two B decay vertices, to about 200 μm which can be resolved by the vertex detector (having a resolution of 100 μm). Due to this boost, the time difference between the two B decay vertices is measured, which is further used in the measurement of the time-dependent CP violation.

The SuperKEKB accelerator is an upgraded version of the KEKB accelerator [31]. Its schematic layout is shown in Fig. 2.3. It consists of a 7 GeV electron ring, a 4 GeV positron ring, and a LINAC with a 1.1 GeV positron damping ring. To reduce beam losses due to Touschek scattering[2] in the LER, lower beam energy asymmetry is set (compared to KEKB). The Lorentz boost in this case is 0.28 in the laboratory frame, which results in the average flight distance of 130 μm of the produced B mesons.

SuperKEKB is designed to achieve the instantaneous luminosity of 8×10^{35} cm^{-2}s^{-1} and the total integrated luminosity of 50 ab^{-1}. To achieve this higher luminosity, significant upgrades have been performed to HER, LER, and the focus system of KEKB (a detailed discussion can be found in [31]). As mentioned earlier, the luminosity can be increased by decreasing the beam sizes and increasing the beam currents. In SuperKEKB, the vertical size of the beam bunch at IP is ~ 60 nm (nano-beam scheme originally developed for the SuperB experiment [32]) and the beam currents are set to be 3.6 A and 2.6 A in LER and HER, respectively. The two beams collide at an angle of 83 mrad and a centre of mass energy corresponding to 10.58 GeV.

This concludes the overview of the essential ingredients of KEKB and SuperKEKB accelerators. In the next section, the other important part of the Belle experiment is discussed, which is the Belle detector (and it will be followed by a discussion on the Belle II detector).

[2]Large transfer of momentum from the transverse planes to the longitudinal planes in the scattering events. For a sufficiently large change in the longitudinal momentum, the energy deviation of the particles can be outside the energy acceptance of the ring, and these particles will be lost from the beam.

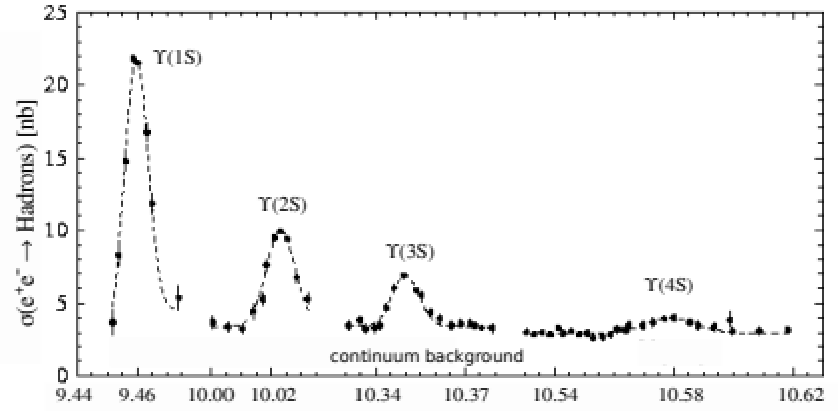


Figure 2.2: Hadronic cross section as a fucntion of e^+e^- CM energy (GeV). [33]

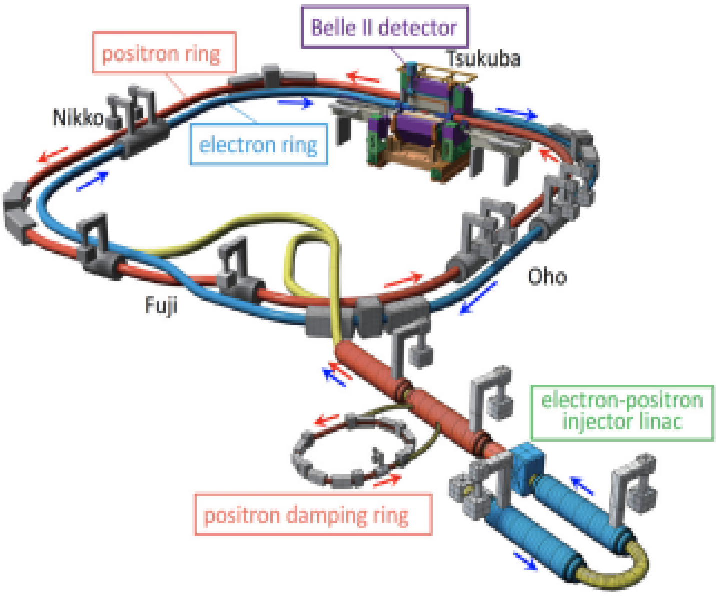


Figure 2.3: SuperKEKB accelerator schematic layout. Credit [31]

2.2 Belle Detector

The Belle detector is a state-of-the-art 4π detector, consists of various subdetectors whose arrangement is shown in Fig. 2.4. Its geometrical center is displaced from the IP in the forward direction to take into account the CM boost in the experiment. The coordinate system used in the detector is the Cartesian coordinate system with IP as the origin. The z-axis points in the direction opposite to the e^+ beam, the x-axis is radially outward and the y-axis is vertically upward. Due to the cylindrical shape of the detector, it is convenient to use the spherical coordinate system in which θ and ϕ are the polar and azimuthal angles, respectively. Based on the polar angle ranges, the detector is divided into three regions: forward end caps ($17^\circ < \theta < 34^\circ$), barrel region ($34^\circ < \theta < 127^\circ$), and backward end caps ($127^\circ < \theta < 150^\circ$).

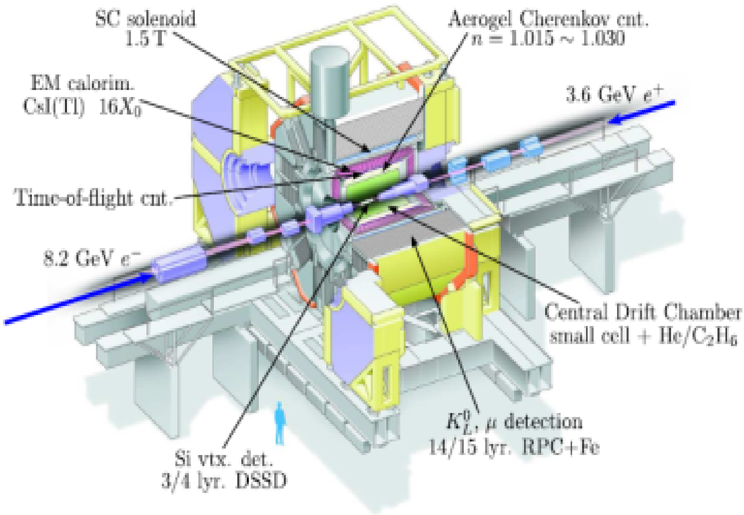


Figure 2.4: Belle detector composition. Credit [35]

The particles produced in the beam collision, i.e. B mesons, τ leptons, etc., decay further into the stable particles (π, K, μ, e, γ, etc.). These particles produce signals in the different sub-detectors through numerous mechanisms. In the following subsections, these different subdetectors are briefly described while a detailed description of the complete Belle detector and its components can be found in [34].

2.2.1 Beam Pipe

The beam pipe is the innermost part of the Belle detector that encloses the IP. The resolution of the vertex varies inversely with the distance from the first detection layer. Therefore, the beam pipe is designed of two thin (5 mm thickness) beryllium cylinders having inner radii of 20 mm and 23 mm. To counteract beam-induced heating (of the order of a few hundred Watts), a 2.5 mm gap between the outer and inner layers of beam pipe is used as a helium gas channel for cooling. To reduce the low-energy X-ray background from the HER, a 20 mm thick gold sheet is attached outside the beryllium cylinders. The cross section of the beam pipe around IP is shown in Fig. 2.5.

2.2.2 Silicon Vertex Detector

The silicon vertex detector (SVD) is located outside the beam pipe. The essential job of SVD is to precisely reconstruct the part of the tracks of the charged particles in that region. An accurate measurement of the separation between the two B vertices is a key parameter for time-dependent CP violation studies in neutral B mesons, and SVD plays a crucial role in it. The resolution of slow charged particle's momentum is also improved by SVD. Belle experiment has used two different SVD

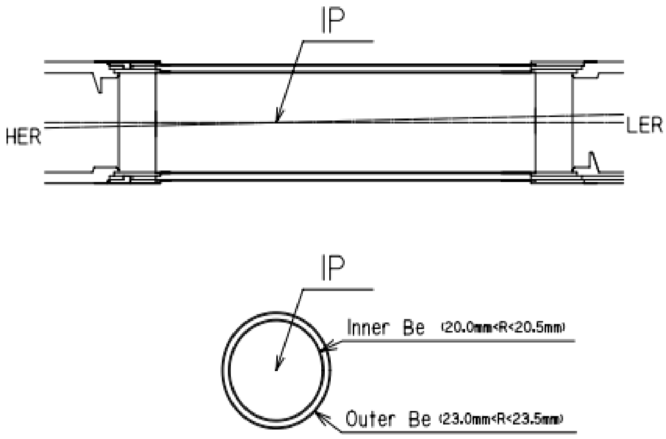


Figure 2.5: Schematic view of beam pipe. Credit [36]

detectors. The first Belle SVD (SVD1) operated from 1999 until the summer of 2003 and contributed to many important physics results, including the discovery of CP violation in the B meson system [37]. However, because of some of its limitations, (e.g. non-negligible dead time, limited radiation tolerance, etc.), it was replaced by a second Belle SVD (SVD2), which started data taking in October 2003 onward. Some important characteristics of SVD1 and SVD2 are summarized in Table 2.1.

detector	layers	radii (mm)	θ coverage	beam pipe diameter (mm)
SVD1	3	$30, 45.5, 60.5$	$23° - 139°$	40
SVD2	4	$20, 43.5, 70, 88$	$17° - 150°$	30

Table 2.1: Key differences between SVD1 and SV2. More technical and performance details are discussed in [36], [38] for SVD2 and SVD1 respectively.

The detailed side view of the SVD2 detector is shown in Fig. 2.6. The layers in SVD consist of ladders, which are composed of double-sided silicon strip detectors (DSSDs). DSSDs are actually reverse-biased P-N junction diodes. Their readout chain is based on the VA1 (Viking architecture) integrated circuit. For background electronics, flash analog-to-digital converters (FADCs) and field programmable gate arrays (FPGAs) are used.

2.2.3 Central Drift Chamber

Central drift chamber (CDC) is the central tracking system of the Belle detector placed inside a solenoidal magnetic field of strength 1.5 T (to bend the charged particles). It provides an efficient reconstruction of the charged particle's tracks and a precise measurement of their momenta (with a momentum resolution of $\frac{\sigma_{p_t}}{p_t} \sim 0.5\%\sqrt{1+p_t^2}$)[3]. In addition to them, it also measures the $\frac{dE}{dx}$ of the charged particles

[3] p_t in GeV/c.

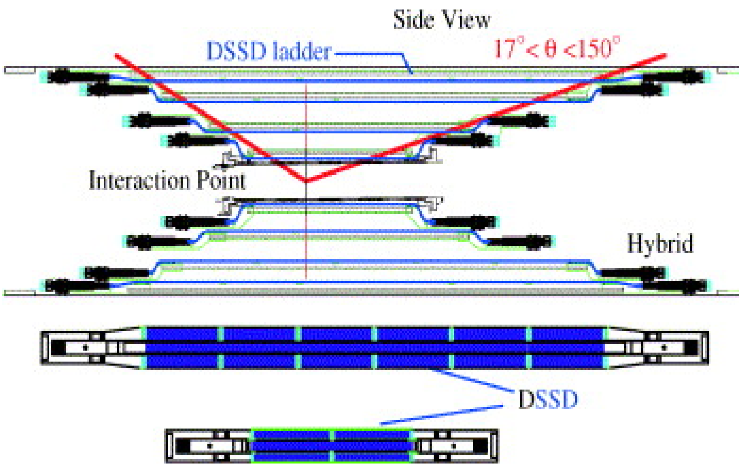


Figure 2.6: SVD2 detector sideview. Credit [36]

and therefore provides key information for the identification of the charged particles (PID).

The structure of CDC is shown in Fig. 2.7. To provide polar angle coverage of $17° < \theta < 150°$, it is designed to be asymmetric around the z-axis. The inner and outer radius of the CDC are 103.5 mm and 874 mm respectively. It has 8400 drift cells, which are the squares (in shape) except the ones in the first three layers. There are 50 cylindrical layers in the chamber, each contain three to six layers of either axial or stereo layers and three cathode strip layers. The axial layers are aligned in the z-axis (in the direction of the solenoid magnetic field) while the stereo layers are skewed at an angle of 50 mrad. To minimize the effects of multiple Coulomb scattering in momentum resolution, the chamber is filled with a mixture of low Z (50 % helium, 50 % ethane) gas. To maximize the drift electric field, sense wires (which are gold plated tungeston wires of 30 µm diameter) surrounded by the eight aluminum field wires of 126 µm are used.

When a charged particle enters the CDC, it produces ionization in the gas mixture. The electrons produced in the ionization drift towards the sense wires (because of the voltage difference between field and sense wires), which leads to the detection of the original charged particle. The sign (positive or negative) of the charged particle is determined by the track curvature. The particle's $r - \phi$ position is determined from the information from both axial and stereo layers, while its z-position is measured from the angle recorded in the stereo layer.

2.2.4 Aerogel Čerenkov Counter System

Aerogel Čerenkov counter system (ACC) performs the crucial task of distinguishing high momenta (1.2 GeV/c to 3.5 GeV/c) $\pi^{\pm}$ from $K^{\pm}$. An efficient distinction between $\pi^{\pm}$ and $K^{\pm}$ leads to effective flavor tagging at Belle. ACC is based on the fact that when a particle moves through a medium with a speed greater than the speed of light in that medium, it emits light (Čherenkov radiation or Čerenkov photon). This condition can be generalized further for a particle of mass m, moving

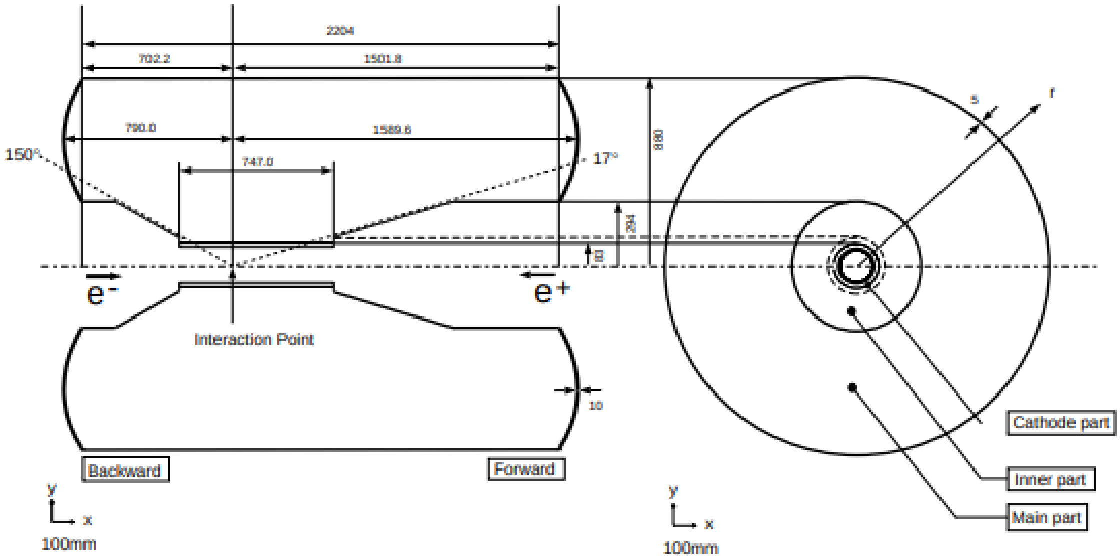


Figure 2.7: CDC structure overview. The lengths are in mm. Credit [34]

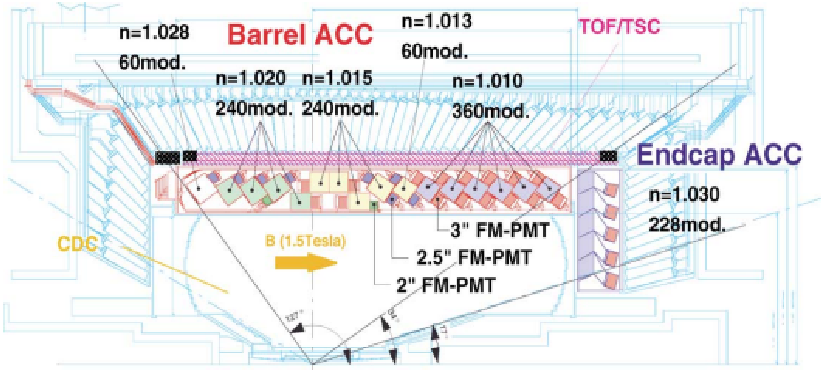


Figure 2.8: ACC configuration. Credit [34]

with momentum p, traveling through a medium with refractive index n will emit Čherenkov radiation only when $m < p\sqrt{n^2 - 1}$ is satisfied. By selecting a specific value of n (choosing a specific material), one can essentially separate $\pi^{\pm}$ from $K^{\pm}$.

The complete structure of the ACC is shown in Fig. 2.8. ACC covers a polar angle range of $17° < \theta < 127°$. Its barrel part consists of 960 silica aerogel counter modules that are further segmented into 60 cells in the ϕ direction. In its forward endcap part, 228 counter modules are arranged in five concentric layers. The refractive indices of the aerogels are kept in the range of 1.01 to 1.03 (depending on the range of the polar angle) to obtain a good $\pi^{\pm}/K^{\pm}$ classification. Aluminum boxes of approximate dimensions $12 \times 12 \times 12$ cm^3 are used to stack the aerogel tiles. To detect the Čherenkov photons, one or two fine mesh-type photomultiplier tubes (FM-PMT) are connected directly to the sides of the boxes.

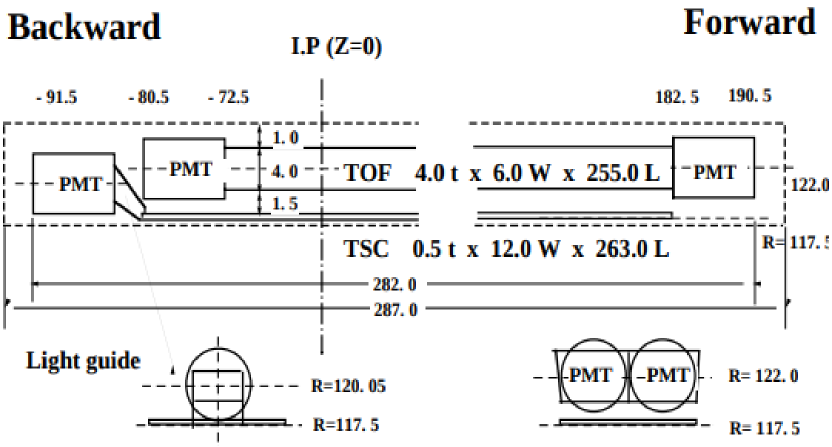


Figure 2.9: TOF module schematic layout. Credit [34]

2.2.5 Time of Flight Counter

In order to provide a clean and efficient b tagging, low momenta (less than 1.2 GeV/c) $\pi^{\pm}$ and $K^{\pm}$ also need to be distinguished. For this purpose, the time of flight counter (TOF) is used. It covers a polar angle range of $34^{\circ} < \theta < 120^{\circ}$ and provides a timing resolution of 100 ps. The TOF measures the mass of the particle using the particle's flight length ($L \sim 1.2$ m) and momentum (p) information from SVD and CDC and flight time (T) from the IP to the TOF using the following formula:

$$m = \frac{p}{c}\sqrt{\frac{c^2T^2}{L^2} - 1} \qquad (\text{GeV/c}^2) \tag{2.1}$$

It consists of plastic scintillation counters that provide fast timing signals for the trigger system. The rate of the TOF trigger signals is kept below 70 kHz (using trigger scintillation counters (TSC)) to avoid pile-up in the trigger queue. There are a total of 128 TOF counters and 64 TSC counters located in a radius of 1.2 m from IP. In order to prevent the $e^- - e^+$ pairs produced in the TSC to enter the TOF a 1.5 cm gap is kept between a TOF and a TSC counter. A typical TOF module is shown in Fig. 2.9. The information from TOF and TSC is readout by FM-PMT attached to the sides and ends of the scintillators. It has provided a 2σ separation for particles ($\pi^{\pm}$ and $K^{\pm}$) with momenta up to 1.25 GeV/c. More in-depth details on TOF detector electronics and performance can be found in [34].

2.2.6 Electromagnetic Calorimeter

To efficiently detect photons from B meson decays, an electromagnetic calorimeter (ECL) is used. The ECL provides good resolution for both the energy and position of the photons. The majority of photons are produced in cascade decays with relatively low energy (below 500 MeV), so good detection performance for these low-energy photons is very important. In some of the B decays ($B \to \pi^0\pi^0, B \to K^*\gamma$), high-energy ($\sim 4$ GeV) photons are also produced, for which high resolution is required

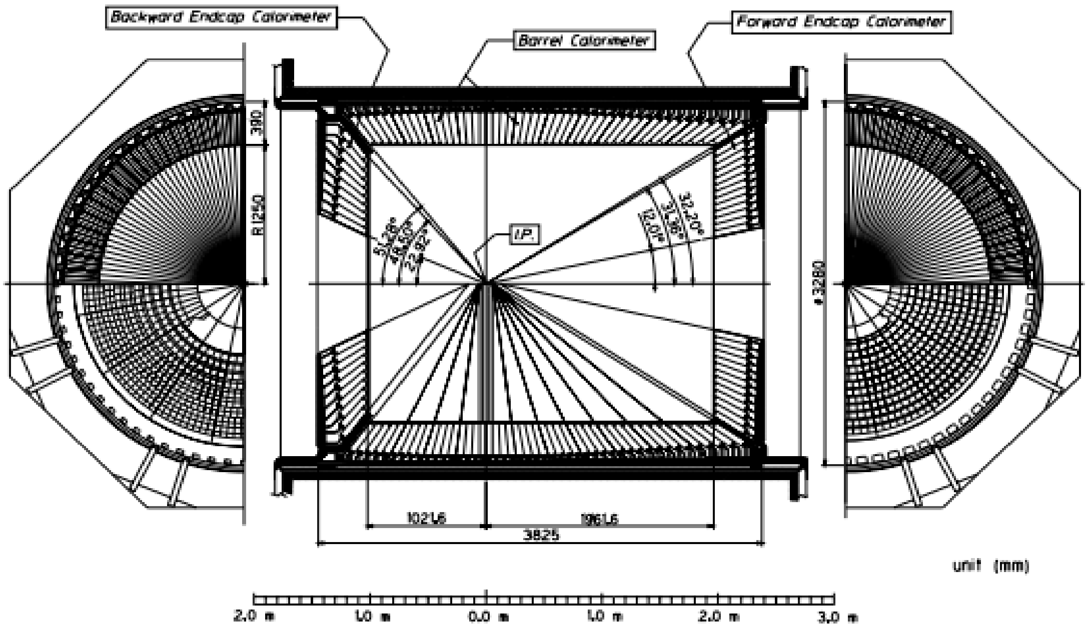


Figure 2.10: The ECL configuration. Credit [34]

to distinguish them from the background. In addition to photon detection, ECL is a key element in electron identification at Belle. Its polar angle coverage range is $17° < \theta < 150°$, which corresponds to 91% of the total solid angle (4π) coverage.

It consists of an immensely segmented array of CsI(Tl) crystals with readout of silicon photodiodes placed inside a magnetic field of 1.5 T produced by the superconducting solenoid magnet. The detailed schematic diagram of the ECL is shown in Fig. 2.10. The length of its barrel section is 3.0 m with an inner radius of 1.25 m. The end caps are placed at z $= +2.0$ m and -1.0 m from the IP. The crystals are in a tower-like shape, and they point towards the IP. To prevent photon escape from the gaps between the crystals, a small tilt angle ($\sim 1.3°$) is kept between them in the $\theta - \phi$ direction in the barrel region. To readout information from CsI(Tl) counter, every individual counter is connected to an independent pair of silicon PIN photodiodes, and charge-sensitive preamplifiers are attached at the end of the crystal.

The ECL achieved an energy resolution of 1.7%, 1.74%, and 2.85% for the barrel, forward, and backward ECL parts, respectively. For the invariant mass distributions of $\pi^0(\pi^0 \to \gamma\gamma)$, it has achieved a mass resolution of 4.9 MeV, and for $\eta(\eta \to \gamma\gamma)$ it has a resolution of less than 10 MeV.

2.2.7 Extreme Forward Calorimeter

For some physics processes (e.g $B \to \tau\nu_\tau$), polar angle coverage of ECL needs to be extended. For this purpose, an extreme forward calorimeter (EFC) is used which increases the angular range from 6.4° to 11.5° in the forward direction and from 163.3° to 171.2° in the backward direction. It also serves as a mask to reduce background for CDC. In the beam monitoring and Belle luminosity measurements, EFC is required.

As the EFC is located in a very high radiation level area, it is made up of a radiation-hard BGO (Bismuth German Oxygen, $Bi_4Ge_3O_{12}$). These crystals are packed inside a bucket-shaped stainless steel container of 1 mm thickness. For better position resolution, every EFC section is further segmented into 32 (azimuthal) and 5 (polar) divisions. More details on EFC can be found in [34].

2.2.8 K_L and Muon Detector

The K_L and muon (KLM) is a dedicated detector located outside of the superconducting solenoid for the detection of K_L and muons which are in a broad momentum range of more than 600 MeV/c. The muon is distinguished from K_L based on the fact that on average the muon travels much further with smaller deflections than strongly interacting K_L. The total polar angle coverage range of KLM is $20° < \theta < 155°$.

It consists of alternating layers of charged particle detectors and iron plates of a thickness of around 4.7 cm. In the barrel region, there are 15 detector layers and 14 iron plates while in the forward and backward end caps there are 14 detector layers. The detailed structure of a superlayer is shown in Fig. 2.11. For a K_L traveling normal to the detector, the iron plates provide a total of 3.9 interaction lengths (mean free paths), and the additional 0.8 interaction length of the material is provided by the ECL. Based on transverse scattering, multiple layers of detector and iron plates allow to distinguish K_L and muon. For the detection of charged particles, glass-electrode-resistive plate counters (RPCs) are used. The RPCs consist of two parallel plate electrodes (with high bulk resistivity $\geq 10^{10}$ Ωcm) separated by a gas-filled gap. When an ionizing particle passes through the gap, it results in a local discharge of the plates. This discharge leads to the induction of a signal on the external pickup strips.

For muons with an energy greater than 1.5 GeV/c, KLM provided a muon identification efficiency of better than 90 % with a fake rate of less than 5%.

2.2.9 Solenoid and Iron Structure

In the Belle detector, a superconducting solenoid magnet is used to provide a strong magnetic field (1.5 T) inside a cylindrical volume of length 4.4 m and diameter of 3.4 m. The solenoid coil is made up of a single layer of Niobium-Titanium-Copper around a highly pure (99.99%) aluminium.

The iron yoke provides support for all detector components and also serves as a return path for magnetic flux. The weight of barrel yoke is 608 tons and end cap yokes are 524 tons.

This concludes the discussion on the different sub detectors used in Belle, and in the following section, the Belle II detector is reviewed.

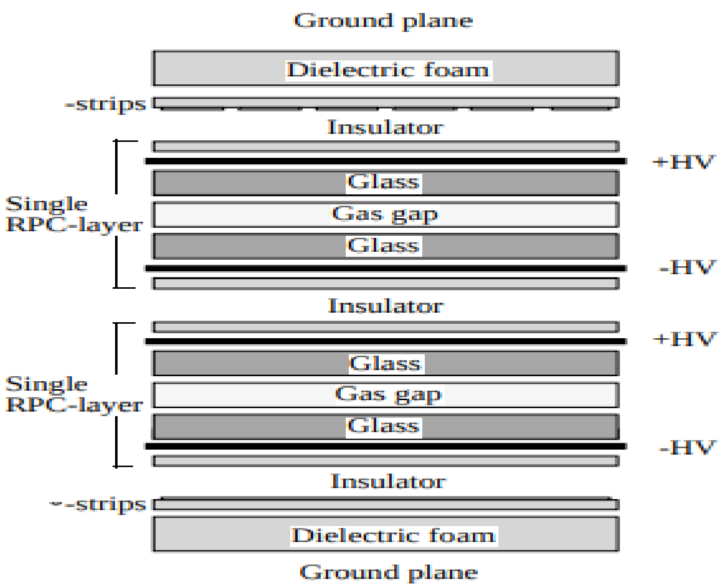


Figure 2.11: The KLM super layer schematic construction. Credit [34]

2.3 Belle II Detector

Like the Belle detector, the Belle II detector is also a hermetic and a 4π detector surrounding the IP. It covers the polar angle range of $17° < \theta < 150°$ and consists of various subdetectors. Its schematic construction is shown in Fig. 2.12. At the heart of the Belle II detector lies the vertex detector, which is discussed in the next subsection.

2.3.1 Vertex Detector

The vertex detector is the innermost part of the Belle II detector and is designed for precise vertex reconstruction. Its schematic view is shown in Fig. 2.13. It consists of two sub-detectors, namely the pixel detector (PXD) and the silicon vertex detector (SVD). The PXD is made up of two layers, which are at the radius of 14 mm and 22 mm around the beam pipe. The sensors used in the PXD are of the depleted field effect transistor (DEPFET) type [40]. The usage of DEPFET reduces the thickness of the sensors, which minimizes the possibility of multiple scattering.

The SVD comprises four layers at radii of 38 mm, 80 mm, 115 mm, and 140 mm. These layers are arranged in a cylindrical shape with increasing lengths towards the outer layers. To maintain a coverage of $17° < \theta < 150°$, the three outermost layers of SVD are tilted (at an angle of 15°) towards the beamline in the forward region. The sensors used in the SVD are DSSD. Compared to the Belle SVD, the Belle II SVD innermost layers (PXD) are closer to the IP and the outermost layer is at a considerably larger radius. These changes lead to a significant improvement in the vertex resolution compared to Belle. The Belle II SVD is discussed in more detail in [41].

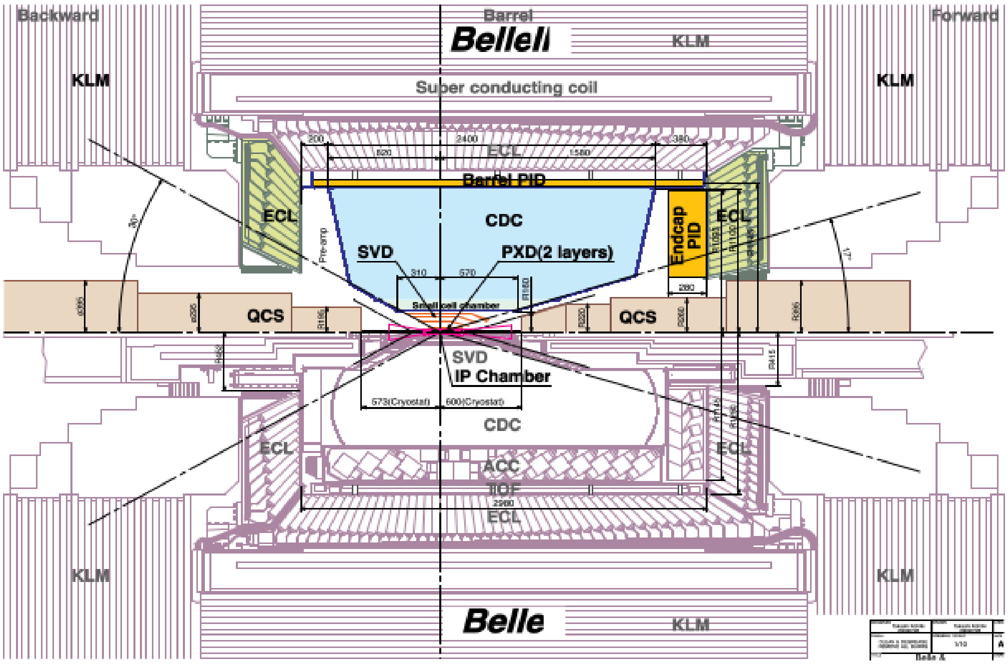


Figure 2.12: Belle II detector schematic construction. Credit [39]

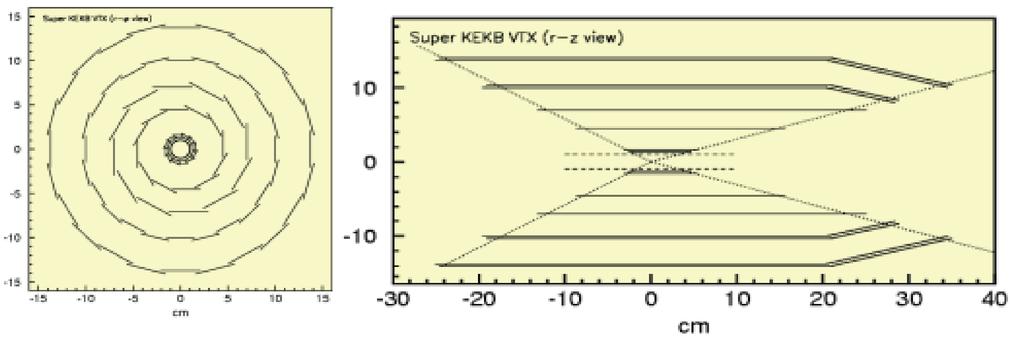


Figure 2.13: Credit [42]

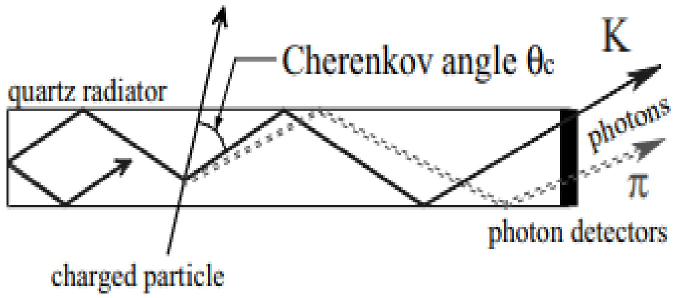


Figure 2.14: Schematic view of θ_c determination . Credit [44]

2.3.2 Central Drift Chamber

The central drift chamber (CDC) is the crucial component of the Belle II detector for the tracking and identification of charged particles. It is a large volume drift chamber composed of small drift cells. The radius of CDC is 1130 mm. There are $14,336$ wires arranged in 56 layers in two orientations (axial and stereo). In the axial orientation, the wires are aligned with the solenoidal magnetic field, and in the stereo orientation, the wires are skewed with respect to the axial wires. To reconstruct a full 3D helix track of the particle, information from both axial and stereo wires is used. The chamber is filled with a helium-ethane $50:50$ mixture. More in depth details on Belle II CDC can be found in [43].

2.3.3 Particle Identification System

The particle identification system in Belle II consists of two sub-detectors: a time of propagation (TOP) counter (in the barrel region) and an ARICH counter (in the forward end cap region). Both sub-detectors use the Čherenkov effect to identify charge particles passing through them. The particle velocities are determined by measuring the angle θ_c (Čherenkov angle)[4]. This information is then combined with the measured momentum information from the tracking devices and the CDC energy loss to determine the mass of the particle.

TOP is a special type of Čherenkov detector. There are a total of sixteen detector modules in TOP, and each module consists of a 2 cm thick, 45 cm wide, and 2.6 m long quartz bar. A mirror is placed on the forward extremity, and on the backward end, micro-channel plate photomultiplier tubes (MCP-PMTs) are installed. When a charged particle passes through the quartz radiator, it emits Čherenkov photons. These photons are collected by the MCP-PMTs, which provide timing and spatial information in the (x, y) plane. It uses the time taken by the Čherenkov photons to pass through a quartz bar (and reflected internally) and the position at which they arrive to determine the θ_c. This is illustrated in Fig. 2.14. TOP covers a range of $31° < \theta < 128°$.

[4] using $cos\theta_c = 1/n\beta$, where n is the refractive index of the material and β is the velocity of the particle.

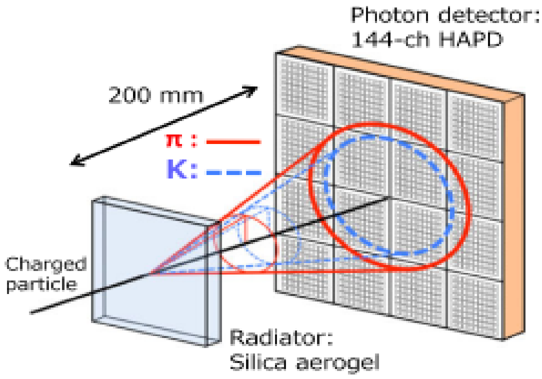


Figure 2.15: π/K separation in the ARICH. Credit [45]

The PID information is provided by the ARICH in the forward end cap region of the detector. To identify particles, ARICH uses ring imaging of the Čherenkov cones. It consists of a 4 cm large aerogel radiator (for particles to emit Čherenkov photons), a 20 cm large expansion volume (for Čherenkov rings to be formed), and a single photon detector based on the hybrid avalanche photodetector (HADP) technology (to detect Čherenkov photons). ARICH covers a range of $14° < \theta < 30°$. As the decays in the Belle II are forward boosted (due to asymmetric beam energies), there is no PID system in the backward end cap region of the Belle II detector. It minimizes the overall detector capability for particle identification; however, only 10% of the CDC coverage falls outside the TOP and ARICH coverage.

2.3.4 Electromagnetic Calorimeter

Approximately 33% of the end products of the B decays are π^0 and other neutral particles that produce photons. For the efficient detection of these photons, the electromagnetic calorimeter is a crucial component. The Belle II ECL reuses the material and structure of the Belle ECL. However, to be able to handle the higher background rates in Belle II, the readout electronics is upgraded. The photodetectors in the ECL are connected with waveform sampling readout electronics. The ADC samples are stored in the FPGA internal buffers to use waveform fitting to differentiate between the signal events and the pileup noise events.

In Belle II, the ECL serves three important purposes. It detects photons and provides information about their energy and angle. It also identifies electrons (using the likelihood ratio) and monitors the luminosity of SuperKEKB. The performance of the Belle II ECL closely resembles that of the Belle ECL. Its working principles and performance (during 2018) are comprehensively reviewed in [46].

2.3.5 K_L and Muon Detector

For the detection of long-lived and high penetration particles (K_L and μ), the K_L and Muon detector (KLM) is used. For muons to reach the KLM, they should have

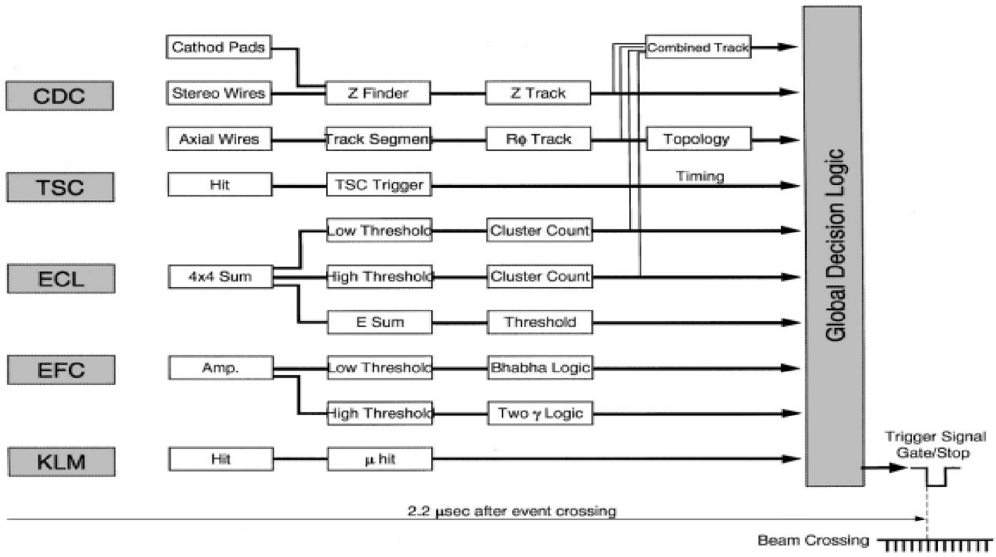


Figure 2.16: The Belle trigger system. Credit [47]

at least a momentum of 600 MeV/c. The K_L can produce hadronic showers in ECL and/or KLM.

The Belle II KLM reuses the same structure as the Belle KLM; however, the detection system in the Belle II KLM is upgraded to deal with higher background levels compared to the Belle[5]. It is located outside the superconducting solenoid and consists of three (barrel, forward endcap, backward endcap) regions. It covers an angular range of $20° < \theta < 155°$. It comprises alternating layers of iron plates (of 4.7 cm thickness) and active detector elements. In the barrel region, there are 15 detector layers and 14 iron plates while in the end cap regions, there are 14 detector layers and 14 iron plates. The iron plates enable K_L to produce hadronic showers.

This ends our discussion on the Belle II detector, and in the next section, we review the trigger and data acquisition system of Belle.

2.4 Trigger and Data Acquisition at Belle

At Belle, $B\overline{B}$ are the primary events of interest. The cross section for these events is small (1.15 nb), so a high trigger efficiency is required with the ability to reject the uninteresting events. For detector calibration and luminosity measurement, the Bhaba and $\gamma\gamma$ events are also important. For these events, the cross section is large (44 nb and 2.4 nb, respectively), so they are prescaled by a factor of 1/100. In the low transverse momentum (p_t) range, two-photon processes are rejected because of their resemblance with the beam background events. After this suppression, the trigger rate is set to $\sim$ 100 Hz for the physics events. The schematic design of the Belle trigger system is shown in Fig. 2.16 and is discussed in detail in [47].

[5] Instead of RPCs, scintillator strips read out by silicon photomultipliers (SiPMs) are used.

The trigger system consists mainly of two parts.
i- Trigger system of individual subdetectors.
ii- Central trigger system (global decision logic or GDL).

It can also be categorized into three levels, namely the level 1 (L1) trigger, level 3 (L3) trigger, and level 4 (L4) trigger. L1 is a hardware level trigger, while L3 and L4 are the software level triggers. In the next subsections, these trigger levels are further elaborated.

2.4.1 Level 1 Trigger

The L1 trigger system is based on the input of all sub-detectors (except SVD). The trigger system of the sub-detectors is based on the two trigger categories which are track triggers and energy triggers. The trigger signals from charged particles are sent by the CDC ($r-\phi$ and $r-z$ trigger signals) and TSC/TOF sub-detectors. Triggers based on the total energy deposit and the cluster counting of crystal hits are provided by the ECL trigger system. Additional information on muons is provided by the KLM trigger system. All the trigger signal processing is done in parallel in the sub-detectors, and final information is provided to the GDL system to characterize the event type.

To match the SVD readout system (with 2.5 µs shaping time), the trigger system latency[6] is set to 2.2 µs. To maintain this set latency, each subdetector trigger signal must reach the GDL at a maximum latency of 1.85 µs. The GDL has a processing time of 350 ns to form the final trigger signal. The hit information from the SVD is available after 25.6 µs to be used in the trigger system.

2.4.2 Level 3 Trigger

The L3 trigger system is a software-level trigger that selects events using the track finding algorithm. It rejects events with no track with a z-impact parameter[7] less than 5 cm and total energy deposited in the ECL less than 3.0 GeV. The hadronic events are selected with high efficiency (>99%) with these conditions mentioned, while it rejects $50-60$ % of the events that were passed by the L1 trigger system. Events that pass the L3 trigger constitute the raw data and are stored for further refinement.

2.4.3 Data Acquisition System

For data acquisition, a distributed parallel system has been developed that provides data acquisition at 500 Hz with a dead-time fraction of less than 10 %. Its design is shown in Fig. 2.17. It consists of seven subsystems that run in parallel to handle the data from each subdetector. Input from every subsystem is combined to form an event builder. The output of the event builder is transferred to a computer

[6] Time taken from the event collision to the decision-making by GDL.
[7] Closest point of approach to the IP along the beam axis.

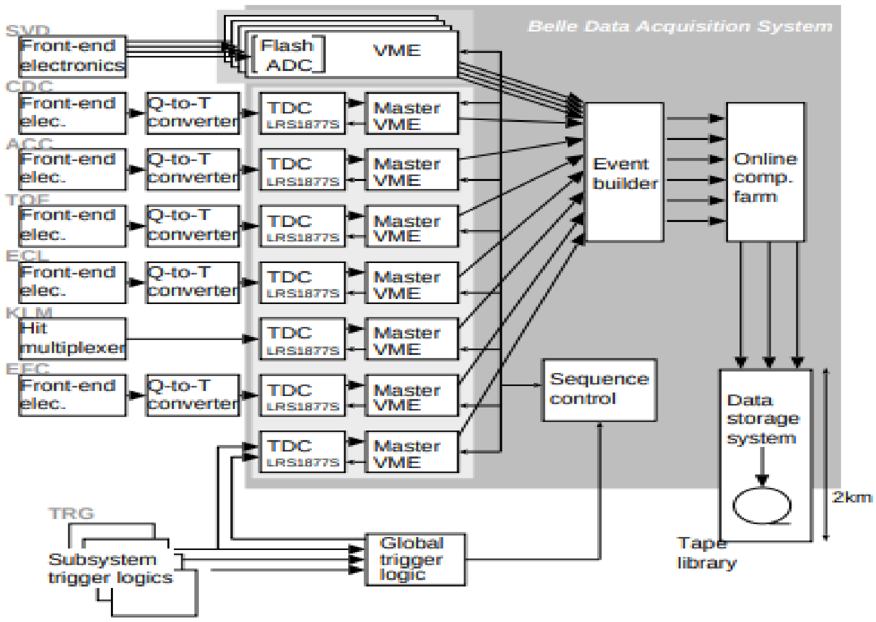


Figure 2.17: Belle DAQ system. Credit [34]

farm, where further event filtering is performed before sending the data to a mass storage system[8]. The stored data contains $B\overline{B}, \tau^+\tau^-, \gamma\gamma$ events alongside the events required for the detector calibration, e.g. $\mu\mu$ events.

2.4.4 Level 4 Trigger

The L4 trigger is an offline trigger that allows events only if they have at least one good charged track. A charged track is qualified as a good charged track when it meets the following requirements:

i- $p_t > 300$ MeV/c

ii- $|dr| < 1$ cm and $|dz| < 4$ cm[9]

After L4 trigger, 26.7% of the initially triggered events remained to be reconstructed and saved in data summary tapes (DST). At this stage, the raw data are converted into physics objects of four vectors x_μ and p_μ. These physics objects are the primary sources for any further physics studies.

2.5 Trigger and Data Acquisition at Belle II

At Belle II, the instantaneous luminosity is higher than that of Belle, so a higher background is expected. To deal with this higher background, an efficient trigger system is required. The Belle II trigger system consists of a hardware-level trigger (L1 trigger) [48] and a software-based high-level trigger (HLT) [49]. The overall

[8] At this stage, a typical data size of a $B\overline{B}$ or $q\overline{q}$ event is about 30 kB.

[9] dr and dz are the radial and longitudinal impact parameters respectively.

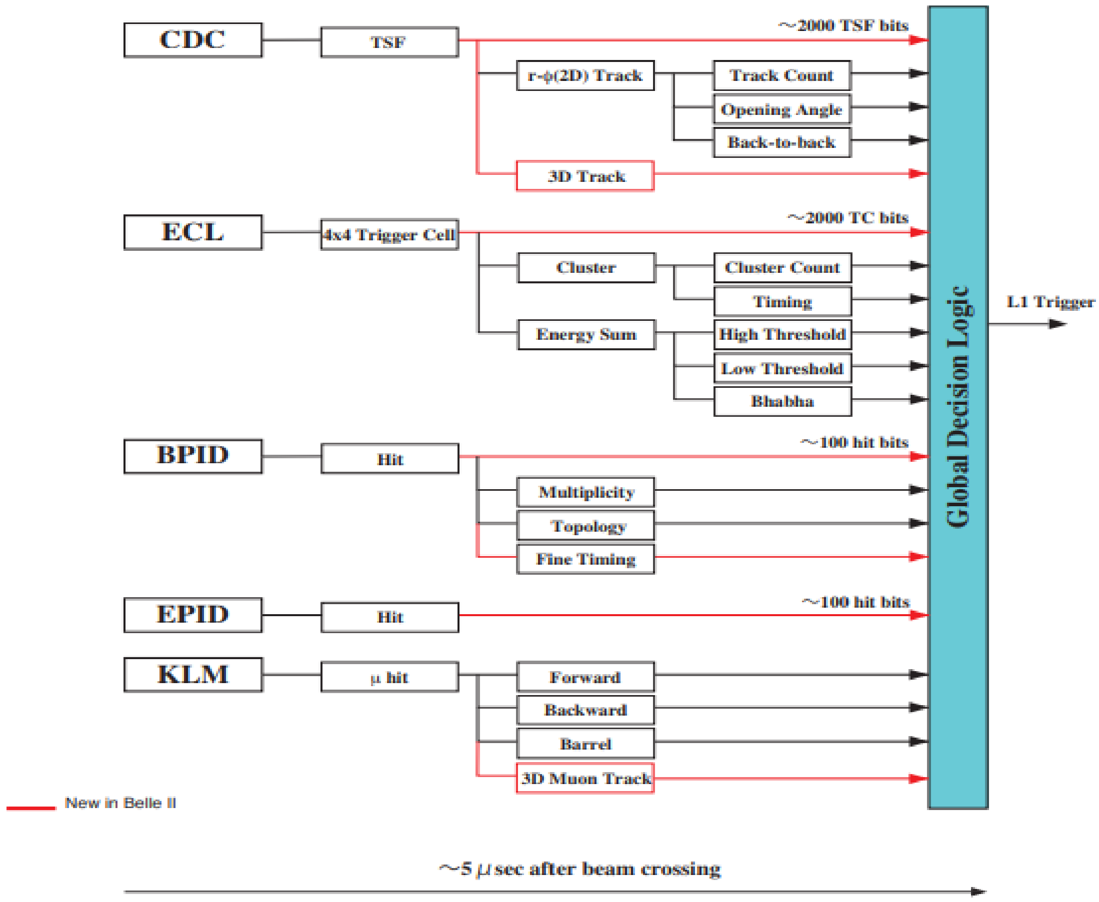


Figure 2.18: Schematic view of Belle II L1 trigger system. Credit [44]

structure of the Belle II trigger system is similar to that of the Belle trigger system, however, with the adoption of new technologies. Input to the GDL is provided by the various sub-triggers (corresponding to various sub-detectors). The GDL makes the final decision to retain or reject the event in the L1 trigger system. In the next sub-section we review the L1 trigger in more detail.

2.5.1 L1 Trigger

The schematic view of the Belle II L1 trigger system is shown in Fig. 2.18. It particularly focuses on rejecting the beam induced background, two photon background, and radiative Bhabha background events. It retain 99.9% of the produced $B\overline{B}$ events. It is designed to provide data output at a maximum rate of 30 kHz. The sub-trigger in CDC uses the track segment finding method to provide information about 2D and 3D charged tracks. The triggers used in ECL select events based on the energy deposited information from ECL clusters. The barrel and end-cap PID (BPID+EPID) sub-triggers provide the precise timing information to GDL. The data related to the muon tracks is provided by the KLM sub-trigger.

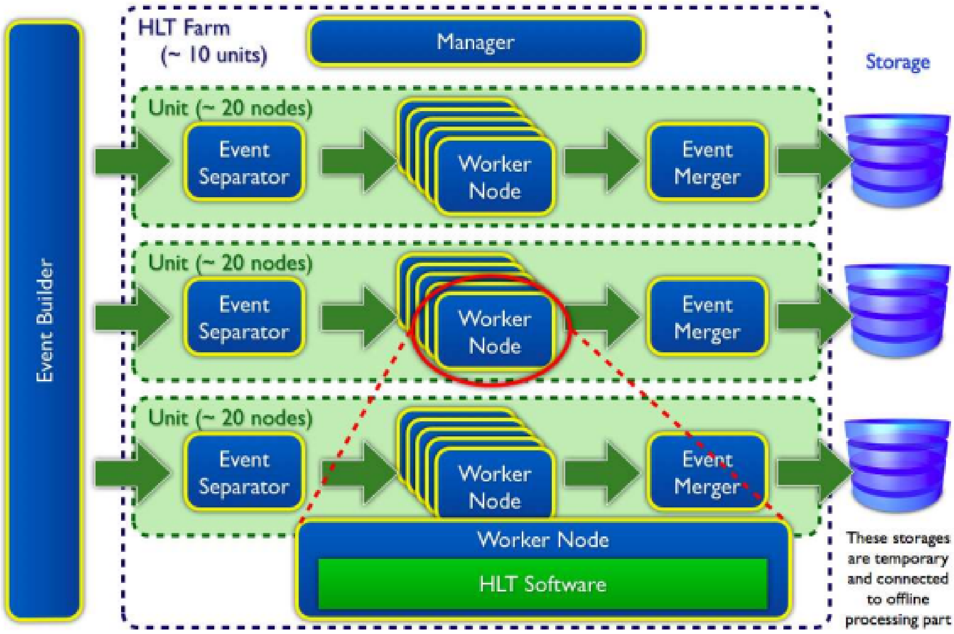


Figure 2.19: Schematic view of Belle II HLT system. Credit [49]

2.5.2 HLT

It is a software level trigger designed to reduce event rates to 10 kHz. Its schematic view is shown in Fig. 2.19. The HLT system consists of ten units in total and there are twenty nodes in each unit. The nodes are further divided into three categories; event separator, worker node, and event merger. The distribution of the data over the network to the worker nodes is performed by the event separator. The actual data reduction is performed by the worker nodes. The processed data are collected by the event merger. A special node (manager node) manages and monitors the entire HLT system. A software framework[10] works on the worker nodes to do the event processing.

2.5.3 Data Acquisition System

In Belle II data acquisition system, there are two data streams. One stream of data (first stream) comes from the PXD, and the other stream (second stream) of data comes from all the other sub detectors. The first stream of data is processed by the PXD readout processor while the second stream of data is processed by the HLT system.

To reduce the data rate in the first stream, information about the events that are reconstructed by the HLT is provided to the PXD readout processor. The PXD readout processor only accepts those hits that are associated with the tracks reconstructed by HLT. The schematic view of the Belle II data acquisition system is shown in Fig. 2.20.

[10]This software framework is called Basf2 which is described later in this chapter.

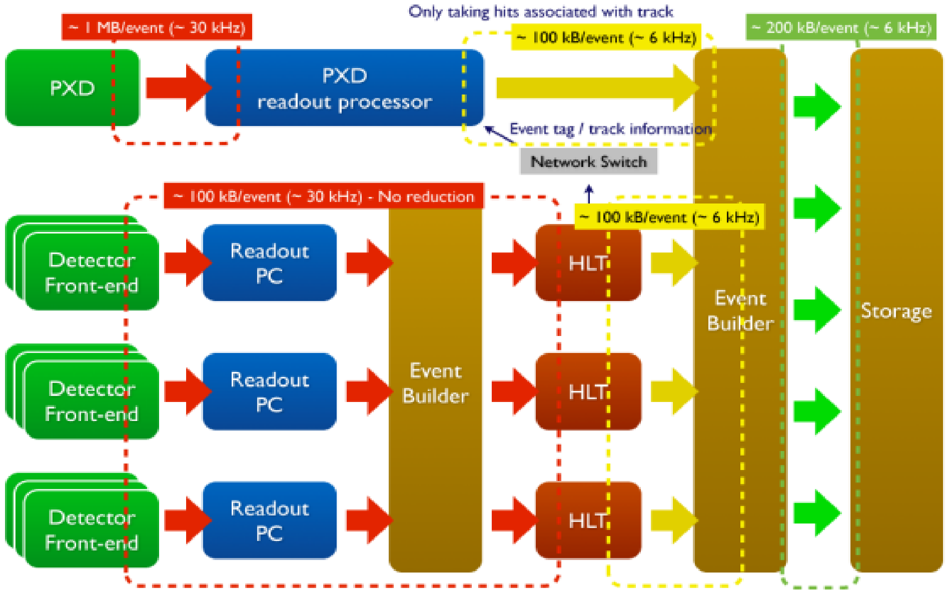


Figure 2.20: Schematic view of Belle II DAQ system. Credit [49]

2.6 Belle Luminosity

The duration of Belle data taking was from 1 June 1999 to 30 June 2010. It reached a total integrated luminosity [11] of 1040 fb^{-1}. Fig. 2.21 shows the integrated luminosity accumulated over the data taking years. Most of the collected data are at the $\Upsilon(4S)$ resonance energy. For 10% of the running time, off-resonance (60 MeV below resonance peak) data was also collected to determine the non $B\overline{B}$ background. The summary of the data collected, the types of resonance energy, and the corresponding resonances are given in Table 2.2. It is worth mentioning that the data in Table 2.2 is after removing the bad runs which cannot be used in physics analysis due to serious detector problems.

Resonance	On-peak luminosity fb^{-1}	Off-peak luminosity fb^{-1}	Number of resonances (10^6)
$\Upsilon(1S)$	5.7	1.8	102
$\Upsilon(2S)$	24.9	1.7	158
$\Upsilon(3S)$	2.9	0.25	11
$\Upsilon(4S)$ SVD1	140	15.6	152 $B\overline{B}$
$\Upsilon(4S)$ SVD2	571	73.8	620 $B\overline{B}$
$\Upsilon(5S)$	121.4	1.7	7.1 $B_s\overline{B}_s$
Scan		27.6	

Table 2.2: Summary of the different types of data collected at Belle. [25]

[11] measure of the total number of potential collisions.

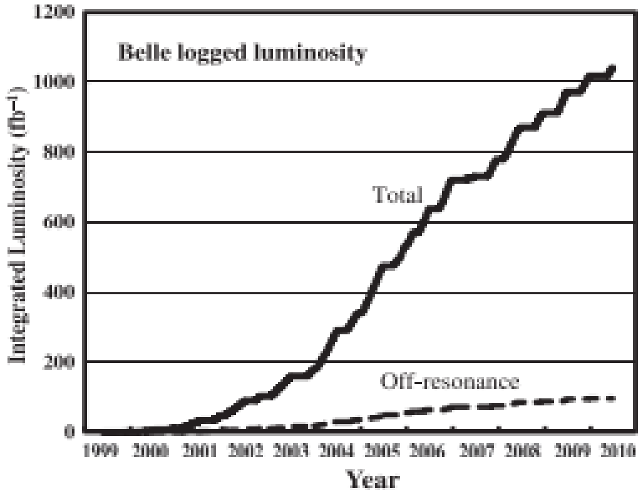


Figure 2.21: Belle integrated luminosity. Credit [25]

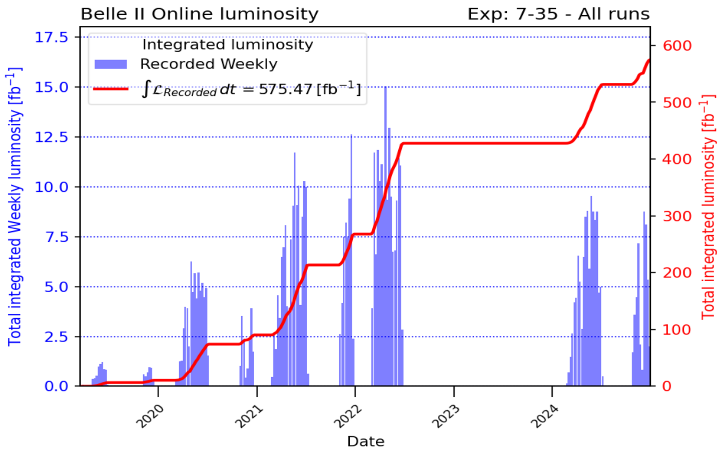


Figure 2.22: Belle II integrated luminosity. Credit (Belle II official web)

2.7 Belle II Luminosity

The Belle II experiment started data taking in 2019, and so far it has accumulated over 500 fb^{-1} integrated luminosity. The integrated luminosity of Belle II is shown in Fig. 2.22. During December 2024, it achieved a peak instantaneous luminosity of 5.1×10^{34} cm^{-2} s^{-1}, which is the current world record.

For the near future, the target is to boost the instantaneous luminosity to 2×10^{35} cm^{-2} s^{-1} and to achieve the target peak luminosity of 8×10^{35} cm^{-2} s^{-1}, an upgrade program is in progress. Over the years, Belle II has an ambitious plan to collect around 50 ab^{-1} data[12]. The luminosity projection for the future is shown in Fig. 2.23. With more data acquisition, more interesting physics results are expected.

[12] for the near future; the expected Belle II collected data will be a few times the total Belle data set.

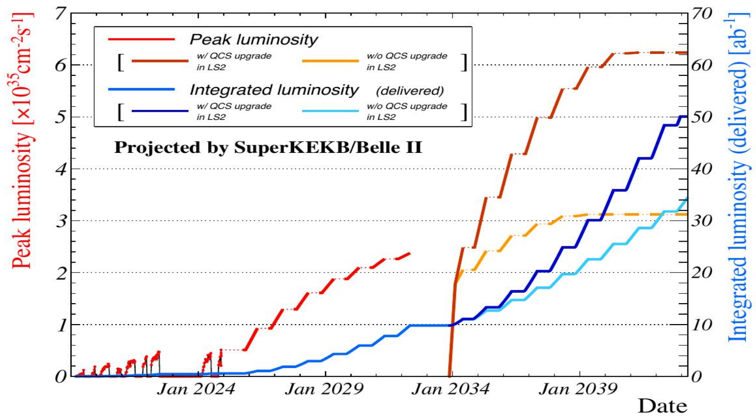


Figure 2.23: Belle II luminoisty projection. Credit (Belle II official web)

2.8 Basf2 Reconstruction Software

Basf2 (Belle II analysis software framework) is used to analyze both Belle and Belle II data. Basf2 is the upgraded version of Basf (Belle analysis software) which was used to analyze the Belle data. It is specifically made for the offline and online data handling of Belle/Belle II. The Basf2 core is written in C++ and provides the python interface. It contains several modules for different purposes. Basf2 is described in more detail in [50]. In this analysis, the Basf2 b2bii module is used, which is overviewed in the next subsection.

2.8.1 b2bii

To analyze the Belle data in the Basf2 framework, the b2bii (Belle to Belle II) module is used [51]. It converts the Belle mDST (mini DST) files from PANTHER [13] format to the ROOT format. After this format conversion, all the modules of Basf2 can be used, and similar reconstruction scripts can be used for both Belle and Belle II data analysis. This tool is an essential part of the analyses that analyze both Belle and Belle II data.

2.9 Tagging Methods

In the B factories, $B\overline{B}$ pairs are produced without additional particles[14]. In the analysis of physics processes (that requires tagging), one B meson is termed the signal B ($B_{\rm sig}$) and the other the tag B ($B_{\rm tag}$). A $B_{\rm sig}$ candidate is formed by combining the track candidates for signal side particles (in this case K, τ, ℓ). There

[13] A custom serialization format which consist of tables compressed by the zlib library. The table consists of multiple rows, called entries. The index of each entry can be used to relate entries to other entries.

[14] In addition to it, the collision energy is known and the detector encloses the IP hermetically.

are various methods to reconstruct the B_{tag}, and the selection of a particular method depends on the objectives of the analysis. The two main tag techniques are the exclusive tag and the inclusive tag, which are discussed briefly in the following subsections, while a detailed discussion on tag approaches can be found in [52].

2.9.1 Exclusive Tag

In the exclusive tag method, B_{tag} is reconstructed in several specific decay modes. Based on the choice of the decays that are specifically used in this reconstruction, it can be either an exclusive hadronic tag (only using the hadronic B decay modes) or an exclusive semileptonic tag (only using the semileptonic B decay modes).

2.9.2 Inclusive Tag

In this tag method, all detected particles (except those assigned to B_{sig}) are used to reconstruct B_{tag} candidate. This B_{tag} candidate is further subjected to different selection criteria to reject the background it contains. Based on the requirement of the number of leptons in this B_{tag} candidate, it can be inclusive hadronic tag (no lepton) or inclusive semileptonic (one lepton).

There are various pros and cons of all the tag methods. In the hadronic tag, the kinematics of B_{tag} are well known, so B_{tag} is pure; it is not very efficient $\sim 0.1\%$. The semileptonic tag is more efficient in reconstructing semileptonic B decays. However, because of the presence of a missing neutrino, it lacks kinematics information, which leads to a less pure B_{tag} candidate. In the case of an inclusive tag, the tag efficiency is one order of magnitude higher than that of a hadronic tag and a semileptonic tag at the cost of a high background.

In this analysis, the inclusive semileptonic tag approach is used, which is more efficient; however, its implementation is more challenging compared to the other tag methods. The next chapter is based on the analysis framework, in which it will be further discussed in more detail.

Chapter 3

Analysis Framework

"Truth is ever to be found in simplicity,
and not in the multiplicity and confusion
of things."
Isaac Newton

In this analysis, a blind analysis strategy [53] is adopted to search for the decay $B^+ \to K^+\tau^\pm(\to \pi^\pm\nu_\tau)\ell^\mp$ using Belle data, with direct applicability to Belle II.

The key challenge of this channel arises from the presence of a neutrino in the τ decay, which prevents full kinematic reconstruction using only visible final-state particles. Instead of relying on explicit hadronic tagging, a dedicated kinematic reconstruction framework is developed that exploits the precisely known initial-state conditions at B factories operating at the $\Upsilon(4S)$ resonance.

The central idea of this method is that the signal-side B meson can be reconstructed using only kinematic constraints, without requiring full reconstruction of B_{tag}. This significantly improves signal efficiency while retaining strong background discrimination. The method is particularly suited for decays with a single missing neutrino and provides a two-fold solution for the neutrino momentum, fully determined by mass-shell constraints.

3.1 Theoretical Framework

At B factories, $B\overline{B}$ pairs are produced through $\Upsilon(4S)$ decays under well-defined kinematic conditions. In the center-of-mass frame, the total four-momentum of the system is precisely known, and the two B mesons are produced nearly at rest and back-to-back. This provides powerful global constraints for decays involving invisible particles.

The signal decay considered in this analysis is

$$B^+ \to K^+\tau^\pm(\to \pi^\pm\nu_\tau)\ell^\mp.$$ [1]

[1] Based on the charge on the lighter lepton, we divide the signal modes into two general categories

Due to the undetected neutrino from the τ decay[2], the event cannot be fully reconstructed from visible particles alone. However, the neutrino momentum is not arbitrary. It is constrained simultaneously by:

- the τ mass-shell condition applied to the (π, ν) system,
- the B mass-shell condition applied to the $(K\ell\pi, \nu)$ system.

Each constraint defines a cone of allowed neutrino momentum directions in momentum space, as illustrated in Fig. 3.1. The intersection of these two cones determines the neutrino momentum up to a two-fold ambiguity. Consequently, two possible solutions for the signal-side B momentum are obtained.

This geometric interpretation leads naturally to the definition of a discriminator variable, $\sin\varphi$, which quantifies the consistency of the two mass constraints. For physically consistent signal events, a geometrical intersection exists and $\sin\varphi$ lies within the physical interval $[-1, 1]$. Background combinations frequently violate the simultaneous constraints, leading to non-physical values.

The explicit form of $\sin\varphi$ is given by[3]

$$\sin\varphi = \frac{\cos(p_\nu, p_\pi)\cos(p_\pi, p_{K\ell\pi}) - \cos(p_\nu, p_{K\ell\pi})}{\sin(p_\nu, p_\pi)\sin(p_\pi, p_{K\ell\pi})}. \tag{3.1}$$

Thus, the method provides both:

- a kinematically constrained reconstruction of B_{sig},
- and a powerful background rejection variable derived from geometric consistency.

After obtaining the two possible B_{sig} momentum solutions, the B_{tag} side is reconstructed using an inclusive semileptonic tagging approach. Unlike full hadronic tagging, this strategy avoids the severe efficiency penalty associated with exclusive reconstruction while still providing global event constraints.

The combination of signal-side kinematic reconstruction with inclusive semileptonic tagging maximizes overall sensitivity for final states with missing energy.

This complete schematic reconstruction is shown in Fig. 3.2.

The back-to-back topology of B_{sig} and B_{tag} in the center-of-mass frame provides an additional global consistency check. For each of the two signal-side momentum solutions, the angle with respect to the visible tag momentum is calculated.

The variable

$$\Delta\cos\theta = \min\{|\cos\theta_1 + \cos\theta_{\text{tag}}|, |\cos\theta_2 + \cos\theta_{\text{tag}}|\} \tag{3.2}$$

that are same sign (SS) modes (B and ℓ have the same charge) and opposite sign (OS) modes (B and ℓ have the opposite charge). The SS modes contain $B^+ \to K^+\tau^-(\to \pi^-\nu_\tau)\ell^+$ modes while the OS modes contain $B^+ \to K^+\tau^+(\to \pi^+\overline{\nu}_\tau)\ell^-$ modes.

[2] In the Belle/Belle II detectors, neutrinos are not detectable, however, because of the very well-known initial conditions of the B mesons, decays with neutrinos can be partially reconstructed

[3] Its complete derivation is provided in the appendix.

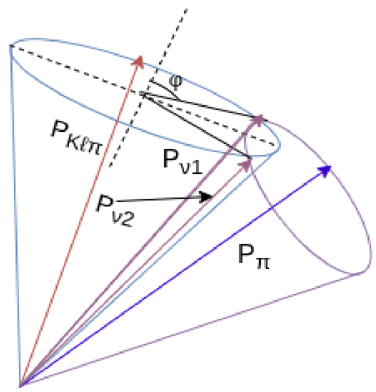


Figure 3.1: The illustration of neutrino cones.

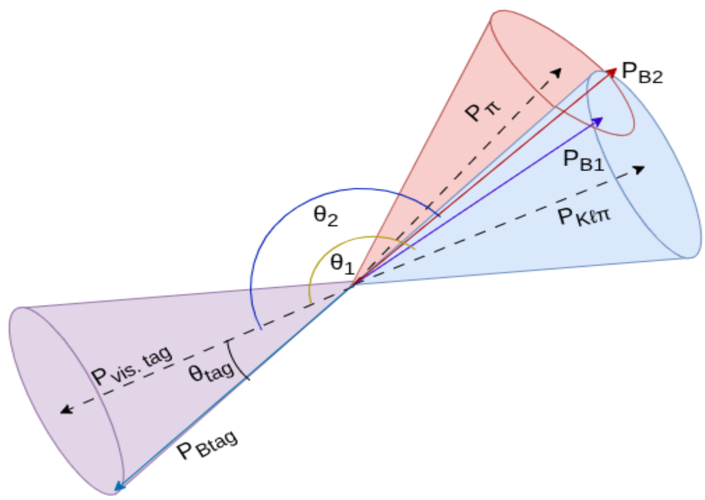


Figure 3.2: Schematic reconstruction of signal and tag side.

selects the solution most compatible with the expected back-to-back configuration. For signal events, this variable peaks near zero[4].

The reconstruction framework presented here is general and can be applied to a broad class of semileptonic or rare decays involving a single invisible particle, making it particularly relevant for future Belle II analyses with large datasets.

Before implementing this framework on the events, first in the next section, we will discuss our signal and background MC generation strategy.

3.2 Monte Carlo Samples

The reconstruction method developed in this analysis relies critically on precise kinematic consistency conditions. It is therefore essential to ensure that the Monte Carlo (MC) samples used for validation do not introduce artificial biases in angular correlations or phase-space distributions that could affect the cone-intersection solutions.

At the time of this study, Belle provided the statistically dominant dataset, and all validation studies are performed using Belle MC and Belle data. The reconstruction framework is fully transferable to Belle II.

Signal MC Generation

Signal MC samples are generated using `EvtGen` [54]. For each decay mode $B^+ \to K^+\tau^\pm(\to \pi^\pm\nu_\tau)\ell^\mp$, 1.0×10^6 events are produced.

The generation proceeds in three steps:

Step 1: $\Upsilon(4S)$ production. $\Upsilon(4S)$ events are simulated using the VSS model, describing the decay of a vector resonance into two scalar B mesons. Half of the generated events contain a signal decay on the B^+ side and generic decay on the B^- side, while in the other half the roles are reversed. This symmetric treatment ensures identical acceptance and detector effects for both charge configurations.

Step 2: Signal decay model. The decay $B^+ \to K^+\tau^\pm\ell^\mp$ is generated using the PHSP model combined with PHOTOS for final-state radiation. The PHSP model provides a uniform population of the available three-body phase space, ensuring that the kinematic reconstruction method is validated over the entire allowed region without model-dependent assumptions about dynamics.

Since the analysis strategy is based purely on kinematic mass constraints and not on angular distributions predicted by specific new-physics operators, the use of a phase-space model does not bias the geometrical reconstruction procedure.

Step 3: τ decay. The decay $\tau \to \pi\nu_\tau$ is simulated using the TAUSCALARNU model. This preserves the correct two-body kinematics of the τ decay and ensures an accurate description of the neutrino–pion angular correlation, which directly enters the cone-intersection method[5].

[4] This distribution is shown later in this chapter, where this behavior can be observed.

[5] More details on the models mentioned in these steps can be found in [55].

Following similar steps, we generated 1.0×10^6, $B^+ \to K^+\tau^\pm(\to \rho^\pm\nu_\tau)\ell^\mp$ and 4.0×10^6, $B^+ \to K^+\tau^\pm(\to \text{generic})\ell^\mp$ samples for each signal decay mode.

Detector Simulation

After event generation, detector response is simulated using the GEANT3 package [56]. This includes tracking efficiency, material interactions, photon reconstruction, and resolution effects. The reconstructed events are stored in mdst format, identical to the format used for data.

Generic Background MC

To study background contributions, centrally produced generic MC samples are used. These include:

- generic $B\overline{B}$ events (B^+B^- and $B^0\overline{B}^0$),
- continuum $e^+e^- \to q\overline{q}$ events with $q = u, d, s, c$.

In these samples, known B-meson decay modes are generated with `EvtGen` [54], while poorly measured or unknown decays and continuum processes are modeled using `PYTHIA` [57]. Final-state radiation is simulated with `PHOTOS` [58], and the detector response is modeled using `GEANT3` [56]. Each MC stream corresponds to the full Belle integrated luminosity ($\mathcal{L}_{\text{int}} = 711\ \text{fb}^{-1}$).

The background samples are particularly important for validating the behavior of the $\sin\varphi$ variable. Since non-signal combinations generally violate the simultaneous mass-shell constraints, they frequently produce non-physical cone intersections. The generic MC therefore provides a direct test of the discriminating power of the geometrical reconstruction.

3.3 Event Selection Strategy

The goal of the event selection is to maximize the sensitivity to signal channels while preserving the geometrical consistency required by the kinematic reconstruction framework.

The selection procedure is designed in three stages:

- removal of poorly reconstructed tracks and obvious combinatorial background,
- particle identification optimized to suppress dominant physics backgrounds,
- control of neutral energy to reduce fake neutrino signatures.

All selection requirements are chosen to balance background rejection and signal efficiency, avoiding overly aggressive cuts that could distort kinematic distributions entering the cone-intersection method.

3.4 Particles Selection

The first step in reconstructing the decay is the selection of the particles involved in the decay. For this purpose, various initial selections are made to reject as much background as possible on this level. In the next subsection, we will summarize the selections of charged particles.

3.4.1 Charged Track Preselection

All charged tracks are required to originate from the interaction point (IP), with transverse and longitudinal impact parameters satisfying $|dr| < 1.0$ cm and $|dz| < 4.0$ cm.

These requirements suppress tracks from secondary interactions, beam-related background, and long-lived particle decays, which would otherwise spoil the vertex reconstruction and bias the χ^2-based candidate ranking.

In addition, tracks are required to have transverse momentum $p_T > 500$ MeV/c in the laboratory frame. This removes low-momentum tracks with poor reconstruction quality and large multiple-scattering uncertainties, which would degrade the precision of the kinematic constraints.

The next step involved the identification of particles based on the different identification likelihood ratios. In the following subsections, we briefly describe the identification of different charged particles.

3.4.2 Particle Identification

Particle identification (PID) plays a crucial role in suppressing dominant background processes such as semileptonic B decays and continuum events. The likelihood-ratio approach combines information from multiple sub-detectors and provides efficient separation between leptons and hadrons.

$e^{\pm}$ Identification:

Electrons are identified using combined information from CDC, ECL, and ACC sub-detectors. The likelihood ratio

$$\mathcal{R}_e = \frac{\mathcal{L}_e}{\mathcal{L}_e + \mathcal{L}_{\bar{e}}} \tag{3.3}$$

is required to satisfy $\mathcal{R}_e > 0.6$. This requirement maintains high signal efficiency while significantly suppressing hadron misidentification, which constitutes a major source of combinatorial background.

$\mu^{\pm}$ Identification

Muon identification relies primarily on KLM information. The likelihood ratio

$$\mathcal{R}_\mu = \frac{\mathcal{L}_\mu}{\mathcal{L}_\mu + \mathcal{L}_\pi + \mathcal{L}_K} \quad (3.4)$$

is required to satisfy $\mathcal{R}_\mu > 0.6$. This suppresses pion and kaon contamination while retaining high muon identification efficiency.

Charged Hadrons Identification

Charged hadrons are separated using combined CDC, TOF, and ACC information. For kaons and pions, the likelihood ratio

$$\mathcal{R}_{K|\pi} = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi} \quad (3.5)$$

is used. Kaon candidates are required to satisfy $\mathcal{R}_{K|\pi} > 0.6$, while pion candidates must satisfy $\mathcal{R}_{K|\pi} < 0.6$.

These requirements suppress cross-feed between kaon and pion hypotheses, which would otherwise broaden the m_{ROE} (mass of the hadronic system on tag side) distribution, add charm decay background contamination and weaken the tagging performance.

3.4.3 Photons Selection

Photon selection is optimized to suppress fake missing-energy signatures arising from beam background and cluster splitting. A multivariate classifier is used to distinguish genuine electromagnetic showers from background clusters [59]. In this approach, a multivariate analysis (MVA) classifier is developed using the shower related variables to distinguish between the good photon candidates and the background photon candidates. It provides two variables (fake photon suppression score (f_{ps}) and beam background photon suppression score (b_{bs})) to select the best photons.

Photon selection is optimized to suppress fake missing-energy signatures arising from beam background and cluster splitting. A multivariate classifier is used to distinguish genuine electromagnetic showers from background clusters.

Two MVA scores are required:

$$f_{ps} > 0.3, \quad b_{bs} > 0.3.$$

In addition to these MVA based selections, we have also required the goodBelleGamma =1 i.e. the energy of the photons in the barrel region is greater than 50 MeV, in the forward end cap photon energy is required to be greater than 100 MeV, and in the backward end cap the energy of photons is required to be greater than 150 MeV. This is particularly important for preventing artificial enhancement of missing momentum, which would bias the neutrino reconstruction.

3.4.4 π^0 Selection

The π^0 candidates are reconstructed from photon pairs with invariant mass in the range $115 < m_{\gamma\gamma} < 155$ MeV$/c^2$.

This requirement suppresses combinatorial photon pairs while preserving genuine neutral pion candidates, which are relevant for the correct modeling of the rest-of-event system on the tag side.

The final selection criteria are summarized in Table 3.1. The chosen working points provide an optimal compromise between signal efficiency and background suppression, while preserving the kinematic distributions entering the cone-intersection reconstruction.

Name	Selections
charged tracks	$\|dr\| <$ 1 cm, $\|dz\| <$ 4 cm, p > 500 MeV/c
$e^{\pm}$	$\mathcal{R}_e > 0.6, \mathcal{R}_\mu < 0.98, \mathcal{R}_{\pi\|e} < 0.98$
$\mu^{\pm}$	$\mathcal{R}_\mu > 0.6, \mathcal{R}_e < 0.98, \mathcal{R}_{\pi\|\mu} < 0.98$
$K^{\pm}$	$\mathcal{R}_\mu < 0.98, \mathcal{R}_e < 0.98, \mathcal{R}_{K\|\pi} > 0.6$
$\pi^{\pm}$	$\mathcal{R}_e < 0.6, \mathcal{R}_\mu < 0.6, \mathcal{R}_{K\|\pi} < 0.6$
γ	$f_{ps} > 0.3$, $b_{bs} > 0.3$, goodBelleGamma $= 1$
π^0	$115 < m_{\gamma\gamma} < 155$ (MeV/c^2)

Table 3.1: Summary of all the selections for the particles.

This concludes the discussion on the particle selections, and in the next section, we will discuss our decay reconstruction.

3.5 Decay Reconstruction

As the $\Upsilon(4S)$ decays to a pair of B mesons, a complete decay reconstruction requires reconstructing B_{sig} and B_{tag}. After the reconstructions of B_{sig} and B_{tag}, they are combined to form an $\Upsilon(4S)$ candidate. In the next subsections, we discuss the reconstruction of B_{sig} first, followed by the reconstruction of B_{tag} and the reconstruction of $\Upsilon(4S)$ candidate.

3.5.1 B_{sig} Reconstruction

The B_{sig} decays to a K^+, $\tau^{\pm}$, and $\ell^{\mp}$, i.e.

$$B^+ \to K^+\tau^{\pm}(\to \pi^{\pm}\nu_\tau)\ell^{\mp}$$

To reconstruct the B_{sig}[6], first we partially reconstruct the τ candidate by selecting a charged π track. Next, we combine the two more charged tracks (K, ℓ) with the

[6]Due to the presence of τ, we have a missing neutrino which leads to the application of mechanism described earlier and we get the p_{Bsig} with two folds ambiguity.

π track to form a B_{sig} candidate. In the formation of the B_{sig} candidate, all three charged tracks are required to originate from the common vertex (IP).

The distributions of sinφ are shown in Fig. 3.3 (SS modes) and Fig. 3.4 (OS modes)[7] showing signal peaking at 1.0 and all events that are outside this physical range (-1.0 to 1.0) are background events and are rejected requiring the selection $-1.0 \leq \sin\varphi \leq 1.0$.

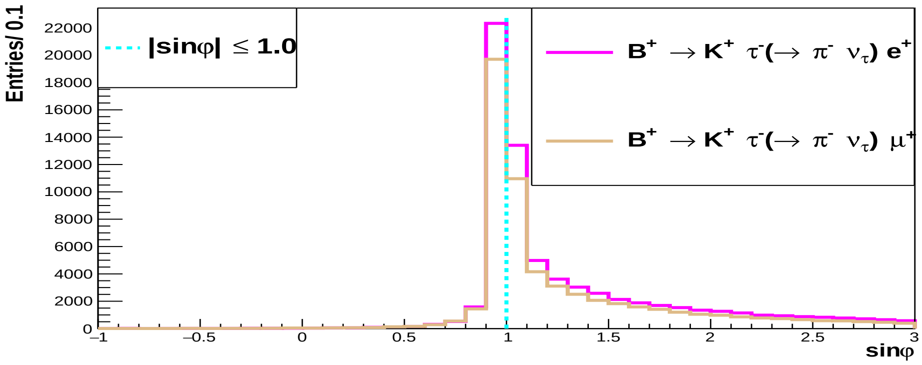


Figure 3.3: The distribution of sinφ for SS decay modes.

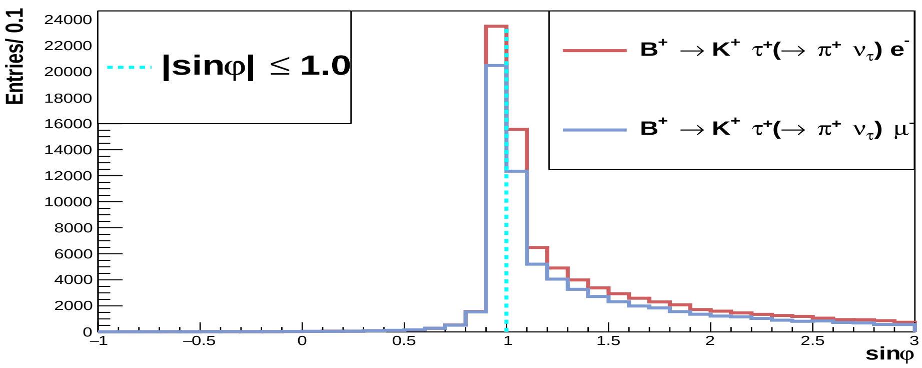


Figure 3.4: The distribution of sinφ for OS decay modes.

3.5.2 B_{tag} Reconstruction

After reconstructing the B_{sig} candidate, we combine all the remaining tracks and clusters (except one prompt lepton e^- or μ^-) to form a rest of event (ROE) candidate. In the next step, this ROE candidate is combined with the prompt lepton to form the B_{tag} candidate i.e.

$$B^- \to \text{ROE} + \ell^- \ , \ (\ell = e, \mu)$$

[7] All the signal mode distributions in this chapter are for 1.0×10^6, $B^+ \to K^+\tau^\pm(\to \pi^\pm\nu_\tau)\ell^\mp$ dedicated signal MC samples.

The ROE part constitutes the hadronic part of B_{tag}, and its invariant mass distributions are shown in Figs. 3.5 (SS modes) and 3.6 (OS modes).The invariant mass distribution of the hadronic ROE system exhibits clear D^0 and D^{*0} structures, validating the semileptonic tagging hypothesis.

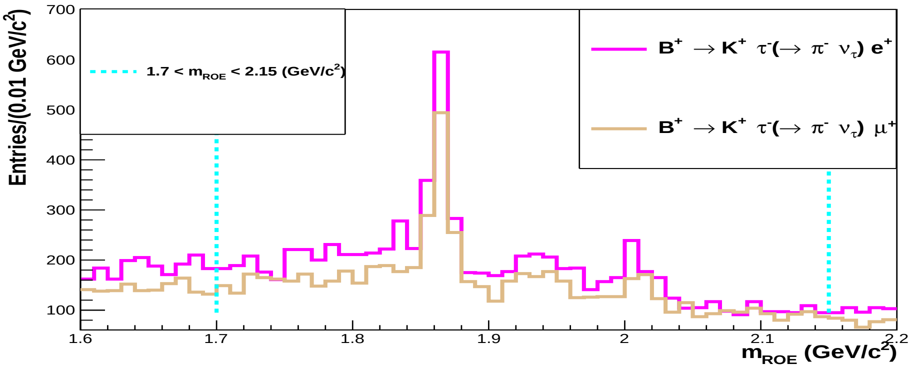


Figure 3.5: The distribution of the invariant mass of the hadronic part of ROE for SS modes. The peak on the left side shows the D^0 peak, the peak on the right side shows the D^{*0} peak, and the peak between them is the D^{*0} with a missing particle (π^0, γ). To reject the lower and higher resonances we select only the $1.7 < m_{\text{ROE}} < 2.15$ GeV/c^2.

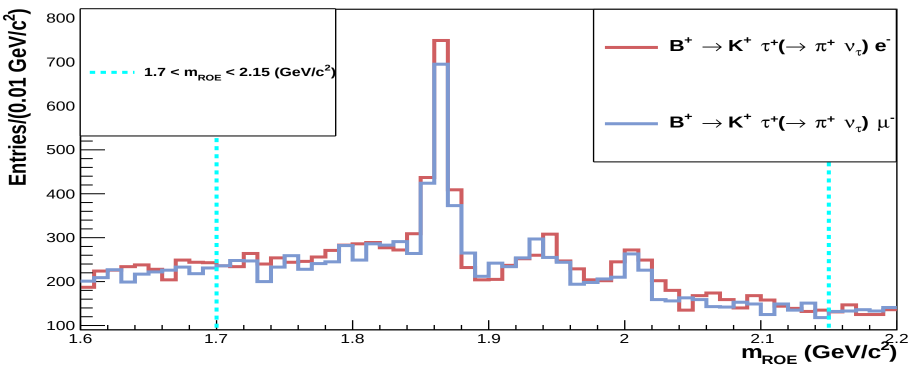


Figure 3.6: The distribution of m_{ROE} for OS modes.

As we also have a missing neutrino on the tag side, a cosine angle between the visible momentum on the tag side ($\overrightarrow{p}^{\text{tag}}_{\text{vis}}$) and the B_{tag} momentum ($\overrightarrow{p}^{\text{tag}}_{B}$), assuming beam energy in the center-of-mass frame (E_B) and the B meson mass (m_B) can be defined as:

$$cos\theta_{tag} = \frac{2E_B E_{\text{vis}} - m_B^2 - m_{\text{vis}}^2}{2\left|\overrightarrow{p}^{\text{tag}}_{B}\right|\left|\overrightarrow{p}^{\text{tag}}_{\text{vis}}\right|} \tag{3.6}$$

where m_{vis} is the invariant mass of the ROE$+\ell$ system. The distributions of $cos\theta_{tag}$ are shown in Figs. 3.7 (SS modes) and 3.8 (OS modes). For the signal events, it lies in the physical region (from -1 to 1). We are making a loose selection on it (from -1.25 to $+1.25$) to also incorporate partially reconstructed events on the tag side.

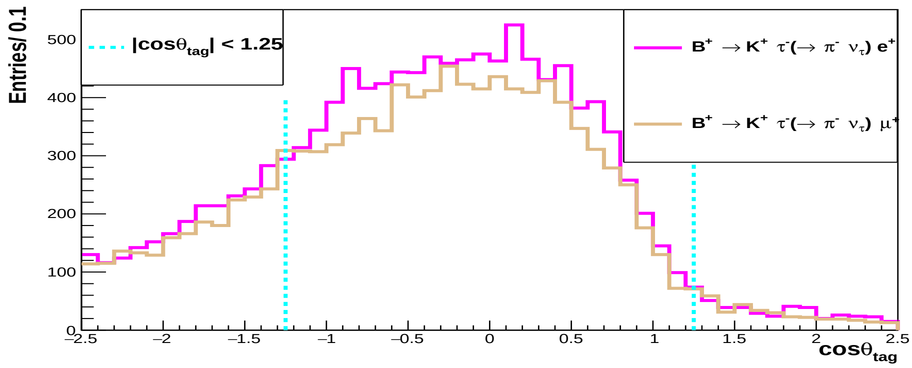


Figure 3.7: The distribution of cosθ_{tag} for SS modes.

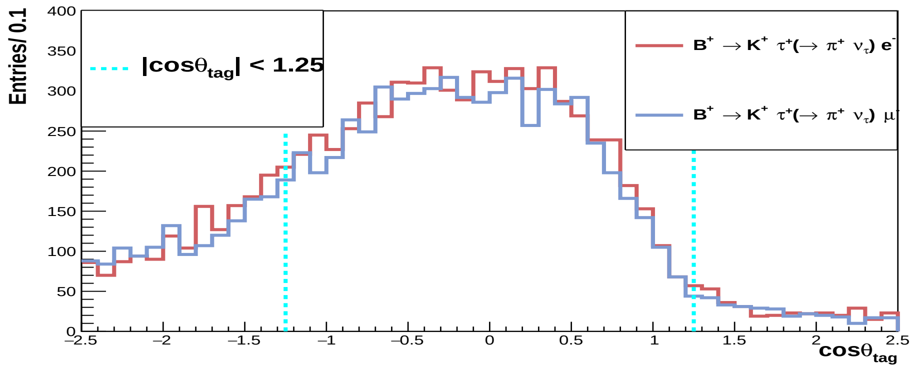


Figure 3.8: The distribution of cosθ_{tag} for OS modes.

3.5.3 $\Upsilon(4S)$ Reconstruction

Now we have the B_{sig} and B_{tag} candidates, so we combine them to form an $\Upsilon(4S)$ candidate and it completes our decay reconstruction. The distributions of the variable $\Delta\cos\theta$, which contains information from both the signal and the tag side, are shown in Figs. 3.9 (SS modes) and 3.10 (OS modes), where we can see that the signal is peaking at zero, as ideally B_{sig} and B_{tag} should be back-to-back.

3.6 Best Candidate Selection

Due to combinatorial ambiguities, more than one B_{sig} candidate can satisfy the initial selection criteria within a single event. This naturally leads to multiple B_{tag} candidates and, consequently, several $\Upsilon(4S)$ candidates per event.

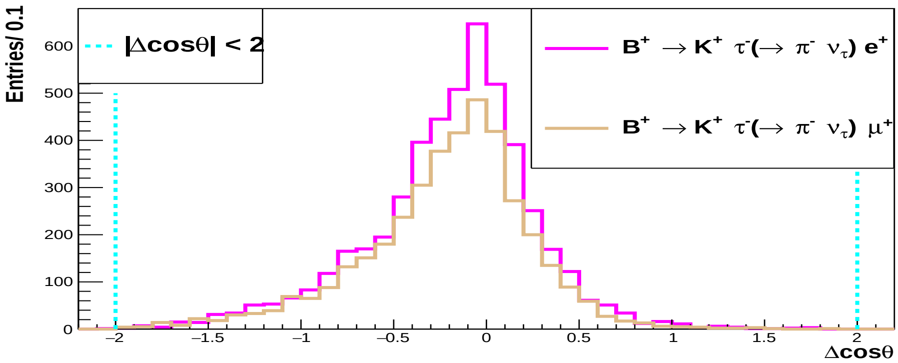


Figure 3.9: The distribution of $\Delta\cos\theta$ for SS modes.

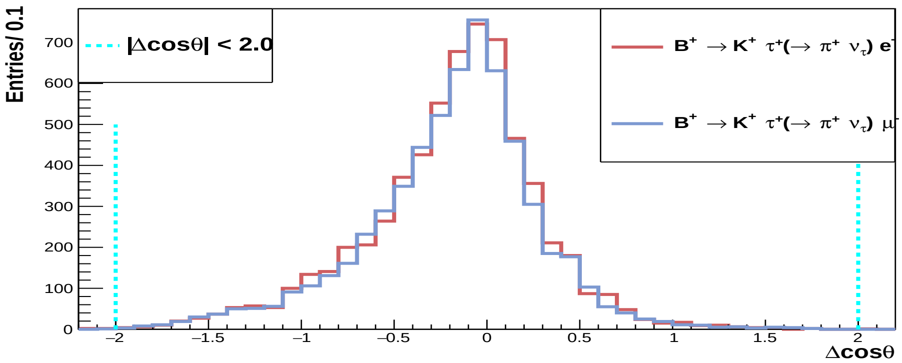


Figure 3.10: The distribution of $\Delta\cos\theta$ for OS modes.

Retaining all candidates would introduce correlations between observables and distort statistical interpretation. Therefore, a unique candidate per event is selected using a ranking procedure designed to favor the most kinematically consistent reconstruction.

Ranking Strategy

The ranking variable is defined as the product of the vertex-fit probabilities of the signal and hadronic part of tag sides:

$$\text{rank} = P_{\chi^2}(B_{\text{sig}}) \times P_{\chi^2}(\text{ROE}). \tag{3.7}$$

Here, P_{χ^2} denotes the vertex-fit probability derived from the χ^2 of the corresponding fit.

This choice is motivated by the following considerations:

- Signal candidates are expected to originate from a common vertex at the interaction point, leading to high vertex-fit probabilities.

- Incorrect track combinations typically produce poorer vertex fits.
- Using probabilities (rather than raw χ^2 values) ensures a statistically meaningful comparison between candidates.

For events with multiple $\Upsilon(4S)$ candidates, the one with the highest ranking value is retained.

The distributions of the number of candidates without any selection are shown in Figs. 3.11 (SS modes) and 3.12 (OS modes). For the signal with final selections, the number of candidates is in the range of $1.10 - 1.72$ (Tables; 4.2 - 4.5).

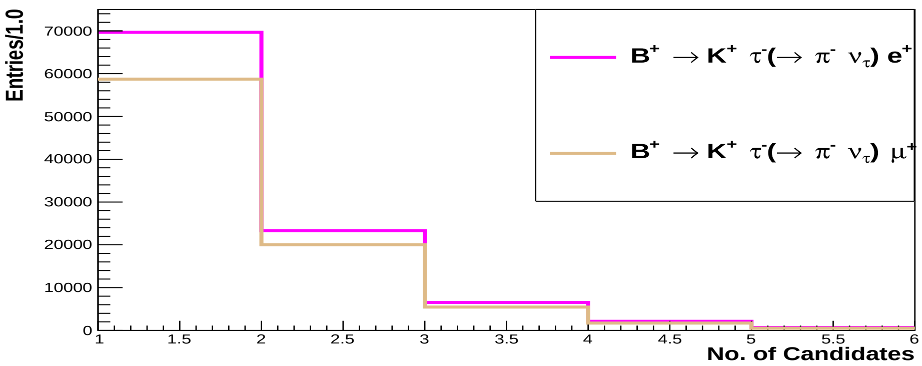


Figure 3.11: The distribution of the ranking of candidates for SS modes without any selection.

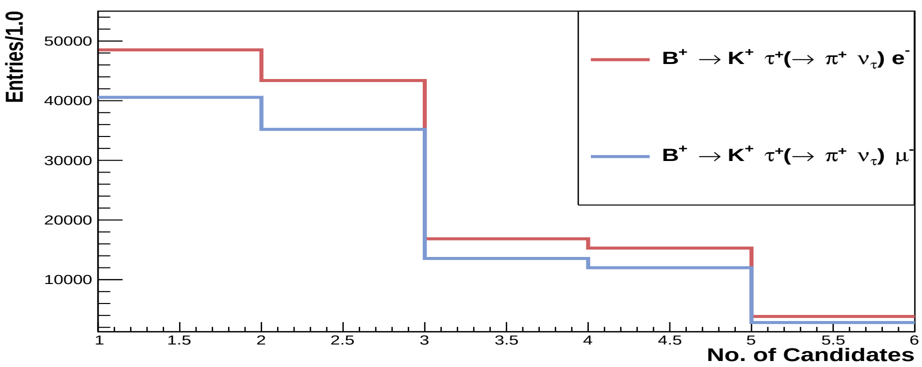


Figure 3.12: The distribution of the ranking of candidates for OS modes without any selection.

Chapter 4

Background Study

"In science there are no shortcuts
to truth."
Karl Popper

The study of different types of backgrounds in the particular analysis is an essential part of high-energy physics analyses. Although most of the background is rejected at the trigger level and the initial particle selection level, some of the background persists, and we have to make special selections to reject it. In this chapter, we discuss in detail the types of background we encounter in this analysis and the different selections to reduce them.

For the background study, we used the Belle generic MC, which was discussed in the previous chapter. The generic MC is divided into two main categories: $B\overline{B}$ and continuum ($e^+e^- \to q\overline{q}$ with $q = u, d, s, c$), which are discussed below.

4.0.1 $B\overline{B}$ MC

The $B\overline{B}$ MC consists of two classes of MC, which are the charged B (B^+B^-) MC and the mixed B ($B^0\overline{B}^0$) MC. It contains all possible decays of both B^+B^- and $B^0\overline{B}^0$. Ten streams of $B\overline{B}$ MC are available. In this analysis, we are using all ten streams of $B\overline{B}$ MC (normalized to Belle luminosity) for the background study.

4.0.2 Continuum MC

The continuum MC is basically the $e^+e^- \to q\overline{q}$[1], processes. Generally, the $c\overline{c}$ events are treated separately[2] from the $u\overline{u} + d\overline{d} + s\overline{s}$ events. There are six streams of continuum MC. The number of events in each stream of $u\overline{u}+d\overline{d}+s\overline{s}$ MC is $2.1\times N_{B\overline{B}}$ or 1621.2 millions events. The number of events in each stream of $c\overline{c}$ MC is $1.3\times N_{B\overline{B}}$

[1] q=u,d,s,c

[2] Because they lead to charmed meson production.

or 1003.6 millions events. We are using all six streams of continuum MC (normalized to Belle luminosity) for the continuum background study.

This ends the discussion on generic MC types, and in the next section, we discuss selections to minimize the different types of background.

4.1 Background Reduction

Background suppression is implemented through a sequence of selections targeting distinct background mechanisms:

- non-physical solutions of the cone-intersection reconstruction,
- semileptonic B decays with real charm resonances,
- charmonium contributions entering through misidentification or wrong combinations,
- continuum events characterized by jet-like event shapes.

Whenever applicable, working points are optimized using the Punzi figure of merit.

4.1.1 Rejection of Non-physical Events

The cone-intersection reconstruction provides a powerful consistency requirement through the variable $\sin\varphi$. For physically consistent solutions, $\sin\varphi$ must satisfy $-1 \leq \sin\varphi \leq 1$. Background combinations frequently violate the simultaneous mass-shell constraints and populate the non-physical region.

Therefore, we apply the requirement

$$|\sin\varphi| \leq 1. \tag{4.1}$$

This selection is applied to both SS and OS decay modes (as $\sin\varphi$ is independent of the underlying charge of the signal side lighter lepton ℓ). The distributions of $\sin\varphi$ are shown in Fig. 4.1 (SS modes) and Fig. 4.2 (OS modes) for all background components. In both distributions, we can see that most of the background lies outside the physical range. With selection of $|\sin\varphi| \leq 1.0$, the majority of the background is rejected from all background components[3].

4.1.2 Rejection of Semi-leptonic B Decays

Semileptonic decays such as $B \to D\ell\nu$ with $D \to K\pi(\text{x})$ can mimic the signal final state when combined with additional tracks. This contribution is visible as a pronounced D^0 structure in the relevant invariant-mass distributions.

[3]The percentage of the rejected background events is given in Table 4.3 - Table 4.5.

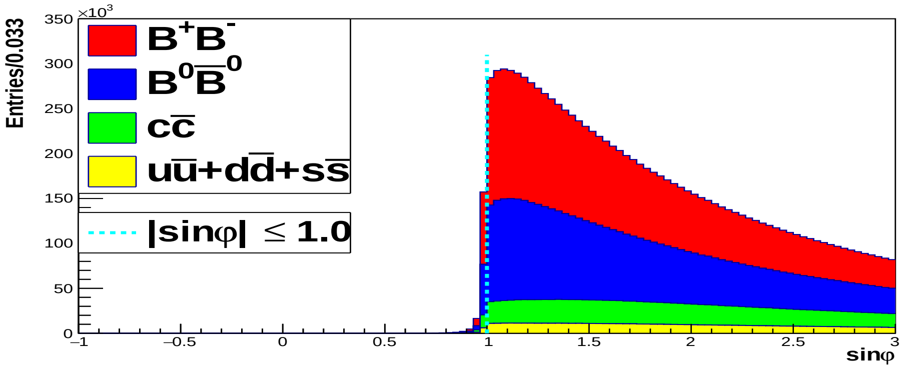


Figure 4.1: The distribution of sinφ of SS modes for all background components.

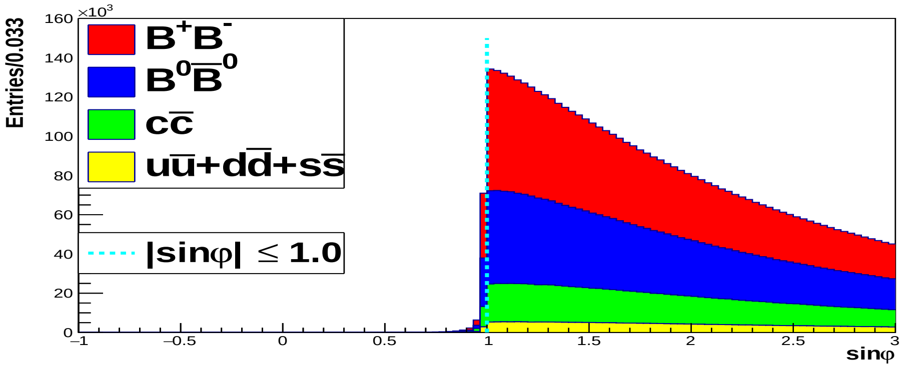


Figure 4.2: The distribution of sinφ of OS modes for all background components.

To suppress this background, we apply a charm-veto requirement: $m_{K\pi} > 1.92$ GeV/c^2 for SS modes and $m_{K\ell} > 1.91$ GeV/c^2 for OS modes.

The thresholds are optimized by maximizing the Punzi figure of merit

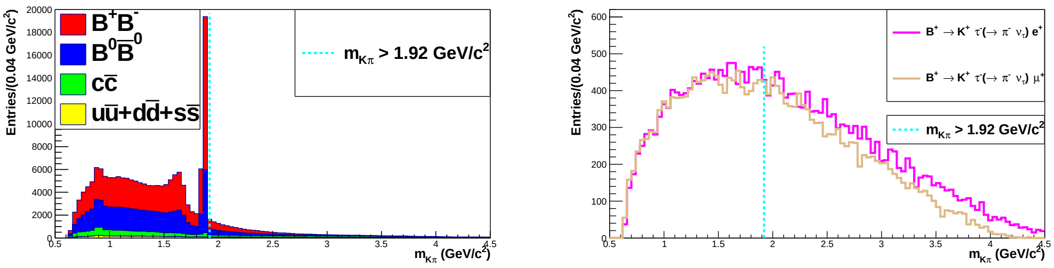
$$\mathrm{FOM}_{\mathrm{Punzi}}(t) = \frac{\epsilon_S(t)}{\frac{\alpha}{2} + \sqrt{B(t)}}, \tag{4.2}$$

with $\alpha = 3$, $\epsilon_S(t)$ is the signal reconstruction efficiency, and $B(t)$ denoting the expected number of background events after applying selection t.

The distribution of $m_{K\pi}$ is shown in Fig. 4.3, where a distinct D^0 peak can be seen in the background events. Similarly, the distribution of $m_{K\ell}$ is shown in Fig. 4.4, where again a D^0 resonance can be seen. This veto selection rejects such types of background events.

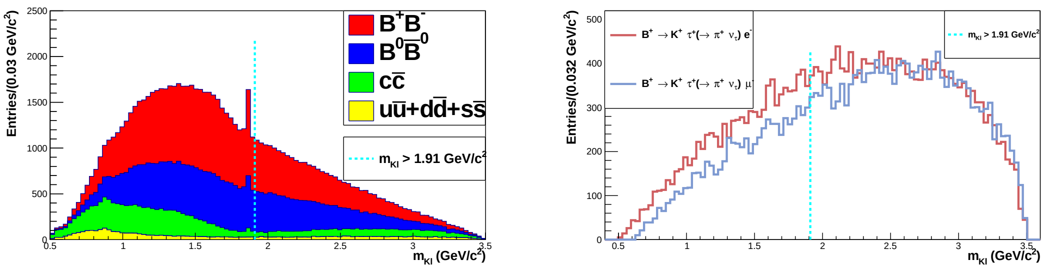
4.1.3 J/ψ and $\psi(2S)$ Vetos in $m_{\ell\pi}$ Distributions

Peaks from J/ψ and $\psi(2S)$ can appear due to misidentification or wrong combinations of tracks. We suppress these contributions by vetoing events consistent with



(a) (b)

Figure 4.3: The distributions of $m_{K\pi}$. (a) Background (b) SS signal modes.



(a) (b)

Figure 4.4: The distributions of $m_{K\ell}$. (a) Background (b) OS signal modes.

charmonium masses:

$$|m_{\ell\pi} - m_{J/\psi}| > 0.05\ \text{GeV}/c^2, \quad (4.3)$$

$$|m_{\ell\pi} - m_{\psi(2S)}| > 0.05\ \text{GeV}/c^2, \quad (4.4)$$

with $m_{J/\psi} = 3.097$ GeV$/c^2$ and $m_{\psi(2S)} = 3.686$ GeV$/c^2$. The distributions of $m_{\ell\pi}$ are shown in Fig. 4.5 (SS modes) and in Fig. 4.6 (OS modes), where these two peaks can be seen. The removal of these peaks further helps us reduce the overall number of background events.

4.1.4 J/ψ and $\psi(2S)$ Vetos in $m_{\ell^+\ell^-}$ Distributions

These background events are the result of the wrong combination of a lepton from the signal side (only in SS modes) and a lepton from the tag side with opposite charges. Such events also lead to form J/ψ and $\psi(2S)$ peaks(Fig. 4.7). The veto selections $|m_{\ell^+\ell^-} - m_{J/\psi}|\&|m_{\ell^+\ell^-} - m_{\psi(2S)}| > 0.05$ GeV/c^2 remove such events.

After all the selections mentioned in the previous subsections, we made some loose selections on a few variables to reject the obvious background events. In the following subsections, we review the list of these loose selections.

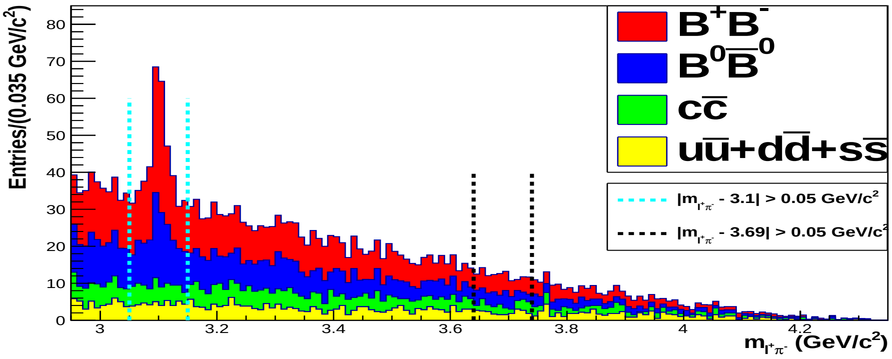


Figure 4.5: The distribution of $m_{\ell\pi}$ for SS modes.

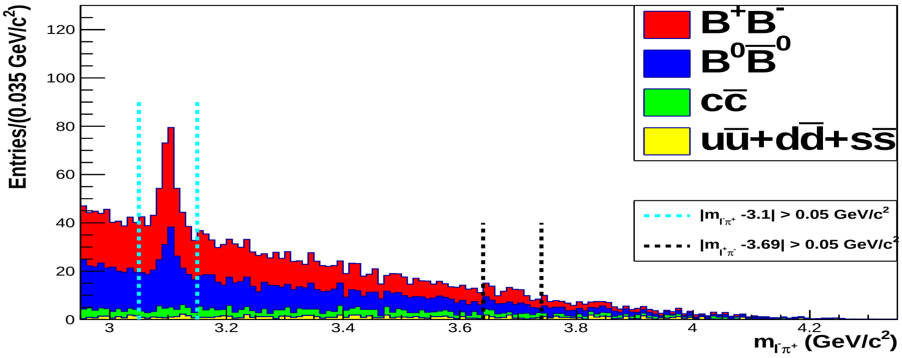


Figure 4.6: The distribution of $m_{\ell\pi}$ for OS modes.

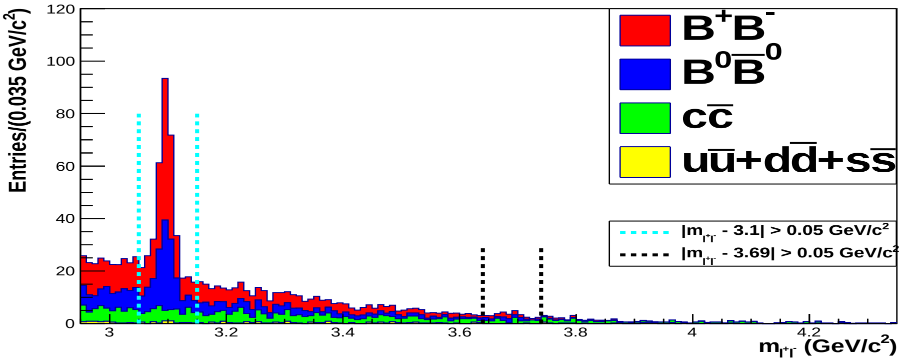


Figure 4.7: The distribution of $m_{\ell\ell}$ for SS modes.

4.1.5 Selection on the Hadronic part of B_{tag}

To select the signal enriched region we make a loose selection of $1.7 < m_{\text{ROE}} < 2.15$ GeV/c^2 on the ROE invariant mass. It reject tag-side hadronic resonance events less than 1.7 GeV/c^2 and greater than 2.15 GeV/c^2. The distributions of m_{ROE} are shown in Fig. 4.8 (SS modes) and Fig. 4.9 (OS modes). This loose selection on

m_{ROE} rejects a substantial number of background events.

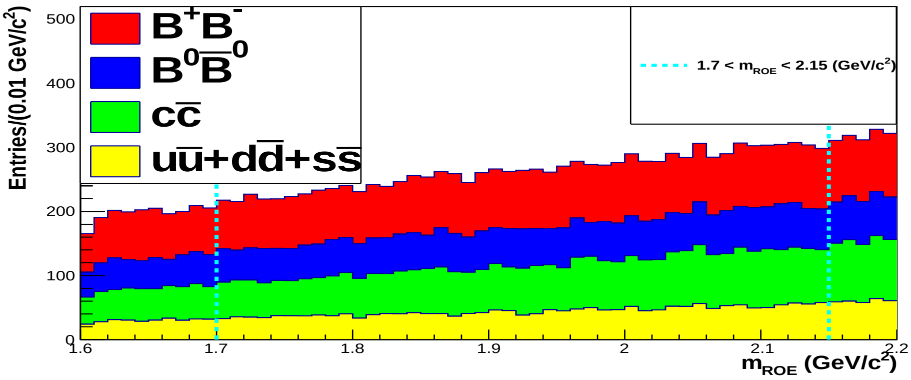


Figure 4.8: The distribution of m_{ROE} for the SS modes.

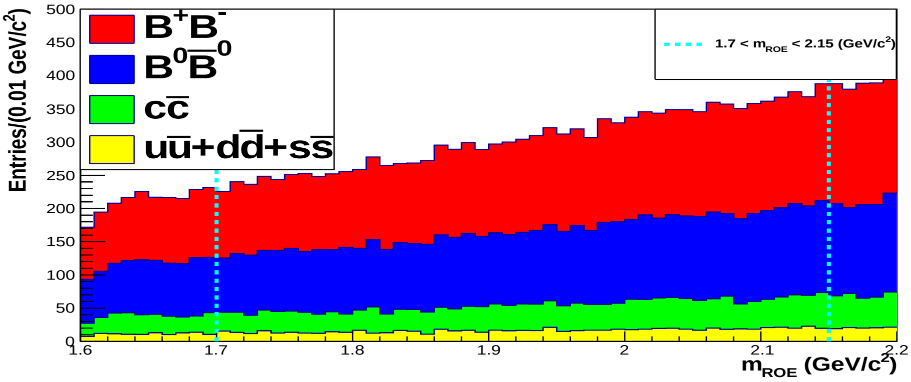


Figure 4.9: The distribution of m_{ROE} for the OS modes.

4.1.6 Selection on the B_{tag} side Lepton Momentum

To reject soft leptons (leptons with a momentum less than 0.2 GeV/c) and background electrons on the tag side, we select only the tag side lepton momentum in the range of $0.2 < p_{l\mathrm{tag}} < 2.5$ GeV/c. The distributions of $p_{l\mathrm{tag}}$ are shown in Fig. 4.10 for the SS modes and Fig. 4.11 for the OS modes. In both distributions, we can see a peak on the right side for the soft leptons. The majority of the rejected soft leptons are electrons from the beam background. By this loose selection, we get rid of these background leptons and further improve the purity of the signal events.

4.1.7 $\Delta E_{B\mathrm{tag}}$ Selection

The variable $\Delta E_{B\mathrm{tag}}$ shows the difference between the nominal energy of the B meson and the energy of the reconstructed B_{tag}. For a perfectly reconstructed B_{tag},

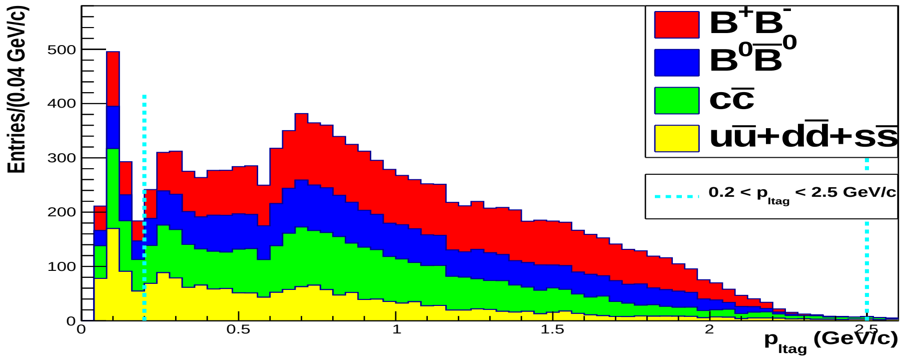


Figure 4.10: The distribution of $p_{l\text{tag}}$ for SS modes.

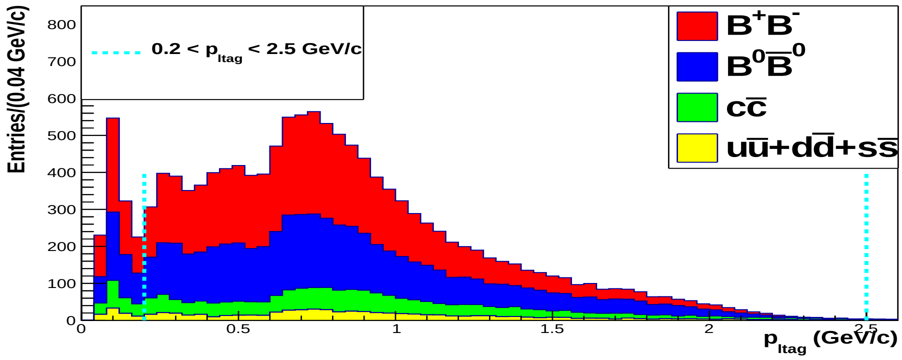


Figure 4.11: The distribution of $p_{l\text{tag}}$ for OS modes.

it should peak at zero. As we are using the semileptonic tagging approach in which we have a missing neutrino, its peak will be shifted toward the negative side (Fig. 4.12 for SS modes and Fig. 4.13 for the OS modes). We make a loose selection of $\Delta E_{B\text{tag}} < 1.0$ GeV, which rejects a few events (no substantial reduction) from all the background components.

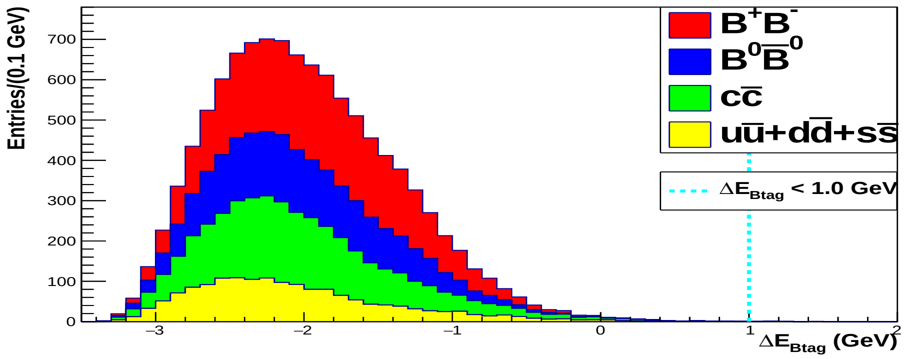


Figure 4.12: The distribution of $\Delta E_{B\text{tag}}$ for SS modes.

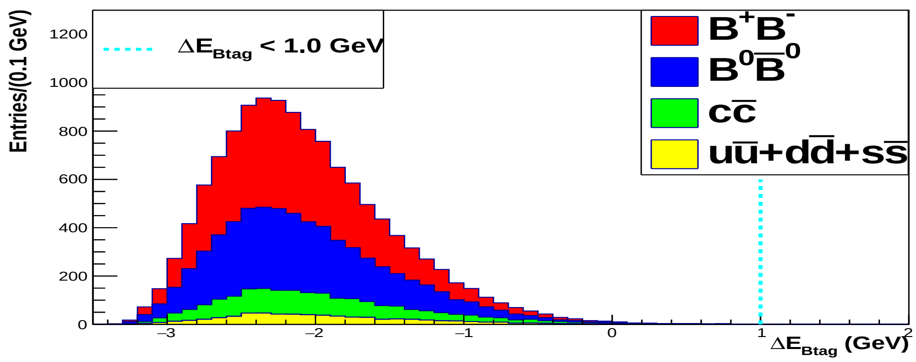


Figure 4.13: The distribution of $\Delta E_{B\mathrm{tag}}$ for OS modes.

4.1.8 $\cos\theta_{\mathbf{tag}}$ Selection

As described in the previous chapter, for true signal events $|\cos\theta_{\mathrm{tag}}| < 1.0$, i.e., in the physical range. We are making this loose selection (from -1.25 to +1.25) to also incorporate the partially reconstructed events on the tag side. The distributions of $\cos\theta_{\mathrm{tag}}$ are shown in Fig. 4.14 (for the SS modes) and Fig. 4.15 (for the OS modes). This loose selection also helps us to reject a substantial number of background events from all the background components.

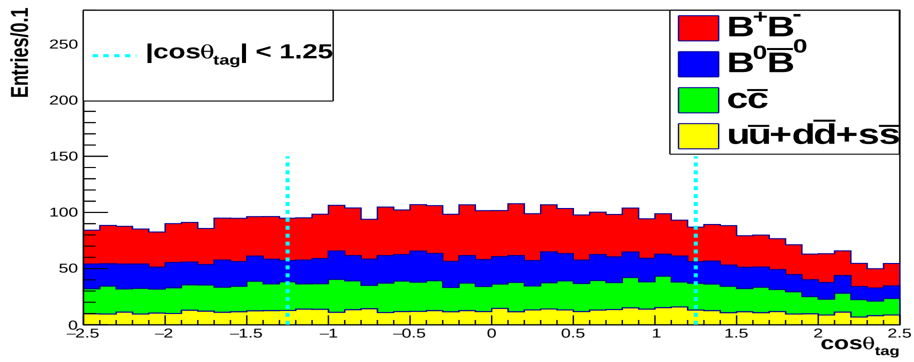


Figure 4.14: The distribution of $\cos\theta_{\mathrm{tag}}$ for SS modes

4.1.9 Continuum Suppression

Continuum $e^+e^- \to q\bar{q}$ events tend to be jet-like, while $B\overline{B}$ events are more spherical. This difference is quantified using the ratio of Fox–Wolfram moments, R_2 [60]. We require $R_2 < 0.38$ for SS modes and $R_2 < 0.45$ for OS modes, where the working points are determined by maximizing the Punzi figure of merit.

Its distributions are shown in Fig. 4.16 for SS modes and Fig. 4.17 for OS modes. These selections rejects mainly the background events of the $c\bar{c}$ and $u\bar{u} + d\bar{d} + s\bar{s}$ components.

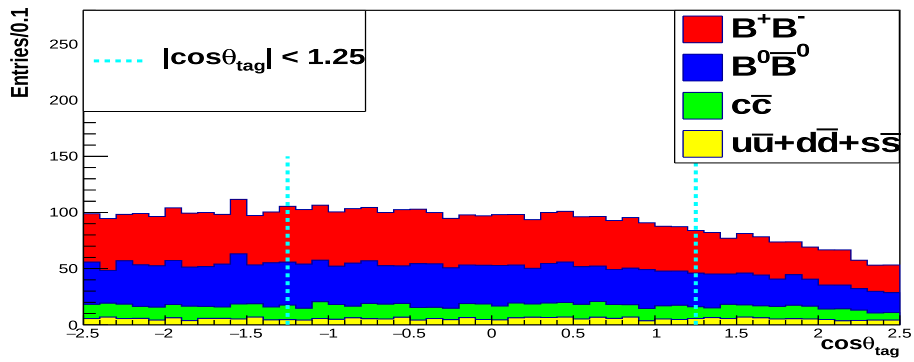


Figure 4.15: The distribution of cosθ_{tag} for OS modes.

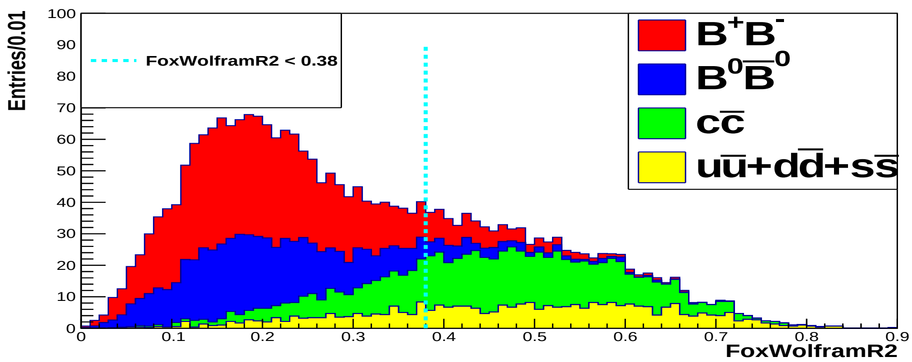


Figure 4.16: The distribution of variable FoxWolframR2 for SS modes.

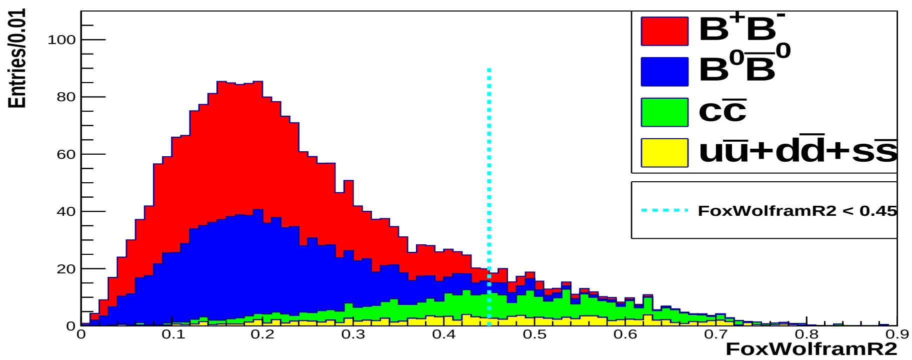


Figure 4.17: The distribution of variable FoxWolframR2 for OS modes.

The conventions for all the selections used are defined in Table 4.1. The cut flow in terms of the number of remaining signal and background events, relative efficiency of the selections for signal, and the relative loss of the background events are given in Table 4.2 to Table 4.5. In these tables and onward in the thesis, the notation SSe denotes $B^+ \to K^+\tau^-e^+$ mode, SSmu denotes $B^+ \to K^+\tau^-\mu^+$ mode, OSe denotes $B^+ \to K^+\tau^+e^-$ mode, and notation OSmu denotes $B^+ \to K^+\tau^+\mu^-$ decay modes.

Moreover the notation uds will be used for the $u\overline{u} + d\overline{d} + s\overline{s}$.

Naming convention	Selection
C01	$\|\sin\varphi\| \leq 1.0$
C02	$m_{K\pi} > 1.92$ GeV/c^2 (SS)
C03	$m_{K\ell} > 1.91$ GeV/c^2 (OS)
C04	$\|m_{\ell\pi} - m_{J/\psi}\| > 0.05$ GeV/c^2 and $\|m_{\ell\pi} - m_{\psi(2S)}\| > 0.05$ GeV/c^2
C05	$\|m_{\ell^+\ell^-} - m_{J/\psi}\| > 0.05$ GeV/c^2 and $\|m_{\ell^+\ell^-} - m_{\psi(2S)}\| > 0.05$ GeV/c^2
C06	$1.7 < m_{\mathrm{ROE}} < 2.15$ GeV/c^2
C07	$0.2 < p_{l\mathrm{tag}} < 2.5$ GeV/c
C08	$\Delta E_{B\mathrm{tag}} < 1.0$ GeV
C09	$\|\cos\theta_{\mathrm{tag}}\| < 1.25$
C10	$R_2 < 0.45$ (OS)
C11	$R_2 < 0.38$ (SS)
C12	Best-candidate selection

Table 4.1: Naming conventions for selections used in this analysis.

After all these selections, the signal efficiencies[4] are 0.44%, 0.36%, 0.61%, 0.58% for the decay modes of SSe, SSmu, OSe, and OSmu, respectively. The survived[5] B^+B^- background component for the SSe, SSmu, OSe, OSmu decay modes are 0.009%, 0.011%, 0.018%, and 0.033%, respectively. The survived percentage of other three ($B^0\overline{B}^0$, $c\overline{c}$, and uds) background components are 0.005%, 0.007%, 0.010%, and 0.001% for the SSe, SSmu, OSe, and OSmu decay modes, respectively. To distinguish the remaining background events from the signal events, we use the machine learning based approach, which is discussed in the next section.

4.2 Boosted Decision Tree Approach

To effectively distinguish between signal and background, the boosted decision tree (BDT) approach is used. BDT is a machine learning based approach that combines multiple decision trees (using a boosting method such as Gradient Boosting or AdaBoost) based on the different input variables to provide better separation between the signal and background [61]. While a simple cut-based selection only selects a specific region of phase space, BDT splits the region into a large number of small signal-like or background-like regions and assigns the corresponding weights. To improve the performance of the classifier and overcome statistical instability, boosting mechanism (modifying the weights) is used.

The overall BDT process consists of two main steps. In the first step, training and testing are performed. In the training process, the splitting criteria for each node are defined. The BDT input variables are ranked depending on how frequently

[4]with respect to the number of generated signal events.

[5]with respect to the number of background events originally reconstructed.

Selection	N_{sig} (ϵ_{rel}%)	$N_{B^+B^-}$ (% loss)	$N_{B^0\overline{B}^0}$ (% loss)	$N_{c\bar{c}}$ (% loss)	N_{uds} (% loss)
Reco. events ($\times 10^6$)	0.10 (100)	5.07 (0)	4.69 (0)	2.50 (0)	1.34 (0)
C01	25390 (24.63)	46352 (99.08)	31099 (99.33)	10785 99.57	5918 (99.56)
C02	13340 (52.54)	3979 (91.41)	2683 (91.37)	3824 (64.54)	3010 (49.14)
C04	12918 (96.84)	3875 (2.61)	2612 (2.65)	3787 (0.97)	2974 (1.20)
C05	12802 (99.10)	3776 (2.55)	2536 (2.90)	3763 (0.63)	2973 (0.03)
C06	8315 (64.95)	2056 (45.55)	1339 (47.20)	1895 (49.64)	1463 1463
C07	8146 (97.96)	1926 (6.32)	1241 (7.32)	1642 (13.35)	1111 (24.06)
C08	8142 (99.95)	1925 (0.05)	1241 (0)	1641 (0.06)	1110 (0.09)
C09	4893 (60.08)	500 (74.02)	285 (77.03)	332 (79.76)	217 (80.45)
C11	4409 (90.10)	444 (11.20)	255 (10.52)	98 (70.48)	57 (73.73)
C12	3946 (90.00)	290 (34.68)	170 (33.33)	71 (27.55)	43 (24.56)

Table 4.2: Cut flow table for SSe mode.

they are used to split a decision tree node. After the first step, it needs to be validated on independent samples of the signal and background. Although the training and testing samples are statistically independent, their agreement is evaluated using standard goodness-of-fit tests. Within statistical precision, no significant overtraining is observed, and the classifier response is stable when applied to independent validation samples. In the following subsections, these steps are discussed in detail.

4.2.1 BDT Training and Testing

As mentioned earlier, BDT uses multiple input variables to discriminate between the signal and background events. In this analysis, we are using six input variables ($\Delta\cos\theta$, m_{ROE}, $p_{l\text{tag}}$, total number of photons, $\Delta E_{B\text{tag}}$, total number of leptons) to train the BDT. For the signal, we used a 1.0 million (for each decay mode) dedicated signal MC $B^+ \to K^+\tau^\pm(\to \pi^\pm\nu_\tau)\ell^\mp$ sample, 0.9 million of this dedicated sample is used for training the BDT, and 0.1 million of this dedicated sample is

Selection	N_{sig} ($\epsilon_{\text{rel}}\%$)	$N_{B^+B^-}$ (% loss)	$N_{B^0\overline{B}^0}$ (% loss)	$N_{c\bar{c}}$ (% loss)	N_{uds} (% loss)
Reco. events ($\times 10^6$)	0.08 (100)	3.92 (0)	3.54 (0)	1.29 (0)	0.39 (0)
C01	22648 (73.91)	47344 (98.79)	31915 (99.09)	8970 (99.30)	2520 (99.36)
C02	10892 (48.09)	3784 (92.00)	2618 (91.79)	2794 (68.85)	1124 (55.40)
C04	10470 (96.12)	3650 (3.54)	2525 (3.55)	2759 (1.25)	1095 (2.58)
C05	10353 (98.88)	3541 (2.98)	2460 (2.57)	2726 (1.19)	1091 (0.36)
C06	6812 (65.79)	1922 (45.72)	1303 (47.03)	1371 (49.70)	533 (51.14)
C07	6666 (97.85)	1813 (5.67)	1216 (6.67)	1250 (8.82)	467 (12.38)
C08	6665 (99.98)	1812 (0.05)	1216 (0)	1250 (0)	466 (0.21)
C09	3897 (58.46)	466 (74.28)	282 (76.80)	281 (77.52)	108 (76.82)
C11	3578 (91.81)	420 (9.87)	256 (9.22)	96 (65.84)	34 (68.52)
C12	3240 (90.55)	275 (34.52)	172 (32.81)	70 (27.08)	25 (26.47)

Table 4.3: Cut flow table for SSmu mode.

used to test the BDT[6]. For background, we used the Belle generic $B\overline{B}$ MC sample ten times the Belle data (90% used for training and 10% for testing). Continuum events are strongly reduced by the R_2 requirement; after the cut-based preselection their contribution becomes subdominant. Training is therefore performed on the dominant $B\overline{B}$ background component, and the classifier response is cross-checked on continuum-enriched samples to confirm stability.

The distributions of the input variables for signal and background are shown in Fig. 4.18, Fig. 4.19, Fig. 4.20, and Fig. 4.21, for SSe, SSmu, OSe, and OSmu modes, respectively. In all these distributions, we can see a discriminating pattern between the signal and the background distributions. The most powerful (relatively) discriminator variable in providing the discrimination between signal and background is m_{ROE} (for SSe mode), $\Delta cos\theta$ (for SSmu mode), and p_{ltag} (for both OSe and OSmu modes).

A classification function optimally classifies the events based on the provided input variables. It contains a discriminating function that is derived from the training

[6] We have limited statistics to train the signal, that's why the train : test is set as 90 : 10.

Selection	N_{sig} (ϵ_{rel}%)	$N_{B^+B^-}$ (% loss)	$N_{B^0\overline{B}^0}$ (% loss)	$N_{c\bar{c}}$ (% loss)	N_{uds} (% loss)
Reco. events ($\times 10^6$)	0.13 (100)	2.57 (0)	2.33 (0)	1.31 (0)	0.40 (0)
C01	26479 (19.78)	19375 (99.24)	13339 (99.43)	6113 (99.53)	1553 (99.61)
C03	17050 (64.39)	6024 (68.90)	4859 (63.57)	1185 (80.61)	403 (74.05)
C04	16630 (97.53)	5895 (2.14)	4769 (1.85)	1164 (1.77)	399 (0.99)
C06	10936 (65.76)	2971 (49.60)	2419 (49.28)	576 (50.51)	204 (48.87)
C07	10706 (97.89)	2670 (10.13)	2199 (9.09)	501 (13.02)	173 (15.19)
C08	10700 (99.94)	2669 (0.03)	2198 (0.04)	501 (0)	172 (0.58)
C09	6306 (58.93)	516 (80.66)	413 (81.21)	91 (81.84)	38 (77.90)
C10	6090 (96.57)	501 (2.90)	396 (4.12)	33 (63.73)	16 (57.89)
C12	3536 (58.06)	147 (70.66)	115 (71.00)	10 (70.00)	6 (62.5)

Table 4.4: Cut flow table for OSe mode.

data. The response of the BDT is known as the BDT score, which is evaluated for the signal events and the background events. We use this BDT score to distinguish between the signal and the background.

The distributions of the BDT score for the training and testing samples are shown in Fig. 4.22, Fig. 4.23, Fig. 4.24, and Fig. 4.25 for SSe, SSmu, OSe, and OSmu decay modes, respectively. In all these distributions, we can observe the discrimination between the signal and the background events due to the different shapes of the BDT score distributions for the signal and background events. As we are limited by statistics, the training and testing sample distributions are not completely superimposed. However, they provide good agreement between the shapes of the training and testing sample distributions for both signal and background. This agreement suggests that this trained BDT can be trusted to classify the independent signal and background samples.

Another important parameter to check the performance of the classifier (BDT) is the receiver operating characteristic (ROC) curve. ROC is basically the Signal efficiency vs Background rejection plot. The area under the curve (AUC) of the ROC provides information about the performance of the classifier. For an ideal classifier, it should be 1.0; however, for the general classifiers, these values lie between 0.50 and 0.98. In our case, the values of AUC are 0.797 (for the SSe mode), 0.802 (for

Selection	N_{sig} (ϵ_{rel}%)	$N_{B^+B^-}$ (% loss)	$N_{B^0\overline{B}^0}$ (% loss)	$N_{c\bar{c}}$ (% loss)	N_{uds} (% loss)
Reco. events ($\times 10^6$)	0.10 (100)	1.76 (0)	1.57 (0)	0.95 (0)	0.37 (0)
C01	23477 (21.67)	19985 (98.86)	14106 (99.10)	7195 (99.24)	2402 (99.35)
C03	16480 (70.19)	6690 (66.52)	5273 (62.61)	2358 (67.22)	1104 (54.03)
C04	16052 (97.40)	6504 (2.78)	5135 (2.61)	2327 (1.31)	1089 (1.36)
C06	10486 (65.32)	3290 (49.41)	2591 (49.54)	1141 (50.96)	537 (50.68)
C07	10287 (98.10)	2979 (9.45)	2357 (9.03)	1019 5385	473 (10.69)
C08	10279 (99.92)	2978 (0.03)	2356 (0.04)	1019 (0)	472 (0.21)
C09	6019 (58.55)	621 (79.14)	455 (80.68)	215 (78.90)	98 (79.23)
C10	5801 (96.37)	600 (3.38)	438 (3.73)	106 (50.69)	50 (48.79)
C12	3375 (58.18)	191 (68.16)	138 (68.49)	31 (70.75)	15 (70.00)

Table 4.5: Cut flow table for OSmu mode.

the SSmu mode), 0.845 (for the OSe mode), and 0.847 (for the OSmu mode). ROC distributions are shown in Fig. 4.26 (for SS modes), Fig. 4.27 (for OS modes).

4.2.2 BDT Validation

As mentioned earlier, after the BDT training and testing step, it should be validated. For this purpose, the trained BDT is validated on an independent 4.0 million dedicated sample (for each decay mode) of $B^+ \to K^+\tau^\pm(\to \text{generic})\ell^\mp$.

For background validation, we used the Belle generic MC sample of the size of the Belle luminosity. The BDT-validation scores distributions are shown in Fig. 4.28 (SSe mode), Fig. 4.29 (SSmu mode), Fig. 4.30 (OSe mode), and Fig. 4.31 (OSmu mode). In all these distributions, we can see the distinct signal and background distributions. Both the signal distribution and the background distribution are normalized to unity. This BDT-validated score for the signal and background will be the input to our fitting model, which is discussed in Chapter 6 (Signal extraction).

This concludes the discussion on the BDT approach and, in the next section, we discuss the cross-feeds in the signal ($\tau \to$ generic) MC sample.

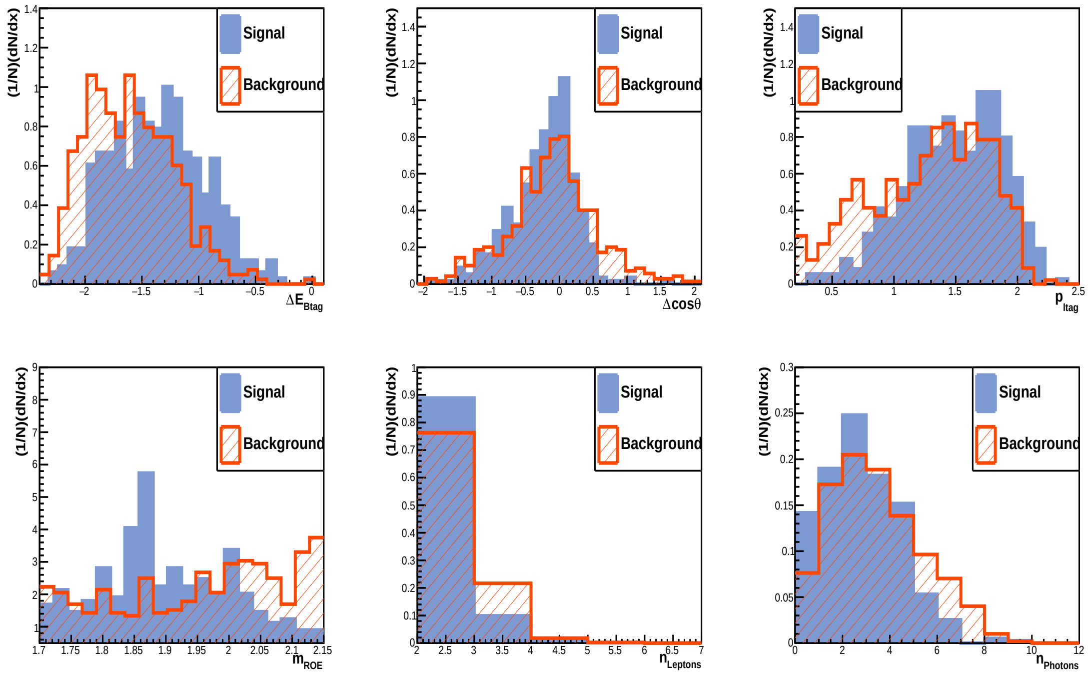


Figure 4.18: BDT input variables for SSe mode.

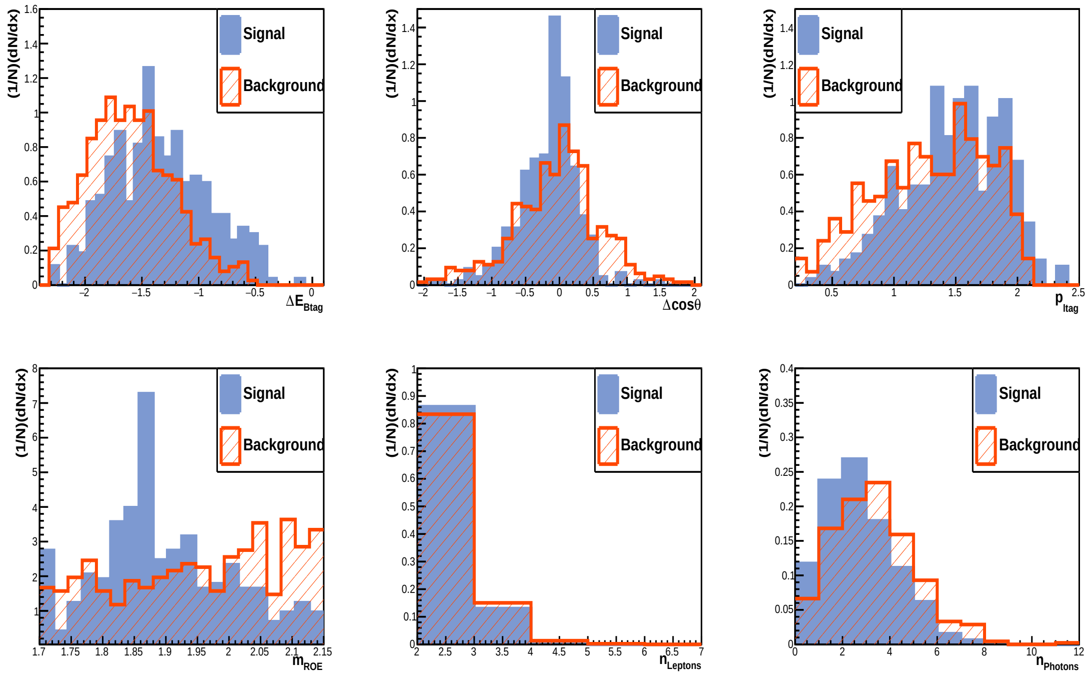


Figure 4.19: BDT input variables for SSmu mode.

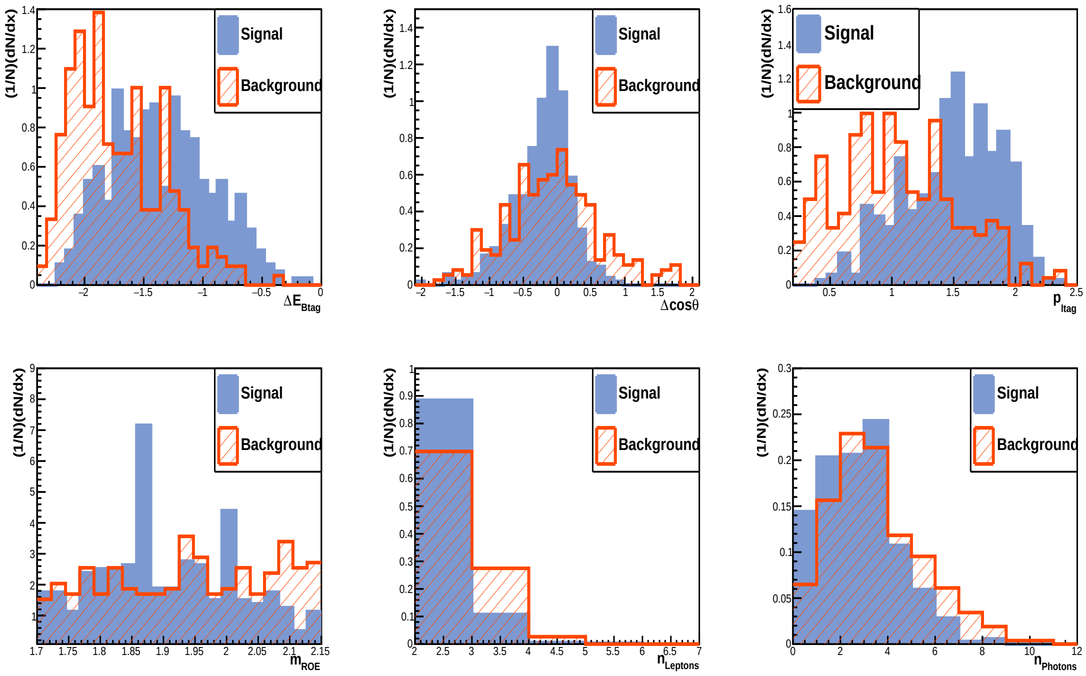


Figure 4.20: BDT input variables for OSe mode.

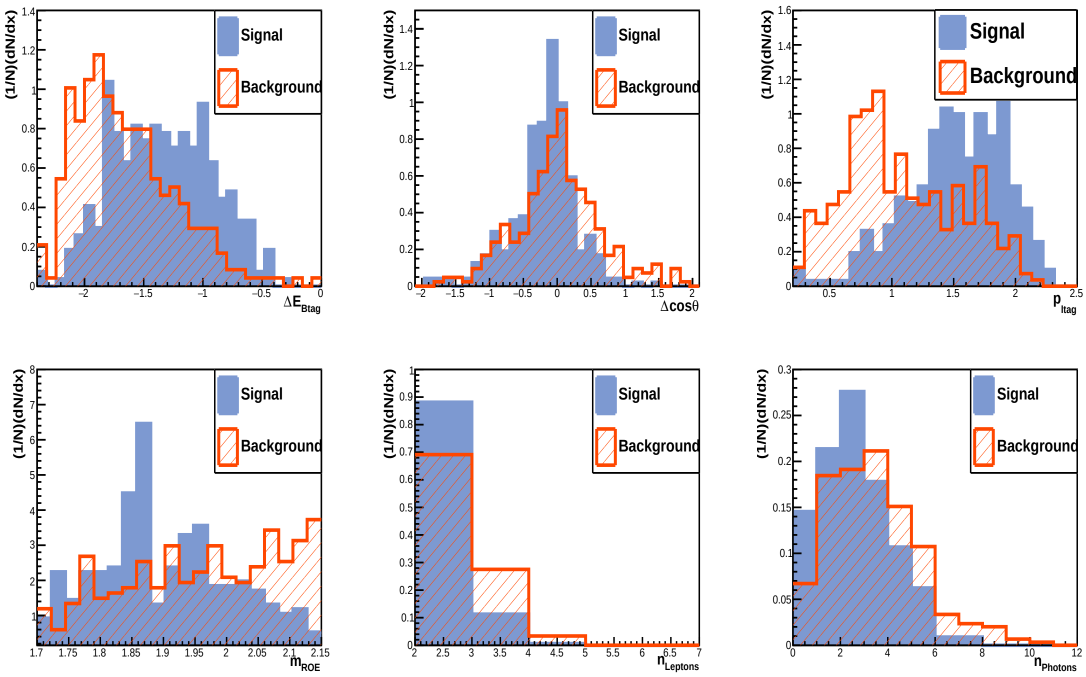


Figure 4.21: BDT input variables for OSmu mode.

4.3 Cross-Feeds

As discussed earlier, we are reconstructing only the $\tau \to \pi\nu_\tau$ mode in this analysis; however, some percentage of other τ decay modes will also contribute to the accepted

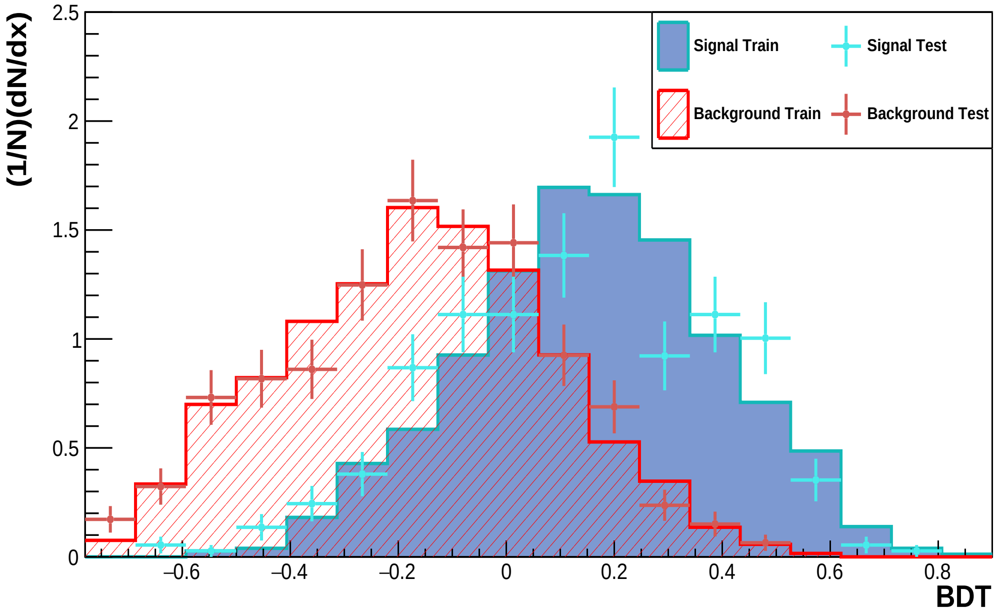


Figure 4.22: The distributions of BDT score for SSe mode.

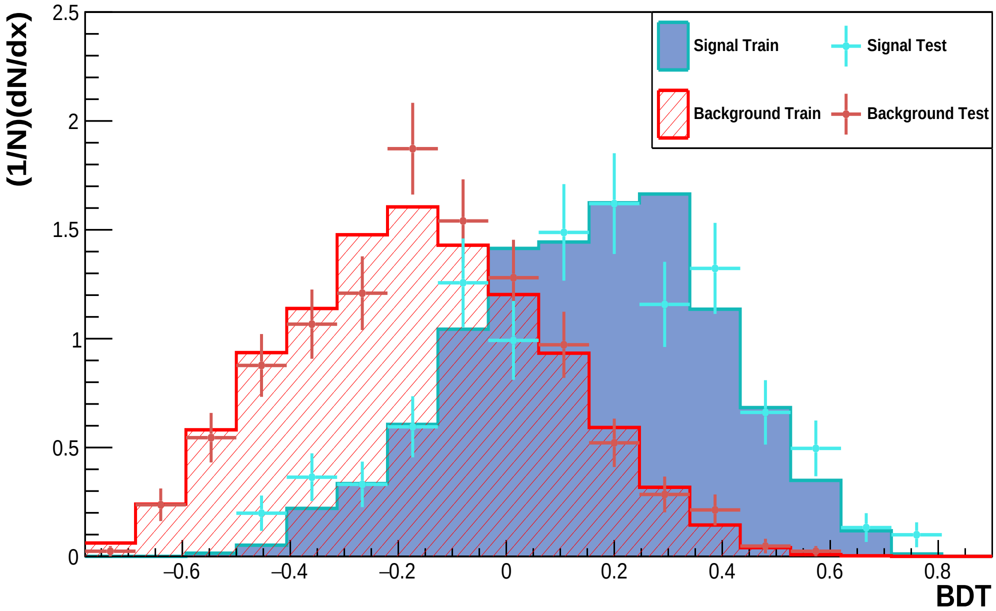


Figure 4.23: The distributions of BDT score for SSmu mode.

signal candidates. These misreconstructed τ decay modes are termed cross-feeds. These are the decays that have passed all of our selection criteria during the reconstruction process. The details of these cross-feeds are given in Table 4.6, where we

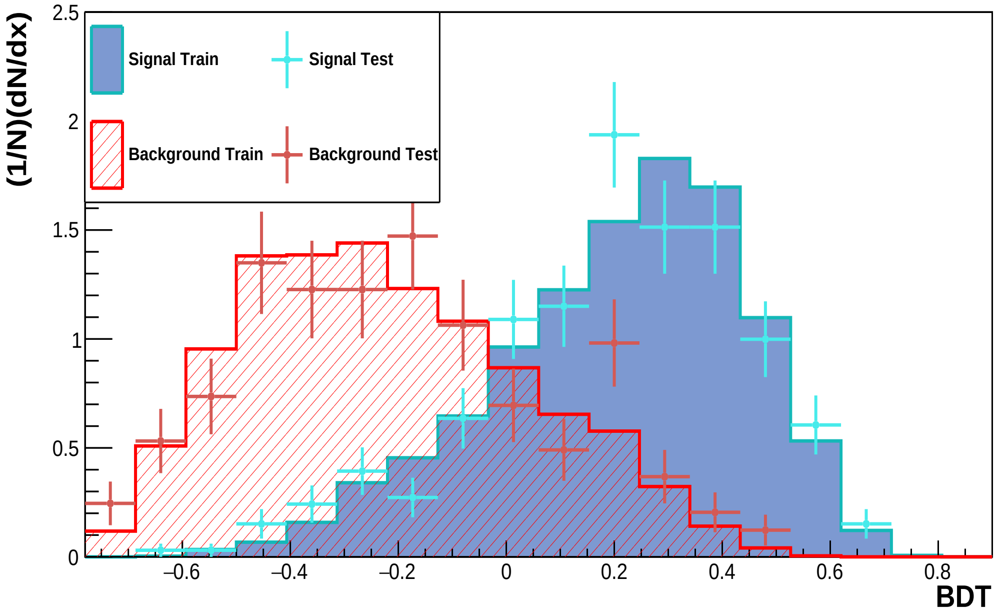


Figure 4.24: The distributions of BDT score for OSe mode.

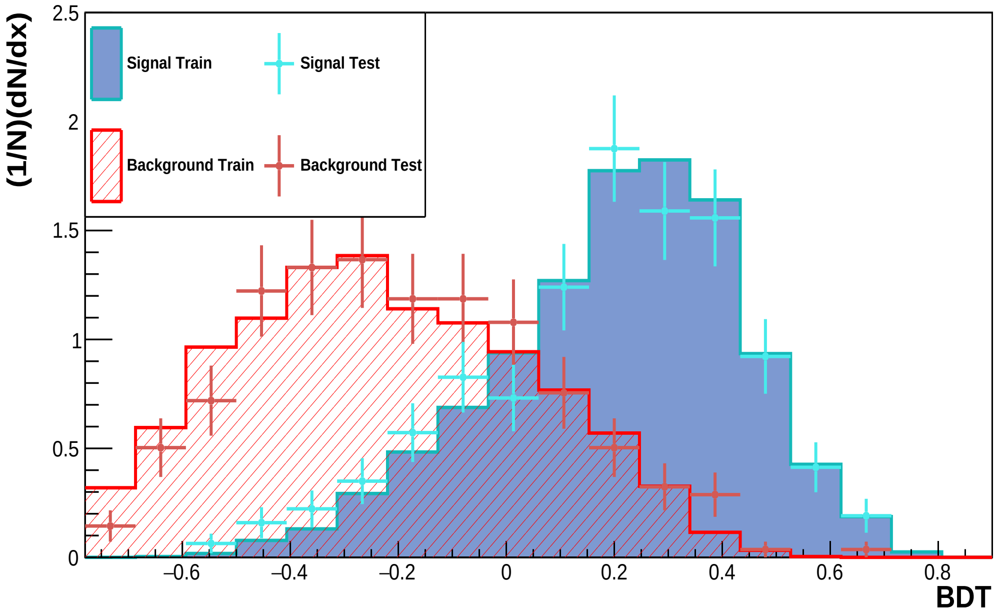
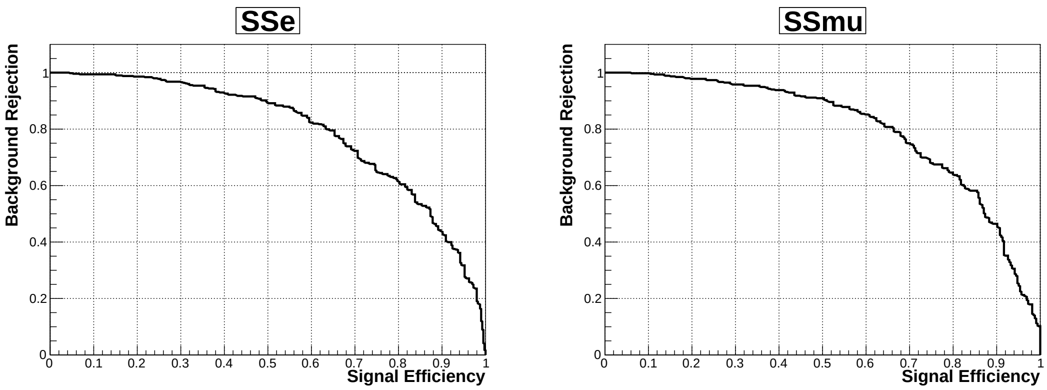


Figure 4.25: The distributions of BDT score for OSmu mode.

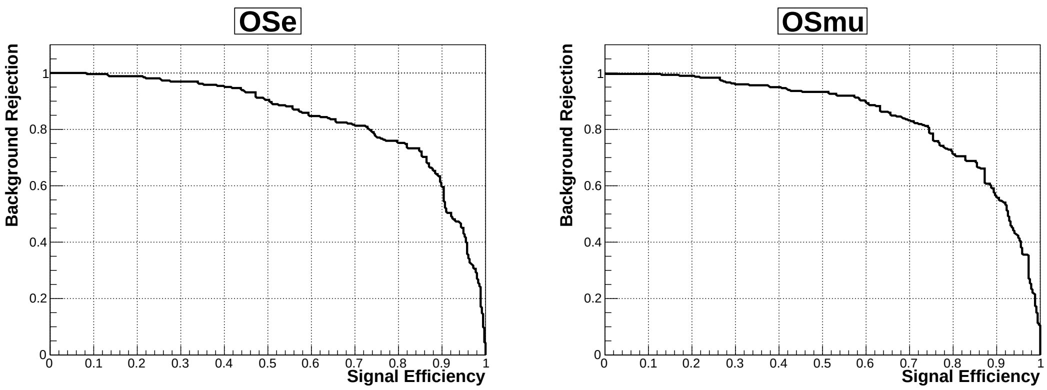
can see that the highest percentage of reconstructed events in all signal decay modes is from the $\tau \to \pi\nu_\tau$ mode.

In all signal decay modes, the most significant cross-feed is from the $\tau \to \rho\nu_\tau$



(a) (b)

Figure 4.26: ROC distributions (a) SSe mode (b) SSmu mode.



(a) (b)

Figure 4.27: ROC distributions (a) OSe mode (b) OSmu mode.

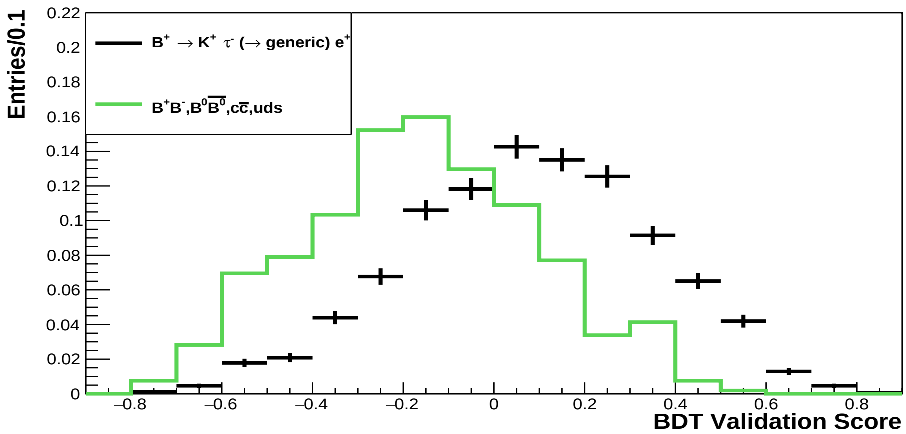


Figure 4.28: BDT validation score for SSe mode.

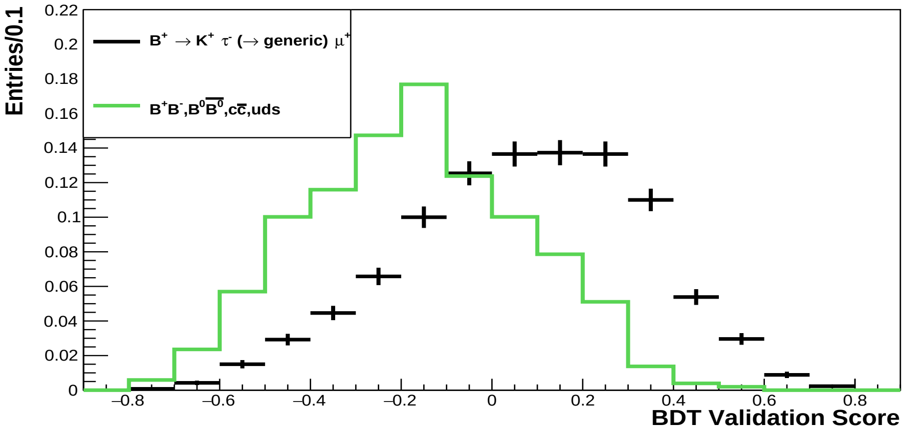


Figure 4.29: BDT validation score for SSmu mode.

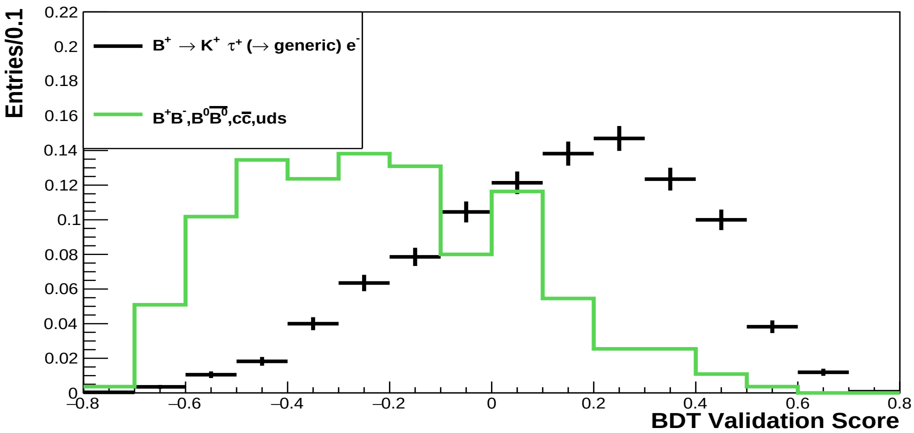


Figure 4.30: BDT validation score for OSe mode.

Mode	$\tau \to \pi\nu_\tau$	$\tau \to \rho\nu_\tau$	$\tau \to \mu\nu_\tau\nu_\mu$	$\tau \to e\nu_\tau\nu_e$	$\tau \to$ other
SSe (%)	55.74	26.81	6.75	4.56	6.14
SSmu (%)	55.04	27.60	6.01	4.39	6.94
OSe (%)	51.51	26.86	8.76	4.63	8.23
OSmu (%)	51.50	27.12	9.23	4.08	8.06

Table 4.6: Cross-feeds details.

mode, which is around 27% in all modes. This is due to the presence of a π in the $\tau \to \rho\nu_\tau$ decay mode as ρ is reconstructed by combining a π and π^0, that is, $\rho \to \pi\pi^0$. This π from the ρ mode can pass our selection criteria and then can be reconstructed as $\tau \to \pi\nu_\tau$. The second most significant cross-feed is from the $\tau \to \mu\nu_\tau\nu_\mu$ mode. This is due to the misidentification of a μ as a π. All these distributions are shown in Fig. 4.32, Fig. 4.33, Fig. 4.34, and Fig. 4.35 for SSe, SSmu, OSe, and OSmu

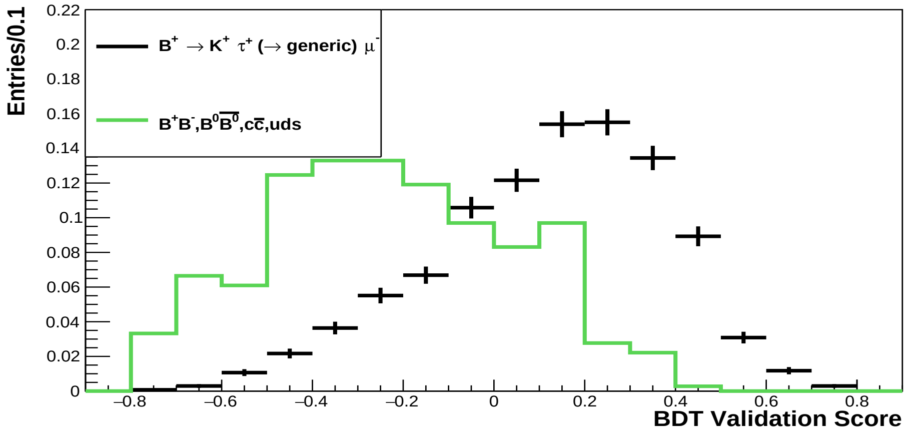


Figure 4.31: BDT validation score for OSmu mode.

modes, respectively. In all these distributions we can observe a slight difference in the shape between $\tau \to \pi\nu_\tau$ BDT validation score and the $\tau \to$ all other (except $\pi\nu_\tau$) BDT validation score. We use this difference in shape when discussing the signal extraction in Chapter 6.

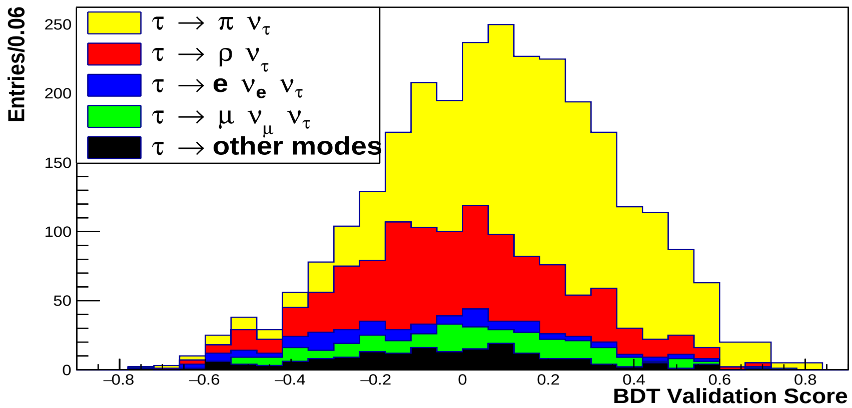


Figure 4.32: Cross feeds for SSe mode.

In addition to these τ cross-feeds, there is also a small percentage of overall misreconstructed signal events. In percentage terms, they are 8.02% for the SSe mode, 4.03% for the SSmu mode, 5.35% for the OSe mode, and 4.9% for the OSmu mode. These misreconstructed events are rejected by requiring selection on the true signal particles' ID.

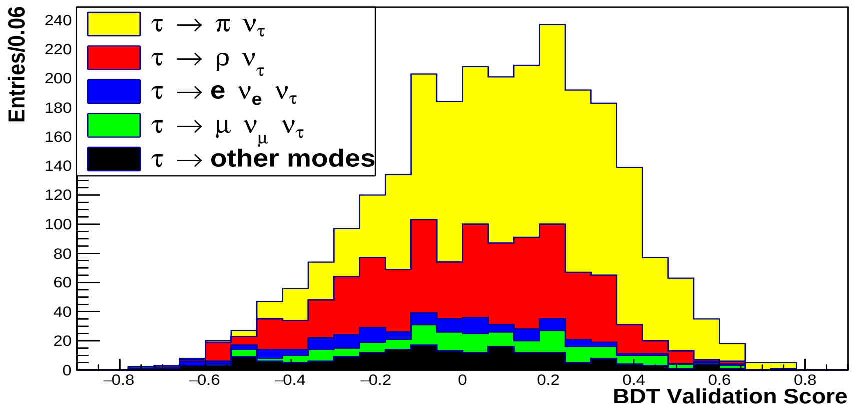


Figure 4.33: Cross feeds for SSmu mode.

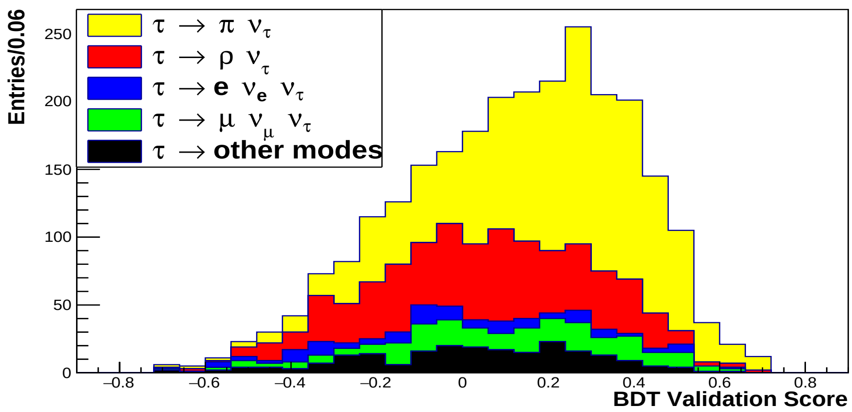


Figure 4.34: Cross feeds for OSe mode.

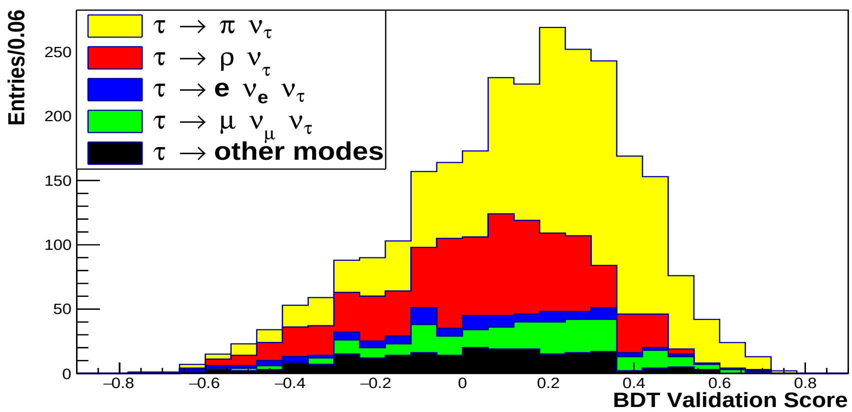


Figure 4.35: Cross feeds for OSmu mode.

4.4 Reconstruction of $\tau \to \rho\nu_\tau$ Mode

As in the previous section, we have seen that the most significant cross-feed is from the $\tau \to \rho\nu_\tau$ decay channel, therefore, we also reconstruct this τ decay mode[7]. As a first step in this reconstruction, a π^0 candidate is formed by combining the two photons and requiring the invariant mass of this π^0 candidate to be in the range of 115 to 150 MeV/c^2. In the second step, this π^0 candidate is combined with a $\pi^\pm$ to form a $\rho^\pm$ candidate. The invariant mass of the $\rho^\pm$ candidate is constrained in the range of 570 to 970 MeV/c^2. The distributions of π^0 invariant mass are shown in Fig. 4.36 and Fig. 4.37 for SS modes and OS modes, respectively. The $\rho^\pm$ invariant mass distributions are shown in Fig. 4.38 and Fig. 4.39 for SS modes and OS modes, respectively. The rest of the reconstruction method and selections are similar to what is described for the $\tau \to \pi\nu_\tau$ mode.

After reconstructing the $\tau \to \rho\nu_\tau$ decay mode, a detailed study of the backgrounds is performed, similar to the one described earlier in this Chapter, and the majority of the background events are rejected. The BDT process is also performed for this decay mode to get the BDT validation score for all the signal decay modes (SSe, SSmu, OSe, and OSmu).

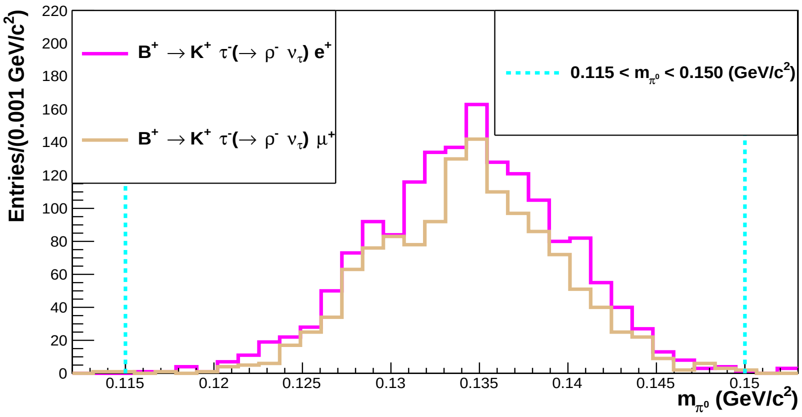


Figure 4.36: The distribution of $m_{\gamma\gamma}$ for SS modes.

The final signal reconstruction efficiencies and the expected number of total background events for all signal decay modes for both τ modes are given in Table 4.7, where we can see that the $\tau \to \pi\nu_\tau$ reconstruction is more efficient than the $\tau \to \rho\nu_\tau$ mode. This is due to the cleaner selection for $\pi^\pm$ as compared to the $\rho^\pm$ candidate. The total number of expected background events is also larger for the $\tau \to \rho\nu_\tau$ mode. Moreover, the overall number of total background events is larger in the SS modes as compared to the OS modes. This is because in the SS modes, secondary leptons[8] can also be selected, which is not the case for the OS modes.

[7]The work described in this section also has the contributions (reconstruction and the background studies) from Dr. Devender Kumar.

[8]Leptons originating from hadrons on the tag side.

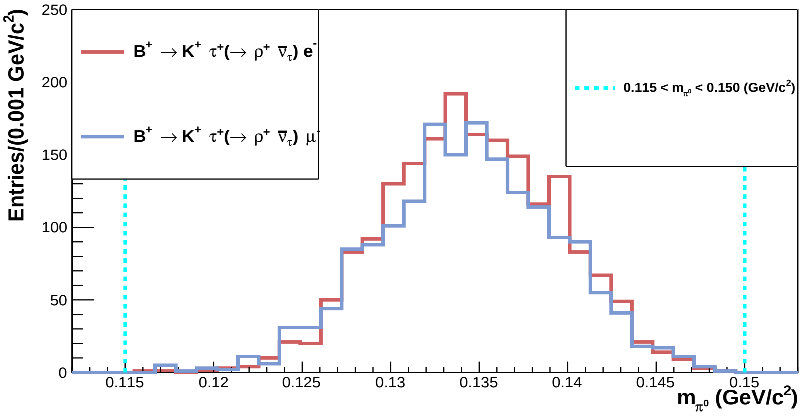


Figure 4.37: The distribution of $m_{\gamma\gamma}$ for OS modes.

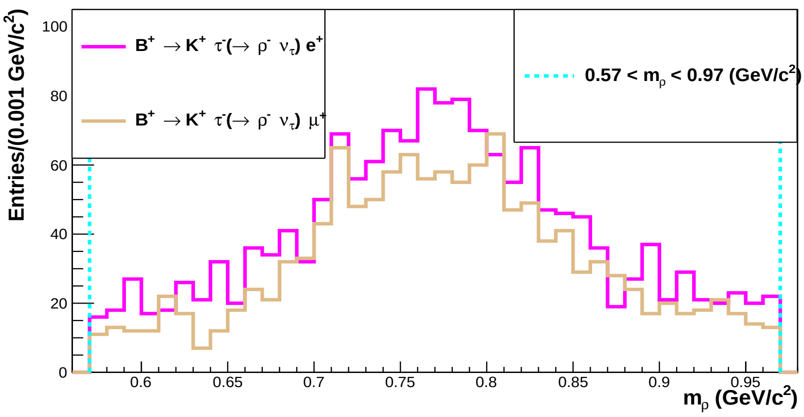


Figure 4.38: The distribution of $m_{\pi^-\pi^0}$ for SS modes.

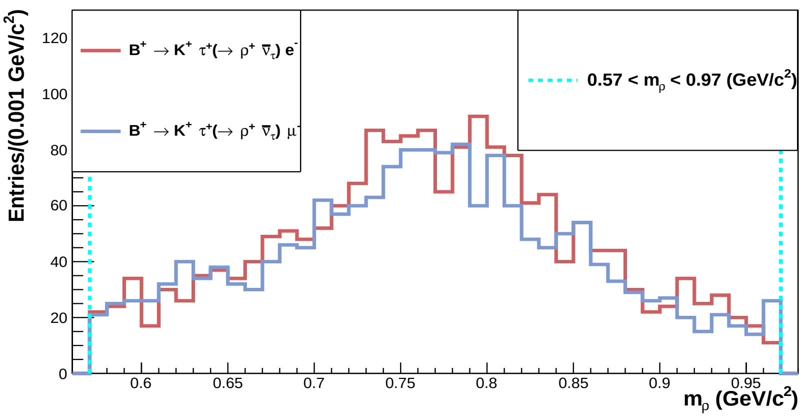


Figure 4.39: The distribution of $m_{\pi^+\pi^0}$ for OS modes.

This completed the study of the background involved in this analysis, and now, in the next chapter, we will perform the checks of our approach on the Belle data using the control channels.

Signal mode	τ mode	Sig. efficiency (%)	N_{bkg}
	$\tau \to \pi\nu_\tau$	0.39	574
SSe	$\tau \to \rho\nu_\tau$	0.15	1276
	$\tau \to \pi\nu_\tau$	0.32	542
SSmu	$\tau \to \rho\nu_\tau$	0.12	1300
	$\tau \to \pi\nu_\tau$	0.35	278
OSe	$\tau \to \rho\nu_\tau$	0.17	574
	$\tau \to \pi\nu_\tau$	0.34	375
OSmu	$\tau \to \rho\nu_\tau$	0.16	728

Table 4.7: Signal reconstruction efficiencies and expected background events.

Chapter 5

Control Modes

"It doesn't matter how beautiful your theory is,
it doesn't matter how smart you are. If it
doesn't agree with experiment, it's wrong."
Richard P. Feynman

So far, the analysis development has relied on Monte Carlo samples (dedicated signal MC and Belle generic MC). Since residual differences between data and simulation can affect both background modeling and the efficiency of kinematic constraints, validation on Belle data is mandatory.

As this is a blind analysis [53], the validation is performed using control modes with well-known properties and kinematics closely related to the signal topology. These channels serve two complementary purposes:

- to validate the reconstruction and selection chain on data, and
- to test the stability of the kinematic-constrained algorithm (cone-intersection method) in data.

In this chapter, two control channels are studied and the data/MC agreement is quantified.

5.1 Selection of Control Modes

For the selection of control channel modes, we need to find some decays that can mimic signal decay and produce results similar to signal decay. For this analysis, we have selected the following two decays as our control channel modes:

$$B^+ \to J/\psi(\to \mu^+\mu^-)K^+ \quad \text{(Control channel 1)},$$
$$B^+ \to \bar{D}^0(\to K^+\pi^-)\pi^+ \quad \text{(Control channel 2)}.$$

Both decays are fully reconstructed with high purity in Belle and provide clean samples in data. To validate the signal-side kinematic reconstruction method, an additional "pseudo-signal" configuration is constructed by removing one charged track from the signal-side reconstruction: a μ^- from the J/ψ decay in Control channel 1 and a π^- from the $\bar{D}^0$ decay in Control channel 2. The removed track is treated as invisible, and the full signal reconstruction framework (including $\sin\varphi$ and $\Delta\cos\theta$) is applied unchanged. In practice, the removed track is excluded from the signal-side four-momentum construction, and the event is processed with the same reconstruction chain as the signal.

All remaining selections (track quality, PID, and tag-side requirements) are identical to those used in the signal analysis.

In the fully reconstructed control samples (without removing any track), two standard Belle variables are used:

$$M_{bc} = \sqrt{E_{\text{beam}}^2 - |\vec{p}_B|^2}, \tag{5.1}$$

$$\Delta E = E_B - E_{\text{beam}}, \tag{5.2}$$

where E_{beam} is the beam energy in the center-of-mass frame and $(E_B, \vec{p}_B)$ is the reconstructed four-momentum of the B candidate. For correctly reconstructed events, M_{bc} peaks at the nominal B mass and ΔE peaks near zero.

In the pseudo-signal configuration (with one missing track), these variables are no longer expected to peak at their nominal positions and are therefore used only for initial sample definition and background suppression, while the primary validation of the signal reconstruction relies on $\sin\varphi$ and $\Delta\cos\theta$.

We use the full Belle data set. The MC samples correspond to three times the Belle integrated luminosity and are scaled to the Belle luminosity by a factor of 1/3. To reduce the event size and suppress obvious background already at the ntuple production stage, the following loose preselection is applied:

$$|\sin\varphi| < 1.25, \qquad |\cos\theta_{\text{tag}}| < 1.5. \tag{5.3}$$

These requirements are intentionally looser than the final offline selections.

In the upcoming sections, we discuss the distributions of important variables and offline selections[1] for these control channel modes.

5.2 Distributions and Selections

The first step in control mode studies is the selection of the J/ψ (D^0) signal in the $\mu^+\mu^-$ ($K^+\pi^-$) invariant mass distributions. We start with these selections in the next subsection, followed by a discussion on other selections.

5.2.1 Invariant Masses Selection

For the first control mode, we reconstruct J/ψ by combining two muon tracks with the opposite charge, and for the second control modes we reconstruct $\bar{D}^0$, by combining a K^+ track with a π^- track. The distributions of invariant masses of $K^+\pi^-$

[1] The selections made after the event reconstruction stage (on the produced ntuples).

and $\mu^+\mu^-$ are shown in Fig. 5.1 and Fig. 5.2, respectively. To reject background events, we require $3.07 < m_{\mu^+\mu^-} < 3.13$ (GeV/c^2) and $1.845 < m_{K^+\pi^-} < 1.883$ (GeV/c^2) for J/ψ and D^0, respectively. For the case of the first control channel, the online selections reject the majority of the background events from the $B^0\overline{B}^0$, $q\overline{q}$ components, and only the B^+B^- background component remains. In the second control channel, subsequent selections will reduce background events from different background components (mainly $B^0\overline{B}^0$, $q\overline{q}$).

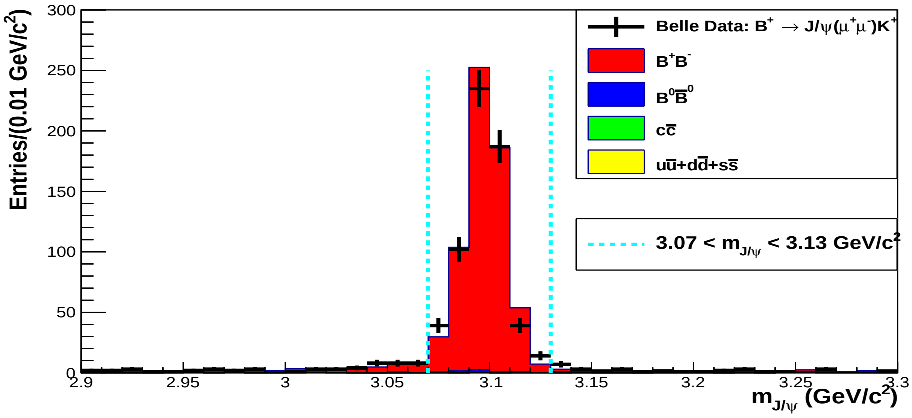


Figure 5.1: Distribution of $m_{\mu^+\mu^-}$ for data and generic MC.

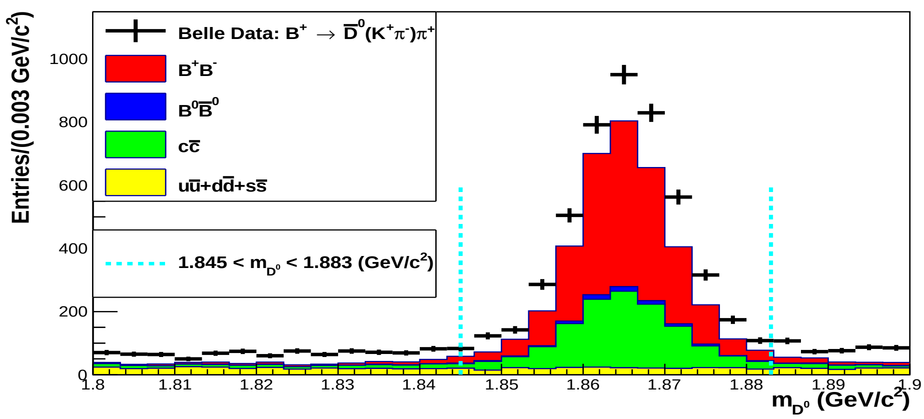


Figure 5.2: Distribution of $m_{K^+\pi^-}$ for data and generic MC.

5.2.2 $M_{bc} > 5.27$ GeV/c^2

As mentioned earlier, in the case of control modes, we can use the variable M_{bc} to suppress the background. The distributions of M_{bc} are shown in Fig. 5.3 and Fig. 5.4 for the first control channel and the second control channel, respectively. All events that are significantly below the nominal B meson mass are background events, and to reject these background events, we take $M_{bc} > 5.27$ GeV/c^2. This selection significantly removes the background. In both distributions, we can see that after making the selection $M_{bc} > 5.27$ GeV/c^2, the majority of the remaining background events are the $B\overline{B}$ background events.

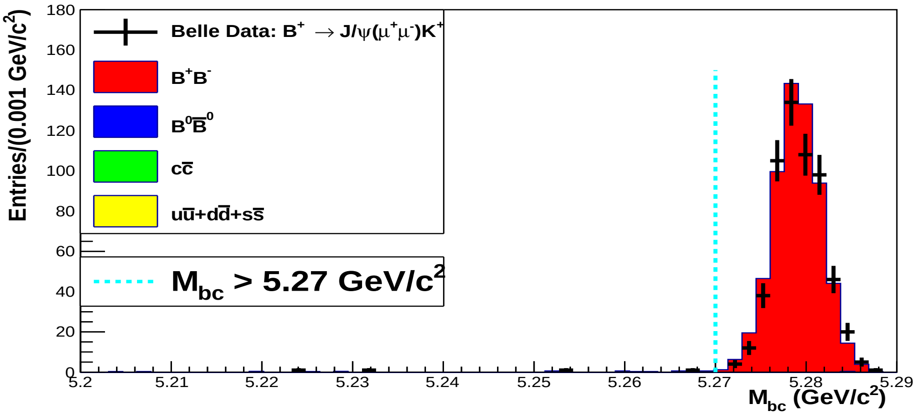


Figure 5.3: The M_{bc} distribution for the first control channel.

5.2.3 $|\Delta E| < 0.050$ GeV

Another important variable to remove background in Belle is ΔE. In the fully reconstructed control samples, correctly reconstructed events lie in the vicinity of $\Delta E = 0$. The distributions of ΔE are shown in Fig. 5.5 (first control channel) and Fig. 5.6 (second control channel). To select the majority of correctly reconstructed $B_{\rm sig}$ candidates, we require $|\Delta E| < 0.050$ GeV. This selection is especially useful for rejecting background events from semileptonic B decays[2] on the signal side.

5.2.4 $|\sin\varphi| \leq 1.0$

As described in the previous chapters, sinφ is an essential variable to select $B_{\rm sig}$ candidates. The distributions of sinφ are shown in Fig. 5.7 for the first control channel and Fig. 5.8 for the second control channel. Events outside the physical range constitute background events and are rejected by selecting $|\sin\varphi| \leq 1.0$. This selection further suppresses the background. This completes our selections on the $B_{\rm sig}$ side variables. To further increase the signal to background ratio, we use the information from the tag side (as we are reconstructing the full $\Upsilon(4S)$ candidate). In the next subsections, we discuss the selections we make on $B_{\rm tag}$ side variables to reject the background events.

[2]These are decays in which a B meson decays to a meson, a lepton, and a neutrino. For such decays, ΔE typically shifts toward negative values.

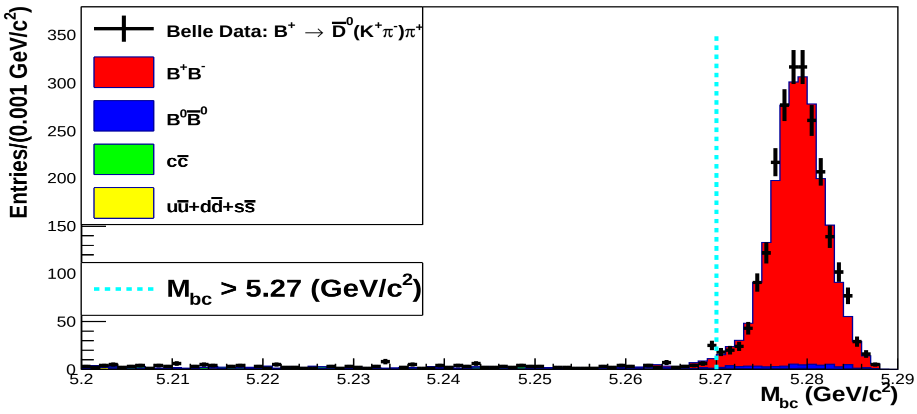


Figure 5.4: The M_{bc} distribution for the second control channel.

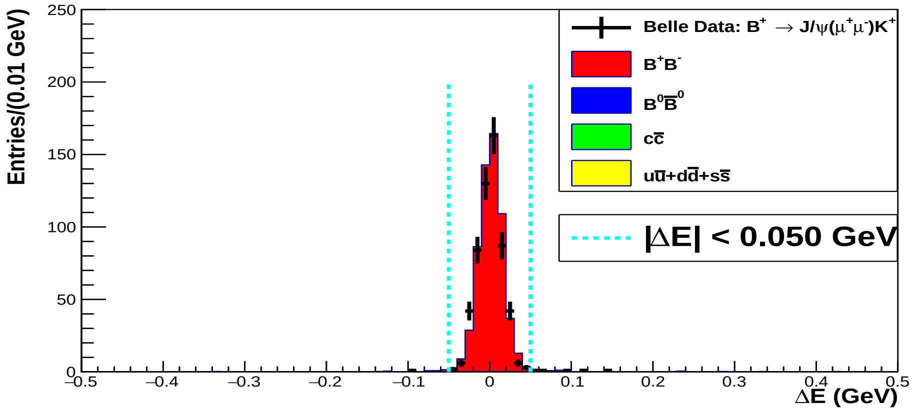


Figure 5.5: ΔE distribution for the first control channel.

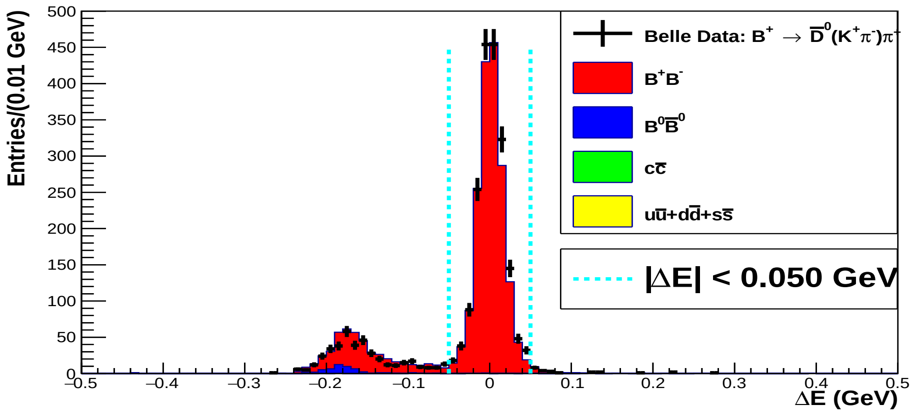


Figure 5.6: ΔE distribution for the second control channel.

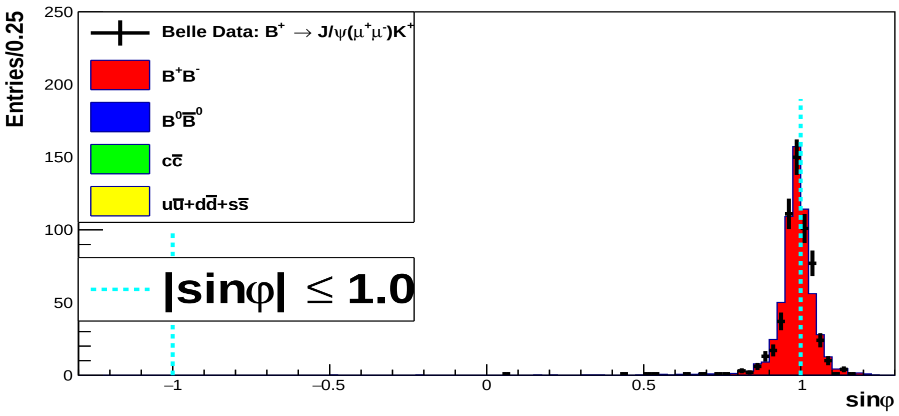


Figure 5.7: The distribution of sinφ for the first control channel.

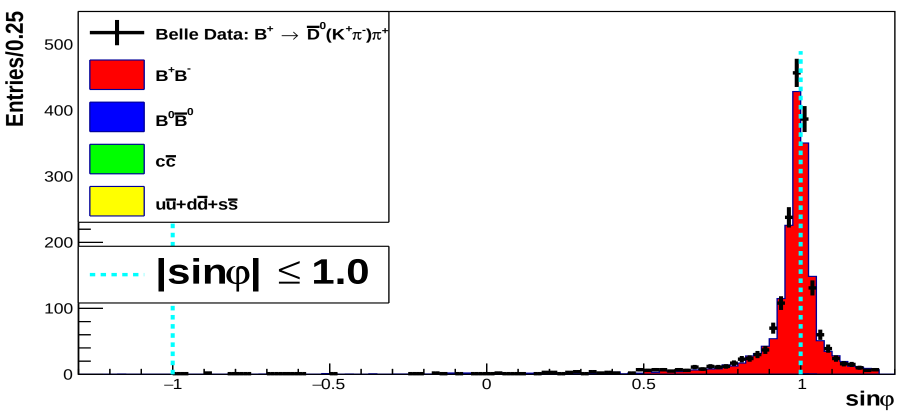


Figure 5.8: The distribution of sinφ for the second control channel.

5.2.5 Loose Selection on cos$\theta_{\mathbf{tag}}$

As we are using the inclusive semileptonic tagging approach to reconstruct tag side, so we use the variables that can help us select efficiently the semileptonic decays on the tag side. One of the key variables for selecting the semileptonic $B_{\rm tag}$ decays is cos$\theta_{\rm tag}$. This variable will help us select the events that have one missing particle (semileptonic B decays with a missing neutrino) on the tag side and reject the other events.

The distributions of cos$\theta_{\rm tag}$ are shown in Fig. 5.9 and Fig. 5.10 for first control channel and second control channel, respectively. As discussed in the previous chapters, for truly reconstructed semileptonic $B_{\rm tag}$ decays, it should lie in the physical range. We make a loose selection of $|\cos\theta_{\rm tag}| < 1.25$ to further suppress the background events.

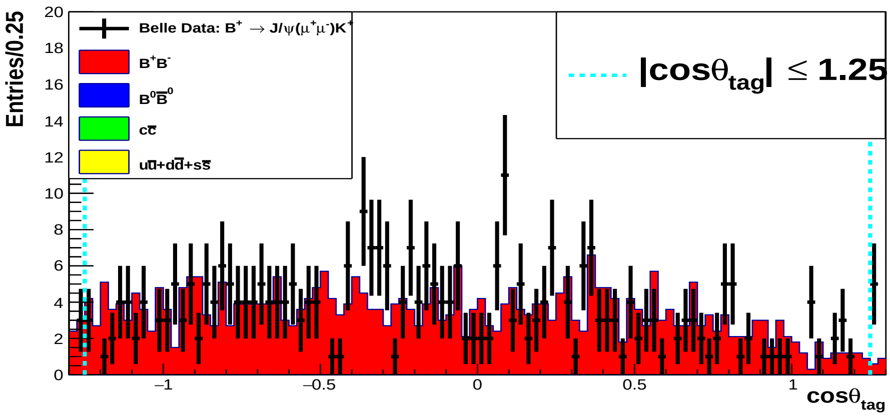


Figure 5.9: The distribution of $\cos\theta_{\mathrm{tag}}$ for the first control channel.

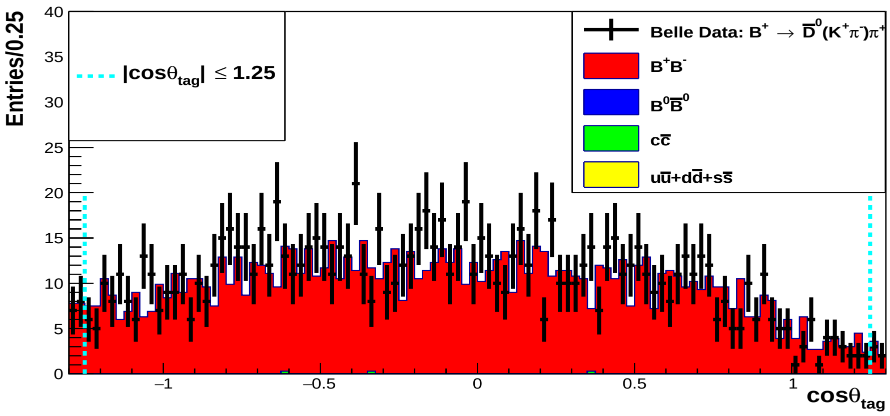


Figure 5.10: The distribution of $\cos\theta_{\mathrm{tag}}$ for the second control channel.

5.2.6 Loose Selection on m_{ROE}

The invariant mass of the rest of the events (m_{ROE}) is another important variable on the B_{tag} side. It shows the structure of the hadronic part of the tag side. The distributions of m_{ROE} are shown in Fig. 5.11 (first control mode), Fig. 5.12 (second control mode). These distributions are similar in structure to the ones that we have seen in the signal MC studies (Chapter 3). We can see a dominant D^0 peak (around 1.86 GeV/c^2), a small D^{*0} peak (around 2.006 GeV/c^2), and a small intermediate structure between them (around 1.96 GeV/c^2). The similarity of the shape between the data and the MC distributions indicates a good agreement between the data and the MC. To select the signal-enriched region, we take $1.7 < m_{\mathrm{ROE}} < 2.15$ GeV/c^2 and reject the lower and higher hadronic resonances to further reduce the background. This further reduces the number of background events.

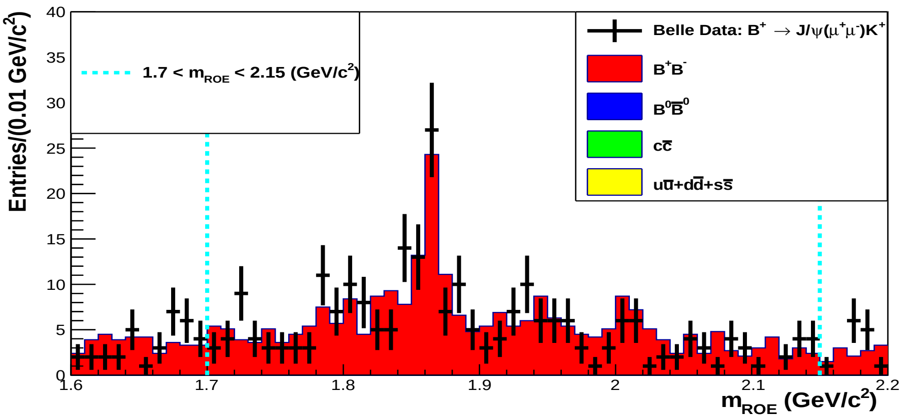


Figure 5.11: The $m_{\rm ROE}$ distribution for the first control channel.

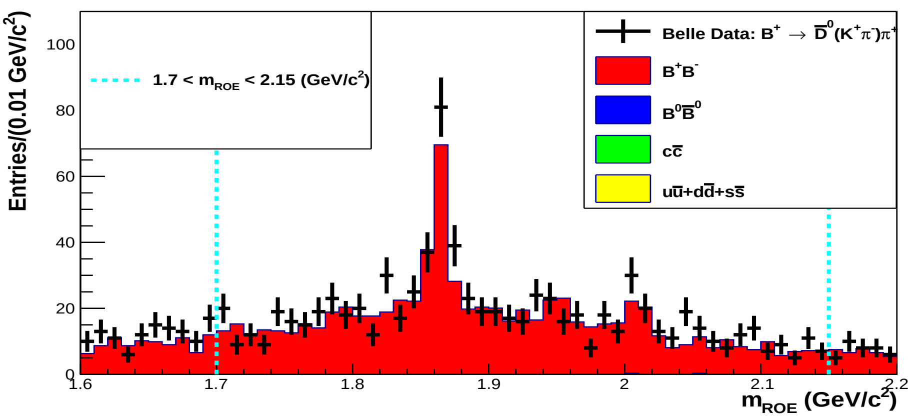


Figure 5.12: The $m_{\rm ROE}$ distribution for the second control channel.

5.2.7 Tag Side Lepton Momentum Selection

The soft leptons and beam background electrons on the $B_{\rm tag}$ side can be rejected by making loose selections on $p_{l\rm tag}$. The distributions of $p_{l\rm tag}$ for the first control channel and the second control channels are shown in Fig. 5.13 and Fig. 5.14, respectively. We make a selection of $0.2 < p_{l\rm tag} < 2.5$ GeV/c to reject the obvious background events.

This concludes the selection strategy for the control channels[3] In the next section, we summarize the corresponding cut-flow tables.

[3]Selection on $\Delta E_{\rm Btag}$ doesn't remove any background event in case of control modes. Also, after all the selections, we don't have $q\overline{q}$ background events remaining, so the FoxWloframR2 selection is also not used in the case of control modes.

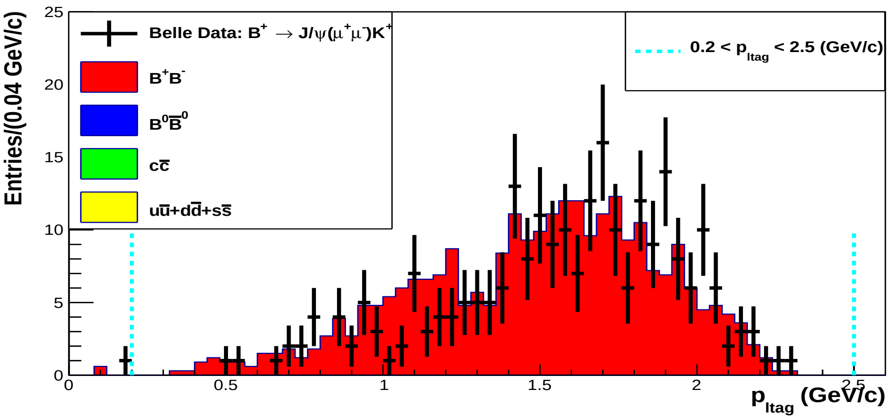


Figure 5.13: The distribution of p_{ltag} for first control channel.

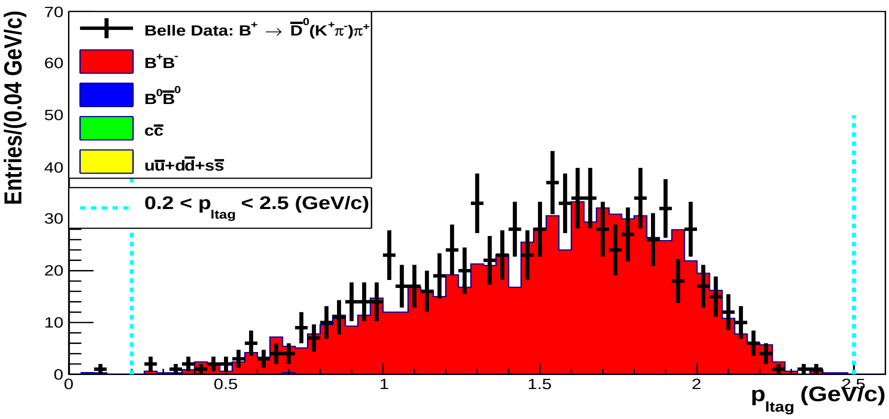


Figure 5.14: The distribution of p_{ltag} for second control channel.

5.3 Summary of the Selections

In this section, we summarize all the selections (described in the previous section) with the number of events for the Belle data and the background components for both control modes. The details of the first control mode are given in Table 5.1 and the details of the second control mode are given in Table 5.2. The background events are normalized to Belle luminosity.

In both tables, we can note that most background events are rejected due to selections on invariant masses, M_{bc}, and ΔE. The only remaining events in both cases in the B^+B^- MC are the correctly reconstructed signal from B^+B^- generic MC. It can also be noted that the number of events in the Belle data is closely comparable to the number of events in the B^+B^- MC in both control modes. In the next section, we discuss this in more detail.

Selection	Belle data	B^+B^-	$B^0\overline{B}^0$	$c\bar{c}$	uds
Reco. events	764	804	35	27	9
$3.07 < m_{J/\psi} < 3.13$ GeV/c^2	616	696	6	1	0
$M_{bc} > 5.27$ GeV/c^2	571	673	0	0	0
$\lvert\Delta E\rvert < 0.050$ GeV	565	661	0	0	0
$\lvert\sin\varphi\rvert < 1.0$	347	413	0	0	0
$\lvert\cos\theta_{\text{tag}}\rvert < 1.25$	315	380	0	0	0
$1.7 < m_{\text{ROE}} < 2.15$ GeV/c^2	246	294	0	0	0
$0.2 < p_{l\text{tag}} < 2.5$ GeV/c	245	293	0	0	0

Table 5.1: Summary of all selections with the corresponding number of events remained after the selection for the first control channel.

Selection	Belle data	B^+B^-	$B^0\overline{B}^0$	$c\bar{c}$	uds
Reco. events	6188	2720	93	1554	688
$1.845 < m_{D^0} < 1.883$ GeV/c^2	4826	2543	78	1335	263
$M_{bc} > 5.27$ GeV/c^2	2282	2435	55	5	4
$\lvert\Delta E\rvert < 0.050$ GeV	1854	1949	0	1	1
$\lvert\sin\varphi\rvert < 1.0$	1159	1201	0	1	0
$\lvert\cos\theta_{\text{tag}}\rvert < 1.25$	1055	1093	0	1	0
$1.7 < m_{\text{ROE}} < 2.15$ GeV/c^2	825	840	0	0	0
$0.2 < p_{l\text{tag}} < 2.5$ GeV/c	824	840	0	0	0

Table 5.2: Summary of all selections with the corresponding number of events remained after the selection for the second control channel.

5.4 Data/MC Agreement

Checking the agreement between data and MC is an essential part of the high-energy physics analyses. To check this agreement, we use the variable $\Delta\cos\theta$, which contains information from both the signal side and the tag side. The distributions of $\Delta\cos\theta$ are shown in Fig. 5.15 for the first control mode and in Fig. 5.16 for the second control mode. In both distributions, we can see a reasonable agreement between the data and the MC distributions.

For the first control channel, we have 245 ± 16 data events corresponding to 293 ± 6 MC events. For the second control mode, there are 824 ± 29 data events corresponding to the 840 ± 10 MC events. For background modeling, Belle generic MC corresponding to three times the Belle integrated luminosity is used and scaled to the Belle luminosity by a factor 1/3. Therefore, the statistical uncertainty of the scaled MC yield is

$$\sigma_{MC} = \frac{\sqrt{N_{MC}^{(3\times)}}}{3}. \tag{5.4}$$

The data uncertainty is taken as $\sigma_{\text{data}} = \sqrt{N_{\text{data}}}$. The data/MC ratio is computed

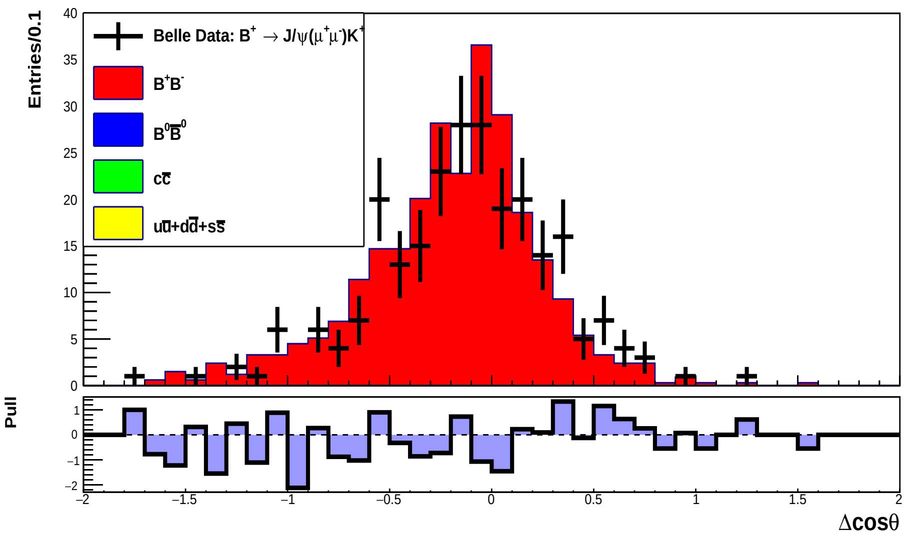


Figure 5.15: The distribution of $\Delta\cos\theta$ for the first control channel.

as $R = N_{\text{data}}/N_{\text{MC}}$ with uncertainty propagated from σ_{data} and σ_{MC}.

The resulting data/MC ratios are compatible with unity within uncertainties, supporting the reliability of the reconstruction and selection chain in data. In the next section, we discuss the signal sideband studies.

5.5 Signal Sideband Studies

A signal-sideband region is defined by $\sin\varphi > 1.2$, adjacent to the signal region $|\sin\varphi| \leq 1$. This region is dominated by background and is therefore suitable to test the background normalization and shape modeling in data. We use the full Belle data set and all available streams of Belle generic MC (normalized to the Belle luminosity) for this sideband study. The distributions of m_{ROE} in the interval $1.2 < \sin\varphi < 1.3$ are shown in Fig. 5.17 (SSe), Fig. 5.18 (SSmu), Fig. 5.19 (OSe), and Fig. 5.20 (OSmu), with all other selections (Chapter 4) applied. The sideband comparison indicates a tendency of MC to overestimate the data in several channels. A normalization factor derived from this sideband will be propagated in the final analysis to the signal extraction as a systematic cross-check.

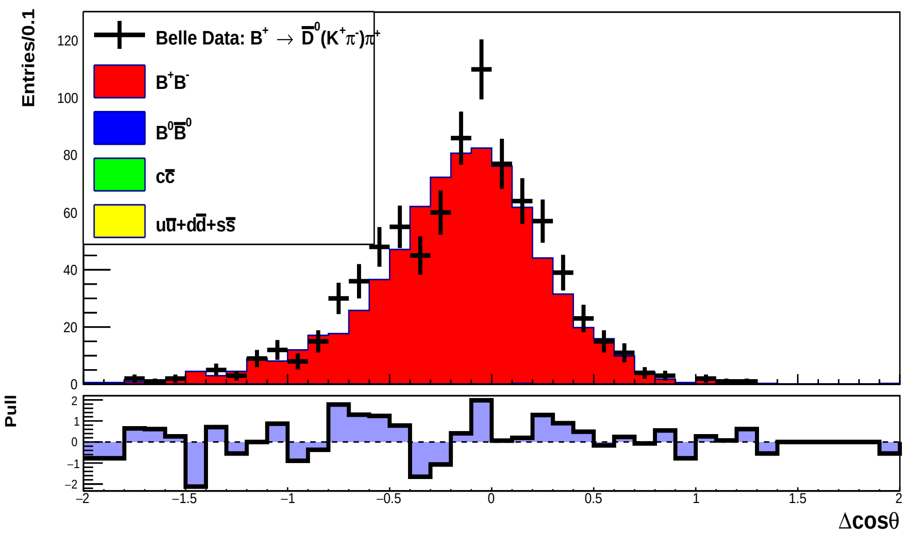


Figure 5.16: The distribution of $\Delta\cos\theta$ for second control channel.

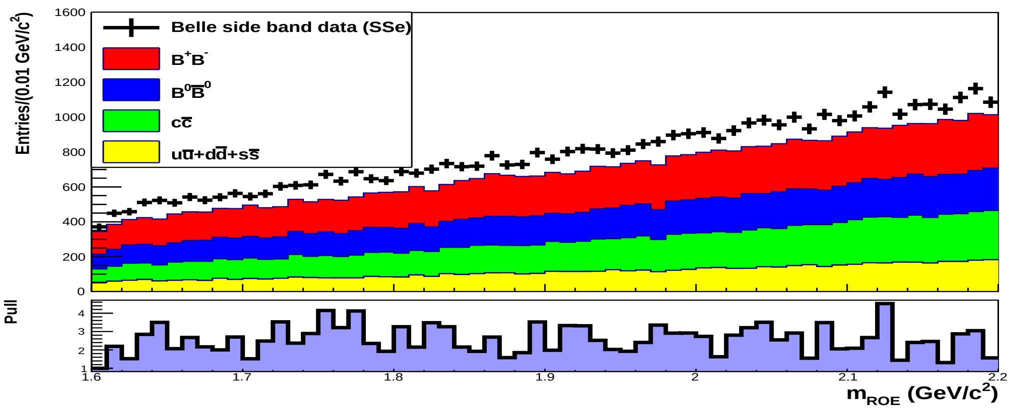


Figure 5.17: The distribution of m_{ROE} for SSe sideband region.

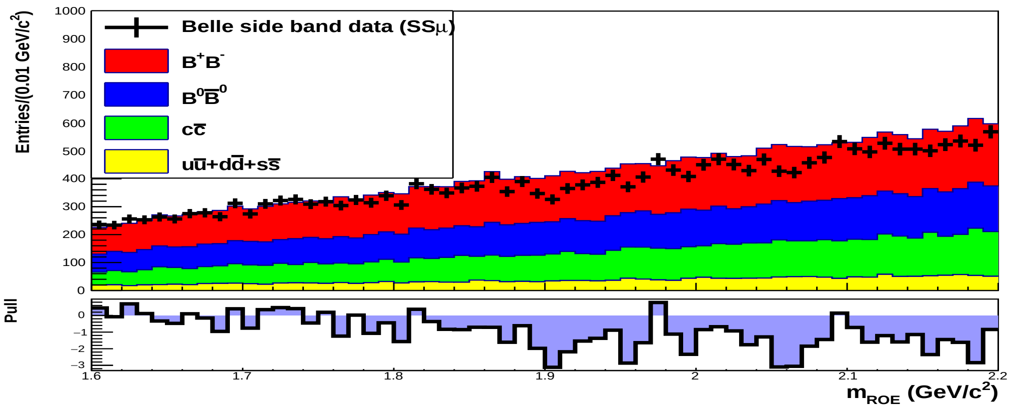


Figure 5.18: The distribution of m_{ROE} for SSmu sideband region.

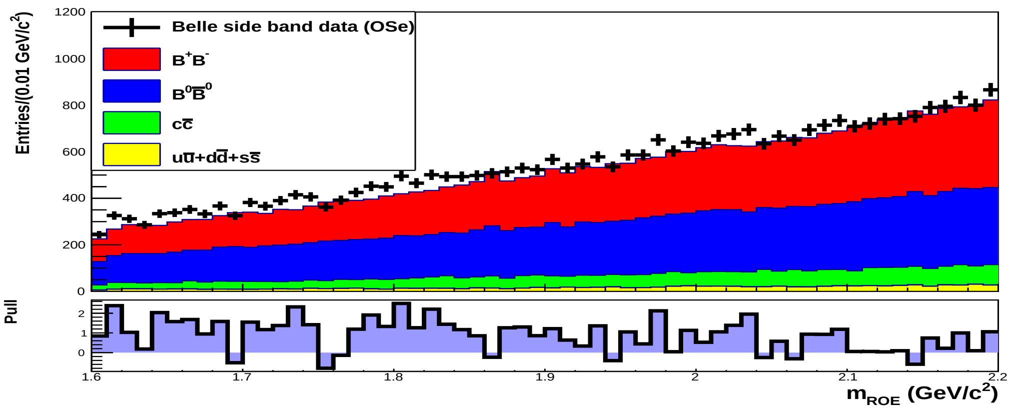


Figure 5.19: The distribution of m_{ROE} for OSe sideband region.

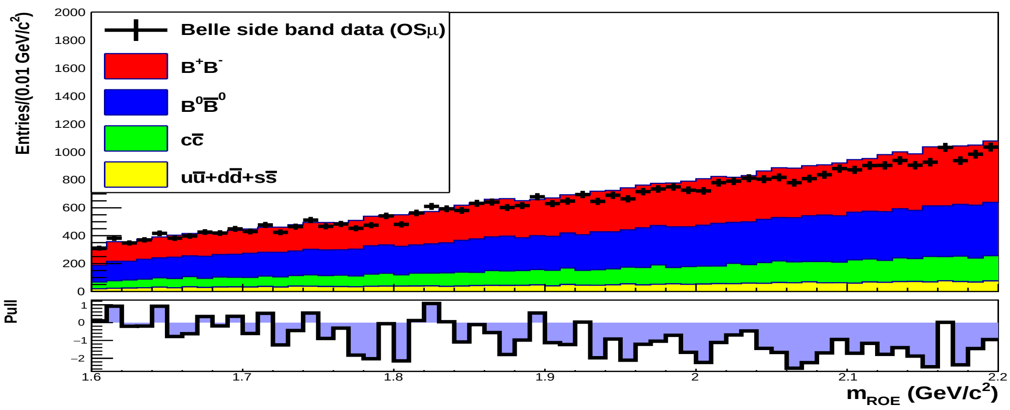


Figure 5.20: The distribution of m_{ROE} for OSmu sideband region.

Chapter 6

Signal Extraction

"In God we trust. All others must bring data."
W. Edwards Deming

In this chapter, we discuss the fitting strategy for signal and background events and briefly review the statistical methods used in this analysis. A detailed discussion of these methods can be found in [62]. The BDT response is used as the final discriminant variable, and the signal extraction is performed through a template-based extended maximum likelihood fit.

The most general method of estimation is the method of maximum likelihood. In the next section, the general concept of this method is described.

6.1 Maximum Likelihood Method

Suppose that we have a set of N independently measured quantities x_i, obtained from a probability distribution function (pdf) $f(x;\theta)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ is a set of n parameters whose values are unknown. The maximum likelihood method takes the estimators $\hat{\theta}$ to be the values of θ that maximize the likelihood function,

$$L(\theta) = \prod_{i=1}^{N} f(x_i;\theta). \tag{6.1}$$

This likelihood function is the joint pdf for the data and is regarded as a function of the parameters. Generally, it is easier to work with the log-likelihood function $\ln L$ (since both are maximized for the same parameter values θ). By solving the following likelihood equations, the maximum likelihood (ML) estimators can be found:

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \quad , \quad i = 1, 2, \ldots, n. \tag{6.2}$$

The ML estimators found in this way are approximately unbiased, and for large data samples, they are efficient. For small data samples, the extended maximum likelihood method is used, which is discussed in the next subsection.

6.1.1 Extended Maximum Likelihood

For a small data sample, the unbinned maximum likelihood method is preferred. In this case, binning may lead to loss of information and larger statistical errors in parameter estimation. In the extended formulation, the observed sample size is treated as a Poisson-distributed variable.

In our case, the unbinned extended likelihood function can be written as:

$$\mathcal{L}(N_s, N_b, \theta) = \frac{e^{-(N_s+N_b)}}{N!} \prod_{i=1}^{N} \Big(N_s\, f_s(x^i;\theta) + N_b\, f_b(x^i;\theta) \Big), \tag{6.3}$$

where N is the number of events in data, N_s and N_b are the expected signal and background yields, and f_s and f_b are the signal and background pdfs (normalized to unity).

In this analysis, the expected signal yield is parameterized in terms of the branching fraction as

$$N_s = \epsilon_{\text{reco}} \times \text{BF} \times N_{B\overline{B}}, \tag{6.4}$$

with

$$\epsilon_{\text{reco}} = \epsilon_{\text{reco.}}^{\tau\to\pi/\rho} + \epsilon_{\text{reco.}}^{\tau\to\text{all other}}, \tag{6.5}$$

where $\epsilon_{\text{reco.}}^{\tau\to\pi/\rho}$ is the reconstruction efficiency for correctly reconstructed signal $\tau \to \pi/\rho\,\nu_\tau$ channel, and $\epsilon_{\text{reco.}}^{\tau\to\text{all other}}$ accounts for cross-feed contributions from other τ decay modes reconstructed in the signal selection. Here, $N_{B\overline{B}}$ is the number of $B\overline{B}$ pairs in the dataset.

The logarithm of the extended likelihood function (dropping the constant term $\ln N!$) is given by:

$$\ln \mathcal{L}(N_s, N_b, \theta) = -(N_s + N_b) + \sum_{i=1}^{N} \ln\Big(N_s f_s(x^i;\theta) + N_b f_b(x^i;\theta) \Big). \tag{6.6}$$

The floating parameters in the nominal fit are BF (through N_s) and N_b, while the shape parameters θ are fixed from template studies, as described in the next section.

6.2 Fitting Model

We use a simultaneous fit in this analysis. In the first step, the signal pdf and background pdf are obtained using template fits to the $\tau \to \pi\nu_\tau$ and $\tau \to \rho\nu_\tau$ samples. In both cases, the signal pdf is composed of two sub-pdfs: one describes the correctly reconstructed signal events, and the second describes the cross-feed signal events. The background pdf is the sum of the pdf for the $B\overline{B}$ background component and the pdf for the $q\overline{q}$ background component. The final fitting model is formed by combining the signal and background pdfs, and a simultaneous fit is performed.

The fit is performed simultaneously in the two τ decay modes $\tau \to \pi\nu_\tau$ and $\tau \to \rho\nu_\tau$, with a common branching fraction parameter BF shared between them. The relative signal contributions in the two modes are determined by their corresponding reconstruction efficiencies entering the definition of N_s.

The fit is unbinned and implemented within the `RooFit` framework. The discriminating variable used in the fit is the response of the Boosted Decision Tree (BDT). The BDT classifiers are trained independently for the $\tau \to \pi\nu_\tau$ and $\tau \to \rho\nu_\tau$ modes, and the corresponding BDT responses are fitted separately within the simultaneous framework.

The signal and background shapes of the BDT response are modeled using `RooHistPdf` objects constructed from Monte Carlo templates. These templates are normalized to unity and describe the expected distribution of the BDT output for signal and background in each τ decay mode.

The background shapes are treated independently in the two τ decay modes, while the branching fraction parameter is common. Only the overall normalizations (N_s through BF and N_b per mode) are floated in the nominal fit.

Initial checks of the fitting model were made on a single toy dataset generated under the background-only hypothesis (taking BF $= 0$) and assuming a luminosity corresponding to five times the Belle integrated luminosity. These distributions are shown in Fig. 6.1 (SSe mode), Fig. 6.2 (SSmu mode), Fig. 6.3 (OSe mode), and Fig. 6.4 (OSmu mode). The model describes the single toy dataset well. As a next step, we test the stability of the fit using an ensemble of toy datasets, discussed in the next section.

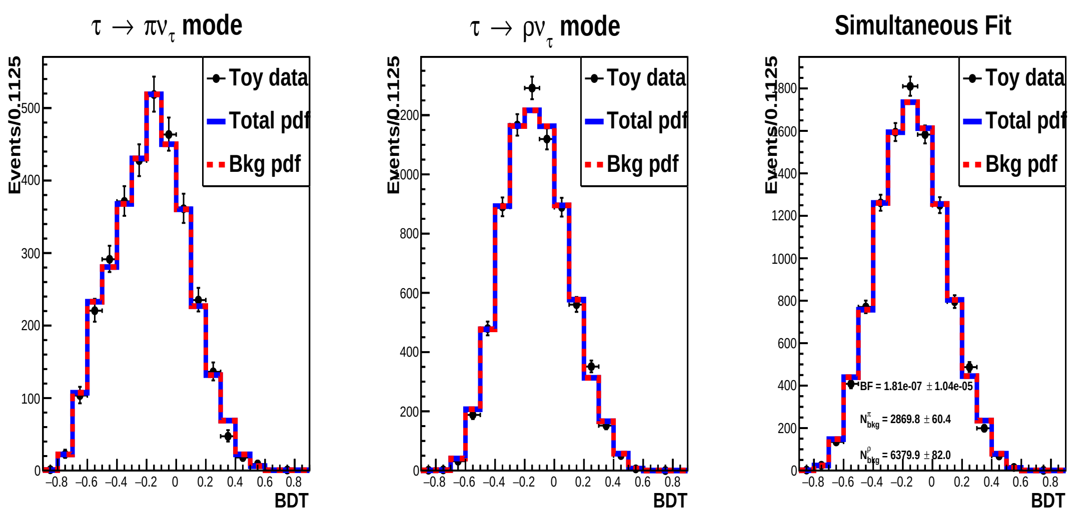


Figure 6.1: SSe mode fitting templates on a single toy dataset.

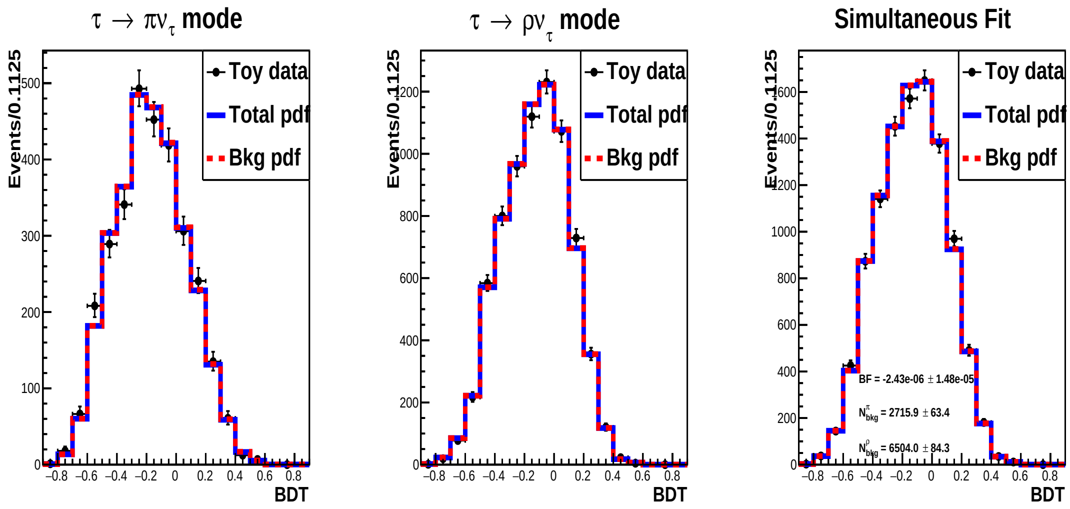


Figure 6.2: SSmu mode fitting template on a single toy dataset.

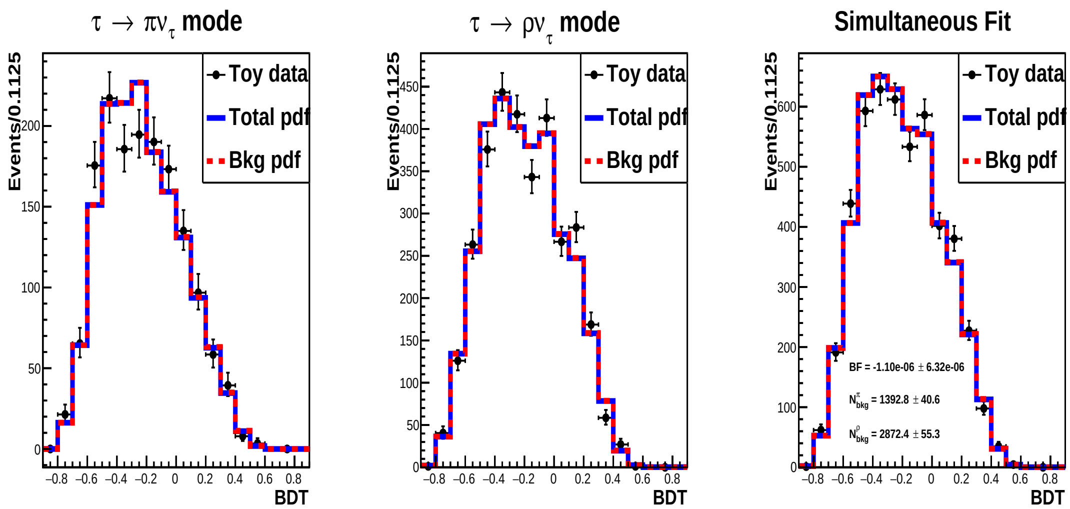


Figure 6.3: OSe mode fitting template on a single toy dataset.

6.2.1 Stability of Fit

To check the stability of the fit, we generated an ensemble of one thousand toy datasets. Each dataset is generated in the same way as described above. The fit is performed on each toy dataset, and we compute the pull distribution for BF:

$$\text{Pull} = \frac{\text{BF}^{\text{toy}}_{\text{gen}} - \text{BF}^{\text{toy}}_{\text{fit}}}{\sigma^{\text{toy}}_{\text{BF}}}, \tag{6.7}$$

where $\text{BF}^{\text{toy}}_{\text{gen}}$ is the generated branching fraction, $\text{BF}^{\text{toy}}_{\text{fit}}$ is the fitted value, and $\sigma^{\text{toy}}_{\text{BF}}$ is the fit uncertainty returned for that toy experiment. The pull distribution should ideally be Gaussian with mean 0 and width 1. Deviations from these ideal values are used to quantify potential fit biases and under/over-coverage.

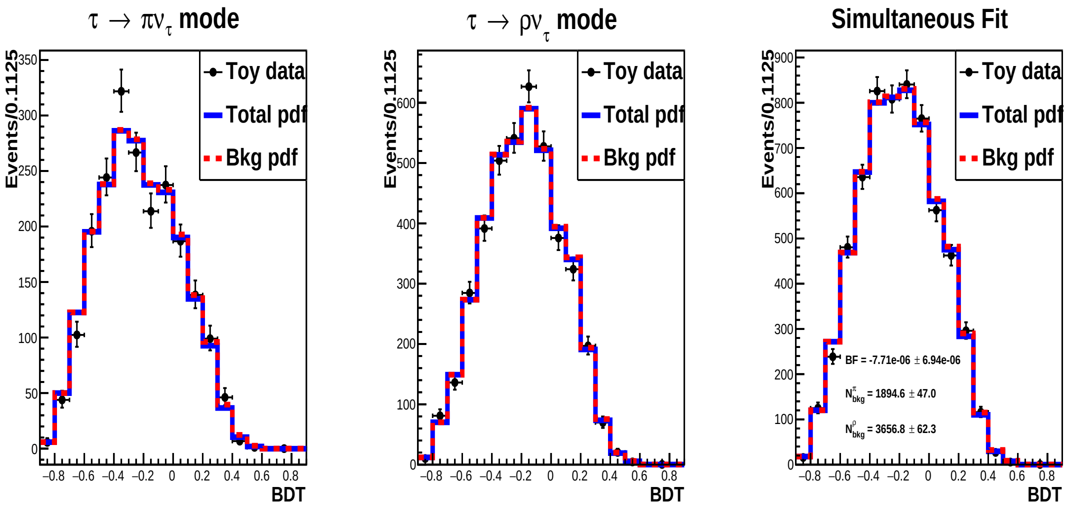


Figure 6.4: OSmu mode fitting template on a single toy dataset.

The distributions of BF (fitted with a double-Gaussian function with a common mean), the BF uncertainty (fitted with a Crystal Ball function), the BF pull (fitted with a single Gaussian function), and the negative log-likelihood are shown in Fig. 6.5 (SSe), Fig. 6.6 (SSmu), Fig. 6.7 (OSe), and Fig. 6.8 (OSmu). All pull distributions have widths close to one, and their means are shifted slightly towards negative values. This shift is treated as a systematic uncertainty. Similar studies are performed for the expected number of background events; the corresponding distributions are given in Appendix B.

As a next step in checking the performance of the fit, we perform the linearity test, which is discussed in the next section.

6.3 Linearity Test

In this test, the fit procedure is repeated for different generated values of the floating parameter. The mean fitted values are then compared to the generated values to check linearity. In this analysis, we repeated the fit (six times) for one thousand toy datasets, taking the generated BF values in the range $(0\text{–}5) \times 10^{-5}$. The mean values of BF and the mean uncertainties are stored and used to produce the linearity plots shown in Fig. 6.9 (SS modes) and Fig. 6.10 (OS modes)[1]. All distributions are consistent with linear behavior, with slopes close to one and intercepts compatible with zero. The fit parameters are given in Table 6.1 for all linearity plot distributions. This further confirms the stability of the fit procedure.

Now that we have established and validated our fitting procedure, we discuss the upper-limit estimation for all signal decay modes in the next chapter.

[1]The errors are small, which is why they are not visible on the plots.

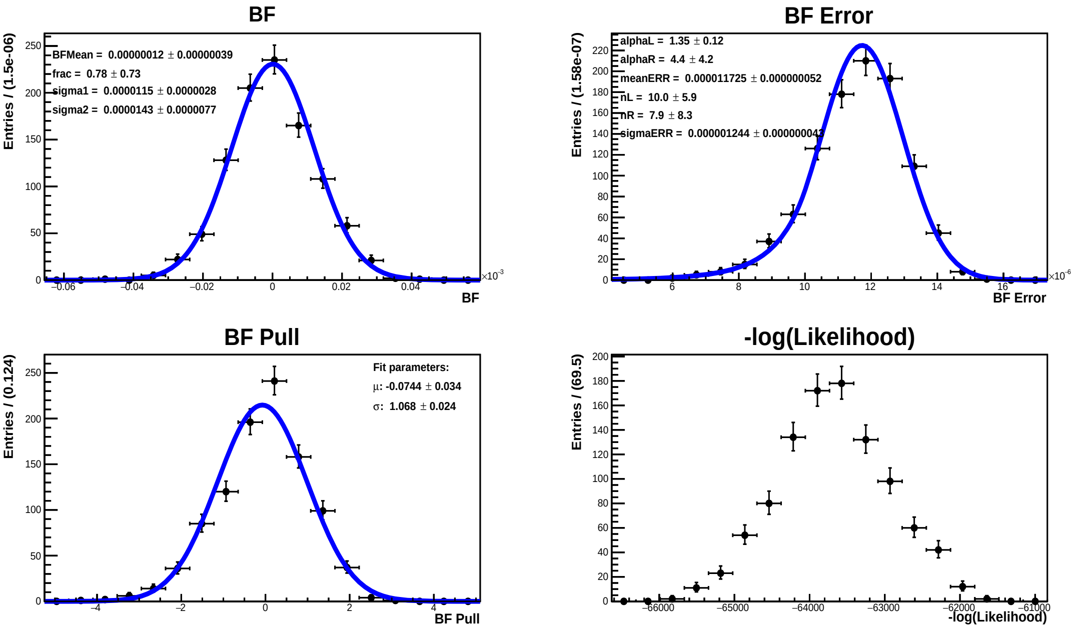


Figure 6.5: The BF distributions for SSe mode.

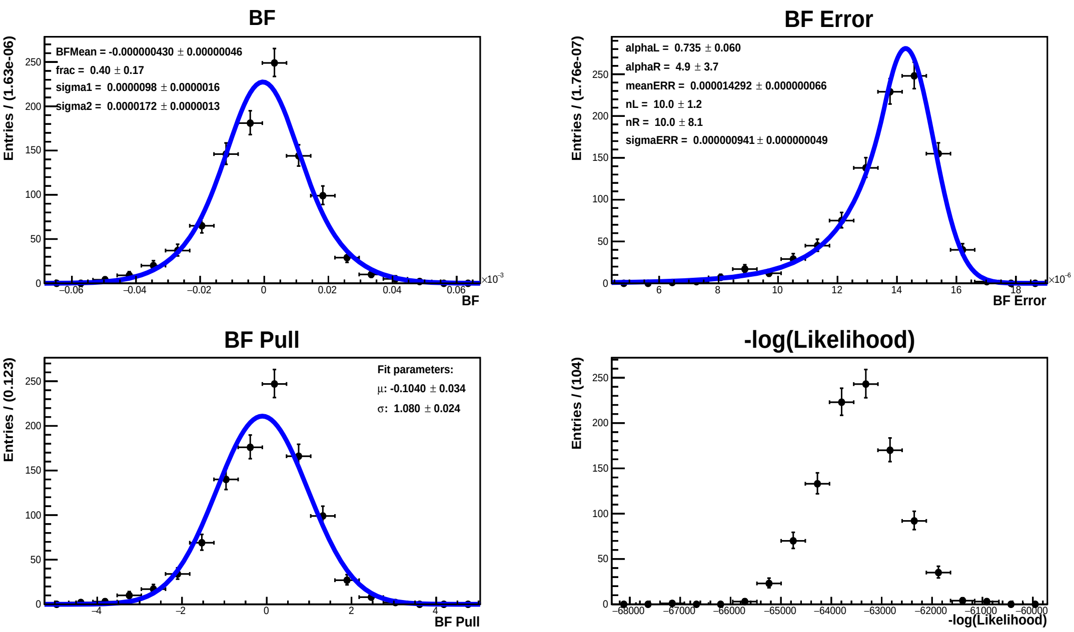


Figure 6.6: The BF distributions for SSmu mode.

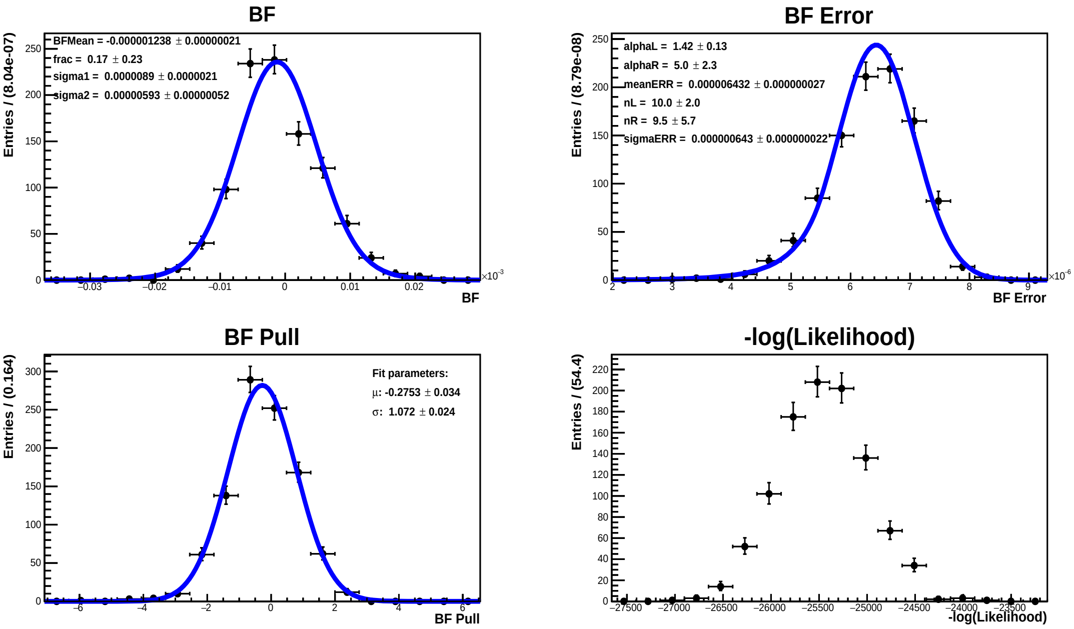


Figure 6.7: The BF distributions for OSe mode.

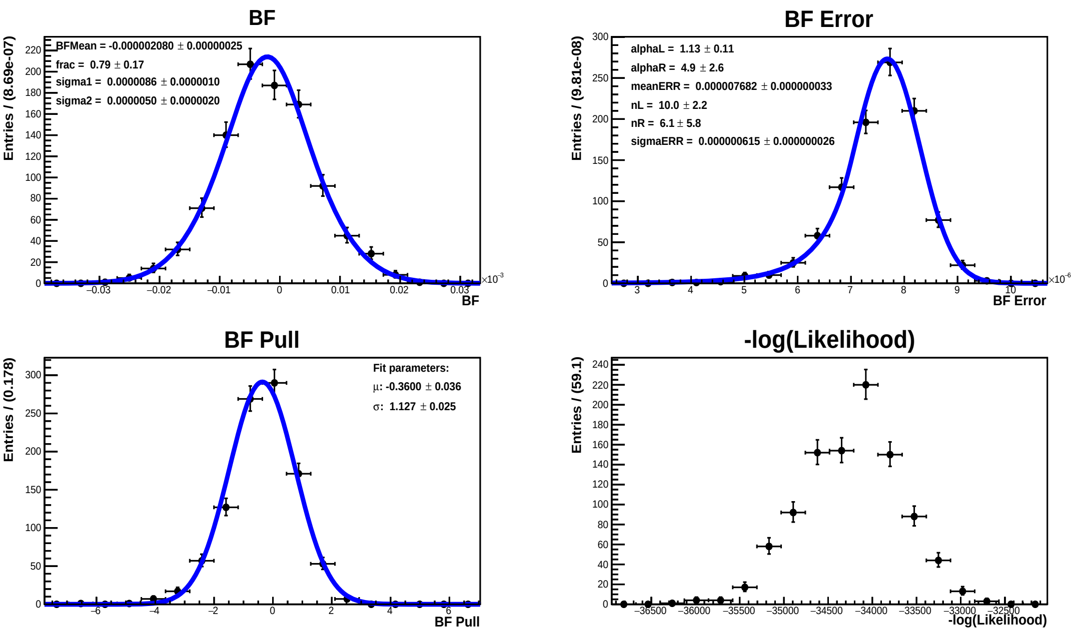
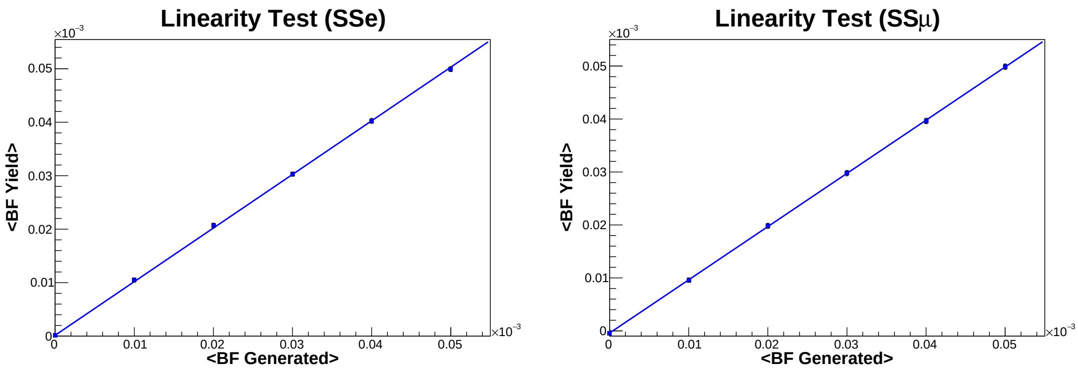
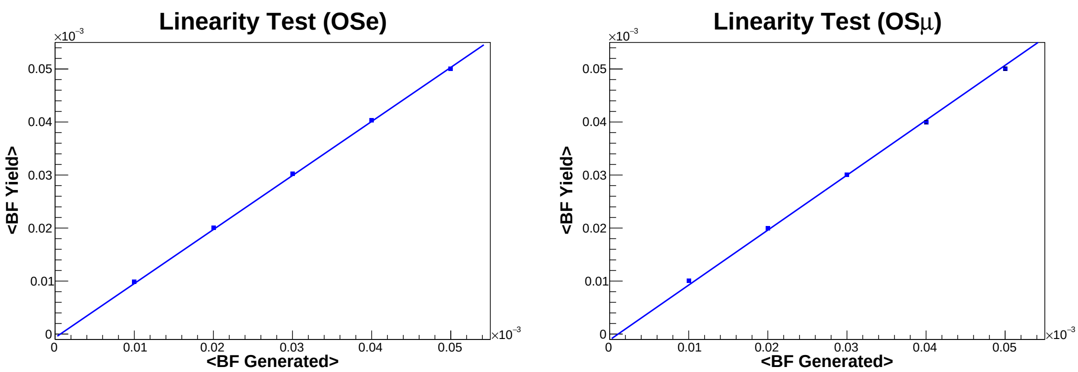


Figure 6.8: The BF distributions for OSmu mode.



(a) (b)

Figure 6.9: Linearity test distributions (a) SSe mode (b) SSmu mode.



(a) (b)

Figure 6.10: Linearity test distributions (a) OSe mode (b) OSmu mode.

Signal Decay Mode	y-intercept ($\times 10^{-7}$)	slope
SSe	-6.57 ± 1.5	1.01 ± 0.004
SSmu	-0.11 ± 1.8	1.03 ± 0.006
OSe	1.66 ± 1.2	1.00 ± 0.006
OSmu	-4.14 ± 3.3	1.00 ± 0.011

Table 6.1: Fit parameters of linearity test.

Chapter 7

Results and Discussion

"A physicist is just an atom's way of looking at itself."
Niels Bohr

In this chapter, we present the statistical upper limit (UL) estimation for all signal decay modes based on the simultaneous unbinned extended maximum likelihood fit to the BDT response described in the Chapter 4.

7.1 Statistical Upper Limit Estimation

The upper limits (ULs) on the signal branching fractions are determined using a Bayesian approach [63]. At this stage, only statistical uncertainties are considered.

For a given dataset, the likelihood function is obtained from the profile likelihood of the branching fraction parameter BF, while all nuisance parameters are fixed at their fitted values. The negative log profile likelihood ratio is defined as

$$-2\ln\left(\frac{L(\mathrm{BF})}{L_{\max}}\right),$$

where $L(\mathrm{BF})$ is the likelihood for a fixed value of BF and $L_{\max}$ is the global maximum likelihood value.

In the Bayesian framework with a uniform prior for $\mathrm{BF} \geq 0$, the upper limit at 90% credibility level (C.L.) is obtained by integrating the likelihood function over the physical region:

$$\int_0^{\mathrm{UL}} L(\mathrm{BF})\, d\mathrm{BF} = 0.90 \int_0^{\infty} L(\mathrm{BF})\, d\mathrm{BF}.$$

The upper bound of this interval defines the statistical UL.

This procedure is illustrated for a single toy dataset in Fig. 7.1. The likelihood curve is integrated from $\mathrm{BF} = 0$ to the value enclosing 90% of the total area. The corresponding upper bound (approximately 5×10^{-5} in this example) represents the UL for this toy dataset.

To evaluate the expected sensitivity of the analysis, the UL estimation is repeated for one thousand toy datasets generated under the background-only hypothesis. The distributions of the resulting UL values are shown in Figs. 7.2–7.5. These distributions exhibit asymmetric shapes characteristic of upper limit estimators near the physical boundary. A Landau function is used to model the distributions, and the most probable value (MPV) of the fit is taken as the expected statistical upper limit for each mode. The resulting values are summarized in Table 7.1 and lie in the range $(0.9 - 2.0) \times 10^{-5}$.

This shows the effectiveness of the analysis approach that we have used in this study.

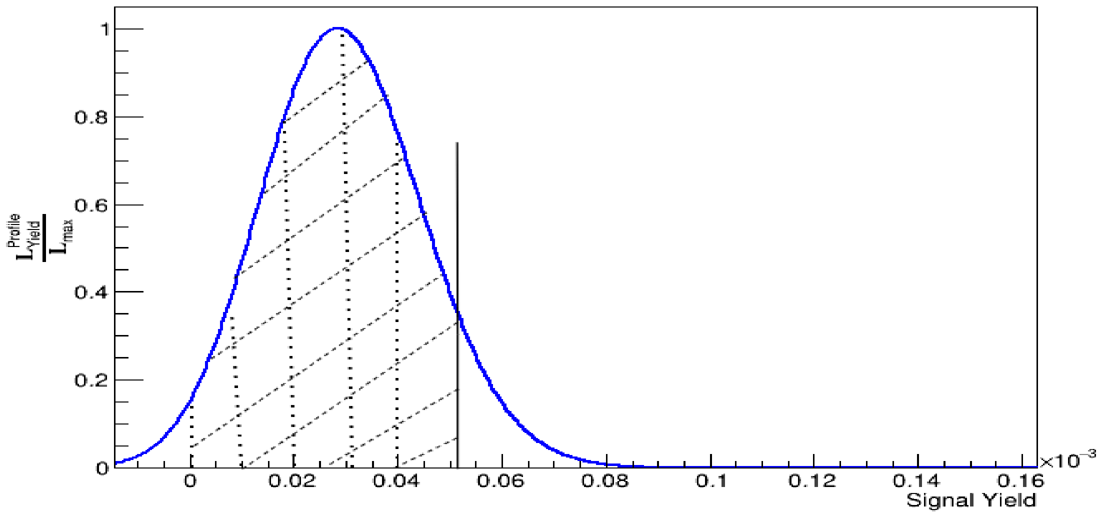


Figure 7.1: UL estimation for a single toy (SSmu mode).

Signal Decay Mode	Estimated Statistical UL ($\times 10^{-5}$)
SSe	1.7
SSmu	2.0
OSe	0.9
OSmu	1.0

Table 7.1: Estimated statistical ULs.

For the case of OS modes, we have relatively fewer background events compared to the SS modes; therefore, the resulting statistical ULs for OS modes are relatively stringent than for the SS modes.

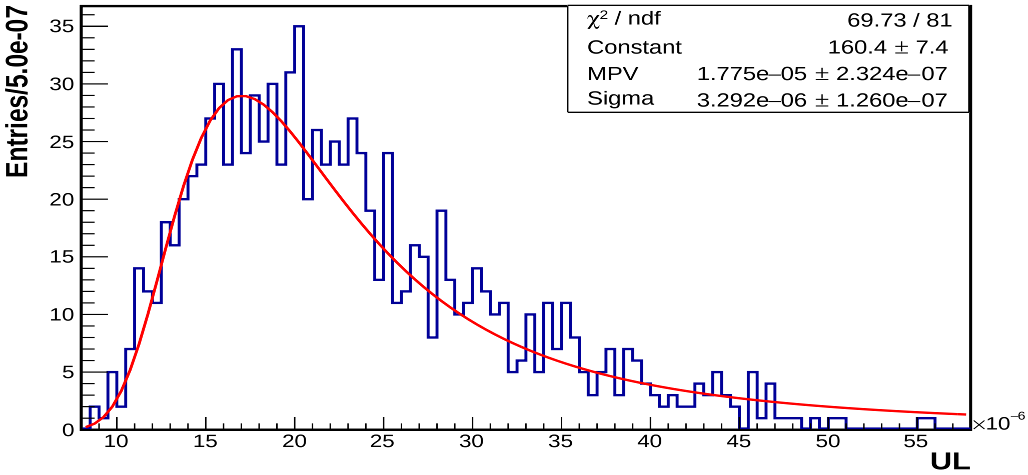


Figure 7.2: The distribution of statistical UL for SSe mode.

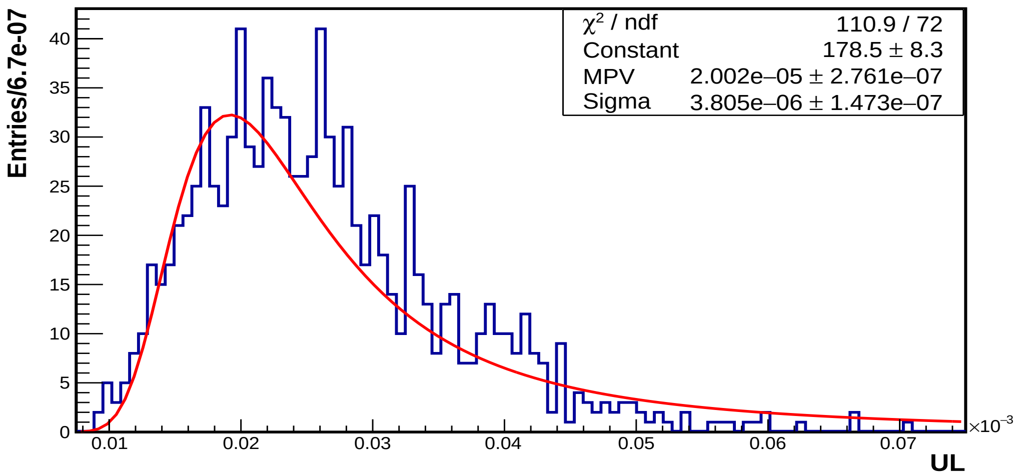


Figure 7.3: The distribution of statistical UL for SSmu mode.

7.2 Discussion

Since the accumulated dataset did not reach the initially anticipated $5\ \mathrm{ab}^{-1}$, the analysis was not extended to the full statistical interpretation stage, and we therefore stop short of performing a final measurement on data. The estimated statistical ULs presented above correspond to the search for $B \to K\tau\ell$ decays using the inclusive semileptonic tagging approach and considering the $\tau \to \pi/\rho\nu_\tau$ decay modes.

These results are complementary to previous studies based on hadronic tagging and leptonic τ decay modes [10]. The inclusion of semileptonic tagging and hadronic τ modes increases the overall signal efficiency and provides an independent sensitivity channel.

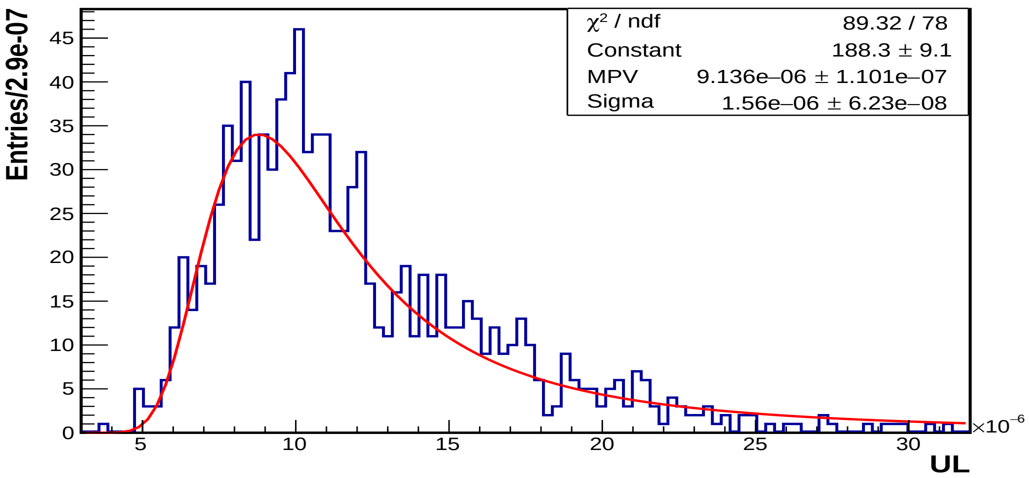


Figure 7.4: The distribution of statistical UL for OSe mode.

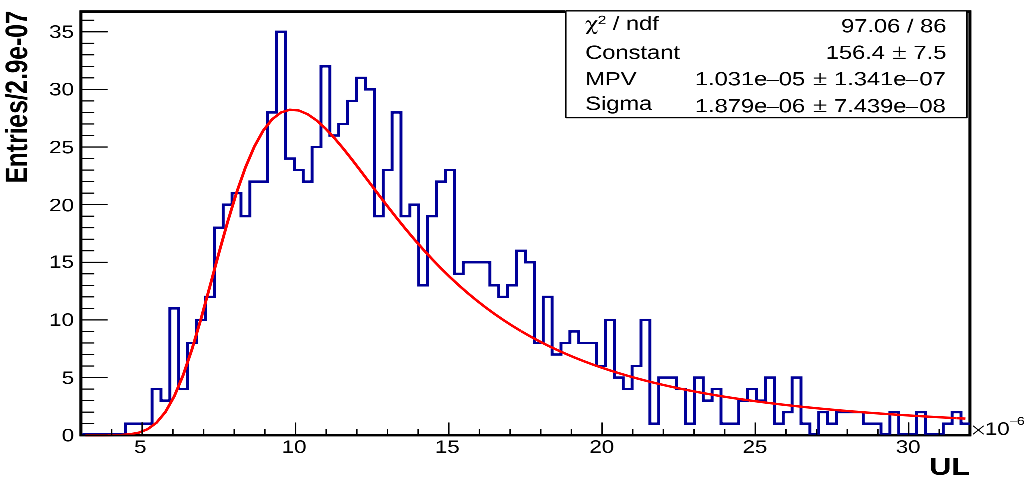


Figure 7.5: The distribution of statistical UL for OSmu mode.

With the significantly larger dataset expected at Belle II, this strategy has the potential to improve the current experimental sensitivity and probe branching fractions well below the present limits. It can be further improved using the following additional steps.

7.2.1 Semileptonic and Hadronic tag Joint Fit

A combined fit of inclusive semileptonic and hadronic tagging strategies is expected to significantly improve the overall sensitivity by increasing signal efficiency and reducing systematic correlations.

7.2.2 Inclusion of Leptonic τ Decay Modes

The addition of leptonic τ decay modes ($\tau \to \ell\nu_\ell\nu_\tau$) will further increase the sensitivity of this analysis.

7.2.3 Oppositely Charged Lepton tag

Another promising extension of this analysis is the implementation of the oppositely charged lepton tagging approach for the reconstruction of B_{tag}. In this configuration, the tag-side B meson is reconstructed as

$$B^- \to D^0(\to X^-\ell^+\nu_\ell)X^-.$$

In this topology, the electric charge of the tag-side lepton is correlated with the charge of the B_{tag}, providing an additional handle for background suppression. This charge correlation introduces further discrimination power between signal and combinatorial background contributions.

The leptons produced in this decay chain are typically softer, reflecting the kinematics of semileptonic charm decays. Preliminary studies of the tag-side lepton momentum distributions are shown in Fig. 7.6 and Fig. 7.7 for SS and OS modes, respectively. These studies indicate that the inclusion of the oppositely charged lepton tag increases the available event sample while maintaining controllable background levels.

The implementation of this tagging strategy is therefore expected to enhance the overall sensitivity of the analysis.

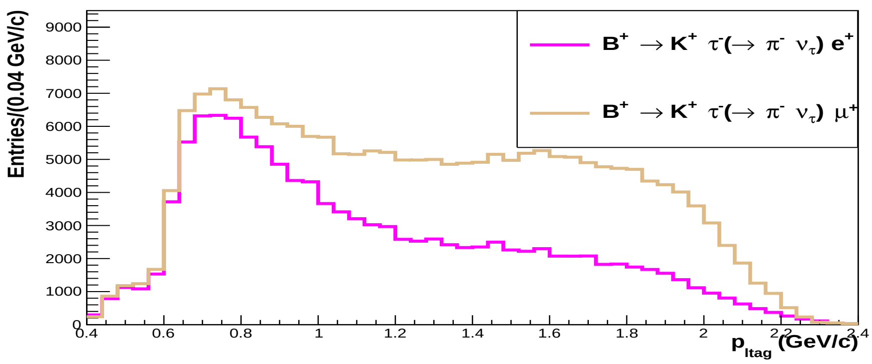


Figure 7.6: The preliminary distribution of p_{tag} for SS decay modes in the case of the oppositely charged lepton tag approach.

7.3 Systematic Uncertainties

The evaluation of systematic uncertainties in this analysis requires detailed studies based on real Belle II data. Since the signal region remains blinded, a complete determination of all systematic effects has not yet been performed.

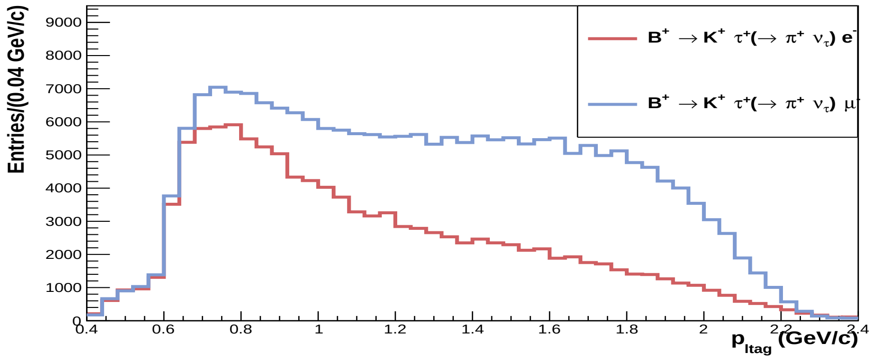


Figure 7.7: The preliminary distribution of p_{tag} for OS modes in the case of the oppositely charged lepton tag approach.

Several sources of systematic uncertainty are expected to contribute to the final result:

- Reconstruction efficiency uncertainties for $\tau \to \pi\nu_\tau$ and $\tau \to \rho\nu_\tau$,
- Data/MC differences in the BDT response modeling,
- Limited Monte Carlo statistics used to construct the `RooHistPdf` templates,
- Uncertainty on the total number of $B\overline{B}$ pairs,
- Possible residual mismodeling of background shapes.

Most of these effects require control samples in the data and dedicated calibration procedures. In particular, corrections to the BDT response and reconstruction efficiencies must be derived from independent data-driven studies.

Because the analysis is performed in a blind framework, the final systematic uncertainty evaluation will only be completed after unblinding. Therefore, the upper limits presented in this chapter correspond to the expected statistical sensitivity of the analysis.

A full systematic uncertainty treatment, including incorporation into the likelihood model through constrained nuisance parameters, will be performed in the final result.

In the next chapter, we summarize this work and discuss the outlook of this analysis for future case scenarios.

Chapter 8

Summary and Outlook

"Somewhere, something incredible is waiting to be known."
Carl Sagan

In this work, a novel strategy has been developed to search for the lepton flavour violating decay $B \to K\tau\ell$ with hadronic τ decays $\tau \to \pi\nu_\tau$ and $\tau \to \rho\nu_\tau$.

The key innovation of this analysis lies in the kinematic reconstruction of the signal B meson using geometrical constraints specific to the B-factory environment. By exploiting the known initial state and the cone-intersection method, the signal-side kinematics can be reconstructed despite the presence of missing neutrinos. This allows the construction of powerful discriminating observables that suppress the background while maintaining signal efficiency.

The analysis combines signal-side kinematic constraints with inclusive semileptonic tagging on the B_{tag} side, significantly increasing usable statistics compared to fully hadronic tagging approaches. Signal extraction is performed through a simultaneous unbinned extended maximum likelihood fit to the BDT response in the $\tau \to \pi\nu_\tau$ and $\tau \to \rho\nu_\tau$ channels, with a common branching fraction parameter. The robustness of the fitting framework has been validated through extensive toy Monte Carlo studies, including pull and linearity tests.

Control channel studies using the full Belle dataset demonstrate good agreement between data and simulation in both kinematic and multivariate distributions. Additional sideband studies confirm the stability of the background modeling. At the current stage, the analysis remains blinded; therefore, the quoted upper limits correspond to the expected statistical sensitivity.

The achieved expected sensitivity demonstrates that inclusive semileptonic tagging combined with hadronic τ decays constitutes a competitive and complementary approach in searches for LFV in B decays.

Outlook

Several directions can further enhance the sensitivity of this search.

First, a full data-driven calibration of reconstruction efficiencies and BDT response shapes will be performed after unblinding, allowing a complete treatment of systematic uncertainties within the likelihood framework.

Second, the inclusion of oppositely charged lepton tagging provides additional charge-correlation discrimination and is expected to improve background suppression while increasing effective statistics.

Finally, with the rapidly growing Belle II dataset, the scalability of this method becomes particularly relevant. The combination of inclusive semileptonic tagging, hadronic tagging, and hadronic τ decays opens a pathway to significantly improved sensitivity to $B \to K\tau\ell$ decays in the near future.

The developed framework is fully compatible with Belle II conditions and provides a solid foundation for future precision searches for lepton flavour violation in the heavy flavour sector.

Appendix A

Neutrino Momentum Calculation

In this analysis, the B meson decays to a K^+ and a pair of leptons, i.e.

$$B^+ \to K^+\tau^\pm(\to \pi^\pm\nu_\tau)\ell^\mp$$

The mass of the B meson in terms of its four-momentum can be written as:

$$m_B^2 = P_B^\mu P_{B\mu} \tag{A1}$$

It can be written in terms of energy and three momentum of B mesons as

$$m_B^2 = E_B^2 - p_B^2, \tag{A2}$$

where $E_B = E_{K\ell\pi} + E_\nu$ and $p_B = |\vec{p}_{K\ell\pi} + \vec{p}_\nu|$. The above equation can be further simplified as follows:

$$m_B^2 = (E_{K\ell\pi} + E_\nu)^2 - (\vec{p}_{K\ell\pi} + \vec{p}_\nu)^2 \tag{A3}$$

After a few steps of simplification, we can obtain the following expression for the cosine angle between the neutrino momentum and the $K\ell\pi$ momentum.

$$cos\angle(\vec{p}_\nu, \vec{p}_{K\ell\pi}) = \frac{2E_{K\ell\pi}E_\nu + m_{K\ell\pi}^2 - m_B^2}{2|\vec{p}_\nu||\vec{p}_{K\ell\pi}|} \tag{A4}$$

As a neutrino has a negligible mass, taking it equal to zero, the above equation can be further written as follows:

$$cos\angle(\vec{p}_\nu, \vec{p}_{K\ell\pi}) = \frac{2E_{K\ell\pi}(E_{\text{beam}} - E_{K\ell\pi}) + m_{K\ell\pi}^2 - m_B^2}{2|\vec{p}_{K\ell\pi}|(E_{\text{beam}} - E_{K\ell\pi})} \tag{A5}$$

Similarly, the cosine angle between the neutrino momentum and the pion momentum is written as:

$$cos\angle(\vec{p}_\nu, \vec{p}_\pi) = \frac{2E_\pi(E_{\text{beam}} - E_{K\ell\pi}) + m_\pi^2 - m_\tau^2}{2|\vec{p}_\pi|(E_{\text{beam}} - E_{K\ell\pi})} \tag{A6}$$

Now for the calculation of neutrino momentum, we consider the two cones with a

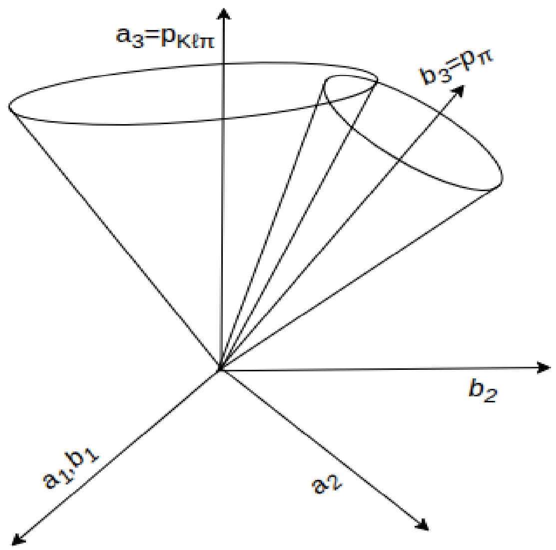


Figure A.1: Two frames of references for the cones.

set of axes, as shown in Fig. A.1. The two cones are further elaborated separately in Fig. A.2. The following two bases are used to parameterize the cones:

$$\mathcal{A} = \{\hat{a}_1, \hat{a}_2, \hat{a}_3\} = \{\frac{\hat{p}_\pi \times \hat{p}_{K\ell\pi}}{|\hat{p}_\pi \times \hat{p}_{K\ell\pi}|}, \frac{\hat{p}_\pi - cos\angle(p_\pi, p_{K\ell\pi})\hat{p}_{K\ell\pi}}{|\hat{p}_\pi \times \hat{p}_{K\ell\pi}|}, \hat{p}_{K\ell\pi}\}^{1} \tag{A7}$$

$$\mathcal{B} = \{\hat{b}_1, \hat{b}_2, \hat{b}_3\} = \{\frac{\hat{p}_\pi \times \hat{p}_{K\ell\pi}}{|\hat{p}_\pi \times \hat{p}_{K\ell\pi}|}, \frac{cos\angle(p_\pi, p_{K\ell\pi})\hat{p}_\pi - \hat{p}_{K\ell\pi}}{|\hat{p}_\pi \times \hat{p}_{K\ell\pi}|}, \hat{p}_\pi\} \tag{A8}$$

where $\hat{a}_3 \equiv \hat{p}_{K\ell\pi}$, $\hat{b}_3 \equiv \hat{p}_\pi$. In these bases, $\hat{a}_1$ and $\hat{b}_1$ are collinear and are orthogonal to the $p_\pi - p_{K\ell\pi}$ plane. For two bases to be orthonormal, $\hat{a}_2$ and $\hat{b}_2$ are chosen in the form mentioned above. Now we can parameterize the two cones for neutrino momentum as follows:

$$[p_\nu(\varphi)]_\mathcal{B} = |p_\nu| \begin{bmatrix} \sin\angle(p_\nu, p_\pi)\cos\varphi) \\ \sin\angle(p_\nu, p_\pi)\sin\varphi) \\ \cos\angle(p_\nu, p_\pi) \end{bmatrix}, \varphi \in [0, 2\pi] \tag{A9}$$

$$[p_\nu(\varphi')]_\mathcal{A} = |p_\nu| \begin{bmatrix} \sin\angle(p_\nu, p_{K\ell\pi})\cos\varphi') \\ \sin\angle(p_\nu, p_{K\ell\pi})\sin\varphi') \\ \cos\angle(p_\nu, p_{K\ell\pi}) \end{bmatrix}, \varphi' \in [0, 2\pi] \tag{A10}$$

These two parametric equations represent the neutrino momentum in two different bases. For the intersection of two cones, the following transformation is required.

$$[p_\nu(\varphi')]_\mathcal{A} = \mathcal{T}_{\mathcal{B}\to\mathcal{A}}[p_\nu(\varphi)]_\mathcal{B} \tag{A11}$$

[1] $cos\angle(\vec{p}_\pi, \vec{p}_{K\ell\pi}) = \frac{\vec{p}_{pi}\cdot\vec{p}_{K\ell\pi}}{|\vec{p}_{pi}||\vec{p}_{K\ell\pi}|}$

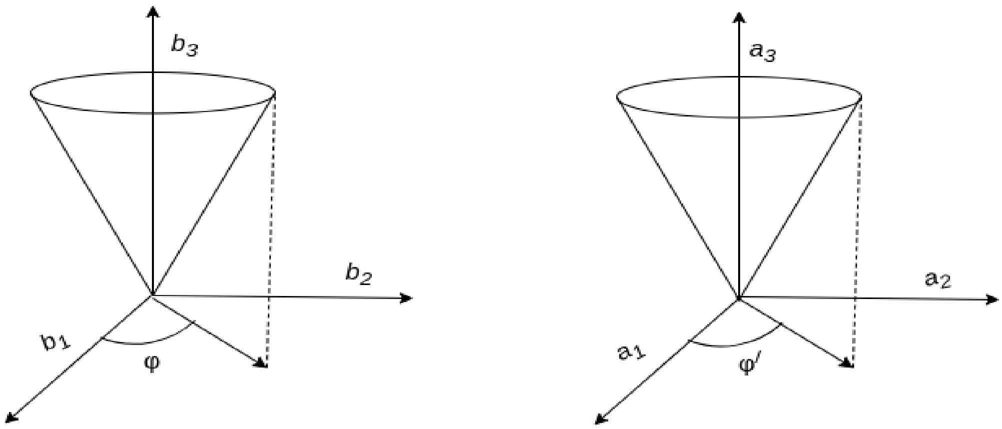


Figure A.2: The schematic description of two axes frame.

$$= |p_\nu| \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\angle(p_\pi, p_{K\ell\pi}) & \sin\angle(p_\pi, p_{K\ell\pi}) \\ 0 & -\sin\angle(p_\pi, p_{K\ell\pi}) & \cos\angle(p_\pi, p_{K\ell\pi}) \end{bmatrix} \begin{bmatrix} \sin\angle(p_\nu, p_\pi)\cos\varphi) \\ \sin\angle(p_\nu, p_\pi)\sin\varphi) \\ \cos\angle(p_\nu, p_\pi) \end{bmatrix} \tag{A12}$$

$$= |p_\nu| \begin{bmatrix} \sin\angle(p_\nu, p_\pi)\cos\varphi \\ \cos\angle(p_\pi, p_{K\ell\pi})\sin\angle(p_\nu, p_\pi)\sin\varphi + \sin\angle(p_\pi, p_{K\ell\pi})\cos\angle(p_\nu, p_\pi) \\ -\sin\angle(p_\pi, p_{K\ell\pi})\sin\angle(p_\nu, p_\pi)\sin\varphi + \cos\angle(p_\pi, p_{K\ell\pi})\cos\angle(p_\nu, p_\pi) \end{bmatrix} \tag{A13}$$

By comparison of Eqn. A10 and Eqn. A13[2], we obtain:

$$\sin\varphi = \frac{\cos\angle(p_\pi, p_{K\ell\pi})\cos\angle(p_\nu, p_\pi) - \cos\angle(p_\nu, p_{K\ell\pi})}{sin\angle(p_\nu, p_\pi)\sin\angle(p_\pi, p_{K\ell\pi})} \tag{A14}$$

Using the basic trigonometric identity, we can find the expression for $cos\varphi$, i.e.

$$cos\varphi = \pm\sqrt{1 - sin^2\varphi} \tag{A15}$$

After finding the value of $sin\varphi$, the neutrino momentum can be written (with two-fold ambiguity) with respect to the basis $\mathcal{B}$ as:

$$\vec{p}_\nu = |p_\nu| \begin{bmatrix} \pm\sin\angle(p_\nu, p_\pi)\sqrt{1 - sin^2\varphi}) \\ \sin\angle(p_\nu, p_\pi)\sin\varphi) \\ \cos\angle(p_\nu, p_\pi) \end{bmatrix} \tag{A16}$$

In the short notation, it can be written as:

$$\vec{p}_{\nu_{1,2}} = |p_\nu|[\pm x\hat{b}_1 + y\hat{b}_2 + z\hat{b}_3] \tag{A17}$$

where $\hat{b}_1, \hat{b}_2, \hat{b}_3$ are the basis introduced in the Eqn. A8 and the parameters $x = \sin\angle(p_\nu, p_\pi)\sqrt{1 - sin^2\varphi}, y = \sin\angle(p_\nu, p_\pi)\sin\varphi), z = \cos\angle(p_\nu, p_\pi)$ are the coordinates of $\vec{p}_{\nu_{1,2}}$ with respect to the basis $\mathcal{B}$ as described in the Eqn. A9. These two possible neutrino momenta[3] are further added to the $\vec{p}_{K\mu\pi}$ to get the two possible $\vec{p}_{B_{\rm sig}}$ solutions. In this way, we obtain the $B_{\rm sig}$ momentum with two-fold ambiguity.

[2] The third rows of the two matrices are compared

[3] After translating to the centre of mass frame.

Appendix B

Stability Check for Background Modelling

As we are performing a simultaneous fit on the $\tau \rightarrow \pi\nu_\tau$ and $\tau \rightarrow \rho\nu_\tau$ samples, therefore we have two different background distributions (one for $\tau \rightarrow \pi\nu_\tau$ sample and second for $\tau \rightarrow \rho\nu_\tau$ sample). The stability of the fit for the background is checked on a thousand toy datasets. The distributions of fitted background samples, their errors, and the pull distributions for both samples are given in this Appendix. The pull distributions are fitted using a Gaussian function. In all pull distributions, the mean value is around zero, and the sigma is around one.

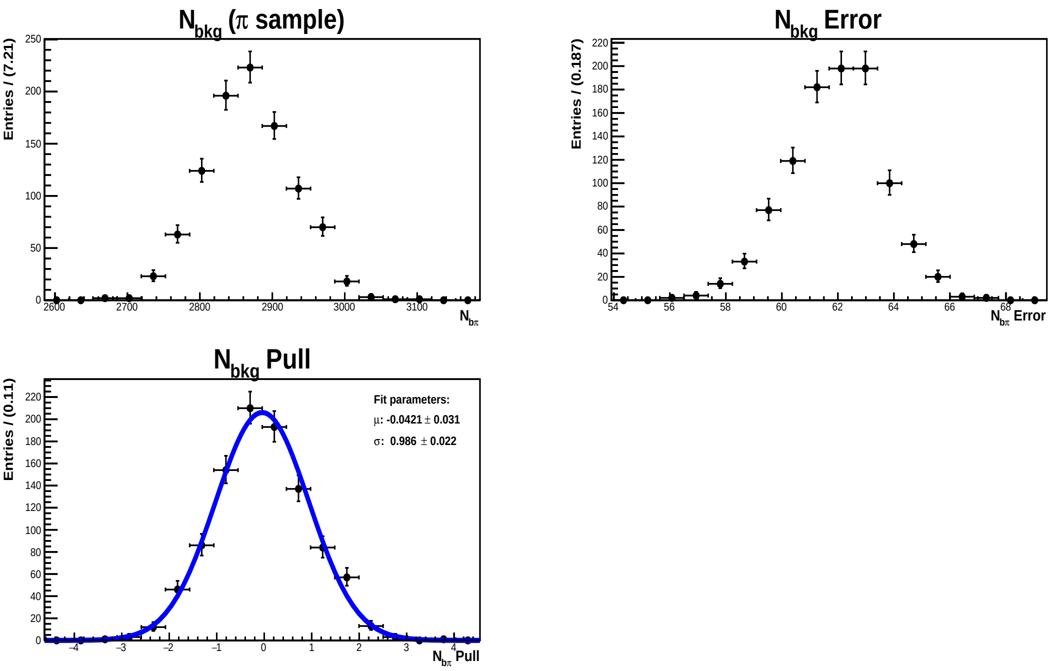


Figure B.1: The expected background distributions ($\tau \rightarrow \pi\nu_\tau$ sample) for SSe mode.

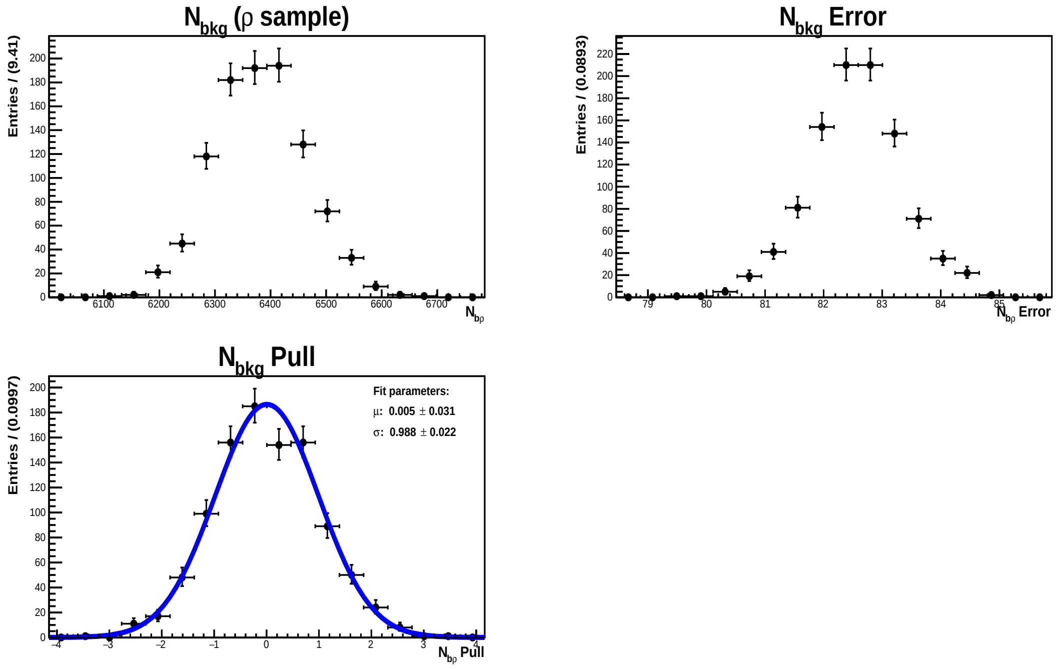


Figure B.2: The expected background distributions ($\tau \to \rho\nu_\tau$ sample) for SSe mode.

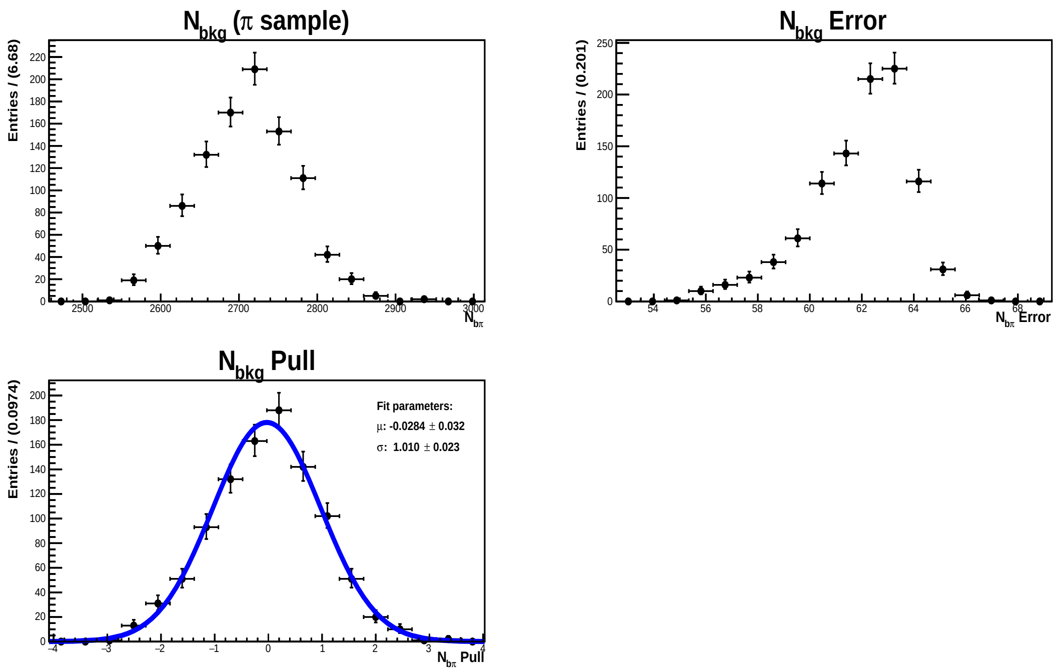


Figure B.3: The expected background distributions ($\tau \to \pi\nu_\tau$ sample) for SSmu mode.

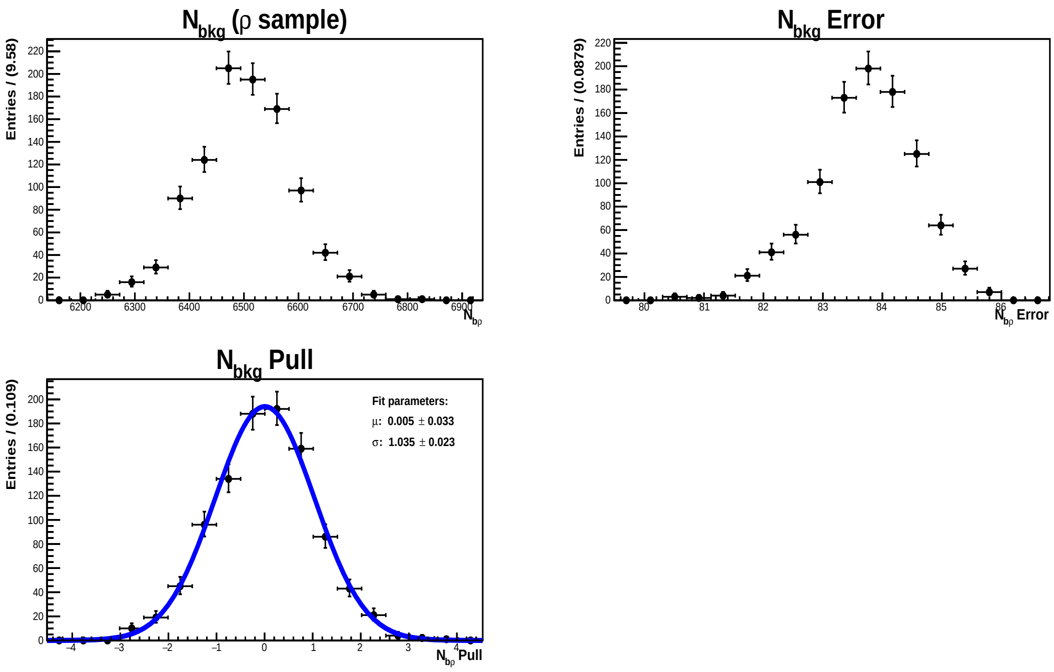


Figure B.4: The expected background distributions ($\tau \to \rho\nu_\tau$ sample) for SSmu mode.

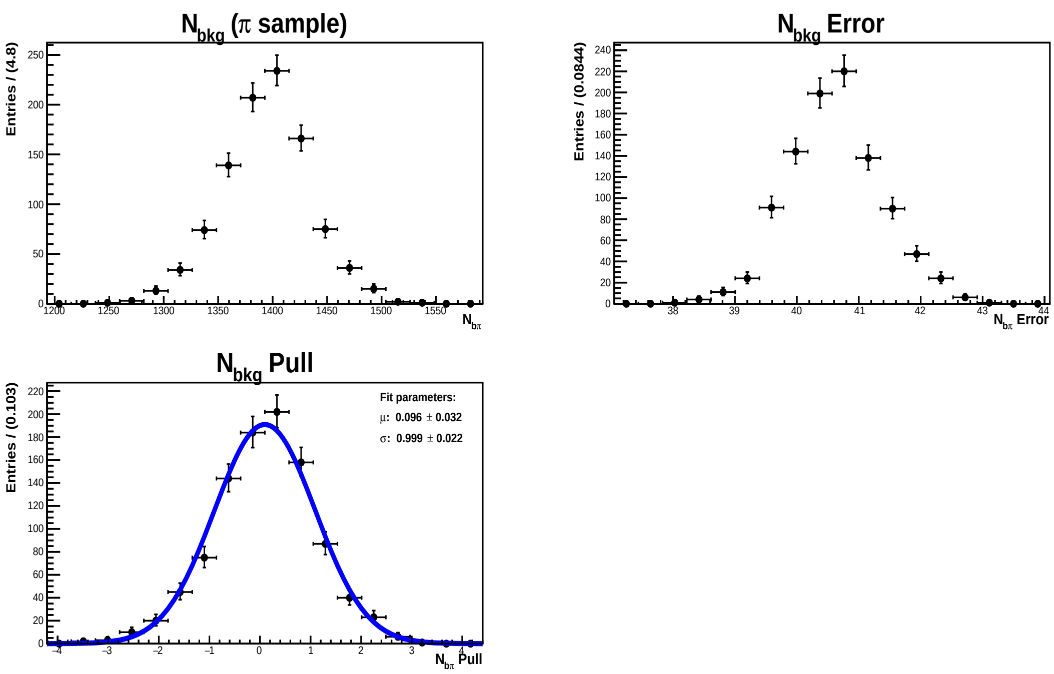


Figure B.5: The expected background distributions ($\tau \to \pi\nu_\tau$ sample) for OSe mode.

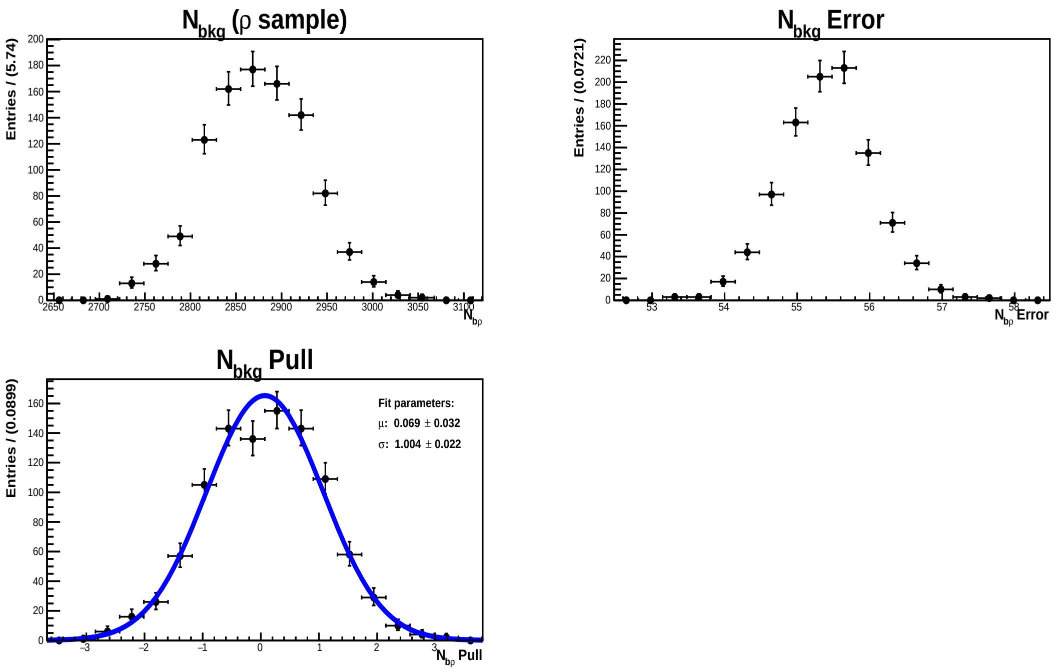


Figure B.6: The expected background distributions ($\tau \to \rho\nu_\tau$ sample) for OSe mode.

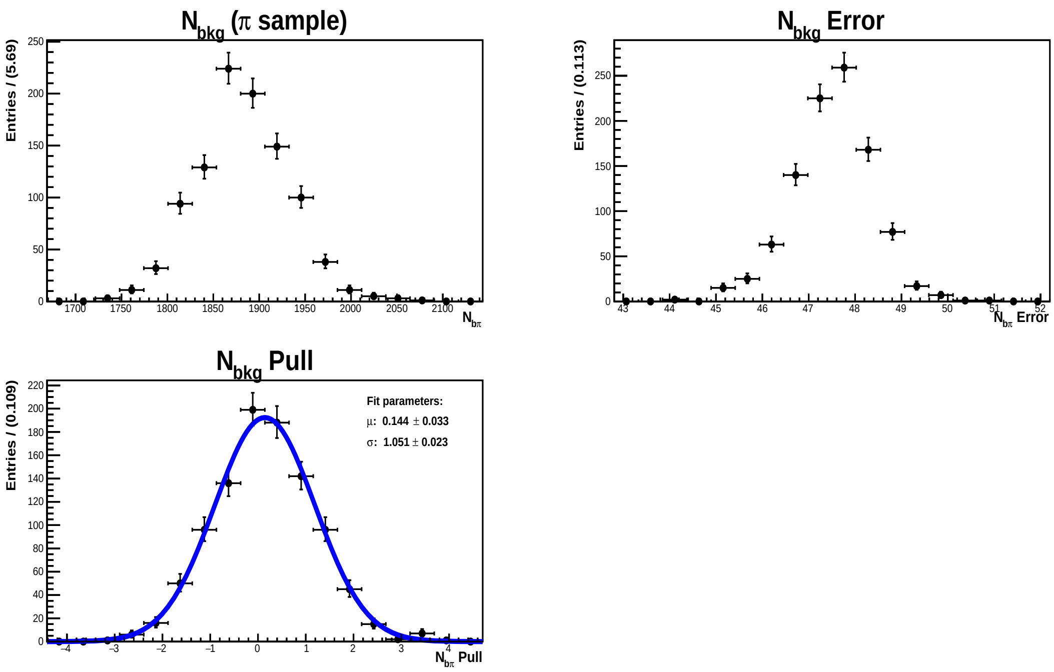


Figure B.7: The expected background distributions ($\tau \to \pi\nu_\tau$ sample) for OSmu mode.

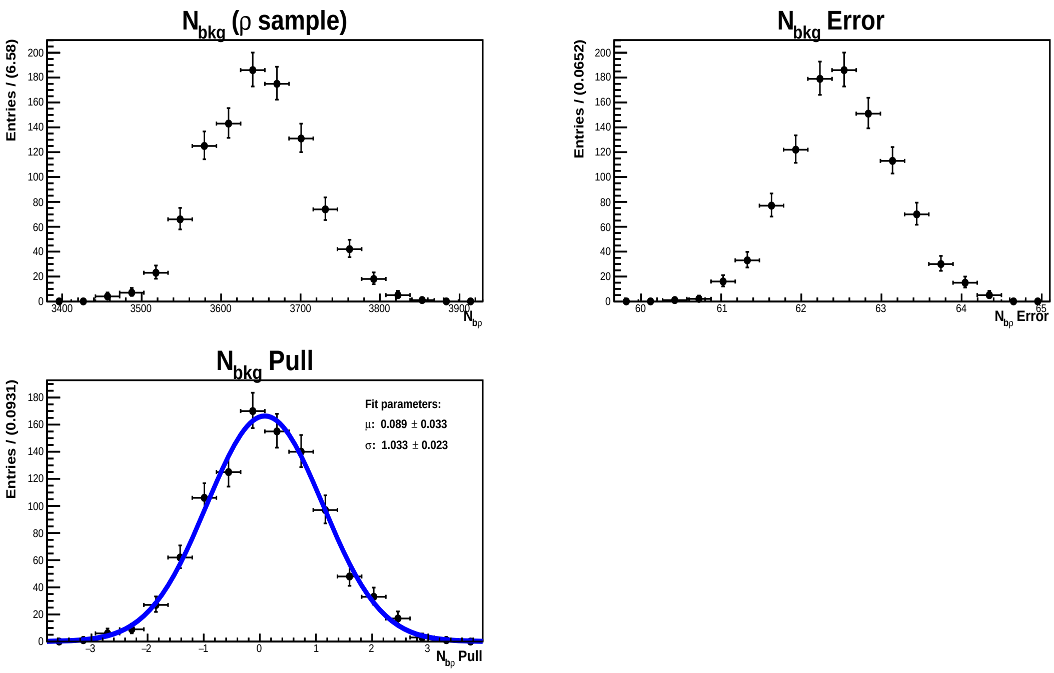


Figure B.8: The expected background distributions ($\tau \to \rho\nu_\tau$ sample) for OSmu mode.

Bibliography

[1] S. Weinberg, "The making of the standard model," *European Physical Journal C*, vol. 34, pp. 5–13, 2004.

[2] P. Higgs, "Broken symmetries and the masses of gauge bosons," *Physical Review Letters*, vol. 13, no. 16, pp. 508–509, 1964.

[3] S. Weinberg, "A model of leptons," *Physical Review Letters*, vol. 19, no. 21, pp. 1264–1266, 1967.

[4] G. 't Hooft and M. Veltman, "Renormalizability of the electroweak interaction," *Nuclear Physics B*, vol. 44, no. 1, pp. 189–213, 1972.

[5] A. Salam and J. Ward, "Gauge theory of elementary interactions," *Physical Review*, vol. 136, pp. B763–B768, 1964.

[6] A. de Gouvêa and P. Vogel, "Lepton flavor and number conservation, and physics beyond the standard model," *Progress in Particle and Nuclear Physics*, vol. 71, pp. 75–92, 2013.

[7] S. Sahoo and R. Mohanta, "Lepton flavor violating B meson decays via a scalar leptoquark," *Physical Review D*, vol. 93, p. 114001, 2016.

[8] Q. R. Ahmad, R. Allen, T. Andersen, J. Anglin, J. Barton, E. Beier, M. Bercovitch, J. Bigu, S. Biller, R. Black, *et al.*, "Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory," *Phys. Rev. Lett.*, vol. 89, p. 011301, 2002.

[9] A. Angelescu, D. Bečirević, D. A. Faroughy, F. Jaffredo, and O. Sumensari, "Single leptoquark solutions to the B-physics anomalies," *Physical Review D*, vol. 104, p. 055017, 2021.

[10] S. Watanuki *et al.*, "Search for the lepton flavor violating $B \to k\tau\ell$ ($\ell = e,\mu$) decays at Belle," *Phys. Rev. Lett.*, vol. 130, p. 261802, 2023.

[11] A. Matyja *et al.*, "Observation of $B^0 \to D^{*-}\tau^+\nu_\tau$ decay at Belle," *Phys. Rev. Lett.*, vol. 99, p. 191807, Nov 2007.

[12] J. Stypula *et al.*, "Evidence for $B^- \to D_s^+ K^- \ell^- \overline{\nu}_\ell$ and search for $B^- \to D_s^{*+} K^- \ell^- \overline{\nu}_\ell$," *Phys. Rev. D*, vol. 86, p. 072007, 2012.

[13] R. Brout and F. Englert, "Broken symmetry and the mass of gauge bosons," *Physical Review Letters*, vol. 13, no. 9, pp. 321–323, 1964.

[14] N. Cabibbo, "Unitary symmetry and leptonic decays," *Physical Review Letters*, vol. 10, no. 12, pp. 531–533, 1963.

[15] M. Kobayashi and T. Maskawa, "CP violation in the renormalizable theory of weak interaction," *Progress of Theoretical Physics*, vol. 49, no. 2, pp. 652–657, 1973.

[16] E. Kou *et al.*, "The Belle II Physics Book," *Progress of Theoretical and Experimental Physics*, vol. 2019, p. 547, 12 2019.

[17] A. Angelescu, D. Bečirević, D. A. Faroughy, and O. Sumensari, "Closing the window on single leptoquark solutions to the B-physics anomalies," *Journal of High Energy Physics*, vol. 2018, no. 10, p. 183, 2018.

[18] J. P. Lees *et al.*, "Evidence for an excess of $\overline{B} \to D^{(*)}\tau^-\overline{\nu}_\tau$ decays," *Phys. Rev. Lett.*, vol. 109, p. 101802, Sep 2012.

[19] R. Aaij *et al.*, "Measurement of the ratio of branching fractions $\mathcal{B}(\overline{b}^0 \to D^{*+}\tau^-\overline{\nu}_\tau)/\mathcal{B}(\overline{b}^0 \to D^{*+}\mu^-\overline{\nu}_\mu)$," *Phys. Rev. Lett.*, vol. 115, p. 111803, Sep 2015.

[20] G. Caria *et al.*, "Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method," *Phys. Rev. Lett.*, vol. 124, p. 161803, Apr 2020.

[21] D. Ghosh, "Explaining the R_K and R_{K^*} anomalies," *The European Physical Journal C*, vol. 77, no. 10, p. 694, 2017.

[22] X.-G. He, G. Valencia, and Y. Wang, "Lepton flavor violating τ and B decays and heavy neutrinos," *Phys. Rev. D*, vol. 70, p. 113011, Dec 2004.

[23] J. P. Lees *et al.*, "Search for the decay modes $B^\pm \to h^\pm\tau\ell$," *Phys. Rev. D*, vol. 86, p. 012004, Jul 2012.

[24] T. Keck *et al.*, "Comput. software big sci.," *Comput. Software Big Sci.*, vol. 1, no. 6, 2019.

[25] J. Brodzicka *et al.*, "Physics achievements from the Belle experiment," *Progress of Theoretical and Experimental Physics*, vol. 2012, p. 04D001, 12 2012.

[26] J. K. (on behalf of the Belle II Collaboration), "The Belle II Experiment," *CERN Proceedings*, vol. 1, 2017.

[27] Y. Han and the Belle & Belle II Collaborations, "Recent highlights from the BELLE and BELLE II experiments," *Nuclear and Particle Physics Proceedings*, vol. 347, pp. 52–61, 2024.

[28] S. Kurokawa and E. Kikutani, “Overview of the KEKB accelerators,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, no. 1, pp. 1–7, 2003. KEK-B: The KEK B-factory.

[29] Wikipedia contributors, “KEKB (accelerator) — Wikipedia, the free encyclopedia,” 2022. [Online; accessed 5-April-2025].

[30] K. Akai *et al.*, “RF systems for the KEK B-Factory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, no. 1, pp. 45–65, 2003. KEK-B: The KEK B-factory.

[31] K. Akai, K. Furukawa, and H. Koiso, “SuperKEKB collider,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 907, pp. 188–199, 2018.

[32] SuperB Collaboration, “SuperB: A High-Luminosity Asymmetric e^+e^- Super Flavor Factory. Conceptual Design Report,” tech. rep., 2007.

[33] T. E. Browder and K. Honscheid, “B mesons,” *Progress in Particle and Nuclear Physics*, vol. 35, pp. 81–219, 1995.

[34] A. Abashian *et al.*, “The Belle detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 479, no. 1, pp. 117–232, 2002.

[35] D. Kumar, *"Search for the Decay $B_s^0 \to J/\psi\pi^0$ at Belle Experiment"*. Ph.D. thesis, Indian Institute of Technology Guwahati, Guwahati, India, 2023.

[36] Z. Natkaniec *et al.*, “Status of the Belle silicon vertex detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 560, no. 1, pp. 1–4, 2006. Proceedings of the 13th International Workshop on Vertex Detectors.

[37] K. Abe *et al.*, “Improved measurement of mixing-induced CP violation in the neutral B meson system,” *Phys. Rev. D*, vol. 66, p. 071102, Oct 2002.

[38] G. Alimonti *et al.*, “The BELLE silicon vertex detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 453, no. 1, pp. 71–77, 2000. Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics.

[39] V. Aulchenko *et al.*, “Electromagnetic calorimeter for Belle II,” *Journal of Physics: Conference Series*, vol. 587, no. 1, p. 012045, 2015.

[40] O. Alonso *et al.*, “DEPFET Active Pixel Detectors for a Future Linear e^+e^- Collider,” *IEEE Transactions on Nuclear Science*, vol. 60, p. 1457–1465, Apr. 2013.

[41] K. Adamczyk *et al.*, "The Design, Construction, Operation and Performance of the Belle II Silicon Vertex Detector," *Journal of Instrumentation*, vol. 17, p. P11042, Nov. 2022.

[42] E. Kou *et al.*, "The Belle II Physics Book," *Progress of Theoretical and Experimental Physics*, vol. 2019, p. 38, 12 2019.

[43] T. Dong *et al.*, "Calibration and alignment of the Belle II central drift chamber," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 930, pp. 132–141, 2019.

[44] T. Abe *et al.*, "Belle II Technical Design Report," tech. rep., Belle II Collaboration, 2010. arXiv:1011.0352 [physics.ins-det].

[45] S. Iwata *et al.*, "Particle identification performance of the prototype aerogel RICH counter for the Belle II experiment," *Progress of Theoretical and Experimental Physics*, vol. 2016, p. 033H01, 03 2016.

[46] A. Kuzmin, "Electromagnetic calorimeter of Belle II," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 958, p. 162235, 2020. Proceedings of the Vienna Conference on Instrumentation 2019.

[47] Y. Ushiroda *et al.*, "Development of the central trigger system for the BELLE detector at the KEK B-factory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 438, no. 2, pp. 460–471, 1999.

[48] Y. Iwasaki *et al.*, "Level 1 Trigger System for the Belle II Experiment," *IEEE Transactions on Nuclear Science*, vol. 58, no. 4, pp. 1807–1815, 2011.

[49] S. Lee, R. Itoh, N. Katayama, and S. Mineo, "Development of High Level Trigger Software for Belle II at SuperKEKB," *Journal of Physics: Conference Series*, vol. 331, p. 022015, dec 2011.

[50] T. Kuhr *et al.*, "The Belle II Core Software," *Computing and Software for Big Science*, vol. 3, p. 1, 2019.

[51] M. Gelb *et al.*, "B2BII: Data Conversion from Belle to Belle II," *Computing and Software for Big Science*, vol. 2, p. 9, 2018.

[52] A. J. Bevan *et al.*, "The Physics of the B Factories," *The European Physical Journal C*, vol. 74, no. 11, pp. 3026,91, 2014.

[53] A. Roodman, "Blind Analysis in Particle Physics," *arXiv preprint physics/0312102*, 2003.

[54] D. J. Lange, "The Evtgen particle decay simulation package," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 462, no. 1-2, pp. 152–155, 2001.

[55] EvtGen Collaboration, "EvtGen User Guide." https://evtgen.hepforge.org/doc/EvtGenGuide.pdf, 2021. Accessed: 2025-05-16.

[56] R. Brun *et al.*, *GEANT3: User's Guide GEANT 3.10, GEANT 3.11; rev. version.* CERN, Geneva, 1987. CERN Program Library Long Writeup W5013.

[57] T. Sjöstrand *et al.*, "High-energy physics event generation with PYTHIA 6.1," *Comput. Phys. Commun.*, vol. 135, pp. 238–259, 2001.

[58] E. Barberio and Z. Was, "PHOTOS: A Universal Monte Carlo for QED radiative corrections. version 2.0," *Comput. Phys. Commun.*, vol. 79, pp. 291–308, 1994.

[59] P. Cheema, "Suppressing Beam Background and Fake Photons at Belle II using Machine Learning," *EPJ Web Conf.*, vol. 295, p. 09035, 2024.

[60] G. Fox and S. Wolfram, "Event shapes in e^+e^- annihilation," *Nuclear Physics B*, vol. 149, no. 3, pp. 413–496, 1979.

[61] A. Hocker, J. Stelzer, F. Tegenfeldt, H. Voss, M. Voss, *et al.*, "TMVA - Toolkit for Multivariate Data Analysis," *PoS*, vol. ACAT, p. 040, 2007.

[62] P. D. Group, "Review of Particle Physics," *Progress of Theoretical and Experimental Physics*, vol. 2022, pp. 660–675, 08 2022.

[63] L. Lista, *Statistical Methods for Data Analysis in Particle Physics*, vol. 909 of *Lecture Notes in Physics.* Springer, 2016.