Doctoral Dissertation
博士論文

Sensitivity studies for $B \to X_s \ell^+ \ell^-$ Decays Branching Fraction Measurement at the Belle II experiment
(Belle II実験における$B \to X_s \ell^+ \ell^-$過程の崩壊分岐比測定の感度評価)

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2025
令和7年12月博士（理学）申請

Department of Physics, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
物理学専攻

WANG Zihan
王 子涵

Abstract

We present sensitivity studies of the branching fraction measurement for the electroweak penguin process $B \to X_s\ell^+\ell^-$, where ℓ is an electron or a muon and X_s is a hadronic system containing an s-quark. This study is based on Belle II run 1 dataset corresponding to an integrated luminosity of 365 fb^{-1}. The X_s hadronic system is reconstructed using the sum-of-exclusive method, combining one $K^{\pm}$ or K_S^0 and up to four pions with at most one pion can be neutral, as well as three kaon modes. In the most interesting $1 < q^2 < 6$ $[(\text{GeV}/c)^2]$ region, where q^2 is the squared invariant mass of di-lepton system, we reached statistical and systematical uncertainties of 17.4% and 9.4% in real data, respectively. The total uncertainty is suppressed by a factor of 1.5 and 2.2 comparing to precedent measurements from *BABAR* and Belle, respectively, achieving a better sensitivity than the current world average. Extrapolating the analysis method developed in this dissertation to the nominal luminosity ($50\,\text{ab}^{-1}$) of Belle II, we can hopefully distinguish the long-standing tensions between observations and the SM predictions with 4.5σ significance.

Acknowledgement

I would like to express my sincere gratitude to all those who have supported me throughout this journey.

First, I am deeply indebted to my supervisor, Prof. Yutaka Ushiroda, for his invaluable guidance, and constant encouragement. He introduced me this valuable chance for exploiting the frontier of Physics at the Belle II experiment.

I greatly appreciate Dr. Akimasa Ishikawa for his helpful advice and bright ideas for the analysis. Without his expertise and insightfulness I could never finish this piece of work.

My sincere thanks also go to all the staffs of the UTokyo/Sokendai B-factory group for their assistance, insightful discussions, and supportive environment. I am especially grateful to Dr. Yo Sato, Dr. Kenta Uno and Dr. Kazutaka Sumisawa for their valuable suggestions and discussions.

I am also grateful for the internal review committee members, Prof. K. Trabelsi, Prof. T. Browder, and Prof. M. Iwasaki for their prompt, careful review and insightful suggestions.

Thanks should also go to UTokyo/Sokendai B-factory group members for their educational and moral support. Discussions with Dr. Xiaodong Shi, Mr. Junewoo Park, Mr. Yuxin Liu, Mr. Shin'nosuke Kato and Mr. Shijie Wang makes my time here both productive and enjoyable.

On a personal note, I extend my thanks to my parents for their unconditional love, endless support, and unwavering belief in me throughout my studies and life. I am forever grateful for their encouragement.

Finally, I would like to express my deepest appreciation to my girlfriend for her understanding, endless patience, and constant emotional support. Her presence has been my source of strength throughout tough times.

Personal Contribution

The sensitivity study reported in this dissertation are based on the joint effort of more than 1000 members in the Belle II collaboration and the SuperKEKB group. The construction and operation of the experimental apparatus, data management by a distributed computing system, data calibration, and development of basic analysis frameworks including vertex fitter and flavor tagger are performed in collaboration. The idea of a sum-of-exclusive measurement strategy follows previous measurements. The author contributes mainly to the improvement of signal modeling in the simulation sample described in Chapter 4 which significantly suppressed related systematic uncertainties described in Chapter 9. The author also improved the background suppression algorithm and the selection criteria described in Chapter 6 which enhances statistical sensitivity. In addition, the author proposed a new algorithm that can significantly improve muon identification performance as described in Appendix G. The details of each effort are described below.

The modeling of the signal process, especially the description of X_s system has been the most significant systematic uncertainty that is comparable to statistical uncertainties. The author proposed improvement in the following four aspects. First, average recent $\mathcal{B}(B \to K^{(*)}\ell^+\ell^-)$ measurements to improve precision on the branching fraction of exclusive processes described in Sec. 9.1.1. Second, use data-driven methods for fragmentation correction of different X_s final states using measurement from $B \to X_s\gamma$ and $B \to X_s J/\psi$ described in Sec. 9.1.6. Third, for the first time introduce five processes via high kaon resonances like $B \to K_1(1270)\ell^+\ell^-$ to improve modeling of the $K\pi$ and $K\pi\pi$ final states in the $m(X_s) > 1.1\,\mathrm{GeV}/c^2$ region described in Sec. 4.1.2. The author made this possible by developing the necessary simulation tools (generator in EvtGen software framework) for these resonances described in Appendix. A. Finally, extrapolate from reconstructed modes to all final states of X_s using a tuned Pythia sample that reproduces the fragmentation of $B \to X_s\gamma$. With all the efforts described above, the author suppressed the systematic uncertainties related to signal modeling by a factor of two.

The author achieved better reconstruction efficiency by utilizing MultiVariate Analysis (MVA) tools for reconstruction, especially K_S^0 and π^0. The author introduced the FT-Transformer method as the MVA tool for background rejection, and performed optimization in 12 different phase spaces that maximized sensitivity in the $1 < q^2 < 6$ $[(\mathrm{GeV}/c)^2]$ region of the most interest.

Finally, the author proposed the idea of improving muon identification performance independently. The author revealed the mechanism of its unsatisfactory performance and proposed a new Deep Neural Network (DNN) based algorithm that suppressed the pion misidentification rate by half while keeping the efficiency of 90% in a simulation sample. Performance improvement in real data is also confirmed. Due to collaboration constraints, this algorithm is not applied to the measurement presented in this dissertation. However, as an important contribution to the analysis tools, it will be widely used in the following studies and other measurements of the collaboration.

Outline

In Chapter 1, we introduce the physics behind the branching fractions measurements and test of lepton flavor universality $R(X_s)$ in the inclusive $B \to X_s \ell^+ \ell^-$ process and review the past measurements. In Chapter 2, we describe the experimental tools of the SuperKEKB accelerator and the Belle II detector with which we recorded the data. In Chapter 3, we give an overview of this study, including the dataset and analysis strategy. Chapter 4 describes the simulation samples used in this analysis. In Chapter 5, we explain the algorithm and criteria for the reconstruction of $B \to X_s \ell^+ \ell^-$ candidates, followed by further background suppression in Chapter 6. Chapter 7 describes the strategy for estimating background contributions that have the same shape as signal events during signal extraction, explained in Chapter 8. Chapter 9 summarizes the systematic uncertainties in this study. We present the results and discuss prospects in Chapter 10.

Contents

Chapter 1

Physics motivation

1.1 Research Background

In the past few decades, the Standard Model (SM) has been the most successful theory in the field of elementary particle physics. However, the SM is considered not to be the ultimate theory, as it is unable to explain several big questions such as the nature of Dark Matter, Dark Energy, and the source of neutrino mass.

Searching for new physics via high-energy collisions is considered to be a promising way. There are two kinds of experiments in this field: energy frontier experiments and intensity frontier experiments. The former searches for new physics by direct production of possible new particles with large masses, represented by the ATLAS and CMS experiments at the Large Hadron Collider (LHC), CERN, working with proton-proton collisions at 13.6 TeV. The latter searches for new physics indirectly by precise measurements of known processes and detecting deviations between data and the SM predictions, such as B-factory experiments of Belle (II) and *BABAR*, and the LHCb experiment at the LHC.

In this dissertation, we focus on the precise measurement of the b-quark to s-quark and dilepton ($\ell^+\ell^-$) transition process: $b \to s\ell^+\ell^-$ at the Belle II experiment, where ℓ can be muons or electrons. This process is a flavor changing neutral current (FCNC) process that is mediated through loop diagrams as shown in Fig. 1.1. Because of this, it is highly suppressed in the SM with decay branching fractions at the order of $O(10^{-6})$ and thus becomes a good probe of contributions from new physics.

In recent years, the LHCb experiment performed several precise measurements on the differential branching fractions against the invariant mass of the di-lepton system q^2 in exclusive processes $B \to K\mu^+\mu^-$ and $B \to K^*(892)\mu^+\mu^-$ [1, 2]. These measurements show tensions with theoretical predictions, especially the differential branching fractions in the low-q^2 region of $1 < q^2 < 6$ $[(\,\mathrm{GeV}/c)^2]$. However, due to the large theoretical uncertainty dominated by the $B \to K^{(*)}$ form factor calculations, it is difficult to claim evidence of new physics beyond the SM. This brings interest to the measurements of the inclusive process that is free from hadron form factors: $B \to X_s\ell^+\ell^-$, where B is either

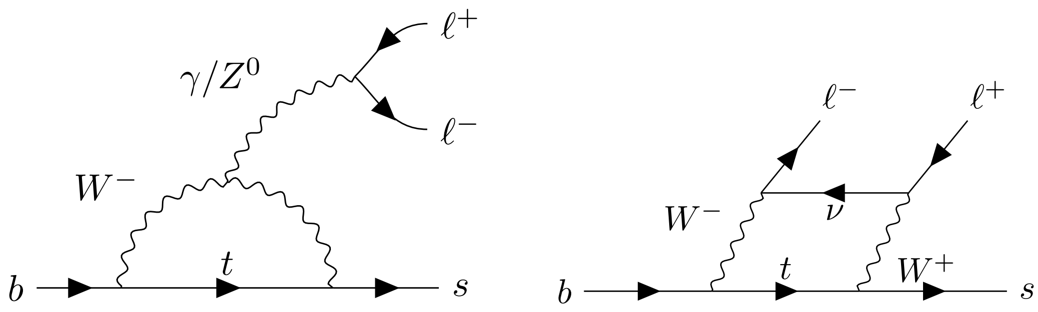


Figure 1.1: Feynman diagrams of $b \to s\ell^+\ell^-$ processes in the SM. The left is penguin diagram and the right is called box diagram.

B^0 or B^+ and X_s is a hadronic system that contains an s-quark.

In the following sections of this chapter, we review the theoretical framework of the $b \to s\ell^+\ell^-$ transition in the SM using the Effective Field Theory, which provides a model-independent way of probing the new physics (Sec. 1.2). Then we go through the physics behind the most interesting q^2 spectrum in this channel (Sec. 1.3), followed by a review of current exclusive measurements (Sec. 1.4) and inclusive measurements (Sec. 1.5), especially observables where anomalies are detected. We finalize this chapter by reviewing the current status of constraints to Wilson Coefficients in Sec. 1.6.

1.2 Effective Hamiltonian

In 1933, E. Fermi developed his four-fermion interaction theory to describe beta decays and introduced the Fermi constant G_F, which represents the strength of weak interaction. Even though this theory is afterwards found to be the first-order approximation of decays via the W boson, it inspires a convenient tool and an excellent example of effective theory for describing physics processes at a lower energy scale by integrating over contributions from large energy scales.

For the $b \to s\ell^+\ell^-$ transition case, in the SM, this process occurs at the energy scale of b-quark mass m_b around $4.2\,\mathrm{GeV}/c^2$, mainly contributed by loops mediated by the W boson ($m_W = 80.4\,\mathrm{GeV}/c^2$) and the t-quark ($m_t \sim 170\,\mathrm{GeV}/c^2$), which are two to three orders of magnitude greater in energy. Given this fact, it is convenient to describe this process by series products of operator $\mathcal{O}_i(\mu)$ and its corresponding strength coefficients $C_i(\mu)$, called Wilson coefficients, at the energy scale μ. The $\mathcal{O}_i(\mu)$s contain the field operators for initial and final states which represent the "long-distance" contributions, while Wilson coefficients are derived by integrating contributions from high energy scale and thus encode the "short-distance" contributions.

The $C_i(\mu)$s are evaluated perturbatively at a high energy scale m_W, and propagated to a low energy scale m_b via the Renormalization Group Equation (RGE). In the meantime, operators also depend on the energy scale μ and RGE guarantees that the dependences of on μ of $C_i(\mu)$ and $\mathcal{O}_i(\mu)$ cancel out, giving a physics result independent of the renormalization scale.

The effective Hamiltonian of the $b \to s$ transition in the SM is expressed as follows [3]:

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \left(\sum_{i=1}^{10} C_i(\mu) \mathcal{O}_i(\mu) \right) \tag{1.1}$$

where $V_{tb}V_{ts}^*$ corresponds to the CKM matrix elements. Ten different dimension-six operators contribute to this process: $\mathcal{O}_{1,2}$ describe the $b \to c \to s$ process with and without gluon exchange; $\mathcal{O}_{3-6}$ denote the QCD penguin process $b \to sq\bar{q},\ q = u, d, s, c$; $\mathcal{O}_7$ is the electro-weak radiative process $b \to s\gamma$ while $\mathcal{O}_8$ is the gluon radiative process $b \to sg$; $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are $b \to s\ell^+\ell^-$ transitions with vector and axial vector lepton current, respectively. The expressions of these operators are:

$$\begin{aligned}
\mathcal{O}_1 &= (\bar{s}_L \gamma^\mu T^a c_L)(\bar{c}_L \gamma_\mu T^a b_L), & \mathcal{O}_2 &= (\bar{s}_L \gamma^\mu c_L)(\bar{c}_L \gamma_\mu b_L), \\
\mathcal{O}_3 &= (\bar{s}_L \gamma^\mu b_L) \sum_q (\bar{q} \gamma_\mu q), & \mathcal{O}_4 &= (\bar{s}_L \gamma^\mu T^a b_L) \sum_q (\bar{q} \gamma_\mu T^a q), \\
\mathcal{O}_5 &= (\bar{s}_L \gamma^\mu \gamma^\nu \gamma^\rho b_L) \sum_q (\bar{q} \gamma_\mu \gamma_\nu \gamma_\rho q), & \mathcal{O}_6 &= (\bar{s}_L \gamma^\mu \gamma^\nu \gamma^\rho T^a b_L) \sum_q (\bar{q} \gamma_\mu \gamma_\nu \gamma_\rho T^a q), \\
\mathcal{O}_7 &= \frac{e}{g_s^2} m_b (\bar{s}_L \sigma^{\mu\nu} b_R) F_{\mu\nu}, & \mathcal{O}_8 &= \frac{1}{g_s} m_b (\bar{s}_L \sigma^{\mu\nu} T^a b_R) G^a_{\mu\nu}, \\
\mathcal{O}_9 &= \frac{e^2}{g_s^2} (\bar{s}_L \gamma^\mu b_L) \sum_\ell (\bar{\ell} \gamma_\mu \ell), & \mathcal{O}_{10} &= \frac{e^2}{g_s^2} (\bar{s}_L \gamma^\mu b_L) \sum_\ell (\bar{\ell} \gamma_\mu \gamma_5 \ell),
\end{aligned} \tag{1.2}$$

where the s, c, b, ℓ are the corresponding quark or lepton field, with subscripts L and R referring to left- and right-handed components, γ^μ is the 4×4 Dirac matrices, T^a are the SU(3) generators of color charge, $F_{\mu\nu}$ and $G_{\mu\nu}$ denote QED and QCD strength tensors, respectively, and e and g_s represent the electromagnetic and strong coupling constants, respectively.

Basically, C_9, C_{10} and C_7 contribute dominantly to our target process, with an additional $\gamma \to \ell^+\ell^-$ vertex in the $\mathcal{O}_7$ diagram. In the meantime, minor contributions from other diagrams to C_7 and C_9 by factorizable loop terms [4] are also possible. For this reason, it is convenient to define the effective $C_7^{\text{eff}} = C_7 - C_5/3 - C_6$ and C_9^{eff} adapting the Naive Dimensional Regularization (NDR) scheme [4]. Table 1.1 summarizes the values of the above three Wilson coefficients in the SM [5]. These values are derived by matching the SM contributions to the Next Next Leading Order (NNLO). Approximately the relationship $C_9 \approx -C_{10}$ holds because the weak interaction is participated by $V - A$ current. In the following, we remove the superscripts for C_7^{eff} and C_9^{eff} for convenience.

Table 1.1: Predicted values in the SM of Wilson coefficients $C_7^{\text{eff,SM}}$, $C_9^{\text{eff,SM}}$ and C_{10}^{SM} at the renormalization scale m_b [5].

$C_7^{\text{eff,SM}}$	$C_9^{\text{eff,SM}}$	C_{10}^{SM}
-0.2923	$+4.0749$	-4.3085

The above Hamiltonian leads to the following free quark decay amplitude:

$$\begin{aligned}\mathcal{M}(b \to s\ell^+\ell^-) = \frac{G_F \alpha}{\sqrt{2}\pi} V_{ts}^* V_{tb} \{ & C_9 (\bar{s}_L \gamma_\mu b_L)(\bar{\ell}\gamma^\mu \ell) \\ & + C_{10}(\bar{s}_L \gamma_\mu b_L)(\bar{\ell}\gamma^\mu \gamma_5 \ell) \\ & - 2\hat{m}_b C_7 \left(\bar{s}_L i\sigma_{\mu\nu} \frac{\hat{q}^\nu}{\hat{s}} b_R \right) (\bar{\ell}\gamma^\mu \ell) \},\end{aligned} \tag{1.3}$$

where $q^\mu = (p_{\ell+}^\mu + p_{\ell-}^\mu)$ is the four-momentum of the di-lepton system, q^2 is the invariant mass of di-lepton, and the hat denotes normalization in terms of the b-quark pole mass, for example $\hat{q}^\mu = q^\mu/m_b$. And we use $\hat{s} = q^2/m_b^2$ to denote normalized q^2. $\sigma_{\mu\nu} = \frac{i}{2}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$.

For exclusive decay cases of $B \to K_J \ell^+\ell^-$, where K_J refers to a kaon resonance with spin J, one needs to operate the free quark field ($\bar{s}_L\gamma_\mu b_L$ or $\bar{s}_L i\sigma_{\mu\nu} b_R$, referred to as $O_{b\to s}$) on the hadron states for describing the $B \to K_J$ transition matrix element:

$$\langle K_J | O_{b\to s} | B \rangle. \tag{1.4}$$

Usually, it is decomposed to the sum of series products of Lorentz covariant kinematic variables and Lorentz invariant form factors as a function of q^2, which encodes the information of non-perturbative QCD. The expansions of this matrix element for various K_J with different spin-parity J^P are summarized in Appendix. A. These form factors are hard to calculate from the first principle, and various model-dependent methods are developed to give numerical results such as Light-Cone Sum Rules (LCSR) [6, 7] and pQCD [8, 9].

The effective Hamiltonian method provides an important tool for testing the SM in a model-independent way. If non-SM physics contributes to the $b \to s\ell^+\ell^-$ process via additional diagrams, it modifies the corresponding Wilson coefficients or even adds new operators. Observables such as decay branching fractions are direct results of the corresponding Wilson coefficient. If any deviation of the measured Wilson coefficient from the SM prediction is confirmed, it becomes an important evidence of new physics.

1.3 Overview of q^2 Spectrum in $b \to s\ell^+\ell^-$ Transition

The q^2 spectrum in the $s\ell^+\ell^-$ final states contains various contributions, as shown in Fig. 1.2. This section gives a qualitative description of the q^2 spectrum and the physics behind it. In this subsection, the units of all figures mentioned are in GeV^2.

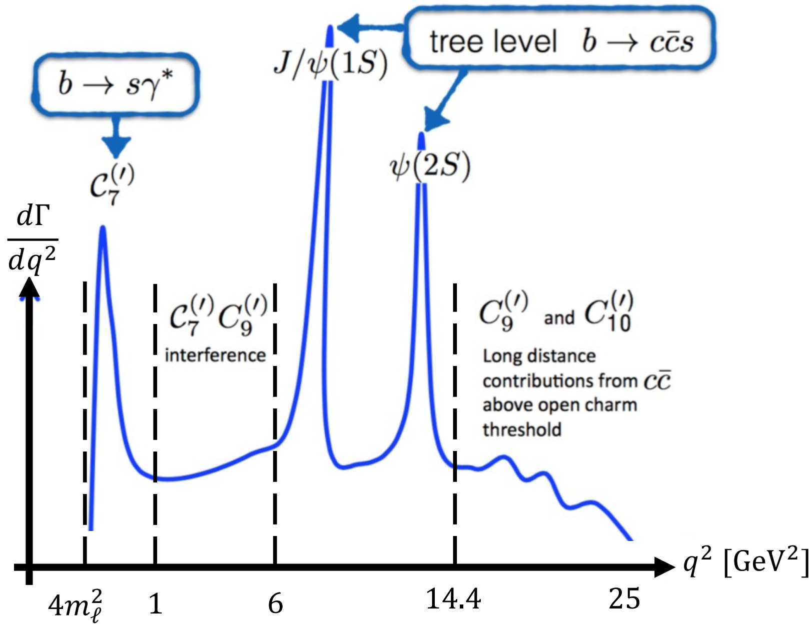


Figure 1.2: Schematic of $d\Gamma/dq^2$ distribution in $b \to s\ell^+\ell^-$ process [10].

Starting from the left side of Fig. 1.2, in the range of $q^2 \in [4m_\ell^2, 1.0]$, it is dominated by the $b \to s\gamma$ diagram controlled by C_7, with an additional vertex converting the photon to di-lepton. Usually, C_7 can be measured precisely in $B \to X_s\gamma$ processes with about 100 times more statistics. It is of less interest in the study of $b \to s\ell^+\ell^-$.

The next bin of $q^2 \in [1.0, 6.0]$ is referred to as the low-q^2 bin in this dissertation unless otherwise specified. This bin is of the most interest, as it is dominated by the electro-weak penguin process, with little influence from on-shell γ conversion via the $b \to s\gamma$ process and tree-level charmonium processes. Also, it is the bin where most anomalies are detected, which are explained in the following sections of this chapter.

The $q^2 \in [6, 14.4]$ bin has two significant peaks from tree-level $b \to c\bar{c}s$ processes where $c\bar{c}$ states subsequently decay to the di-lepton system, with relatively flat contributions from electro-weak penguin diagrams buried underneath. Usually in measurements, peaks from charmonium processes are vetoed out, but it is still interesting to measure the area in between the two peaks to study the interference between the charmonium diagram and the penguin diagram.

The last bin from 14.4 to the end of phase space is contributed dominantly by the penguin diagram with small contributions from higher charmonium resonances. The penguin diagram in this region is mostly contributed by C_9 and C_{10} terms. Still, the measurement in this bin is of special interest for studying interference between these processes.

1.4 Exclusive $B \to K^{(*)}\ell^+\ell^-$ Decays

Exclusive decays of $B \to K^{(*)}\ell^+\ell^-$ are reconstructed from long-lived particles that can be detected directly (such as $K^\pm$) or indirectly by reconstructing from their decay products ($K^0_S \to \pi^+\pi^-$, $K^*(892) \to K\pi$). In these channels, the B-factory experiments of Belle (II) and BABAR have advantages in reconstruction efficiency, especially in the electron modes. Hadron collider experiments like LHCb, CMS, ATLAS and CDF feature in their high B-meson production cross section, leading to high statistics. This enables them to perform precise measurements, especially in the $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ channel.

1.4.1 Differential Branching Fractions

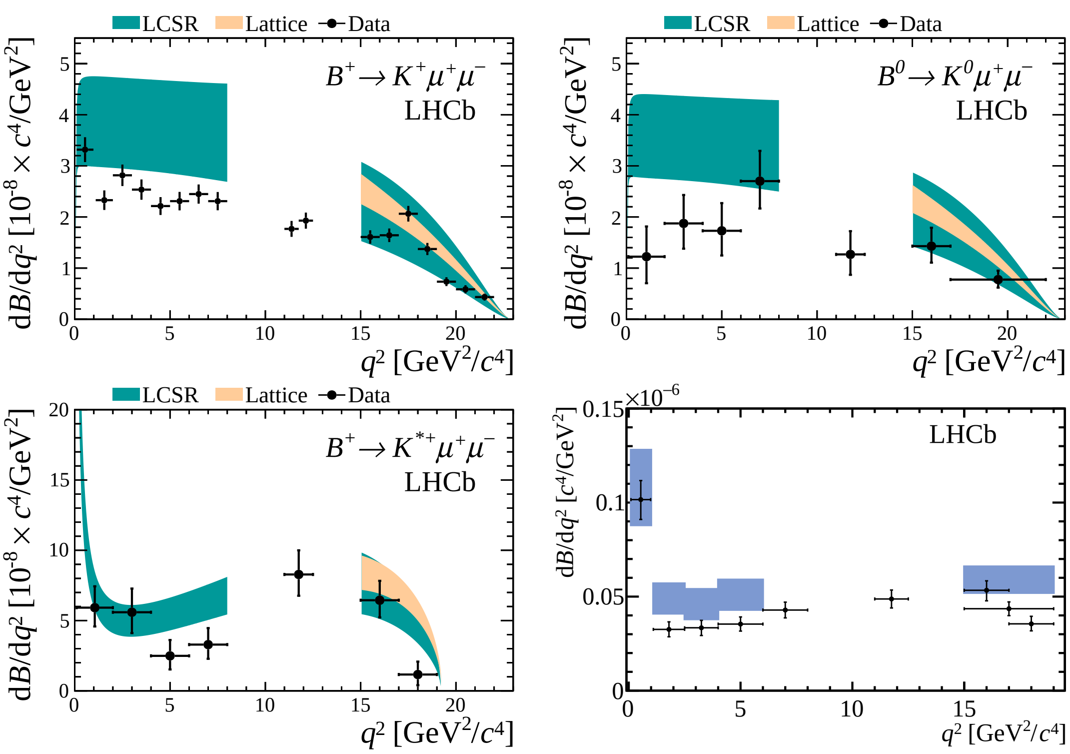


Figure 1.3: Differential branching fractions $d\mathcal{B}/dq^2$ of exclusive $B \to K^{(*)}\mu^+\mu^-$ processes measured by LHCb experiment. From upper left to bottom right are $B^+ \to K^+\mu^+\mu^-$, $B^0 \to K^0\mu^+\mu^-$, $B^+ \to K^{*+}\mu^+\mu^-$ [1], and $B^0 \to K^{*0}\mu^+\mu^-$ [2] channels. In the bottom right plot, dots with error bars are data, and the colored bands represent theoretical prediction.

In the past ten years, the LHCb experiment performed several precise measurements on the differential branching fractions against the invariant mass of the dilepton system q^2 in exclusive processes $B \to K\mu^+\mu^-$ and $B \to K^*(892)\mu^+\mu^-$ [1, 2] and their results are shown in Fig. 1.3. These measurements give lower differential branching fractions than theoretical predictions, especially in the low-q^2 region of $1 < q^2 < 6$ $(\mathrm{GeV}/c)^2$. However, due to the large theoretical uncertainties of about 20% which mainly come from the form factor calculation, it is difficult to claim that new physics beyond the SM has been discovered. And this brings us the interest and importance of performing measurements in the inclusive $B \to X_s \ell^+ \ell^-$ process, whose theoretical prediction is free from form factor calculations and thus has higher precision.

1.4.2 Angular Decomposition

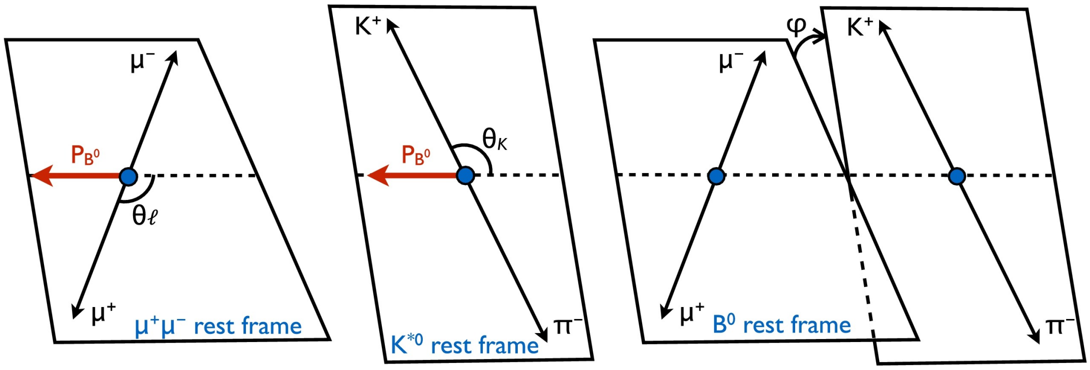


Figure 1.4: Illustration of the angular variables θ_ℓ (left), θ_K (middle), and ϕ (right) for the decay $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$. [11]

In addition to differential branching fractions against q^2, angular decomposition of the decay width is also of special interest. Three angular variables are defined with an illustration shown in Fig. 1.4: θ_ℓ is the angle between the positive (negative) muon momentum and the direction opposite to the B^0 ($\overline{B}^0$) momentum in the dimuon rest frame, θ_K is the angle between the kaon momentum and the direction opposite to the B^0 ($\overline{B}^0$) momentum in the K^{*0} ($\overline{K}^{*0}$) rest frame, and ϕ is the angle between the plane containing the two muons and the plane containing the kaon and the pion in the B^0 rest frame. The angular distribution is then expressed as [12]:

$$\begin{aligned}\frac{1}{d\Gamma/dq^2}\frac{d^4\Gamma}{dq^2\,d\cos\theta_\ell\,d\cos\theta_K\,d\phi} = \frac{9}{32\pi}[&\frac{3}{4}(1-F_L)\sin^2\theta_K + F_L\cos^2\theta_K \\ &+\frac{1}{4}(1-F_L)\sin^2\theta_K\cos 2\theta_\ell \\ &-F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi \\ &+S_4\sin 2\theta_K\sin 2\theta_\ell\cos\phi + S_5\sin 2\theta_K\sin\theta_\ell\cos\phi \\ &+\frac{4}{3}A_{FB}\sin^2\theta_K\cos\theta_\ell + S_7\sin^2\theta_K\sin\theta_\ell\sin\phi \\ &+S_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi + S_9\sin^2\theta_K\sin^2\theta_\ell\sin 2\phi],\end{aligned} \tag{1.5}$$

where F_L is the fraction of the longitudinal polarization of $K^*(892)$ meson, A_{FB} is the forward-backward asymmetry defined later in Eq. 1.14, and S_i are angular observables.

To cancel uncertainties from the $B \to K^*(892)$ form factors, a different notation of observable is defined:

$$P'_{4,5,7,8} = \frac{S_{4,5,7,8}}{\sqrt{F_L(1-F_L)}} \tag{1.6}$$

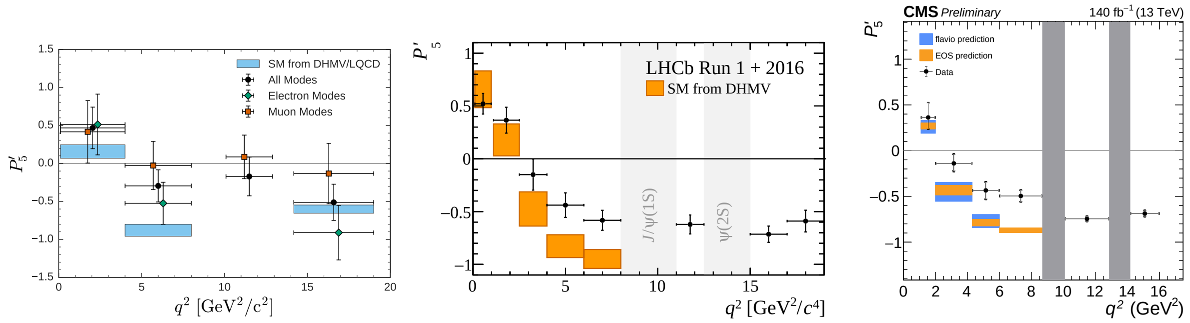


Figure 1.5: Measurement of P'_5 as a function of q^2 by Belle (left) [13], LHCb (middle) [14], and CMS (right) [15, 11]. Points with error bars are data and colored bands are theoretical predictions. Systematic exceeds from SM predictions are observed in all three experiments.

Since 2017, Belle, LHCb, and CMS [13, 14, 15, 11] collaboration presented their measurements on P'_5 as shown in Fig. 1.5. Every measurement result shows a systematic exceeding from the SM prediction, especially at the low-q^2 region, which is the same as where the anomalies in branching fraction measurements are observed. Combining all measurements, one can conclude a very significant tension greater than 4σ. However, recent theory studies reveal that possible long-distance non-perturbative QCD effects such as charm-rescattering can also explain this tension [16, 17, 18], leaving these observations an unsolved puzzle.

1.4.3 Test of Lepton Flavor Universality

Another observable that can cancel out uncertainties from form factors is the Lepton Flavor Universality (LFU) variable $R(K^{(*)})$, whose definition is given below:

$$R(K^{(*)}) = \frac{\mathcal{B}(B \to K^{(*)}\mu^+\mu^-)}{\mathcal{B}(B \to K^{(*)}e^+e^-)}. \tag{1.7}$$

LFU indicates that electroweak coupling strengths are the same for processes including e, μ, and τ leptons, is a key property of the SM. Any violation of this property would become direct evidence of several scenarios of new physics [19, 20]. On the experiment side, various measurements are performed from the Babar, Belle, LHCb, and CMS collaborations [21, 22, 23, 24, 25, 26]. In 2022, LHCb presented their result showing a deviation from 2.5σ of the SM prediction, but it was later corrected in 2023 due to an underestimated background caused by misidentification from pions to electrons, giving SM-consistent results with the best precision so far as shown in Fig. 1.6.

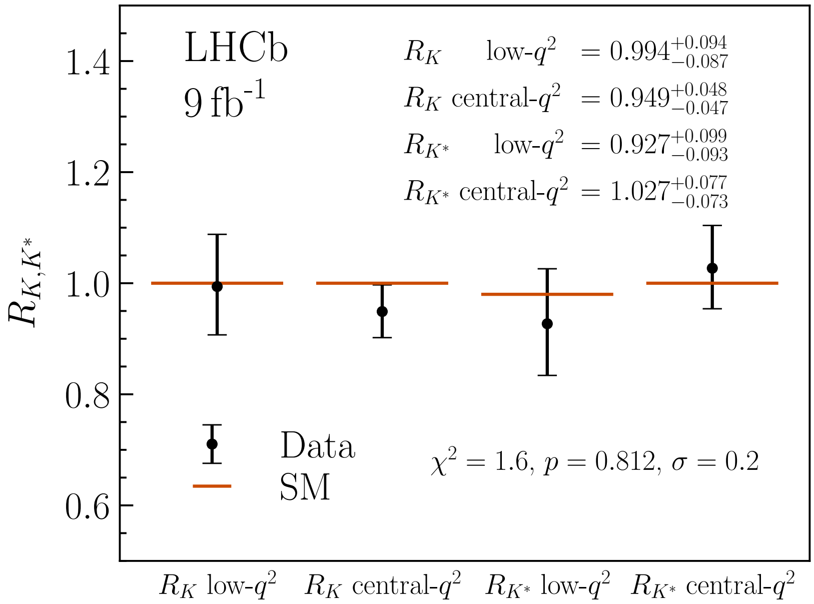


Figure 1.6: Measurement of the LFU observables $R(K)$ and $R(K^*)$ from LHCb [25]. In this plot, low-q^2 refers to $0.1 < q^2 < 1.1$ $[(\text{GeV}/c)^2]$ and central-q^2 refers to $1.1 < q^2 < 6.0$ $[(\text{GeV}/c)^2]$.

1.5 Inclusive $B \to X_s \ell^+ \ell^-$ Decays

Inclusive $B \to X_s \ell^+ \ell^-$ decays, where X_s denotes the hadron system containing an s-quark, are free from hadron transition form factors and thus relatively clean in theoretical calculations. Measurements in the inclusive way are important inputs for understanding the anomalies observed in exclusive decays, although they are experimentally challenging. Inclusive measurements usually require a clean experiment environment or clear constraints on the initial kinematic status, and thus can only be performed in B-factory experiments and are very difficult on hadron colliders.

At the B-factory experiments of Belle (II) and *BABAR*, B mesons are produced in pairs from $\Upsilon(4S)$ decays and thus embrace two ways of inclusive measurements: full inclusive and sum-of-exclusive.

In the full inclusive method, two leptons with the same flavor are reconstructed and the other B meson (denoted as B_{tag}) is reconstructed using hadronic channels without missing energy, by which the full kinematic information of B_{tag} is available. Together with the initial four-momentum of $\Upsilon(4S)$ given by beam parameters, the X_s system is recoiled:

$$m(X_s) = \sqrt{(p^\mu(\Upsilon(4S)) - p^\mu(B_{\text{tag}}) - q^\mu)^2}, \tag{1.8}$$

and we can measure differential branching fractions as a function of q^2 and $m(X_s)$. However, this method has not been applied yet due to its low reconstruction efficiency ($O(0.01\%)$) by the requirement of reconstructing the other B meson and the great challenge in rejecting background from $B \to X_s J/\psi$, where J/ψ subsequently decays to di-leptons. Even if a veto could be applied on the q^2 spectrum, a great amount of charmonium events could escape the veto due to detector resolution or bremsstrahlung in the case of an electron.

The sum-of-exclusive method, however, reconstructs X_s system explicitly by summing various exclusive hadron states, for example, one kaon and several pions, and has been widely adopted in past measurements by Belle [27, 28] and *BABAR* [29, 30]. In this way, the kinematic information of the target B meson is available by summing the four-momentum of X_s system and di-lepton system. It provides an additional key variable compared to the full inclusive method: ΔE, which is defined to be the difference between the reconstructed energy and the beam energy in the center-of-mass frame of $\Upsilon(4S)$: $\Delta E = E_B - E_{\text{beam}}$. Using this variable, one could significantly suppress the $B \to X_s J/\psi$ background without expanding the q^2 veto range. The reconstruction efficiency is also higher than the full inclusive method, usually at the order of $O(1\%)$. However, this method has a significant drawback of large systematics coming from the modeling of the X_s system, concerning the detection efficiency of reconstructed modes and extrapolation from reconstructed modes to all decay modes of X_s. A detailed discussion is given in the following subsections.

1.5.1 Differential Branching Fractions

One of the most important observables in inclusive measurements is the differential branching fraction, which is linked to Wilson coefficients in the following way [31]:

$$\frac{d\Gamma}{dq^2} = \Gamma_0 m_b^3 (1-\hat{s})^2 \left[\left(|C_9|^2 + |C_{10}|^2\right)(1+2\hat{s}) + \frac{4}{\hat{s}}|C_7|^2(2+\hat{s}) + 12\Re\left(C_7^* C_9\right) \right], \quad (1.9)$$

where

$$\Gamma_0 = \frac{G_F^2}{48\pi^3}\frac{\alpha^2}{16\pi^2}|V_{tb}V_{ts}^*|^2 \quad (1.10)$$

and α is the fine-structure constant.

The latest $d\mathcal{B}/dq^2$ measurements by Belle [28] and *BABAR* [30] are presented in Fig. 1.7, performed with 152 and 471 million of $B\overline{B}$ pairs, respectively. The Belle measurement reconstructs X_s with 18 different final states consisting of one $K^\pm$ or K_S^0 and up to four pions, where at most one pion can be neutral. A cut of $m(X_s) < 2.1\,\text{GeV}/c^2$ is applied to reject continuum backgrounds. For the *BABAR* measurement, they reconstruct X_s similarly to the Belle method, except that they only use 10 hadronic modes with at most 2 pions and cut $m(X_s)$ below $1.8\,\text{GeV}/c^2$ to enhance the signal-to-noise ratio. In the most interesting low-q^2 region of $1 < q^2 < 6$ $[(\text{GeV}/c)^2]$, Belle and *BABAR* measured the branching fractions to be $(1.493 \pm 0.504^{+0.411}_{-0.321}) \times 10^{-6}$ and $(1.60^{+0.41+0.17}_{-0.39-0.13} \pm 0.18) \times 10^{-6}$, respectively. For the Belle analysis, the second and third terms are statistical and systematic uncertainties, respectively. For Babar measurements, the first uncertainty is statistical, the second and third uncertainties are systematic uncertainties irrelevant and relevant to the X_s modeling. Although their center values are smaller than the

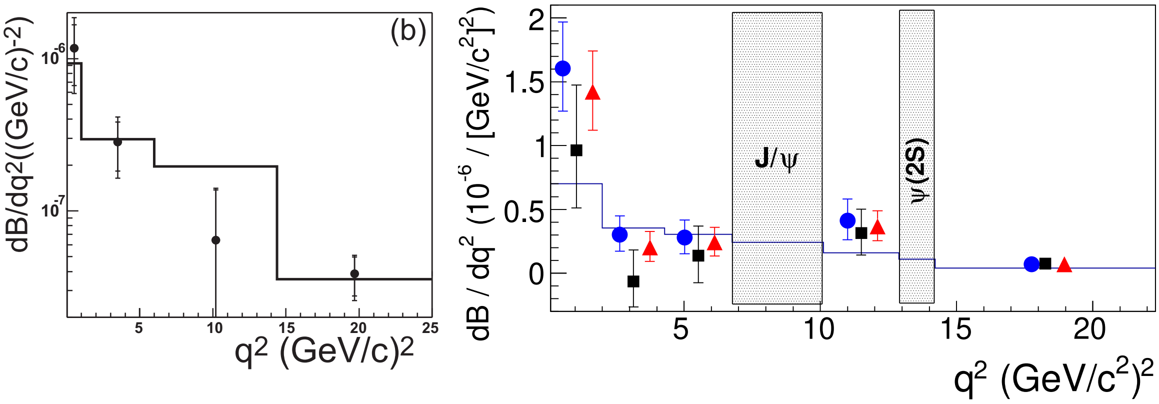


Figure 1.7: $d\mathcal{B}/dq^2$ of inclusive $b \to s\ell^+\ell^-$ process measured by Belle [28] (left) and *BABAR* [30] (right). In the left plot, the points and error bars are data averaging over electron and muon results, while the histogram is the simulation sample normalized to data statistics. In the right plot, blue circles and black squares are electron and muon channel results, respectively. Red triangles are the average. The histogram represents the SM expectation.

theoretical predictions of $(1.75 \pm 0.13) \times 10^{-6}$ [32], showing the same trend as exclusive measurements, they suffer from large statistical and systematic uncertainties that prevent us from reaching a conclusion.

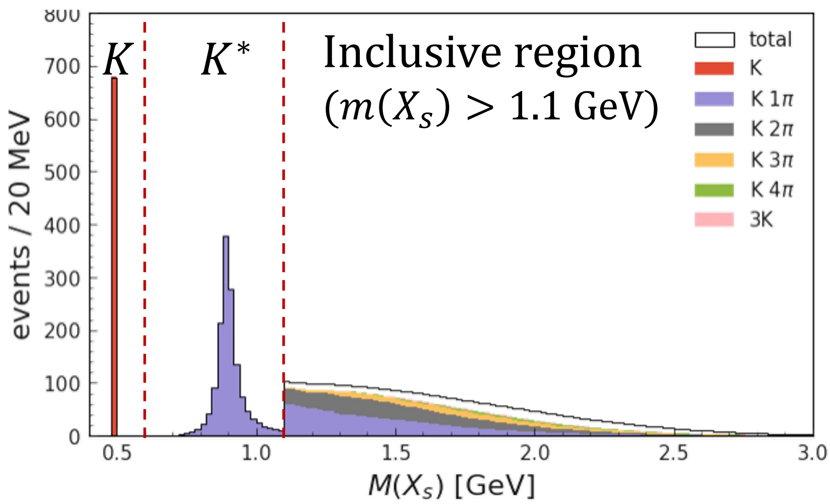


Figure 1.8: Illustration of X_s modeling in Belle analysis.

In both analyses, systematic uncertainties are at the same level as statistical uncertainties, and both are dominated by the modeling of X_s, which is roughly 18% for Belle and 11% for *BABAR*. These analyses rely on Monte Carlo (MC) simulation samples of the signal process for the estimation of reconstruction efficiency and different final states of X_s have significantly different efficiency, and that is why modeling is the dominant systematic uncertainty. Take the Belle analysis as an example, and an illustration of its signal modeling is shown in Fig. 1.8. This modeling is composed of two parts: the exclusive region and the inclusive region. The exclusive region is composed of $B \to K^{(*)}\ell^+\ell^-$ processes modeled with the decay amplitude in Eq. 1.3, hadron transition matrix element (Eq.1.4) decomposition in Appendix A.1.1 and Appendix A.2.1, with form factors taken from the LCSR model [6]. The inclusive region, defined as $m(X_s)$ greater than $1.1\,\mathrm{GeV}/c^2$, is generated with the Fermi motion model [33, 34] to simulate the inclusive $m(X_s)$ lineshape and uses JETSET [35] for fragmentation simulation. Samples of exclusive and inclusive regions are merged according to their theoretically predicted branching fractions with large uncertainties. Estimation of systematics from fragmentation is performed simply by varying all events containing π^0 and K^0_S by 22%. As one of the key components of this dissertation, several methods are studied to improve the modeling of X_s and more elaborate ways are proposed to estimate the corresponding systematics with an outline shown in Chap. 3.

1.5.2 Angular Decomposition

Similarly, as in the exclusive cases, the angular variable θ_ℓ can also be defined in the inclusive decays, and a decay width decomposition is given by the literature [36]:

$$\frac{d^2\Gamma}{dq^2 d\cos\theta_\ell} = \frac{3}{8}\left[(1+\cos^2\theta_\ell)H_T(q^2) + 2\cos\theta_\ell H_A(q^2) + 2(1-\cos^2\theta_\ell)H_L(q^2)\right], \quad (1.11)$$

where the angular observables are linked to Wilson coefficients in the following way:

$$\begin{aligned} H_T(q^2) &= \Gamma_0 m_b^3 2(1-\hat{s})^2\hat{s}\left[\left|C_9 + \frac{2}{\hat{s}}C_7\right|^2 + |C_{10}|^2\right], \\ H_A(q^2) &= -4\Gamma_0 m_b^3(1-\hat{s})^2\hat{s}\Re\left[C_{10}\left(C_9 + \frac{2}{\hat{s}}C_7\right)\right], \\ H_L(q^2) &= \Gamma_0 m_b^3(1-\hat{s})^2(|C_9 + 2C_7|^2 + |C_{10}|^2). \end{aligned} \quad (1.12)$$

For the H_L term the hadronic current is longitudinally polarized; therefore, it is proportional to $\sin^2\theta_\ell = 1 - \cos^2\theta_\ell$, while for the H_T and H_A terms, the hadron current is transversely polarized. H_T and H_L contain only contributions from pure vector lepton current (C_9 and C_7 since the photon field only couples to vector lepton current) or axial vector lepton current (C_{10}). On the other hand, H_A represents the interference between vector and axial vector lepton current; therefore, there is an interference term $\Re\left[C_{10}\left(C_9 + \frac{2}{\hat{s}}C_7\right)\right]$ and it causes asymmetry in the forward ($\cos\theta_\ell > 0$) and backward ($\cos\theta_\ell < 0$) directions.

By integrating out $\cos\theta_\ell$, one can also derive the relationship between angular observables and $d\Gamma/dq^2$ and Forward-Backward asymmetry variable A_{FB}:

$$\frac{d\Gamma}{dq^2} = H_T(q^2) + H_L(q^2), \quad (1.13)$$

$$\frac{dA_{FB}}{dq^2} = \left(\int_0^1 - \int_{-1}^0\right) d\cos\theta_\ell \frac{d^2\Gamma}{dq^2 d\cos\theta_\ell} = \frac{3}{4} H_A(q^2). \tag{1.14}$$

In 2014, Belle presented their measurement on A_{FB} as a function of q^2 using the full dataset with 772×10^6 $B\overline{B}$ pairs [37], whose result is shown in Fig. 1.9. This measurement is consistent with the SM predictions, but is limited by large statistical uncertainty. So far, measurement of the other two angular variables H_T and H_L has not been performed yet due to lack of statistics.

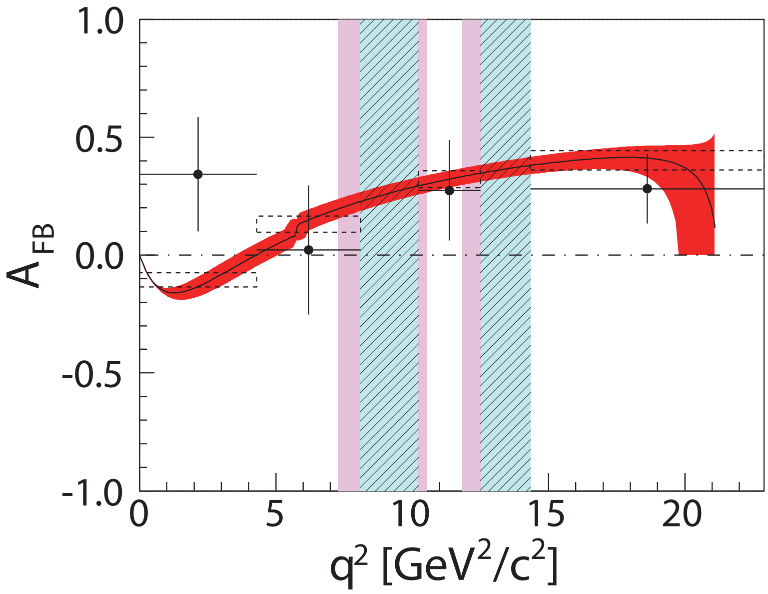


Figure 1.9: Forward-Backward asymmetry A_{FB} as a function of q^2 measured by Belle [37]. Black dots with error bars represent the measurement and curve (black) with the band (red) and dashed boxes (black) shows the SM prediction. The backgrounds from $J/\psi \to \ell^+\ell^-$ and $\psi(2S) \to \ell^+\ell^-$ events have been vetoed by the teal hatched regions. For the electron channel, the pink shaded regions are added to the veto regions due to the large bremsstrahlung effect and worse momentum resolution induced by multiple scattering.

1.5.3 Test of Lepton Flavor Universality

Similar to the exclusive case, one can define the LFU variable also for the inclusive case:

$$R(X_s) = \frac{\mathcal{B}(B \to X_s\mu^+\mu^-)}{\mathcal{B}(B \to X_s e^+e^-)}. \tag{1.15}$$

And the prediction of this variable in the SM is unity with high precision. Past measurements from Belle [28] and *BABAR* [30] reported branching fractions from muon and electron channels separately but did not give an official result on this variable. By assuming that all systematics are canceled out when taking the ratio, one can derive the results to be 1.02 ± 0.42 and $0.57^{+0.18}_{-0.17}$ for Belle and *BABAR*, respectively. The Belle result is consistent with the SM prediction, though it suffers from high statistical uncertainty. The

BABAR result deviates from unity by 2.4σ. However, it should be noticed that *BABAR* observed 0-consistent muon signals at low-q^2 areas because of its issue in the muon detector and for this reason we tend to interpret this discrepancy as statistical fluctuation.

1.6 Constrains on the Wilson Coefficients

Combining all existing measurement results to get their constraints on the values of C_9 and C_{10} is a powerful way to test the existence of new physics. In addition to the measurements mentioned above, other channels such as $B_s \to \mu^+\mu^-$ [38, 39, 40] and $B_s \to \phi\mu^+\mu^-$ [41] are also included in the global fitting.

In the study carried out by B. Capdevila [5], global fitting under various scenarios of new physics assumptions is discussed, including violation of LFU and existence of right-handed quark current in the Effective Hamiltonian. In this section, we discuss the fitting result based on the minimum extension of the SM, which means that we do not admit violation of LFU or right-handed quark current.

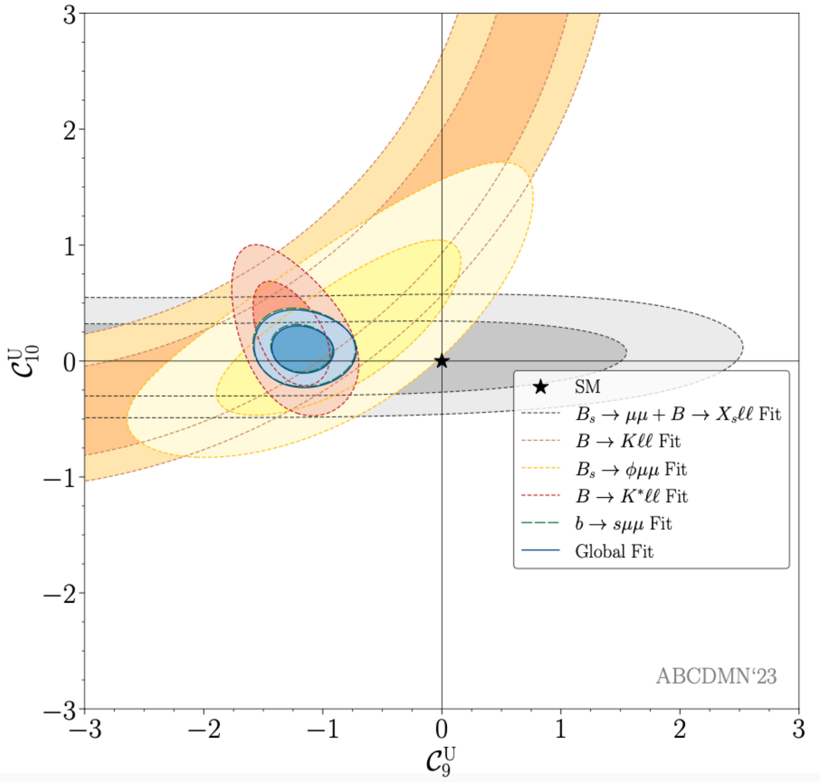


Figure 1.10: 1σ (dark-shaded) and 2σ (light-shaded) confidence regions for (C_9^U, C_{10}^U). Distinct fits are performed separating each of the $b \to s\ell^+\ell^-$ modes (short-dashed contours), and global fit (solid contours). [5]

The result for the global fit is shown in Fig. 1.10. In this plot, both vertical and horizontal axes are Wilson coefficients, subtracting the SM predictions: $C_{9(10)}^U \equiv C_{9,(10)} -$

$C^{\mathrm{SM}}_{9,(10)}$, where the superscript U denotes the LFU assumption. The best fit point and standard deviations are summarized in Tab. 1.2. So far, all existing measurements support the conclusion that $C^{U}_{10} = 0$, while all exclusive measurements indicate that C_9 is smaller than the SM prediction by 1. Still, the inclusive measurements, represented by the grey contour, show no sensitivity to C_9 detection due to limitations on measurement precision. This global fit result favors new physics models that only couples to vector lepton current such as the Z' model [42, 43], which is a possible mediator between the SM particles and dark matter candidates [44].

Table 1.2: Best fit values of C^U_9, C^U_{10} and their projected 1σ interval, pull to the SM predictions (0,0), and p-value of the fitting.

	Best fit	1σ interval	$\mathrm{Pull}_{\mathrm{SM}}$	p-value
C^U_9	-1.18	$[-1.35,\ -1.00]$	5.5	39.1%
C^U_{10}	$+0.10$	$[-0.04,\ +0.23]$		

Chapter 2

The Belle II Experiment

The Belle II experiment [45] is a B-factory experiment being pursued at the SuperKEKB [46] electron-positron collider at KEK, Tsukuba, Japan. The beam energies of electrons and positrons are set to 7 GeV and 4 GeV, respectively. This asymmetric beam energy results in the Lorentz boost of the e^+e^- center-of-mass system with $\beta\gamma = 0.28$. The center-of-mass energy $\sqrt{s}$ is set at the $\Upsilon(4S)$ resonance for most of the time, which subsequently decays into $B\overline{B}$ pair. In addition, continuum processes of $e^+e^- \to q\overline{q}$, where q refers to u, d, s, c quarks and tau-lepton pairs are also produced at this energy, bringing a rich field of physics. Data at off-resonance (60 MeV below $\Upsilon(4S)$ resonance) and $\Upsilon(5S)$ resonance energies are also taken for continuum process and quarkonium studies, respectively.

In this chapter, we describe the components and operation status of the SuperKEKB accelerator and the Belle II detector.

2.1 SuperKEKB

A schematic drawing of SuperKEKB is shown in Fig. 2.1. It consists of a linear accelerator (LINAC), positron damping ring, Beam Transfer (BT), and main storage rings. The injector part of LINAC produces low-emittance, high-current electron beams by employing a photocathode RF gun. Positron beams are generated by shooting electron beams into a tungsten target, with high-current primary electrons generated by a thermionic gun. The LINAC then accelerates electron and positron beams up to their designed energies, and the beams are then injected into the High Energy Ring (HER) and Low Energy Ring (LER) via the BT, respectively. The damping ring near the LINAC reduces the emittance of the positron beam before injecting it into the storage ring. The Belle II detector is located at the crossing point of the two storage rings where the two beams collide. There are two different injection modes in the SuperKEKB collider: "normal injection" mode and "continuous injection" mode. The former is used to accumulate the beam current fast with the Belle II detector turned off, while the latter is used to keep the beam current around a fixed value, with Belle II detector keeping data taking. These

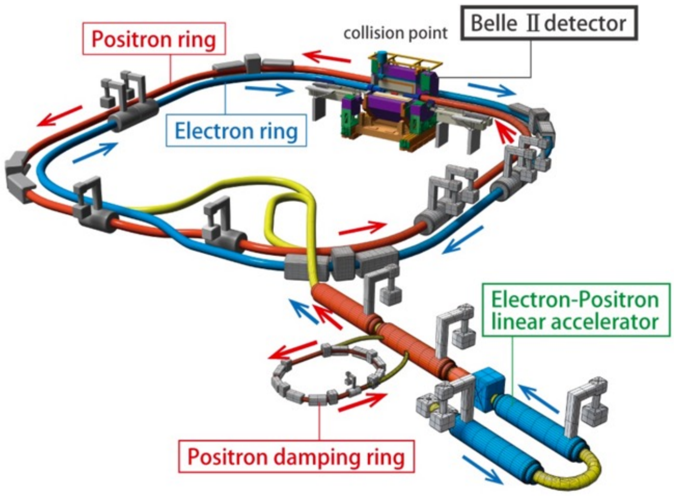


Figure 2.1: The schematic drawing of the SuperKEKB. All key components labeled in the plot with the yellow pipes being the beam transfer.

flexible injection modes guarantee high data collection efficiency.

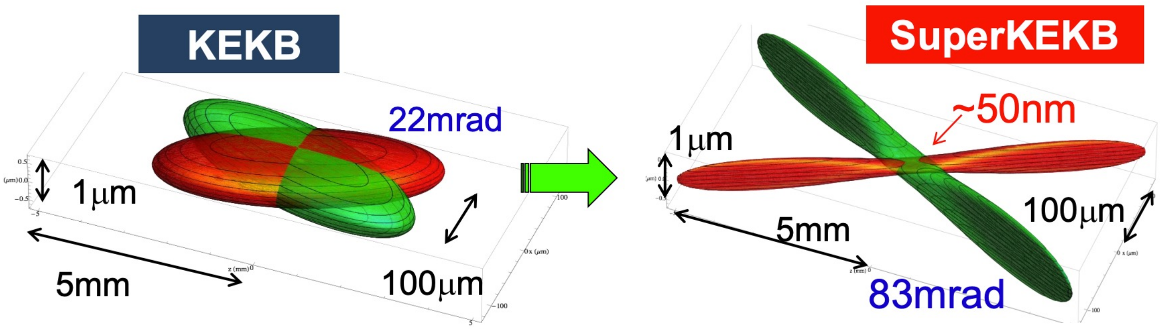


Figure 2.2: Illustration of colliding beams at the IP for KEKB (left) and SuperKEKB (right). Compared with KEKB, in nano-beam scheme of SuperKEKB, the beams are squeezed to a much smaller size at the IP and the crossing angle of two beams are much larger to reduce the overlap region of two beams.

The Belle II experiment aims to accumulate $50\,\mathrm{ab}^{-1}$ of data. To make this target achievable in a reasonable time span, the KEKB accelerator was upgraded to the SuperKEKB accelerator, aiming to achieve an instantaneous luminosity of 6×10^{35} cm^{-2}s^{-1}. To achieve the instantaneous luminosity goal, the so-called "nano-beam" scheme [47] is applied to SuperKEKB. The instantaneous luminosity $\mathcal{L}$ of the e^+e^- collider at the beam-beam limit is proportional to the following components:

$$\mathcal{L} \propto \gamma_\pm \frac{I_\pm \xi_{y\pm}}{\beta_y^*} \tag{2.1}$$

where $(+)$ and $(-)$ stands for positrons and electrons, γ is the Lorentz factor, I is the beam current, ξ_y represents the beam-beam effect in the vertical direction, and β_y^* is the vertical beta function at the Interaction Point (IP). From this formula, increasing the beam current and squeezing the vertical beta function at IP are effective methods to increase the collision luminosity. For this reason, in the upgrade of KEKB to SuperKEKB, the beam currents of LER and HER are increased from 1.6 A to 2.8 A and from 1.2 A to 2.0 A, respectively. And β_y^* is squeezed from 5.9 mm to about 0.3 mm, accompanied by the “Hourglass effect” which means that the smaller the beam size at the IP is squeezed, the larger the beam size before and after the IP steeply expands. For this reason, in the nano-beam scheme, the traditional head-on collision used in the predecessor KEKB accelerator is abandoned, and the two beams form a large crossing angle to avoid the additional beam-beam effect outside the IP that prevents luminosity improvement as shown in Fig. 2.2. A comparison of the beam parameters achieved by KEKB and the target parameters of SuperKEKB is shown in Tab. 2.1.

Table 2.1: Beam parameters achieved by KEKB (left) and the target parameters of SuperKEKB (right). The $\beta_{x(y)}^*$ is the beta function in horizontal (vertical) direction at the IP. I is beam current. θ is the crossing angle of LER and HER beams. L is instantaneous luminosity.

Beam parameters	KEKB achieved	SuperKEKB target
β_{y}^* (LER/HER) [mm]	5.9/5.9	0.27/0.3
β_{x}^* (LER/HER) [mm]	1200/1200	32/25
I (LER/HER) [A]	1.64/1.19	2.80/2.00
number of bunches	1584	1761
θ [mrad]	22	83
$\mathcal{L}\,[\times 10^{35}\ \mathrm{cm}^{-2}\ \mathrm{s}^{-1}]$	0.21	6

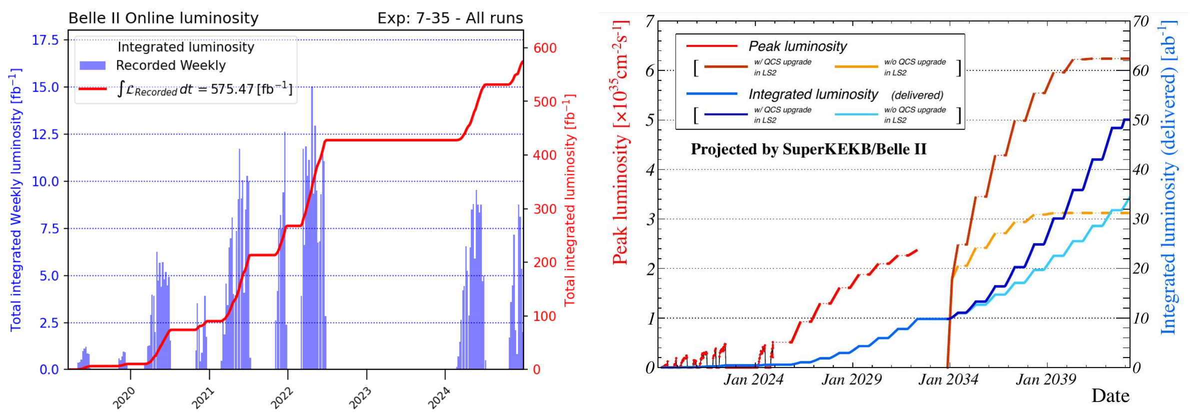


Figure 2.3: Left: weekly collected data luminosity of SuperKEKB operation. Right: luminosity projection of SuperKEKB operation.

The first data taking period, called Run I, started from 2019 and lasted until 2022 summer. In total, $424\,\mathrm{fb}^{-1}$ of data are taken, including $365\,\mathrm{fb}^{-1}$ taken at the $\Upsilon(4S)$ resonance. During Long Shutdown I from 2022 summer to 2024, several accelerator and detector maintenance tasks were performed. The second data taking period, Run II, started in 2024, and $151\,\mathrm{fb}^{-1}$ data have been collected by the end of 2024. The collected luminosity and projected luminosity of SuperKEKB are shown in Fig. 2.3.

2.2 Coordinate System

The global coordinate system of the Belle II detector is shown in Fig. 2.4. The origin point is defined as the nominal interaction point. The z axis is along the bisector of the angle between the LER and HER beam pipes in the horizontal plane. The x axis is in the horizontal plane and points outside of the main storage ring. The y axis is vertical and points upward. The $x-y-z$ axes form a right-handed Cartesian coordinate system. The forward direction is defined as along the z axis direction and the backward direction is defined as against the z axis direction. In addition, in this thesis, the radius r is defined to be $\sqrt{x^2+y^2}$, the polar angle $\theta = \arccos(z/\sqrt{x^2+y^2+z^2})$ and the azimuth angle ϕ is defined to be $\arctan(y/x)$.

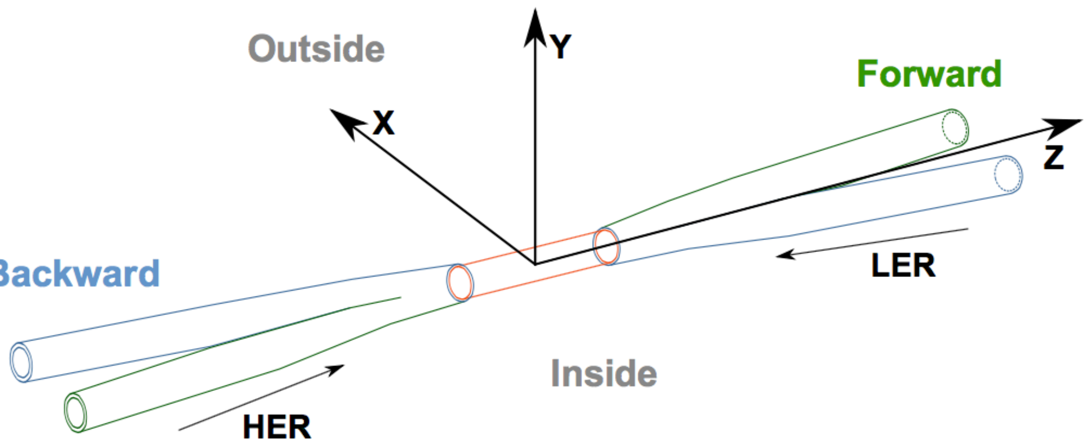


Figure 2.4: The global coordinate system of Belle II.

2.3 Belle II detector

The Belle II detector is a general purpose detector which consists of 7 sub-detectors surrounding the IP. A schematic drawing of the Belle II detector is shown in Fig. 2.5. Starting from the innermost to outside, there are the Vertex Detector (VXD) system, consisting of the Pixel Detector (PXD) [48] and the Silicon Vertex Detector (SVD) [49], the

Central Drift Chamber (CDC) [50], the Time of Propagation counter (TOP) [51], Aerogel Ring Imaging Cherenkov counter (ARICH) [52], and the Electromagnetic Calorimeter (ECL) [53], and K_L^0 and μ detector (KLM) [54]. A superconductor solenoid is placed between the ECL and the KLM, generating a 1.5 T magnetic field along the z axis, which bends the trajectory of charged particles for momentum measurement. This detector system covers a polar angle interval of $17° < \theta < 150°$, having larger geometry acceptance in the forward direction than the backward, due to the asymmetrical energy of the beams.

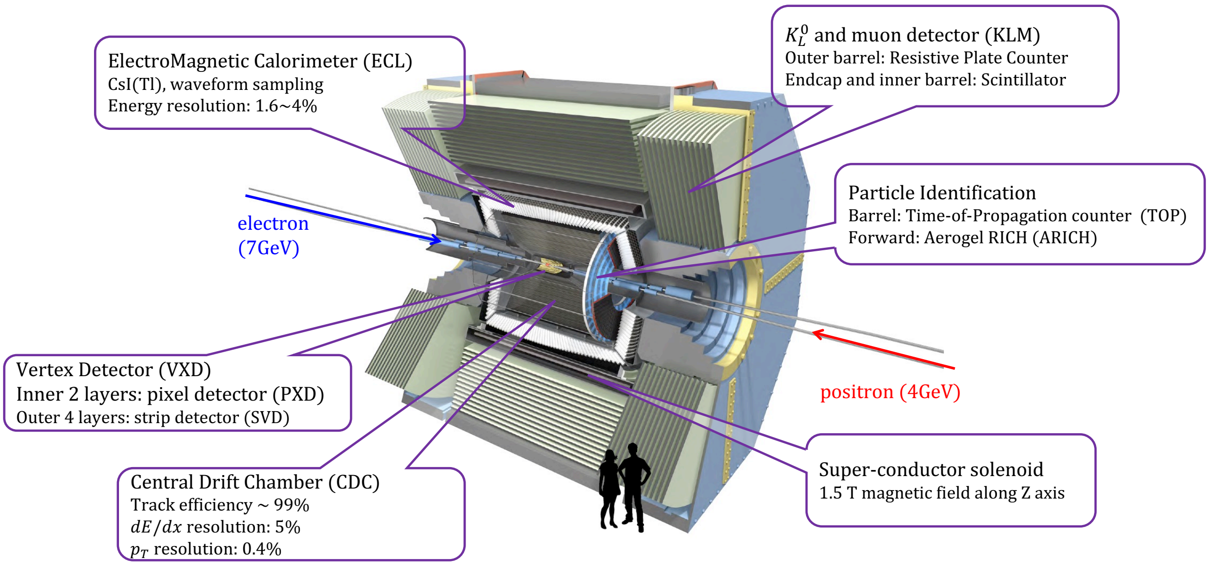


Figure 2.5: Schematic drawing of Belle II detector.

2.3.1 Vertex Detector

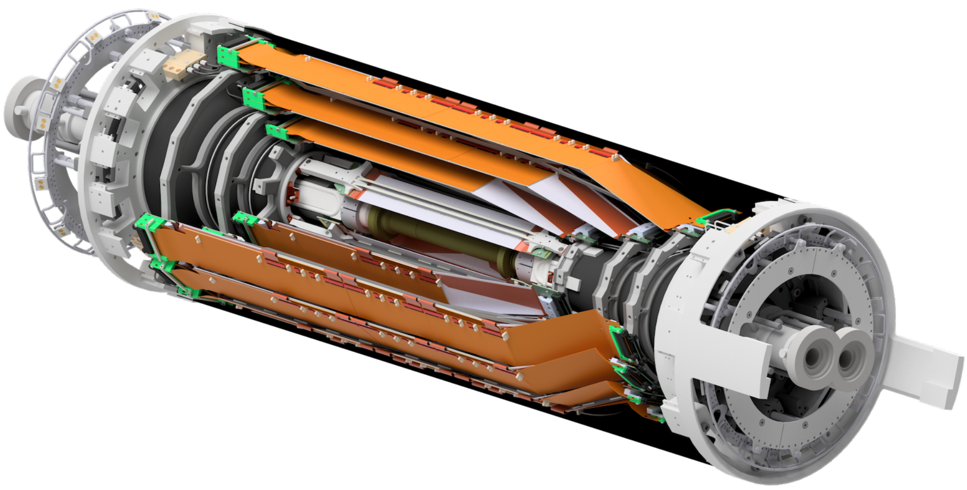


Figure 2.6: Belle II Vertex Detector (VXD).

The VXD is placed closest to the IP for a good vertex reconstruction. The schematic view of VXD is shown in Fig. 2.6. It is a six-layer silicon detector, with the inner two layers being the PXD and the outer four layers being the SVD. To achieve a better resolution for the measurement of the vertex position, the first layer is placed only 14 mm away from the IP, which is 6 mm closer than the innermost layer of the Belle silicon vertex detector. This is achieved by reducing the radius and thickness of the beam pipe. In addition, SVD is also essential in tracking charged particles, especially low-momentum ones, and K_S^0 particle reconstruction. Particle identification by measuring the characteristic energy loss $\mathrm{d}E/\mathrm{d}x$ inside the SVD sensors is also applied in Belle II. The geometry information of VXD is summarized in Tab. 2.2.

Table 2.2: Geometry of the VXD sensors. Strip pitch in r-ϕ plane and thickness are smaller for sensors placed in the forward direction in layer 4–6.

Subsystem	Layer	Radius (mm)	Pitch (z) (μm)	Pitch (r-ϕ) (μm)	Thickness (μm)
PXD	1	14	55–60	50	75
	2	22	70–85	50	75
SVD	3	39	160	50	320
	4	80	240	50–75	300–320
	5	104	240	50–75	300–320
	6	135	240	50–75	300–320

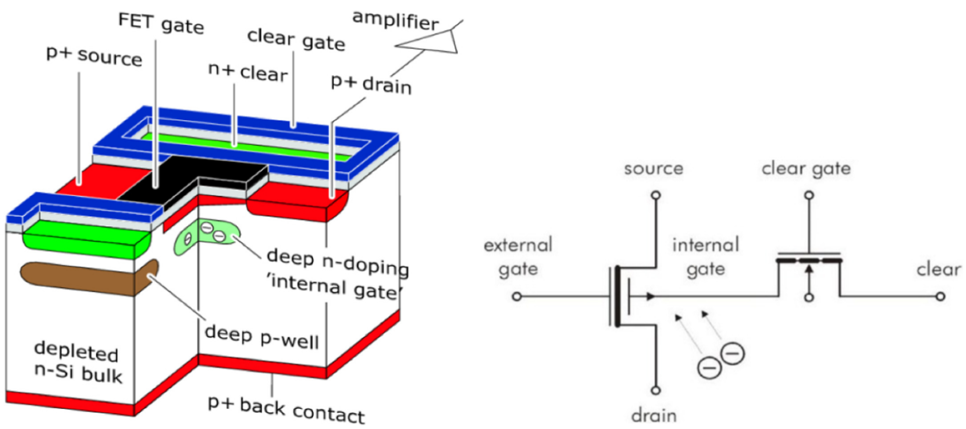


Figure 2.7: Left: structure of the DEPFET pixel. Right: Equivalent circuit of DEPFET. [55]

The PXD uses DEpleted P-channel field effect transistor (DEPFET) [55] technology. The structure of the DEPFET pixel and its equivalent circuit are shown in Fig. 2.7. It has a deep n-doping internal gate beneath the p-channel field effect transister (FET), a deep p-well under n^+ doped clear implant, fully depleted n^- doped Si bulk and p^+ backside. A high voltage of around -70 V is applied to the electrode that is connected to a front side p^+ implant to achieve the depletion. Electrons created in the bulk drift to the internal gate and collected. By applying a negative voltage at the external gate, the

MOSFET is open and the magnitude of source current is proportional to the collected charge at the internal gate. Charges in the internal gate is cleared after readout.

The relatively large area and fine pitches of the PXD sensors result in nearly 4 million pixel units, making it impossible to readout every pixel in one readout cycle. For this reason, only pixels within the Rigon of Interest (ROI) are readout, which is determined by tracks extrapolated to PXD using hit information from the SVD that surrounds the PXD. During Run I of the Belle II operation, only 1/6 of sensors are installed for the outer layer of the PXD covering around 1/6 of azimuth angle. During the Long Shutdown I in 2023 summer shutdown, a new PXD detector with full sensors is installed replacing the old PXD detector.

The resolution of the impact parameter of the track in the x-y plane (d_0) is measured to be 13.64 ± 0.08 μm and the resolution in the z direction is 14.92 ± 0.07 μm. These measurements are performed with Bhabha events collected by the Belle II detector, which require tracks fulfilling $|\theta - \pi/2| < 0.5$, $p_T > 1\,\mathrm{GeV}/c$, $p\beta \sin^{3/2}\theta > 2\,\mathrm{GeV}/c$, where p_T is the transverse momentum of the track, θ is the polar angle, and β is the velocity normalized by the speed of light. These resolutions are better than the Belle/*BABAR* detector by a factor of two.

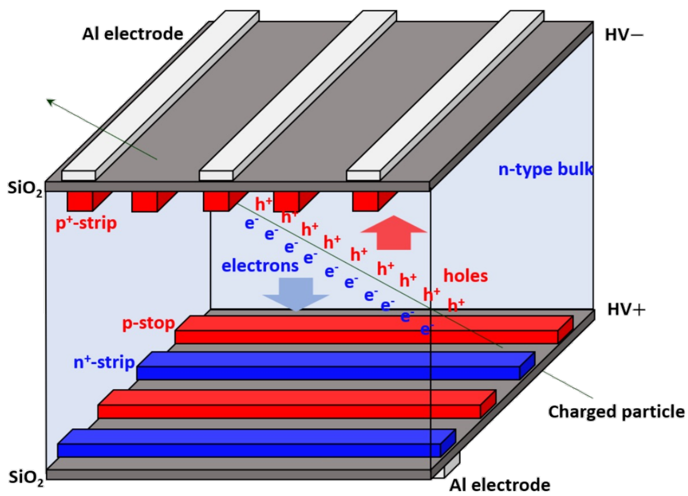


Figure 2.8: Structure of DSSD. p^+-strip and n^+-strip are more deeply doped compared with n-type bulk.

The SVD adopts the Double-sided Silicon Strip Detector (DSSD) technology features in placing perpendicular sensitive strips to two sides of one sensor and small strip pitch for high-resolution position detection. An illustration of the DSSD structure is shown in Fig. 2.8. It is composed of p^+- and n^+- strips on the two surfaces and n^--doped silicon bulk. Inverse bias voltage of 100 V is applied between the p^+-strips and n^+-strips, under which the bulk is fully depleted. When charged particles pass through the DSSD, electron-hole pairs are created and drift to n^+-strip and p^+-strip, respectively, along with the electrical field created by inverse bias voltage. The electric signals are readout via aluminum electrodes that are AC coupled to the n^+ and p^+ strips. In order to improve the spatial resolution, floating strips are placed between every two readout strips. The charge collected by these floating strips is read out by the neighboring readout strips, and the effective strip pitch is reduced to half of the readout pitch. The silicon oxide contains fixed positive oxide charges, which attract electrons in the n-type bulk and strips and thus create a conductive layer in the silicon, shorting all n-type doped areas. For this reason, p-type stop strips (p-stop) are placed between every two adjacent n^+-strips to prevent the divergence of collected charges on each strip and suppress the noise level. The thickness of the DSSD is 300 μm and 320 μm, with a material budget corresponding to approximately 0.4% of radiation length X_0.

2.3.2 Central Drift Chamber

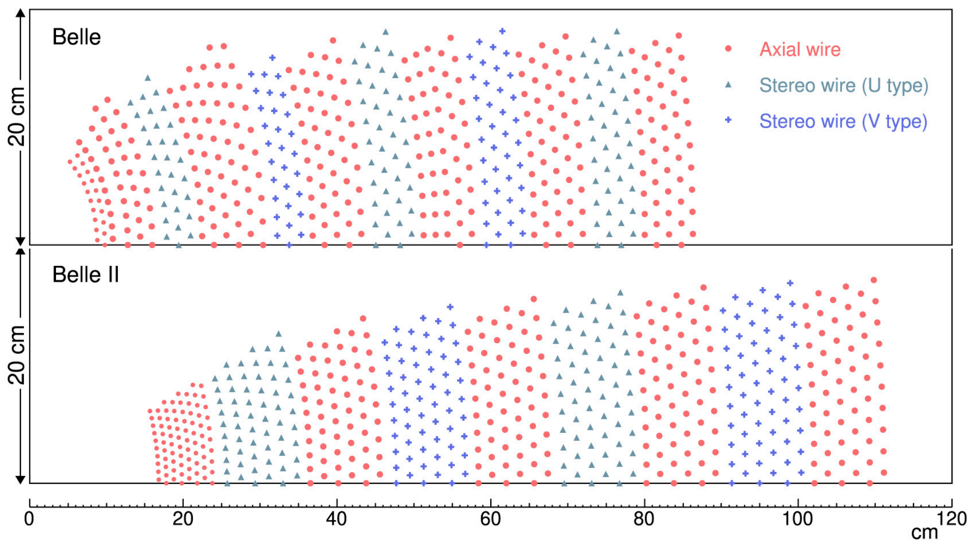


Figure 2.9: Layer configuration of the Belle and Belle II CDCs.

The CDC is the main tracking detector for Belle II. It is a wire chamber filled with a gas mixture of 50% He and 50% C_2H_6. Helium is a low-Z noble gas and is notable for its stable chemical properties, longer radiation length, and smaller multiple scattering effect with charged particles to measure. C_2H_6 plays the role of quencher gas that has small rotation energy levels to absorb the de-excitation photons produced in avalanches near the wire. The CDC contains 14336 sense wires arranged in 56 layers, grouped into nine superlayers shown in Fig. 2.9 with different colors. Five of the superlayers are axial wires along the z axis and thus only sensitive to position measurement in the r-ϕ plane. Four superlayers are stereo wires that are inclined by 60–80 mrad to the z axis, making detection in the z direction possible by combining information from axial wires. Compared to Belle, Belle II CDC has a larger volume with an outer radius extended from 880 mm to 1130 mm, resulting in better momentum resolution. To compensate for a high hit rate from the beam-induced background, a finer cell size is adopted in the innermost superlayer.

Using cosmic rays, the transverse momentum resolution of CDC is measured to be $(0.127 \pm 0.001)\frac{p_T}{\mathrm{GeV}/c} \oplus (0.321 \pm 0.003)$ [%].

The CDC also contributes to particle identification by measuring the characteristic energy loss of charged particles. Fig. 2.10 shows the energy loss $\mathrm{d}E/\mathrm{d}x$ in the CDC as a function of track velocity ($\beta\gamma$) and momentum, with a $\mathrm{d}E/\mathrm{d}x$ resolution to be approximately 5%. This is powerful in separating protons, kaons and electrons from pions, particularly for particles with momentum below 1 GeV/c. The time of each collision (event T0) is measured by averaging the time of all hits in CDC, with a resolution smaller than 3 ns.

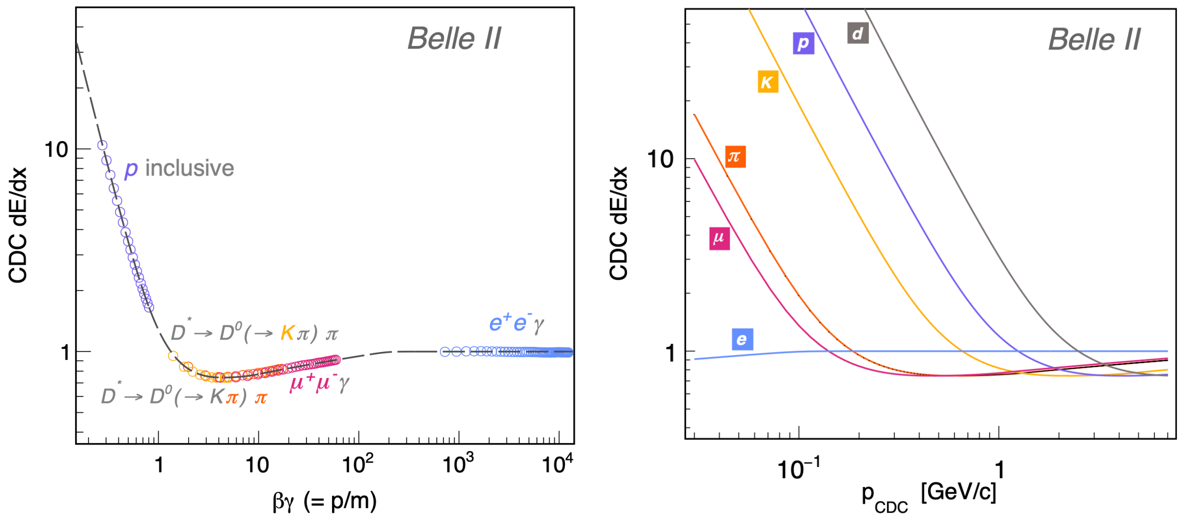


Figure 2.10: Characeristic energy loss of charged particles inside cdc as a function of $\beta\gamma$ (left) and track momentum (right).

2.3.3 Cherenkov Detectors

When charged particles penetrate Cherenkov detectors faster than the speed of light in the detector materials, Cherenkov lights are emitted with an opening angle called the Cherenkov angle θ_C expressed below:

$$\cos\theta_C = \frac{1}{n\beta}, \tag{2.2}$$

where n is the refraction index of the material. In Belle II, this feature is mostly used for particle identification between charged $K^\pm$ and $\pi^\pm$ and there are two Cherenkov detectors in Belle II: TOP and ARICH detectors placed in the barrel and forward end-cap, respectively.

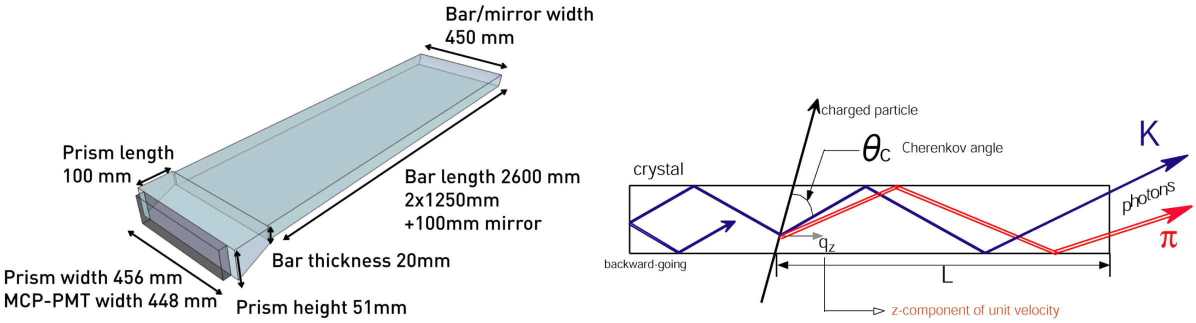


Figure 2.11: Left: Schematic of TOP module. Right: Illustration of time of propagation of Cherenkov photons inside the radiator.

The TOP detector is composed of 16 modules, each made with a 2.6 m long quartz radiator ($n = 1.47$) with a focusing mirror on one side and micro channel plate photomultipliers (MCP-PMT) on the other side, as shown in Fig. 2.11. The MCP-PMT has a time resolution of 30–40 ps. Typically, when a charged particle penetrates the radiator, 20–40 photons are emitted and they propagate inside the radiator until it hits the MCP-PMT. The time interval from Cherenkov photons generation until when they hit the sensor is called time of propagation. This propagation time is smaller for a pion compared to a kaon because of pion's larger Cherenkov angle, which leads to a longer propagation length inside the radiator. At the same time, the time of flight for pions is also smaller than that of kaon, making it possible to perform particle identification based on the hit time of Cherenkov photons, which is basically the sum of the time of flight and the time of propagation. Given the track information from CDC (track momentum, incident position, and incident angle to the quartz bar), one can calculate the probability density function (PDF) of each particle hypothesis and compare it with observed hit pattern to calculate the likelihood. Fig. 2.12 shows an example of particle identification with TOP. Based on the arguments above, it is easy to figure out that with fixed track parameters, the hit time relationship for pion, kaon, and proton should be: pion < kaon < proton, considering their mass relationships. For this special case, the hit time of observed data (crossing point of magenta lines) matches the kaon hypothesis in the middle the best, and thus it has the greatest likelihood. Using only TOP, $K^{\pm}$ selection efficiency of 85% is achieved with $\pi^{\pm}$ fake rate of 10% [56].

ARICH uses Aerogel as the radiator and detects Cherenkov photons using Hybrid Avalanche Photon Detector (HAPD). The radiator consists of two layers of aerogel with different refractive indices (1.045 in the upper stream and 1.055 in downstream) to increase the number of radiated Cherenkov photons. Compared to the TOP detector, material with a relatively small refraction index is chosen to gain more resolution in separating high-momentum K/π tracks up to $4\,\mathrm{GeV}/c$. The photon detector is placed about 200 mm away from the radiator as illustrated in Fig. 2.13. The likelihood of different particle hypotheses is given by comparing the hit pattern with the expected hit patterns. To put it in a mathematical way, suppose that the expected photon yield of certain par-

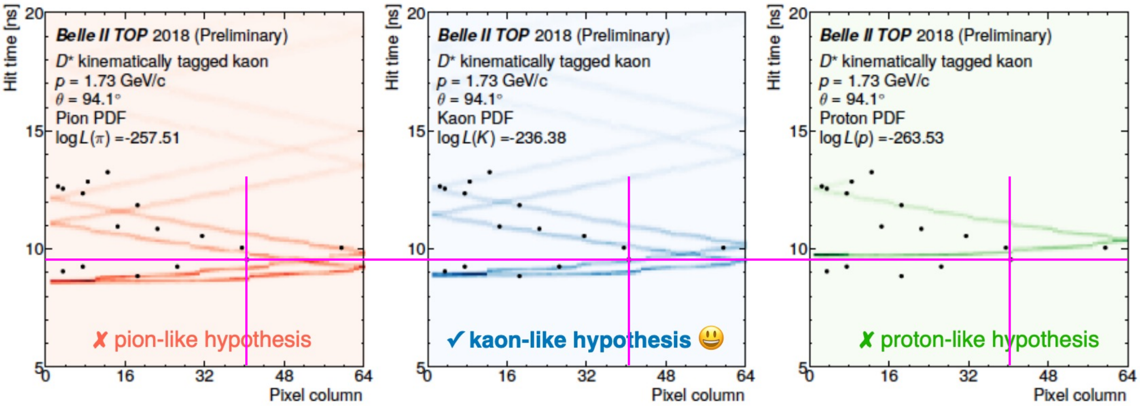


Figure 2.12: Comparison of observed hit pattern (black dots) with PDFs of pion (left, orange), kaon (middle, blue), and proton (right, green). Three verticle magenta lines marks a pixel of MCP-PMT at the same column number, while the horizontal line marks a fixed hit time. Below the three crossing points of magenta lines, there are three black data points correspondingly.

ticle hypothesis (p) in each pixel is $n_{p,i}$, and that each pixel only gives a binary readout, the likelihood is

$$\log \mathcal{L}_p = \sum_{\text{pixels w/o hit}} (-n_{p,i}) + \sum_{\text{pixels w/ hit}} \log(1-e^{-n_{p,i}}) = -N_p + \sum_{\text{pixels w/ hit}} \left[n_{p,i} + \log(1-e^{-n_{p,i}})\right], \tag{2.3}$$

where $N_p = \sum n_{p,i}$ is the expected total photon yield of hypothesis p. ARICH standalone $K^\pm$ identification efficiency is $(93.5 \pm 0.6)\%$ with $\pi^\pm$ a fake rate of $(10.9 \pm 0.9)\%$ using the decay of $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$.

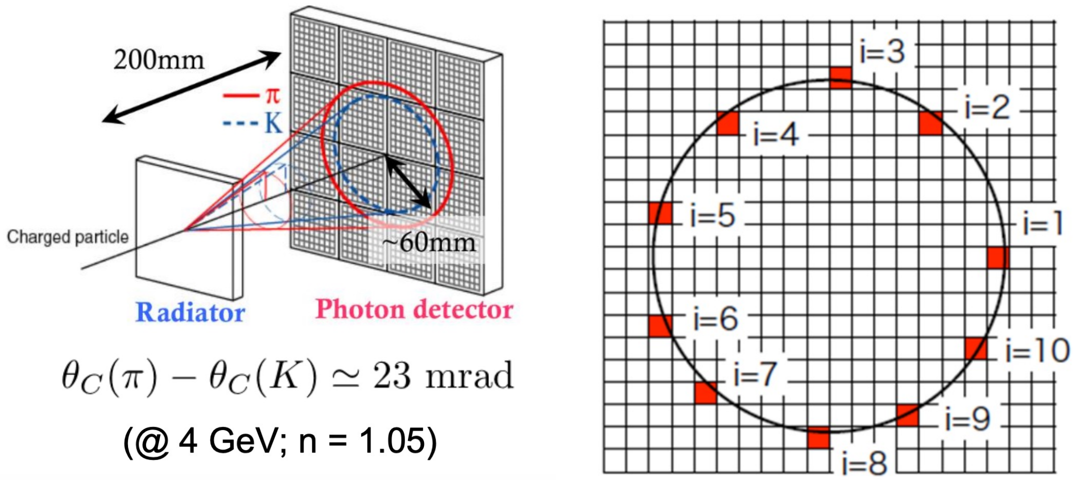


Figure 2.13: Left: Illustration of radiator and photon detector. Right: Illustration of photon hits on the photon detector.

2.3.4 Electron-magnetic Calorimeter

The ECL is a collection of 8736 CsI(Tl) scintillation crystal detectors with scintillation photons collected with the PIN photodiodes, covering 90% of the solid angle in the center-of-mass frame. The crystals have the shape of a truncated pyramid with a length of 30 cm and a 6×6 cm^2 cross section, equivalent to 16.1 radiation length (X_0). It measures energy of photon and electron particles that create electro-magnetic showers inside the scintillator. Other charged tracks penetrating ECL can also lose energies due to ionization energy loss or hadronic interactions.

The scintillation crystals and readout photodiodes were kept unchanged from the Belle experiment, but the backend electronics was upgraded to compensate for the high background environment accompanying with high luminosity operation. The energy resolution is $\sigma_E/E = 4\%$ at 100 MeV, 1.6% at 8 GeV, and the angular resolution is 13 mrad (3 mrad) at low (high) energies.

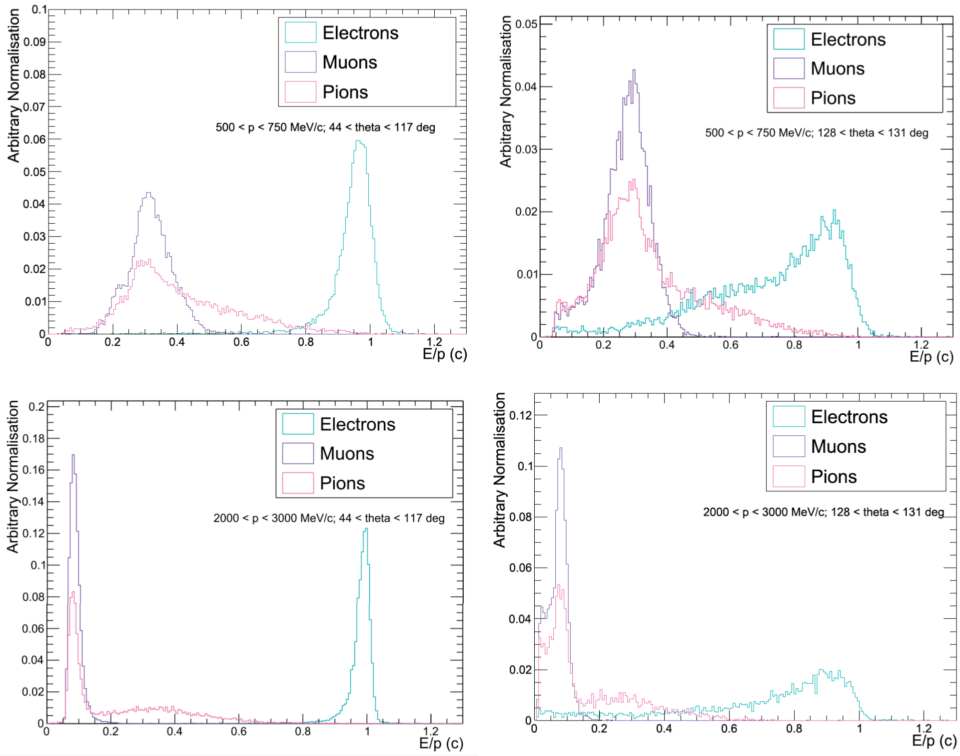


Figure 2.14: The E/p distributions for low momentum ($0.5 < p < 0.75$ [GeV/c], top) and high momentum ($2.0 < p < 3.0$ [GeV/c], bottom) ranges and two polar angle ranges: barrel region of $44° < \theta < 117°$ (left) and backward endcap region of $128° < \theta < 131°$ (right). Excellent identification of electron is achieved when $p > 1$ GeV/c in the barrel region.

The ECL contributes to the identification of electrons primarily based on the ratio between the energy deposited in the ECL and the momentum of the track (E/p), as shown in Fig. 2.14. Given the fact that only electrons deposit all energies inside the

ECL, its E/p peaks at 1 while this is not true for other particles. In addition, charged-hadron particles like pions create strong interaction with crystals and thus leave more energy than muons, who only lose energies via ionization. This explains the tail of pion distributions in the plot.

Recent electron identification algorithms significantly improved electron identification performance compared to likelihood-based algorithms based on E/p by making use of deposit energy distributions in different ECL crystals (called cluster shape) such as Boost Decision Tree (BDT) based algorithm using Zernike moments and Convolution Neural Network (CNN) based algorithms [57].

2.3.5 K_L^0-Muon Detector

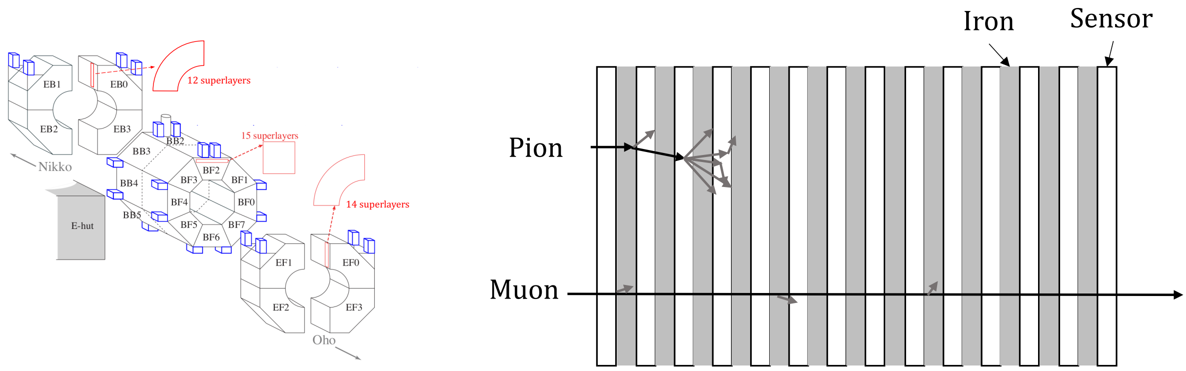


Figure 2.15: Left: KLM schematic structure. Right: Sandwich structure of KLM. The gray arrows represents the secondary particles created in the interaction with KLM materials.

The KLM sub-detector consists of 4.7 cm thick iron plates alternated with 4.4 cm thick active layers. The octagonal barrel KLM (BKLM) is made of 14 iron plates and 15 detector layers, where the sensors in the inner two layers are plastic scintillator and those in the outer 13 layers are Resistive Plate Chamber (RPC). The forward end-cap KLM (EKLM) consists of 14 iron plates and 14 detector layers of plastic scintillator, while there are 12 detector layers in the backward EKLM. A schematic plot of BKLM and EKLM is shown in the left plot of Fig. 2.15. The iron plates in total have a thickness equivalent to more than 3.9 interaction lengths (λ) of pion inelastic hadronic scattering, while the ECL contributes with 0.8λ. Each detection layer is composed of two planes of strip sensors arranged orthogonally to give 2-dimensional coordinates. In KLM, one hit is defined as the overlap region of two strip clusters in the two perpendicular sensor planes. The KLM covers the polar angle range $20° < \theta < 155°$ with respect to the z axis.

Given the similar mass of pion ($139.6\,\mathrm{MeV}/c^2$) and muon ($105.7\,\mathrm{MeV}/c^2$), it is impossible to identify muon with inner detectors that rely on $\mathrm{d}E/\mathrm{d}x$ measurement or Cherenkov radiation. By the sandwich structure of the iron plate and the sensor layers shown in the right plot of Fig. 2.15, the muon identification is carried out based on the difference in penetration ability of muons and pions in the detector. A 90% muon selection efficiency is achieved with a 5% pion misidentification rate. A more detailed description of the muon identification algorithm is available in Appendix. G.1, where we discuss its possible improvements. Muons with a momentum above 0.7 GeV/c penetrate the first layer of the KLM and the majority of muons traverse it completely if their momentum exceeds about 1.5 GeV/c.

2.3.6 Trigger and Data Acquisition System

The Level-1 (L1) trigger system of the Belle II experiment plays an important role in identifying events of interest. The occurrence of the physics events is mainly detected using the CDC and the ECL detectors, with input from the KLM detector for muon identification. These signals are combined in Global Reconstruction Logic (GRL) and then examined in Global Decision Logic (GDL) to judge whether the event should be read out or not. Tab. 2.3 shows the production cross section and rate of physics processes at the target luminosity of the SuperKEKB. The maximum L1 trigger rate is required to be smaller than 30 kHz, limited by the data transition bandwith of the Data AcQuisition (DAQ) system. The trigger decision time is limited within 5 μs required by the buffer depth of the SVD front-end ASIC.

Table 2.3: Estimated trigger rate of different processes at the target luminosity of SuperKEKB at the $\Upsilon(4S)$ resonance. The Bhabha rate is prescaled by a factor of 1/10.

Physics process	Cross section [nb]	Trigger rate [Hz]
$\Upsilon(4S) \to B\bar{B}$	1.1	660
$e^+e^- \to q\bar{q}$	2.8	1680
$e^+e^- \to \mu^+\mu^-$	0.8	480
$e^+e^- \to \tau^+\tau^-$	0.8	480
Bhabha	44	2640
$e^+e^- \to \gamma\gamma$	2.4	1440
Two-photon	12	7200
Total	64	$\sim$ 15000

An overview schematic of the Belle II DAQ system is shown in Fig. 2.16. The front-end readout electronics (FE dig) for each sub-detector are first connected to COPPER via Belle2link. The data from COPPER are first merged in readout PCs (R/O PC) and then merged in the event builders. In Event Builder 1, data from all sub-detectors except for PXD are merged event-by-event. A high level trigger (HLT) selections is then applied to suppress background-like events. In Event Builder 2, the PXD data in ROI is added to the event, which is subsequently stored to online disks.

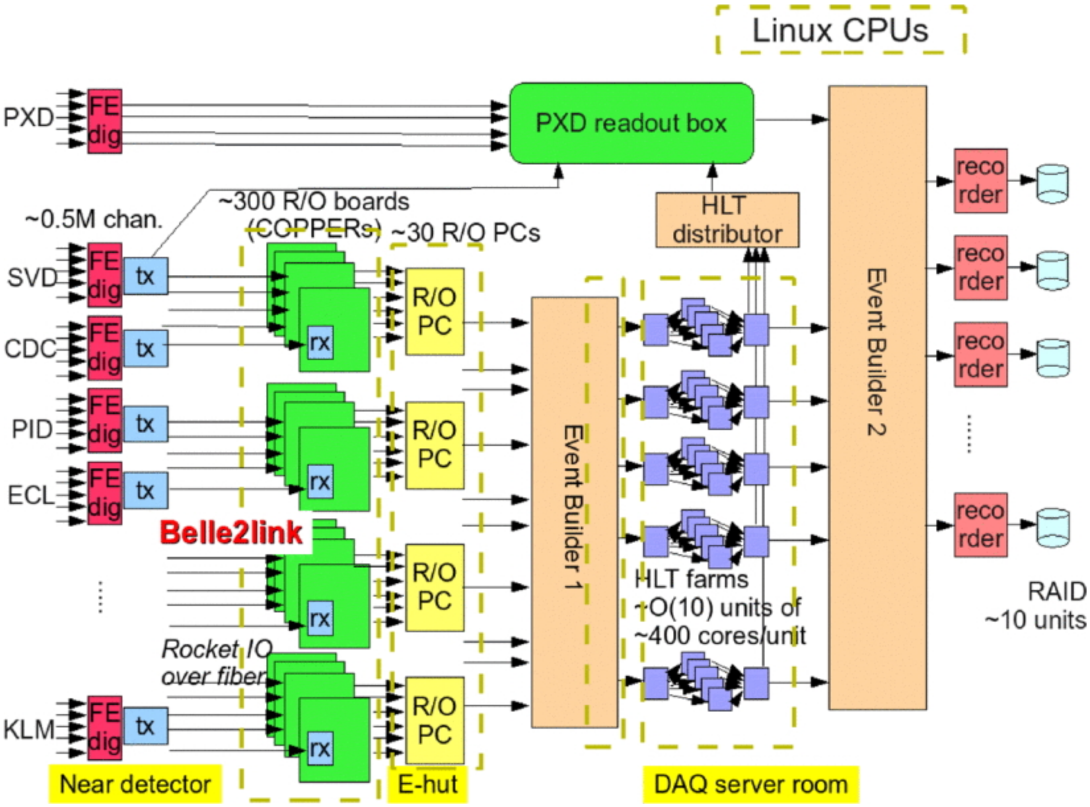


Figure 2.16: Belle II DAQ system.

Chapter 3

Analysis Strategy

In this dissertation, we carry out the sensitivity studies of differential branching fractions of the inclusive $B \to X_s \ell^+ \ell^-$ channel with the sum-of-exclusive method, using the Belle II run I dataset taken at the $\Upsilon(4S)$ resonance, corresponding to an integrated luminosity of $365\,\mathrm{fb}^{-1}$.

We reconstruct inclusive $B \to X_s \ell^+ \ell^-$ decay with two leptons (electrons or muons) and hadronic system X_s. X_s is then reconstructed with 20 hadronic modes, which takes up to 84% of all X_s final states considering K^0_{L} contributes equally to K^0_{S} events. Here, the 20 hadronic modes consists of one $K^{\pm}$ or K^0_{S} and up to four pions (at most one pion can be neutral) or three kaons ($K^+K^-K^+$ and $K^0_{\mathrm{S}}K^+K^-$). Table 3.1 gives a complete list of reconstructed modes. In this dissertation, we define the signal events whose X_s mode listed in Tab. 3.1 to be visible event, and those signal events whose X_s mode not listed in Tab. 3.1 (e.g. $K\eta(\to \gamma\gamma)$, $K\pi^0\pi^0$) to be invisible event. Decay modes not in Tab. 3.1 are called missing modes. It should be noticed that decay modes that replace K^0_{S} with K^0_{L} in Tab. 3.1 are considered as visible as well.

We select signal events by selections on several discriminative variables, followed by a Multi-Variate Analysis (MVA) based method. The main background sources are double semi-leptonic events where two semi-leptonic decays occur within one event, such as $B \to \ell\nu D(\to \ell\nu K)$ and $\Upsilon(4S) \to B(\to D\ell\nu)\bar{B}(\to D\ell\nu)$. These kinds of events are accompanied by generating at least two neutrinos and thus has a signature of a large amount of missing energy. Another contribution comes from the continuum events of $e^+e^- \to q\bar{q},\ q = u, d, s, c$ that can be rejected using the event-shape variables. A detailed description of signal reconstruction and selection is given in Chap. 5.

After event selection, there are three types of background that have the same peaking shape as signal events in the B-meson invariant mass spectrum and cannot be extracted by fitting. These three peaking backgrounds are double mis-ID events, swap mis-ID events, and charmonium events. A detailed description of them is given in Chap. 7. The first two categories are estimated using a pure data-driven method, while the latter one is estimated with simulation.

Table 3.1: Reconstruction modes of X_s

	B^0, $\overline{B}^0$		$B^{\pm}$	
K		K^0_S	$K^{\pm}$	
$K1\pi$	$K^{\pm}\ \pi^{\mp}$	$K^0_S\ \pi^0$	$K^{\pm}\ \pi^0$	$K^0_S\ \pi^{\pm}$
$K2\pi$	$K^{\pm}\ \pi^0\ \pi^{\mp}$	$K^0_S\ \pi^{\pm}\ \pi^{\mp}$	$K^{\pm}\ \pi^{\pm}\ \pi^{\mp}$	$K^0_S\ \pi^0\ \pi^{\pm}$
$K3\pi$	$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}$	$K^0_S\ \pi^0\ \pi^{\pm}\ \pi^{\mp}$	$K^{\pm}\ \pi^0\ \pi^{\pm}\ \pi^{\mp}$	$K^0_S\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}$
$K4\pi$	$K^{\pm}\ \pi^0\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}$	$K^0_S\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}$	$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}$	$K^0_S\ \pi^0\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}$
$3K$		$K^0_S\ K^{\pm}\ K^{\mp}$	$K^{\pm}\ K^{\mp}\ K^{\pm}$	

The differential branching fractions of the specific lepton flavor in the kinematic variable $k,\ k \in \{q^2, m(X_s)\}$ are given in the following way:

$$\begin{pmatrix} \mathcal{B}^1_k \\ \vdots \\ \mathcal{B}^n_k \end{pmatrix} = \frac{1}{2N_{B\overline{B}}} \left(\mathrm{diag}(f_1, \ldots, f_n)\right)^{-1} \mathcal{E}_\ell^{-1} \begin{pmatrix} N^1_{\mu(e)} \\ \vdots \\ N^n_{\mu(e)} \end{pmatrix}, \tag{3.1}$$

where $\mathcal{B}^i_k$ stands for the integration of differential branching fraction to kinematic variable k in bin $[k_{i-1}, k_i)$:

$$\mathcal{B}^i_k := \int_{k_{i-1}}^{k_i} \mathrm{d}k \frac{\mathrm{d}}{\mathrm{d}k} \mathcal{B}(B \to X_s \ell^+ \ell^-), \tag{3.2}$$

f_i is the visible fraction (ratio of visible events over all final states of X_s) in bin number i, $\mathcal{E}_\ell$ ($\ell = \mu, e$) is the signal unfolding matrix expanded to be:

$$\mathcal{E}_\ell = \begin{pmatrix} \varepsilon^\ell_{11} & \cdots & \varepsilon^\ell_{1n} \\ \vdots & \ddots & \vdots \\ \varepsilon^\ell_{n1} & \cdots & \varepsilon^\ell_{nn} \end{pmatrix}, \tag{3.3}$$

with each individual matrix element ε^ℓ_{ij} representing the probability (efficiency) of a certain event in the bin number j to be reconstructed in the bin number i. $N^r_{\mu(e)}$ represents the measured yields in the bin number r and $N_{B\overline{B}}$ is the number of $B\overline{B}$ pairs in the dataset. The boundaries of the q^2 binning are $[0.04, 1.0, 6.0, 14.4, 25.0]$ $(\mathrm{GeV}/c)^2$ with the motivation explained in Sec. 1.3. And for $m(X_s)$ the binning boundaries are $[0.4, 0.6, 1.0, 1.4, 1.8]$ (GeV/c^2), following the BABAR's convention. Additional studies in two more q^2 bins $[1,\ 7.3]$ and $[10.5,\ 11.8]$ $(\mathrm{GeV}/c)^2$ are performed based on the requirements of theorists. Three boundaries of these two bins are set inside the charmonium veto region so that there will be no signal cross-feed over the boundary. The cross-feed over the 1 $(\mathrm{GeV}/c)^2$ boundary is negligibly small. The definition of each measurement bin and their names are given in Tab. 3.2. The development and validation of fitting tools are described in Chap. 8.

Reconstruction efficiency (Eq. 3.3) is estimated based on Monte Carlo (MC) simulation sample. In this work, the reconstruction efficiency is given by the ratio of number of

Table 3.2: List of bins for differential branching fraction measurements.

Bin name	Bin definition
q_1^2	$0.04 < q^2 < 1.0$ [(GeV/c)^2]
q_2^2	$1.0 < q^2 < 6.0$ [(GeV/c)^2]
q_3^2	$6.0 < q^2 < 14.4$ [(GeV/c)^2]
q_4^2	$14.4 < q^2 < 25.0$ [(GeV/c)^2]
q_5^2	$1.0 < q^2 < 7.3$ [(GeV/c)^2]
q_6^2	$10.5 < q^2 < 11.8$ [(GeV/c)^2]
$m_1(X_s)$	$0.4 < m(X_s) < 0.6$ [GeV/c^2]
$m_2(X_s)$	$0.6 < m(X_s) < 1.0$ [GeV/c^2]
$m_3(X_s)$	$1.0 < m(X_s) < 1.4$ [GeV/c^2]
$m_4(X_s)$	$1.4 < m(X_s) < 1.8$ [GeV/c^2]

correctly reconstructed candidates over the number of visible signal events. The incorrectly reconstructed signal events (referred to as self cross-feed events) do not participate in efficiency calculation, but they participate in the fitting procedure for signal extraction (Sec. 8.2.2). In past measurements, the most significant systematic uncertainty, which is at the same level as statistical uncertainty, comes from the modeling of the signal process as explained in Sec. 1.5.1. In this study, we improved the signal modeling in the following four aspects, and the corresponding systematics are significantly suppressed:

- average recent $\mathcal{B}(B \to K^{(*)}\ell^+\ell^-)$ measurements to improve precision on the branching fraction of exclusive processes;
- develop data-driven methods for fragmentation correction of visible modes;
- introduce processes via high kaon resonances like $B \to K_1(1270)\ell^+\ell^-$ to improve modeling of $K\pi$ and $K\pi\pi$ final states in inclusive region; and
- extrapolate from visible modes to all decay modes of X_s using a tuned Pythia sample that reproduce $B \to X_s\gamma$ fragmentation.

Details about the estimation of systematic uncertainties are available in Chap. 9.

Due to constrains from the Belle II collaboration, we only present the uncertainties of each observable, blinding the center values or forcing them to the MC or theoretical predictions. We present the sensitivities of lepton flavor specific and lepton flavor averaged differential branching fractions of each bin in Tab. 3.2. The inclusive branching fraction is given by summing up the results in $q_1^2, \ldots, q_4^2$ bins and we present its estimated uncertainties in this dissertation.

The Belle II software framework (BASF2) [58, 59] is used for event reconstruction and simulation sample generation.

Chapter 4

Simulation Samples

In this work, MC simulation samples are used to optimize event selection criteria, calculate signal reconstruction efficiencies, and study background sources. There are two different kinds of MC samples used in this analysis: signal MC and generic MC. The signal MC simulates the $\Upsilon(4S) \to B\overline{B}$ events with one of the B-mesons decays to the signal process $B \to X_s \ell^+\ell^-$, while the other B-meson decays according to branching fractions written in the Review of Particle Physics by the Particle Data Group (PDG) [60]. The Generic MC, however, simulates all possible physics processes that occur at the center-of-mass energy $\sqrt{s} = 10.58\,\mathrm{GeV}$, including the $e^+e^- \to q\bar{q}$, $q = u, d, s, c$ and also the $\Upsilon(4S) \to B\overline{B}$ process with both B-mesons decay according to PDG.

In these samples, B-meson production and decay are generated by EvtGen [61]; $e^+e^- \to q\bar{q}$ background is generated with Phythia8 [62] and KKMC [63]; photon radiation from charged particles is simulated by PHOTOS [64]; Geant4 [65] is used to simulate the detector response to the passage of particles.

4.1 Signal Monte Carlo (MC) samples

In this analysis, the signal MC samples are generated in two different batches. The first batch, described in Sec. 4.1.1, is named as traditional batch as this is what has been used in past analysis. It includes two exclusive $B \to K\ell^+\ell^-$, $B \to K^*(892)\ell^+\ell^-$ processes and a non-resonant $B \to X_s^{\mathrm{non-resonant}}\ell^+\ell^-$ sample to describe the inclusive X_s region defined as $m(X_s) > 1.1\,\mathrm{GeV}/c^2$. The second batch is new to the inclusive $B \to X_s\ell^+\ell^-$ analysis. It includes five exclusive processes $B \to K_J\ell^+\ell^-$ that exist in the inclusive region. It is described in Sec. 4.1.2. Both of these two samples are used in this analysis and the combination strategy is described in Sec. 4.1.3.

4.1.1 Traditional batch of samples

The resonant processes $B \to K\ell^+\ell^-$ and $B \to K^*(892)\ell^+\ell^-$ are produced with the BTOSLLBALL generator, and the non-resonant $B \to X_s\ell^+\ell^-$ is generated with the BTOXSLL generator with a cutoff of X_s invariant mass above $1.1\,\mathrm{GeV}/c^2$. This $1.1\,\mathrm{GeV}/c^2$ boundary is motivated by the previous analysis of $B \to X_s\gamma$ decay [66] in which the differential branching fraction as a function of $m(X_s)$ is measured as will be explained in Sec. 9.1.5. For the non-resonant MC, the heavy quark effective formula smeared with the Fermi motion model is applied to determine the lineshape of X_s. The fragmentation of X_s is simulated with Pythia8 and corrected with a data-driven method described in Sec. 9.1. MC sample size for each decay mode is summarized in Table 4.1.

Table 4.1: Signal MC of traditional batch

Channel	**MC sample size**	$\mathcal{B} \times 10^6$
$B^+ \to K^+\mu^+\mu^-$	0.5 M	0.45
$B^+ \to K^+e^-e^-$	0.5 M	0.45
$B^0 \to K^0\mu^+\mu^-$	1.0 M	0.41
$B^0 \to K^0e^-e^-$	1.0 M	0.41
$B^+ \to K^{*+}\mu^+\mu^-$	0.5 M	0.99
$B^+ \to K^{*+}e^-e^-$	0.5 M	0.99
$B^0 \to K^{*0}\mu^+\mu^-$	0.5 M	0.92
$B^0 \to K^{*0}e^-e^-$	0.5 M	0.92
$B^+ \to X_{su}\mu^+\mu^-$	2.0 M	2.72
$B^+ \to X_{su}e^-e^-$	1.4 M	2.77
$B^0 \to X_{sd}\mu^+\mu^-$	2.0 M	2.82
$B^0 \to X_{sd}e^-e^-$	1.4 M	2.87

The samples in this traditional batch are merged according to their expected production at $365\,\mathrm{fb}^{-1}$ using the following equation:

$$N^{gen}(\mathcal{L} = 365\,\mathrm{fb}^{-1}) = N_{\Upsilon(4S)} \cdot 2 \cdot f_{\pm,0} \cdot \mathcal{B} \tag{4.1}$$

where $N_{\Upsilon(4S)} = (387.1 \pm 5.6) \times 10^6$ is the number of $\Upsilon(4S)$ in run I dataset, $f_\pm = 0.5113 \pm 0.0108$ and $f_0 = 0.4861 \pm 0.0080$ are the branching fractions of $\Upsilon(4S)$ decaying to B^+B^- and $B^0\bar{B}^0$ [67], respectively. $\mathcal{B}$ is the branching fraction of the corresponding decay mode also summarized in Tab. 4.1. The branching fractions of the $B \to K^{(*)}\ell^+\ell^-$ modes are calculated by averaging all available measurements performed by LHCb, Belle, BABAR and the CDF experiment described in Sec. 9.1. The branching fraction of the inclusive region is calculated by subtracting the contributions of $B \to K^{(*)}\ell^+\ell^-$ from the inclusive branching fractions given by the prediction of SM, which are 4.15×10^{-6} and 4.2×10^{-6} for the $B \to X_s\mu^+\mu^-$ and $B \to X_se^-e^-$ modes, respectively [68, 69]. $m_{\ell^+\ell^-} > 0.2\,\mathrm{GeV}/c^2$, whose motivation is explained in Sec. 6.1, is assumed in all the branching fractions mentioned above.

4.1.2 New batch of samples

In recent LHCb measurements of the $B^0 \to K^+\pi^-\mu^+\mu^-$ [70] and $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ [71] channels, peak structures in the $m(K\pi)$ and $m(K\pi\pi)$ spectrum are observed around $1.3\,\mathrm{GeV}/c^2$ and $1.4\,\mathrm{GeV}/c^2$, respectively. To simulate the observation well in our modeling, in this analysis five different kaon resonances $K_1(1270)$, $K_1(1400)$, $K^*(1410)$, $K_0^*(1430)$ and $K_2^*(1430)$ are considered, and are forced to decay to the $K\pi$ and $K2\pi$ final states. Due to the conservation of angular momentum and parity, $K_1(1270)$ and $K_1(1400)$ do not decay to $K\pi$, while $K_0^*(1430)$ does not decay to the final state $K2\pi$, and therefore one can derive that there are seven possible channels. These samples are generated using a private EvtGen package whose development is described in Appendix. A.

These samples are merged according to their branching fractions. Since there are no reliable measurements of their branching fractions, we use the following equation for the estimation of branching fractions except for $K_0^*(1430)$:

$$
\begin{aligned}
\mathcal{B}(B^+ \to K_J^+\ell^+\ell^-) =& \frac{\Phi_3(B^+ \to K_J^+\ell^+\ell^-)}{\Phi_3(B^+ \to K^*(892)^+\ell^+\ell^-)} \cdot \frac{\Phi_2(B^+ \to K^*(892)^+\gamma)}{\Phi_2(B^+ \to K_J^+\gamma)} \\
& \cdot \frac{\mathcal{B}(B^+ \to K_J^+\gamma)}{\mathcal{B}(B^+ \to K^*(892)^+\gamma)} \cdot \mathcal{B}(B^+ \to K^*(892)^+\ell^+\ell^-),
\end{aligned} \tag{4.2}
$$

where $\mathcal{B}(B^+ \to K_J^+\gamma)$ and $\mathcal{B}(B^+ \to K^*(892)^+\gamma)$ are taken from PDG and $\mathcal{B}(B^+ \to K^*(892)^+\ell^+\ell^-)$ are taken from Tab. 4.1. $\Phi_{2,3}$ stands for the phase space volume of two body and three body decays[1] and it is introduced to compensate the phase space difference caused by different kaon masses. For $K_0^*(1430)$, since $B^+ \to K_0^*(1430)^+\gamma$ is forbidden by the conservation of angular momentum, we estimate its branching fraction from the point of view of spin:

$$
\begin{aligned}
\mathcal{B}(B^+ \to K_0^*(1430)^+\ell^+\ell^-) =& \frac{\Phi_3(B^+ \to K^*(892)^+\ell^+\ell^-)}{\Phi_3(B^+ \to K^+\ell^+\ell^-)} \cdot \frac{\mathcal{B}(B^+ \to K^+\ell^+\ell^-)}{\mathcal{B}(B^+ \to K^*(892)^+\ell^+\ell^-)} \\
& \cdot \mathcal{B}(B^+ \to K^*(1410)^+\ell^+\ell^-),
\end{aligned} \tag{4.3}
$$

where $\mathcal{B}(B^+ \to K^+\ell^+\ell^-)$ and $\mathcal{B}(B^+ \to K^*(892)^+\ell^+\ell^-)$ are taken from Tab. 4.1, and $\mathcal{B}(B^+ \to K^*(1410)^+\ell^+\ell^-)$ from Eq. 4.2 above. We also adapt the following convention in the determination of each exclusive channel:

- Lepton flaovr universality: $\mathcal{B}(B^+ \to K_J^+\mu^+\mu^-) = \mathcal{B}(B^+ \to K_J^+e^+e^-)$;

[1] Φ_2 for two body $B \to K_J\gamma$ decay is proportional to $1-(m_K/m_B)^2$.
The phase space volume for three body $B \to K_J\ell^+\ell^-$ is calculated ignoring lepton mass:

$$
\begin{aligned}
\Phi_3(B \to K_J\ell^+\ell^-) &= \frac{1}{(2\pi)^7 \cdot 8m_B^2} \int_0^{(m_B-m_K)^2} \lambda^{1/2}(m_B^2, m_K^2, q^2)dq^2 \\
&= \frac{1}{(2\pi)^7 \cdot 16m_B^2} \cdot \left[m_B^4 - m_K^4 + 4m_B^2m_K^2\log(m_K/m_B)\right],
\end{aligned}
$$

where $\lambda(a^2,b^2,c^2) = (a^2-b^2-c^2)^2 - 4b^2c^2$ is the Källén function.

- Isospin symmetry: $\Gamma(B^+ \to K_J^+ \ell^+ \ell^-) = \Gamma(B^0 \to K_J^0 \ell^+ \ell^-)$.

Table. 4.2 gives the list of samples and their corresponding branching fractions.

Table 4.2: Signal MC of new batch

Resonance K_J	J^P	$\mathcal{B}(B^+ \to K_J^+ \ell^+ \ell^-) \times 10^6$	$\mathcal{B}(K_J \to K\pi)$ [%]	$\mathcal{B}(K_J \to K\pi\pi)$ [%]
$K_1(1270)$	1^+	0.96 ± 0.17	$(-)$	100
$K_1(1400)$	1^+	0.21 ± 0.10	$(-)$	100
$K^*(1410)$	1^-	0.55 ± 0.17	6.6 ± 1.3	40 ± 10
$K_0^*(1430)$	0^+	0.22 ± 0.07	93 ± 10	$(-)$
$K_2^*(1430)$	2^+	0.28 ± 0.08	49.9 ± 1.2	24.7 ± 1.5

4.1.3 Combination of two batches of signal MC samples

These two batches of samples are found to have comparable reconstruction efficiency and this comparison is described in Appendix. B.

The principle when combining two batches is to keep the total number of $K\pi$ and $K2\pi$ events (denoted $N_{K\pi}$ and $N_{K2\pi}$, respectively) unchanged in each $m(X_s)$ bin, defined to be:

$$[1.1, 1.3, 1.4, 1.5, 1.7, 1.8, 1.9, 2.0, 2.1, 2.4, 2.7, 3.0] \ \text{GeV}/c^2.$$

Under this principle, the fraction of invisible modes and the fraction of each visible mode are kept unchanged. This is because we estimate the fraction of invisible modes by Pythia, and the fraction of each visible mode is determined by a data-driven method.

In each bin, if the number of resonance events (N_{reso}) is greater than $N_{K(2)\pi}$, we replace all Pythia events with resonance events after applying a scaling factor $N_{K(2)\pi}/N_{reso}$ to the resonance samples. In contrast, if $N_{reso} < N_{K(2)\pi}$, we use all resonance samples and apply a scaling factor of $1 - N_{reso}/N_{K(2)\pi}$ to the pythia samples.

The fraction of each source in the combined $K(2)\pi$ sample is summarized in Tab. 4.3.

Table 4.3: Fraction (in %) of each source in the combined $K\pi$ and $K2\pi$ signal MC sample in inclusive region.

	Pythia	$K_1(1270)$	$K_1(1400)$	$K^*(1410)$	$K_0^*(1430)$	$K_2^*(1430)$
$K\pi$	31.7	0	0	7.0	37.8	23.6
$K2\pi$	10.6	57.7	12.8	14.1	0	4.8

The distributions of $m(X_s)$ and q^2 are shown in Fig. 4.1.

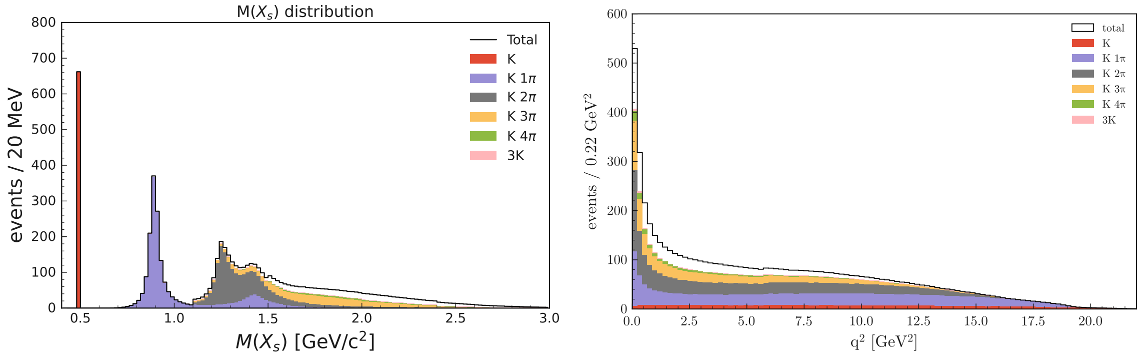


Figure 4.1: $m(X_s)$ (left) and q^2 (right) at the generator level corresponding to a luminosity of $365\,\mathrm{fb}^{-1}$. Black line is the distribution of inclusive X_s while the colored histograms shows the visible modes covered in this analysis.

4.2 Generic MC samples

For generic MC, we prepared a run-dependent sample (referred to as MCrd) corresponding to an integrated luminosity of $4 \times 365\,\mathrm{fb}^{-1}$ for background analysis and optimization of selection criteria. This sample is simulated with detector configurations following the data-taking records at each run. The background hits from beam-induced particles are taken from real-data with a dedicated trigger configuration. In addition, we prepared another run-independent sample corresponding to $1\,\mathrm{ab}^{-1}$ for the continuum process and $3\,\mathrm{ab}^{-1}$ for $B\overline{B}$ processes used to train the MVA classifier. These samples are generated with uniform detector configuration and simulated beam-induced backgrounds for easier and massive production.

In the following, generic MC samples of the continuum process are referred to as $q\overline{q}$ samples, samples of $\Upsilon(4S) \to B^+B^-$ as charged samples, and $\Upsilon(4S) \to B^0\overline{B}^0$ as mixed samples, because B^0 and $\overline{B}^0$ mesons can mix with each other. Signal events $B \to X_s\ell^+\ell^-$ are subtracted from charged and mixed samples using MC truth information.

Chapter 5

Signal Reconstruction and Selection

5.1 General Principles for Final State Particle Reconstruction

The final state particles refer to the particles that have a long lifetime and thus interact directly with the detector. It includes charged particles such as charged pions and muons, and neutral particles such as photons.

The charged particles are selected from tracks reconstructed from the tracking system of PXD, SVD, and CDC. Only tracks originating from IP are selected by requiring $\mathrm{d}r < 0.5$ cm and $|\mathrm{d}z| < 2.0$ cm, where $\mathrm{d}r$ and $\mathrm{d}z$ are the distance from the Point of Closest Approach (POCA) of a track to the IP in the $x-y$ plane and along the z axis, respectively. The energy of each charged particle is determined by $\sqrt{m^2+p^2}$, where p is the track momentum measured by CDC and m is the PDG mass of the assumed particle type.

The identification of the charged particle type (PID) is performed using a likelihood-based variable called particleID. In the Belle II experiment, there are six kinds of long-lived charged particles that can be detected directly: pion, kaon, electron, muon, proton, and deuteron. Each subdetector system d gives a likelihood of each particle type hypothesis t: $\mathcal{L}_t^d$, based on $\mathrm{d}E/\mathrm{d}x$ measurements (SVD, CDC), Cherenkov angle (TOP, ARICH), ratio of deposit energy over track momentum E/p (ECL) and penetration ability (KLM). And the total likelihood of a certain particle type, t, is given by the direct product of likelihoods from each subsystem:

$$\mathcal{L}_t = \prod_{d\in\{\mathrm{SVD,CDC,TOP,ARICH,ECL,KLM}\}} \mathcal{L}_t^d. \tag{5.1}$$

Based on Bayes' theorem, assuming that each particle appears with equal probability, the particleID variable of hypothesis t is defined to be:

$$\mathrm{particleID}_t = \frac{\mathcal{L}_t}{\sum_{j\in\{\pi,K,e,\mu,p,d\}} \mathcal{L}_j}. \tag{5.2}$$

In this work, we use the particleID variable for the identification of muons, kaons, and pions. For convenience, they are referred to as, for example, muonID for the muon hypothesis.

For the identification of electrons, we use an algorithm based on the Boost Decision Tree (BDT) referred to as $e\mathrm{ID}_{\mathrm{BDT}}$. The training utilizes features derived from the ECL, comprising observables such as E/p and cluster shape variables, among others. It also uses likelihoods from subdetectors other than ECL. In this way, it improves the pion misidentification rate from 3% to 0.2% at 90% electron selection efficiency.

The only neutral final state particle that is used in this work is the photon, whose candidates are reconstructed using electro-magnetic showers (called clusters) in ECL that are not associated with a track. The photon energy is set to be the measured cluster energy in the ECL, with momentum direction set to be from IP to the center point of the cluster.

5.2 Particle Selection Criteria

For muons, a momentum requirement greater than $0.7\,\mathrm{GeV}/c$ is applied, and the most strict PID cut of muonID>0.99 is determined to reject misidentification from pions as much as possible. The $0.7\,\mathrm{GeV}/c$ momentum requirement is slightly higher than the threshold for a track to reach the KLM. This cut preserves 80% of muons and rejects 98% of mis-identified pions.

Electrons are required to have a momentum larger than $0.4\,\mathrm{GeV}/c$ and $e\mathrm{ID}_{\mathrm{BDT}}$ greater than 0.9. Similarly, the $0.4\,\mathrm{GeV}/c$ requirement is slightly higher than the threshold for a track to reach the ECL detector. The electron selection efficiency is 90% and it rejects more than 99.9% of mis-identified pions. To recover the electron energy lost in its interaction with beam pipe or detector materials, we find all photons within the 0.05 rad cone along the electron direction at POCA and add their 3-momenta to the electron candidate. Photons used for this bremsstrahlung recovery are required to have an energy deposit in the ECL larger than 75, 50 and 100 MeV for clusters in the forward, barrel, and backward region, respectively, to avoid using beam-induced photons. Typically, 20% of electrons' energies are corrected and for most of the corrected electrons, only one photon is found in the cone.

Loose PID requirements of kaonID > 0.6 and pionID > 0.4 are applied to select kaons and pions, respectively. In addition, they are required to have at least one hit in CDC (nCDCHits > 0) to reject tracks with relatively poor PID quality. These selections preserve about 90% of signal kaons and pions, respectively.

The K_S^0 candidates are reconstructed from $K_S^0 \to \pi^+\pi^-$ decay. Two tracks without PID selection criteria are used to combine a K_S^0 candidate followed by a vertex fit using TreeFit [72] toolkit. A good candidate should have a mass in the range from $0.45\,\mathrm{GeV}/c^2$ to $0.55\,\mathrm{GeV}/c^2$ and not fail in the vertex fit. After that, the ksSelector developed in B^0

$\to K^+ K^- K_S^0$ analysis [73] is applied for further suppression of background. It has two FastBDT (FBDT) based selectors. The first discriminator `V0Selector` separates long-lived true K_S^0 from fake ones originating from IP, making use of variables such as the difference between the pointing direction of the K_S^0 vertex and its momentum. The flight distance, the significance of the distance (flight distance divided by its uncertainty), and the helicity angles are also included in this discriminator. The second discriminator `LambdaVeto` is used to separate K_S^0 candidates from $\Lambda \to p^+\pi^-$ decays using variables that include different mass hypotheses for K_S^0 daughter $\pi^\pm$. The standard selection criteria of `V0Selector` > 0.90 and `LambdaVeto` > 0.11 are applied with a nominal K_S^0 efficiency of 90% while rejecting more than 99% of fake K_S^0 candidates.

π^0 candidates are reconstructed from two photon candidates following the $B^0 \to \rho^+\rho^-$ analysis [74, 75]. Photons in the end-cap and barrel regions of ECL are required to have a cluster energy larger than 90 MeV and 60 MeV, respectively, to reject beam-background photons. The time of the cluster is required to be within 200 ns range from the event T0. To reject electro-magnetic clusters from hadronic clusters and shower splitting off from hadronic clusters, a FBDT based classifier called photonMVA is applied. Cluster energy, ratio of cluster energy of the center cluster and surrounding clusters (clusterE1E9), and cluster Zernike moments are used in the training of the classifier. We required the photonMVA output to be larger than 0.1 for each photon, which removes 58% of mis-reconstructed photons while retaining 98% of signal photons. The photon energy bias from the leakage of electro-magnetic shower outside the crystals is corrected using a data-driven method. As for the selections on π^0, its reconstructed mass is required to be in the window from 120 MeV/c^2 to 150 MeV/c^2, and its momentum should be greater than 210 MeV/c. In addition, the angle between momenta of two daughter photons ($\alpha_{\gamma\gamma}$) is required to be less than 2.5 rad and the azimuth angle difference ($|\Delta\phi_{\gamma\gamma}|$) is required to be less than 1.3 rad in the laboratory frame to suppress random combinations of two photons. This procedure keeps 40% of π^0.

The s-quark system X_s is reconstructed from the twenty channels as explained in Tab. 3.1. And finally we form a B-meson candidate by combining the X_s candidate with a pair of leptons of the same flavor. We keep all possible combinations of final state particles (also referred to as B candidates) for further selections. The beam-constrained mass $M_{bc} \equiv \sqrt{(\sqrt{s}/2)^2 - p_B^2}$ of the B-meson, where p_B is the 3-momentum in the center-of-mass frame, is required to be within the range of 5.2 to 5.3 GeV/c^2 to reject backgrounds from random combination. This is a very loose requirement, since the distribution of M_{bc} for signal events is a Gaussian distribution with a center value of 5.28 GeV/c^2 and a standard deviation smaller than 3 MeV/c^2, originating from the energy spread of the beams.

Chapter 6

Background Suppression

6.1 Pre-selection

A vertex fit to B-meson is performed using the TreeFit [72] toolkit, with π^0 mass constrained to its PDG mass during this fitting to improve the resolution of the kinematic variable ΔE to be mentioned in the following. The χ^2 probability (B_chiProb) which reflects the quality of the fitting is shown in Fig. 6.1. We required the events to have a chiProb greater than zero to ensure that the vertex fitting is successful.

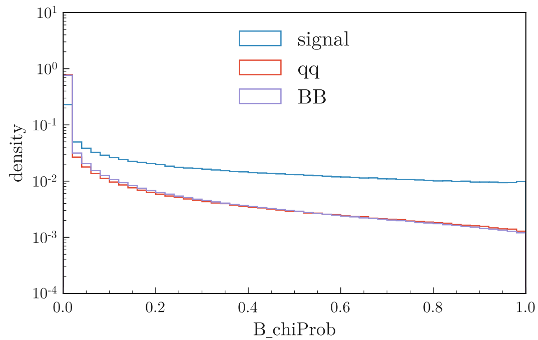


Figure 6.1: Normalized distributions of χ^2 probability in signal and background samples.

Fig. 6.2 shows the distribution of $m(X_s)$. To further reject the combinatorial background, we require that $m(X_s)$ be no larger than $1.8\,\mathrm{GeV}/c^2$ to guarantee a fair signal-to-noise ratio.

For the same reason, we applied selection criteria in ΔE of the B-meson, which is defined to be the difference between the reconstructed B energy E_B in the center-of-mass frame and half of the center-of-mass energy $\sqrt{s}/2$: $E_B - \sqrt{s}/2$. Fig. 6.3 shows the distribution of ΔE of $B \to X_s\mu^+\mu^-$ and $B \to X_s e^+e^-$. The upper bound is set to be $0.05\,\mathrm{GeV}/c^2$ and the lower bound is $-0.10\,\mathrm{GeV}/c^2$ and $-0.05\,\mathrm{GeV}/c^2$ for electron and muon channels, respectively. Electron channels have a wider window because of energy loss via bremsstrahlung.

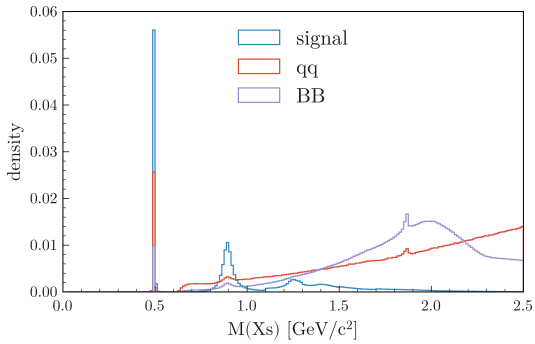


Figure 6.2: Normalized distributions of $m(X_s)$ in signal and background samples.

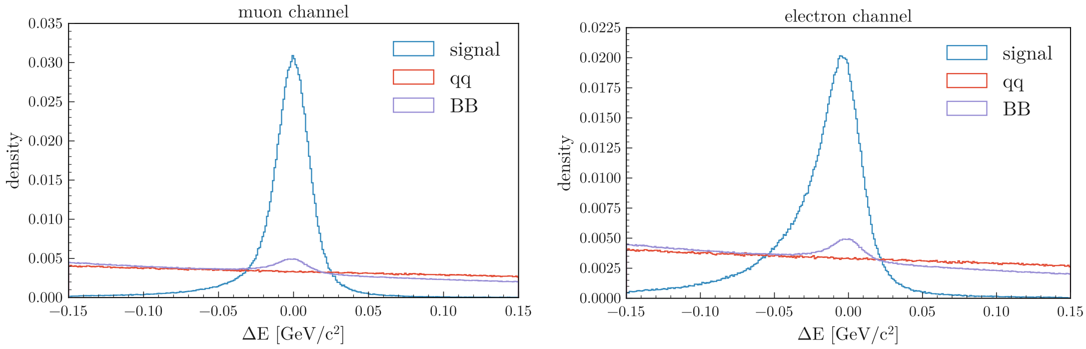


Figure 6.3: Normalized distributions of ΔE in signal and background samples. Left is $B \to X_s\mu^+\mu^-$ channel and right is $B \to X_s e^+e^-$ channel.

To suppress background from gamma conversion $\gamma \to e^+e^-$, especially the $B \to X_s\gamma(\to e^+e^-)$ process with on-shell photon, and three-body π^0 decay: $\pi^0 \to e^+e^-\gamma$, we applied a requirement of $m_{e^+e^-} > 0.2\,\mathrm{GeV}/c^2$ to electron channels.

One of the main background processes is the charmonium process $B \to X_s\psi(\to \ell^+\ell^-)$, as shown in Fig. 6.4 and 6.5. To reject such a kind of background, a veto on the invariant mass of dilepton system is imposed:

$$\begin{aligned}
-0.40\mathrm{GeV} &< M_{e^+(\gamma)e^-(\gamma)} - M_{J/\psi} < 0.15\mathrm{GeV},\\
-0.25\mathrm{GeV} &< M_{e^+(\gamma)e^-(\gamma)} - M_{\psi(2S)} < 0.10\mathrm{GeV},\\
-0.25\mathrm{GeV} &< M_{\mu^+\mu^-} - M_{J/\psi} < 0.10\mathrm{GeV},\\
-0.15\mathrm{GeV} &< M_{\mu^+\mu^-} - M_{\psi(2S)} < 0.10\mathrm{GeV}.
\end{aligned} \tag{6.1}$$

It should be noticed that the invariant mass of electron channels is calculated both with and without bremsstrahlung correction: e^+e^-, $e^+\gamma e^-$, $e^+e^-\gamma$, $e^+\gamma e^-\gamma$. This is because of the possibility that the charmonium process may survive the veto when a wrong bremsstrahlung photon is added to the electron.

D mesons decay to $K\pi$ and $K\pi\pi$ can be reconstructed as X_s candidate as indicated by the peak at $1.8\,\mathrm{GeV}/c^2$ in Fig. 6.2. In addition, events such as $B \to D(\to K\pi)\pi$ have a similar distribution to signal events in the M_{bc} and ΔE spectrum if the two

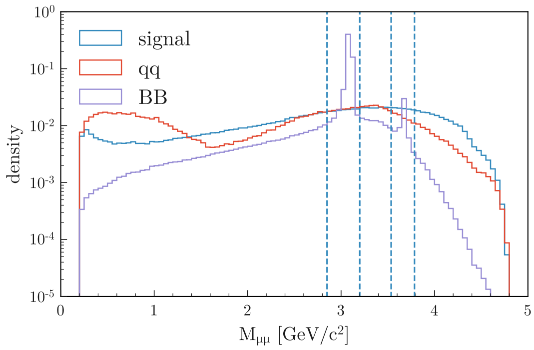


Figure 6.4: Normalized distributions of $m(\mu^+\mu^-)$ of $B \to X_s\mu^+\mu^-$ channel in signal and background samples. The dashed vertical lines corresponds to the boundary of charmonium veto.

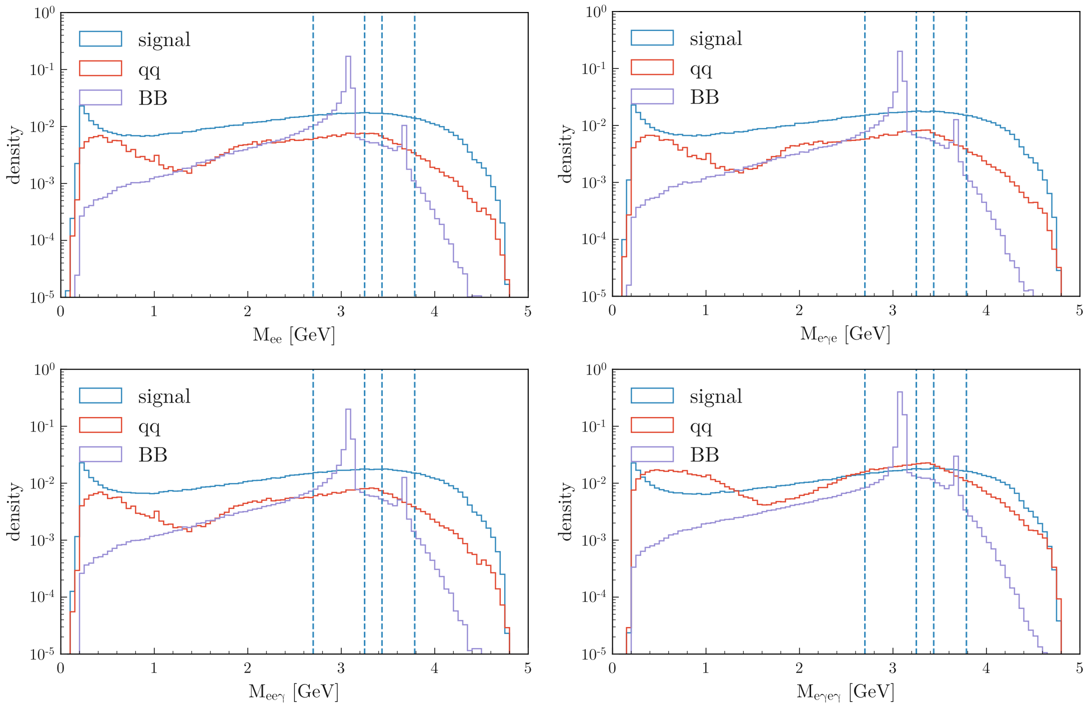


Figure 6.5: Normalized distributions of $m(e^+e^-)$ of $B \to X_se^+e^-$ channel in signal and background samples. From top left to bottom right are plots made with invariant mass spectrum of e^+e^-, $e^+\gamma e^-$, $e^+e^-\gamma$, $e^+\gamma e^-\gamma$, where the appending γ indicates that the corresponding electron/positron energy is corrected by bremsstrahlung recovery algorithm. The dashed vertical lines corresponds to the boundary of charmonium veto.

pions are misidentified as muons. To veto the possible D meson, the most D-like mass of daughters in one B candidate is calculated with the following steps: First, find all combinations of the final-state daughters of the B candidate, with K_S^0 and π^0 considered as final state particles as well. If the B candidate has n final state particles, there are $2^n - n - 1$ different combinations with at least two daughters. Second, only combinations

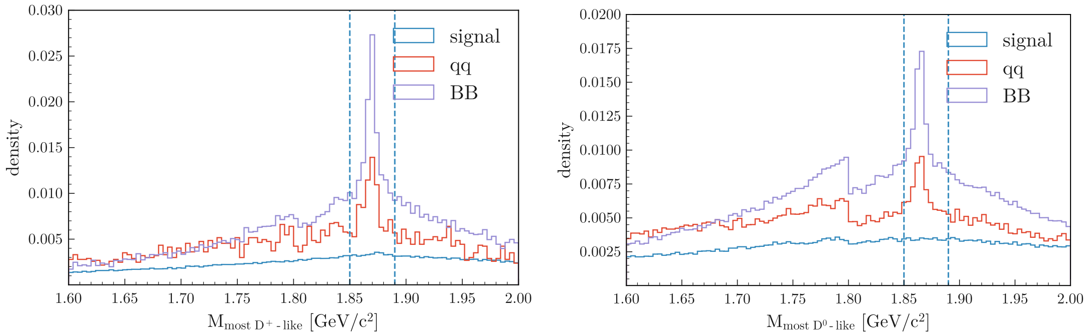


Figure 6.6: Normalized distributions of $m_{\text{most}-\text{D like}}$ in signal and background samples (Left for charged D^+ and right for neutral D^0). Only the candidates whose K meson come from a D meson are plotted for the background samples using the MC truth information. The vertical dashed line indicates the boundary of D-veto. The drop at $1.8\,\text{GeV}/c^2$ is because of the $m(X_s) < 1.8\,\text{GeV}/c^2$ cut.

that satisfy the following criteria are kept:

- The combination should not be the full set of B daughters.
- The combination should not be the di-lepton system.
- The total charge of the combination should be $0, \pm 1$.
- If there is one and only one $K^\pm$ meson in the combination and the total charge is ± 1, the charge of $K^\pm$ should be opposite to the total charge. This is because the decay of $B \to D \to K$ is dominated by the Cabibbo-allowed process.

Third, we calculate the invariant mass of the remaining combinations, with a mass hypothesis of leptons switched to charged pions. Finally, the invariant mass of the remaining combinations closest to the nominal mass of $D^\pm$ (D^0) is assigned as the most-$D^\pm$ like (most-D^0 like) mass of the corresponding B meson, depending on the total charge of the corresponding combination. Fig. 6.6 shows the normalized distribution of most-D^+ (D^0) like mass of the signal and background samples, where only candidates whose K meson comes from a D meson are plotted for the background samples. Clear peaks of D meson are spotted in these two distributions and thus a veto is implied to reject candidates inside the mass window below:

$$1.85\,\text{GeV}/c^2 < m_{\text{most}-\text{D like}} < 1.89\,\text{GeV}/c^2 \tag{6.2}$$

A summary of pre-selection is presented in Tab. 6.1.

6.2 Background analysis after pre-selection

After pre-selection, the expected signal yield is around 350 while the expected number of background is at the order of 10^4 in the region under signal peak $M_{bc} > 5.27\,\text{GeV}/c^2$,

Table 6.1: Summary of pre-selection criteria.

Objects	Selection criteria
B meson vertex fit	chiProb > 0
X_s mass	$m(X_s) < 1.8\,\mathrm{GeV}/c^2$
Energy difference	$-0.05\,\mathrm{GeV} < \Delta E < 0.05\,\mathrm{GeV}$ $(\mu^+\mu^-)$
	$-0.10\,\mathrm{GeV} < \Delta E < 0.05\,\mathrm{GeV}$ (e^+e^-)
γ conversion veto	$m(\ell^+\ell^-) > 0.2\,\mathrm{GeV}/c^2$
charmonium veto	$-0.40\,\mathrm{GeV}/c^2 < m_{e^+(\gamma)e^-(\gamma)} - m_{J/\psi} < 0.15\,\mathrm{GeV}/c^2$
	$-0.25\,\mathrm{GeV}/c^2 < m_{e^+(\gamma)e^-(\gamma)} - m_{\psi(2S)} < 0.15\,\mathrm{GeV}/c^2$
	$-0.25\,\mathrm{GeV}/c^2 < m_{\mu^+\mu^-} - m_{J/\psi} < 0.10\,\mathrm{GeV}/c^2$
	$-0.15\,\mathrm{GeV}/c^2 < m_{\mu^+\mu^-} - m_{\psi(2S)} < 0.10\,\mathrm{GeV}/c^2$
D-veto	$1.85\,\mathrm{GeV}/c^2 < m_{\mathrm{most-D\ like}} < 1.89\,\mathrm{GeV}/c^2$

making it impossible to extract the signal yield by fitting the M_{bc} spectrum. The distribution of M_{bc} for signal and background MC is shown in Fig. 6.7. For this reason, a multivariate analysis to further suppress the background is necessary and in this section we describe a background analysis performed to understand the sources of the remaining background, especially the source of the two leptons in the reconstructed B candidates.

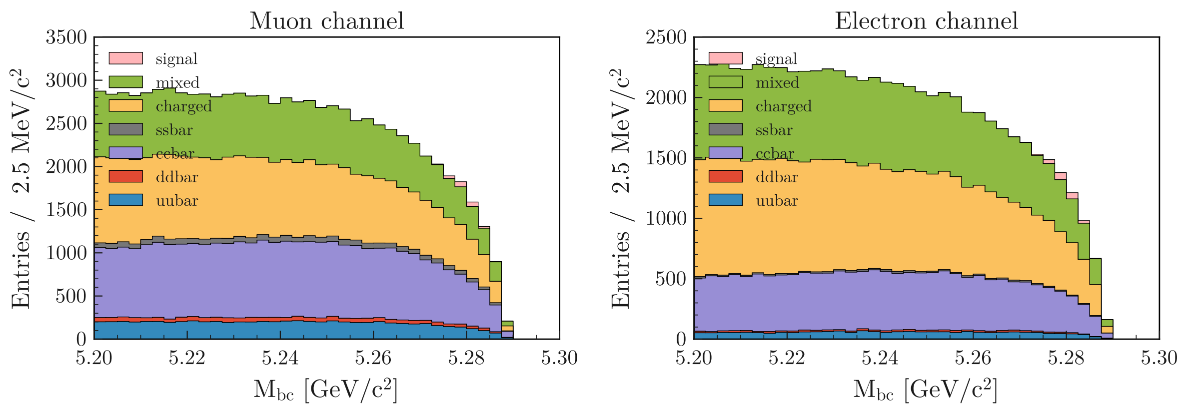


Figure 6.7: M_{bc} distribution of signal and generic MC sample after pre-selection.

We study the source of background events in the generic MC sample using the MC truth information. In this analysis, the $B\overline{B}$ and $c\overline{c}$ background events are studied in detail, while the other continuum background $u\overline{u}$, $d\overline{d}$, $s\overline{s}$ are not analyzed, as they contribute less significantly and can be easily rejected using event shape variables mentioned in the following Sec. 6.3.1. We only trace the decay chain of the two reconstructed leptons using the truth information of the MC while ignoring the reconstruction of X_s at this stage. This is because it is not common to have two leptons in one B decay after vetoing charmonium events.

The results of the $B \to X_s\mu^+\mu^-$ channel and the $B \to X_s e^+e^-$ channel are shown in Fig. 6.8 and 6.9. In these plots, only the mother of the lepton is shown while other daughters of the mother particle are missing. All mother particles are marked in their ground state (e.g. D^{*+} is marked as D), and transitions between excited states and ground states are ignored. If the two leptons come from different mother particles (B or D meson), the decay chain of one lepton is enclosed in parentheses. If the lepton comes from the misidentification of the hadron h, the lepton is remarked as $\ell(h)$. Backgrounds with expected contribution smaller than 50 candidates are merged in "others".

For $B \to X_s\mu^+\mu^-$, the major contributions come from:

- Semi-leptonic process with two leptons come from two different B mesons from one $\Upsilon(4S)$: $\Upsilon(4S) \to B(\to X\mu^-\bar{\nu})\overline{B}(\to X\mu^+\nu)$
- Semi-leptonic process with two leptons come from the chain semi-leptonic process of one B meson: $B \to \mu^-\bar{\nu}D(\to X\mu^+\nu)$.
- Semi-leptonic process with two leptons come from two different D mesons from e^+e^- collision: $e^+e^- \to D(\to X\mu^-\bar{\nu})\overline{D}(\to X\mu^+\nu)$
- Continuum light quark pair production processes: $e^+e^- \to q\bar{q}$, $(q = u, d, s, c)$.
- One lepton comes from a semi-leptonic process, while the other comes from misidentification of pion to muon.

For $B \to X_se^+e^-$, the main source is similar to that of muon channels, with one more significant contribution from the $B \to XJ/\psi(\to e^+e^-)$ process that escapes the charmonium veto.

6.3 MultiVariate Analysis (MVA)

In this section, we describe the suppression of semi-leptonic background events ($B\overline{B}$) and continuum events ($q\bar{q}$) using MVA-based algorithm. We describe the variables that are used to train the MVA classifiers and present their distributions in signal and background events in Sec. 6.3.1. In Sec. 6.3.2, we describe the corrections on the distributions of two key variables in signal MC sample. After describing the data samples in Sec. 6.3.3, we present the details about training and selection optimization in Sec. 6.3.4 and Sec. 6.3.5, respectively. There can be multiple B meson candidates remaining in one event after all selections and we describe the strategy for selecting the best candidate in Sec. 6.3.6. We study the importance of each training variable in the classifiers in Sec. 6.3.7.

6.3.1 Input variables

To reject the continuum process, the event shape variables are used as input of MVA. Compared to the $\Upsilon(4S) \to B\overline{B}$ process where the mass of two B-mesons is close to the mass of $\Upsilon(4S)$, leaving little phase space for B-meson momenta, the collision products of continuum processes are light quarks with high back-to-back momentum at the center-of-mass frame. For this reason, the envelope of the momenta of final state particles is

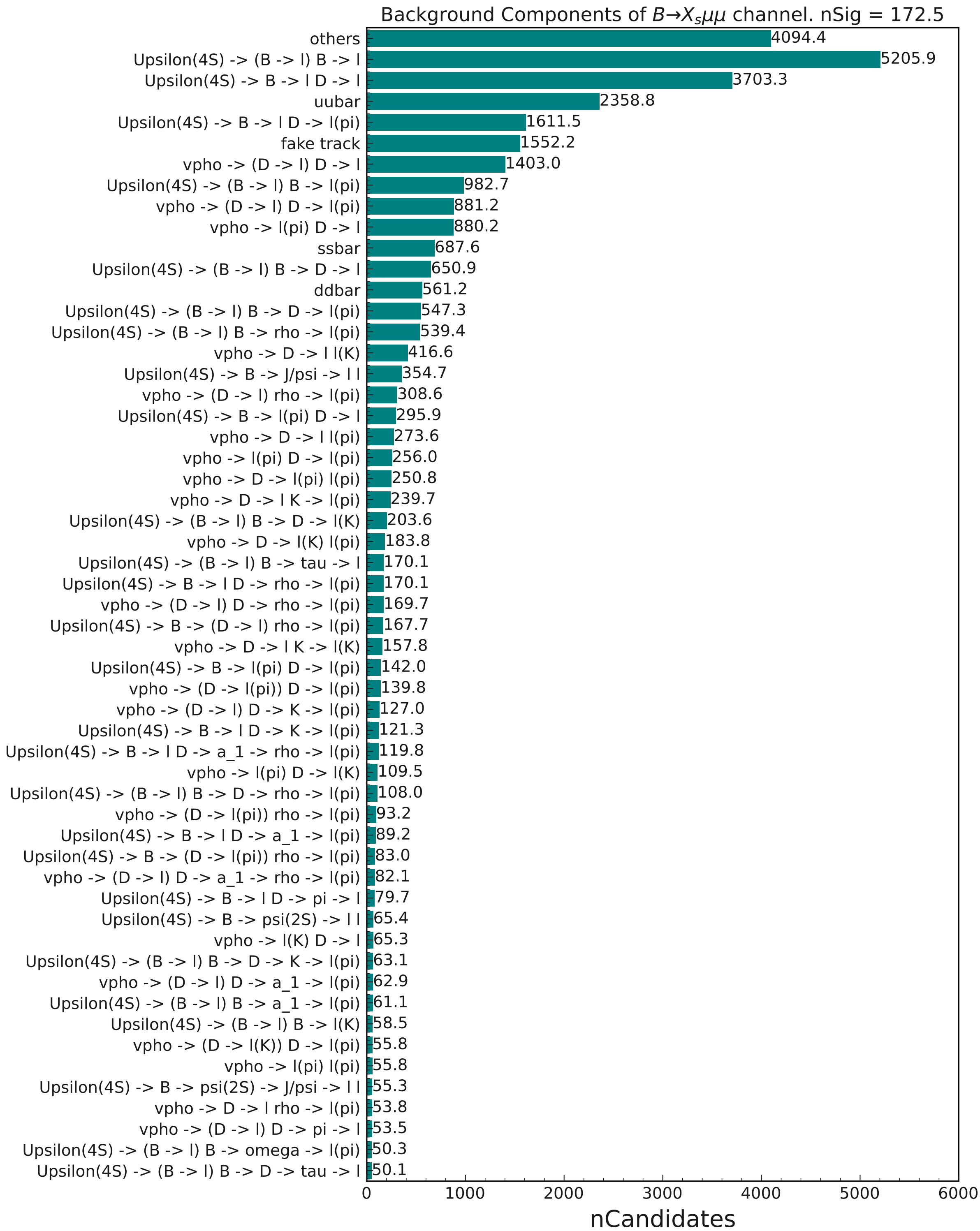


Figure 6.8: Background analysis after pre-selection of $B \to X_s \mu^+ \mu^-$ samples. The expected number of signal event is 172.5. The `vpho` refers to the virtual photon generated by e^+e^- collision.

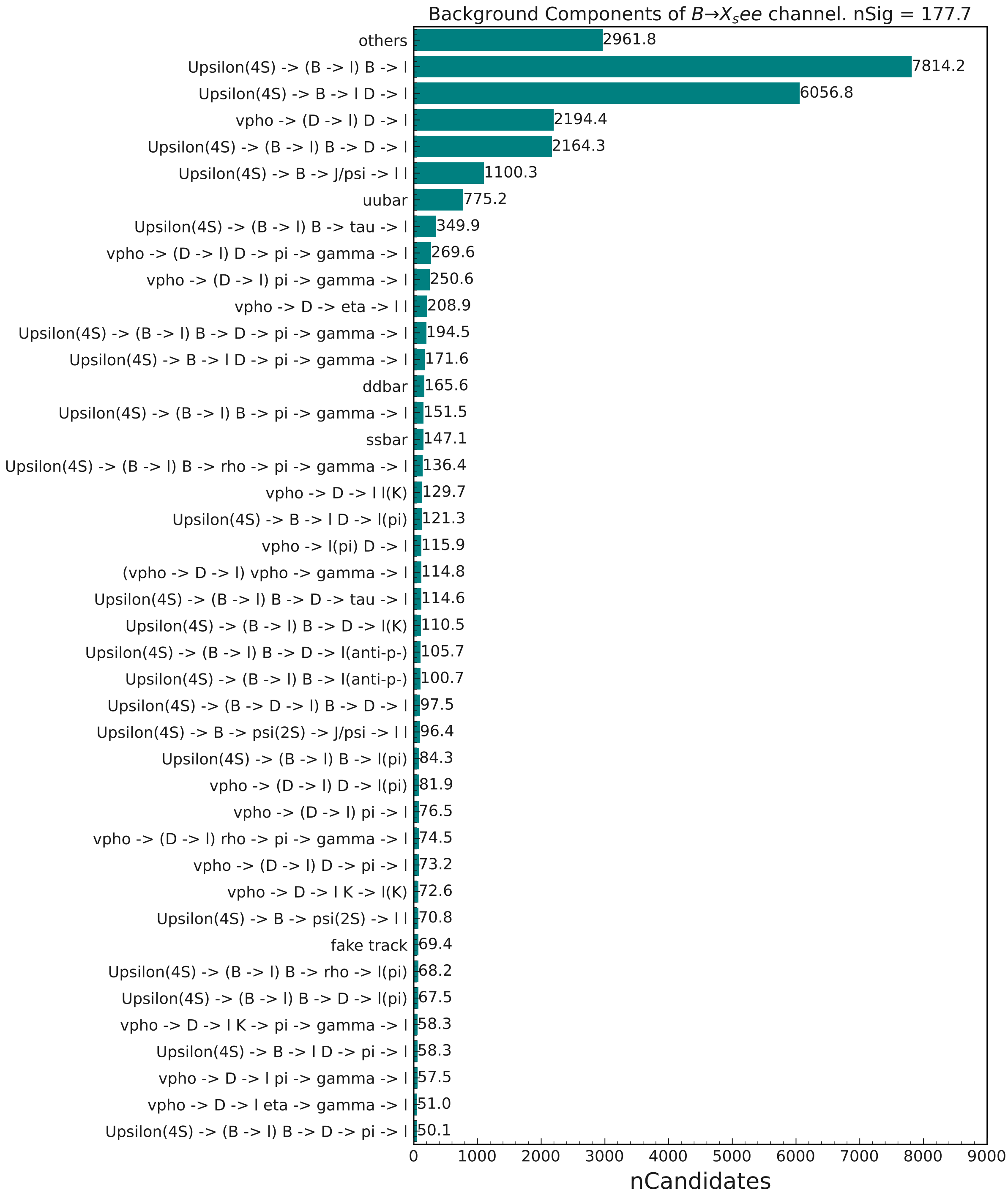


Figure 6.9: Background analysis after pre-selection of $B \to X_s e^+ e^-$ samples. The expected number of signal event is 177.7. The `vpho` refers to the virtual photon generated by e^+e^- collision.

“jet-like” for continuum process and is “sphere-like” for $B\overline{B}$ processes as illustrated in Fig. 6.10.

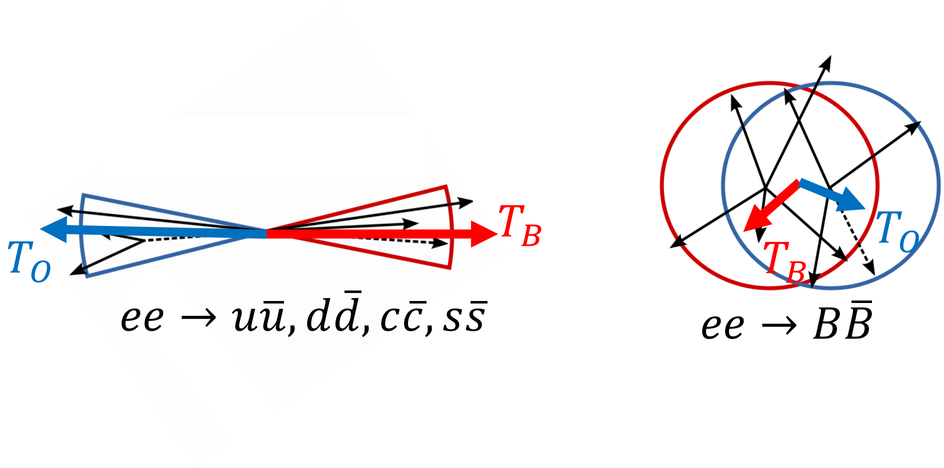


Figure 6.10: Illustration of event shape of $q\overline{q}$ and $B\overline{B}$ events. The black arrows are final state particles. The red and blue arrows represent the direction and magnitude of the thrusts of signal B candidate and ROE, respectively.

To describe the shape of the entire event, tracks and clusters not used in B candidate reconstruction, referred to as Rest Of Event (ROE), are also necessary. We reconstruct them with the following requirements:

- Tracks: $p < 3.2\,\mathrm{GeV}/c$ in the center-of-mass frame; the number of hits left in CDC is greater than 0; mass is assumed to be the pion mass.
- Clusters: $p < 3.2\,\mathrm{GeV}/c$ in the center-of-mass frame; $p > 0.05\,\mathrm{GeV}/c$ at the laboratory frame to reject noise; mass is assumed to be zero.

Here we set the requirement of track and cluster momenta in the center-of-mass frame to be smaller than their maximal value over $3.2\,\mathrm{GeV}/c$, which is confirmed from simulation sample.

Three sets of event shape variables are used in this analysis:

- Thrust variables.
  The thrust is a vector, whose direction in the signal side (T_B) or ROE side (T_O) is determined by maximizing the momentum projection on the thrust. And its magnitude is defined to be the normalized magnitude of the projection:

$$T = \max_{||n||=1} \frac{\sum_i |p_i \cdot n|}{\sum_i |p_i|} \tag{6.3}$$

  where n is the unit vector along the thrust and p_i is the momentum of number i particle in the signal or ROE side. Magnitude of T_B (B_thrustBm), cosine of the angle between T_B and z axis (B_cosTBz), and cosine of the angle between T_B and

T_O (B_cosTBTO) are used as input. Magnitude of T_O is not adapted due to its poor consistency between data and MC. Distributions of thrust variables are shown in Fig. 6.11. We use the prependix "B_" to denote that these variables are calculated based on the reconstructed B meson.

- 15 Kakuno Super Fox-Wolfram (KSFW) moments [76].
  The Fox-Wolfram moments is used to describe the phase space distribution of momenta and energies using spherical harmonics. In B-physics, the modified version called KSFW moments are more commonly used which calculate Fox-Wolfram moments depending on whether the particles are from the signal B candidate or ROE, and particle types (charged, neutral, or missing). In this calculation, missing momentum is included as an additional virtual particle. The moments are defined as:

$$\begin{aligned}
\text{B_hso}\{x,l\} &= \sum_i \sum_j |\mathbf{p}_j| Q_i Q_j P_l(\cos\theta_{i,j}) \quad (x=0; l=1,3), \\
\text{B_hso}\{x,l\} &= \sum_i \sum_j |\mathbf{p}_j| P_l(\cos\theta_{i,j}) \quad (x=0,1,2; l=0,2,4), \\
\text{B_hoo}\{l\} &= \sum_j \sum_k Q_j Q_k |\mathbf{p}_j||\mathbf{p}_k| P_l(\cos\theta_{j,k}) \quad (l=1,3), \\
\text{B_hoo}\{l\} &= \sum_j \sum_k |\mathbf{p}_j||\mathbf{p}_k| P_l(\cos\theta_{j,k}) \quad (l=2,4),
\end{aligned} \tag{6.4}$$

  where i runs over particles of signal B candidates, j,k run over the ROE particles, x defines the subset of ROE particles (0, 1, 2 represents charged, neutral, and missing, respectively) that j runs over, $P_l(\cos\theta_{i,j})$ represents the l-th order Legendre polynomial of the cosine of the angle between momenta of i-th and j-th particle, and $Q_i = 0, \pm 1$ is the charge of i-th particle. B_hso01 is not used in this analysis for its poor data/MC agreement. Additional information of transverse energy (et) and missing mass squared (mm2) are also adapted:

$$\begin{aligned}
\text{B_et} &= \sum_m |(p_T)_m|, \\
\text{B_mm2} &= \left(\sqrt{s} - \sum_m E_m\right)^2 - c^2 \left|\sum_m \vec{p}_m\right|^2, \quad (\sqrt{s} > \sum_m E_m), \\
\text{B_mm2} &= -\left(\sqrt{s} - \sum_m E_m\right)^2 - c^2 \left|\sum_m \vec{p}_m\right|^2, \quad (\sqrt{s} < \sum_m E_m),
\end{aligned} \tag{6.5}$$

  where m runs over all particles in one event, p_T is the transverse momentum, E_i and p_i are calculated at the center-of-mass frame.
  We adapt 15 KSFW moments defined above with good data/MC consistency. Distributions of the 15 variables are shown in Fig. 6.12.
- CLEO Cone [77].

Nine CLEO Cone variables are used as well. They are calculated as the scalar sum of momenta of all charged tracks and neutral showers in nine spaces, with each space cover 10° interval of polar angle from T_B. For example, B_cleo5 covers the polar angle range from 40° to 50°. Distributions of CLEO variables are shown in Fig. 6.13.

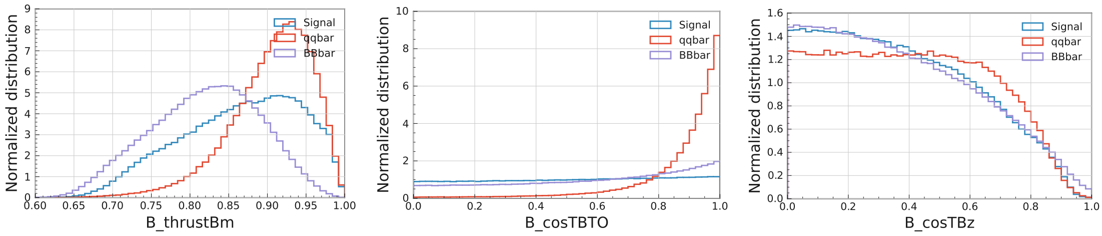


Figure 6.11: Distribution of thrust variables for signal and background samples.

Compared to the signal process where two leptons come from the same B meson decay position, the semi-leptonic processes feature in creating two leptons from different mother particles whose decay positions could be separate from each other. For this reason, it tends to have a smaller χ^2 probability in vertex fitting of the di-lepton system (jpsi_chiProb) and also the distance between POCA of the two lepton tracks along the z axes $\Delta z_{\ell^+\ell^-}$ (jpsi_deltaZll) tends to be larger. In addition, due to the two accompanying neutrinos, the visible energy of the event $E_{\text{visi}} = \sum_i E_i^*$ in the center-of-mass frame would become smaller than that of signal events and the missing mass squared $M^2_{\text{miss}} = (\sqrt{s} - \sum_i E_i^*)^2 - |\sum_i \vec{p}_i^*|^2$ of the event would become nonzero, where $(E_i^*, \vec{p}_i^*)$ are the reconstructed 4-momenta in the center-of-mass frame of all tracks assumed to be pions and all photons in the event. There are other variables that are useful for rejecting background candidates, including the χ^2 probability of B vertex fitting, ΔE, and the momentum direction of the B meson in the center of mass frame (B_cosTheta_CMS). The flavor tagger output (B_qrGNN) is also used as one of the input variables, which was originally developed for time-dependent CP violation analysis. Candidates with flavor tagger output close to +1 are $B^+(B^0)$-like, output close to -1 are $B^-(\overline{B}^0)$-like, and output close to 0 means that it is hard to identify its flavor. Figure 6.14 shows the distributions of the variables mentioned above.

Furthermore, given that the distribution of deltaE differs depending on whether a π^0 is used to reconstruct the signal, the number of π^0 daughters (Xs_nDaughters_pi0) is also used as a training variable. We regard this variable as an indicator of the event type, instead of a discriminator of signal and background. The distribution for this variable is also presented in Fig. 6.14.

The 36 training variables are summarized in Tab. 6.2. Figure 6.15 shows the correlations among them in the simulation samples. The correlations between the training variables and the fitting variable M_{bc}, as well as the MVA output, are also presented in the plots.

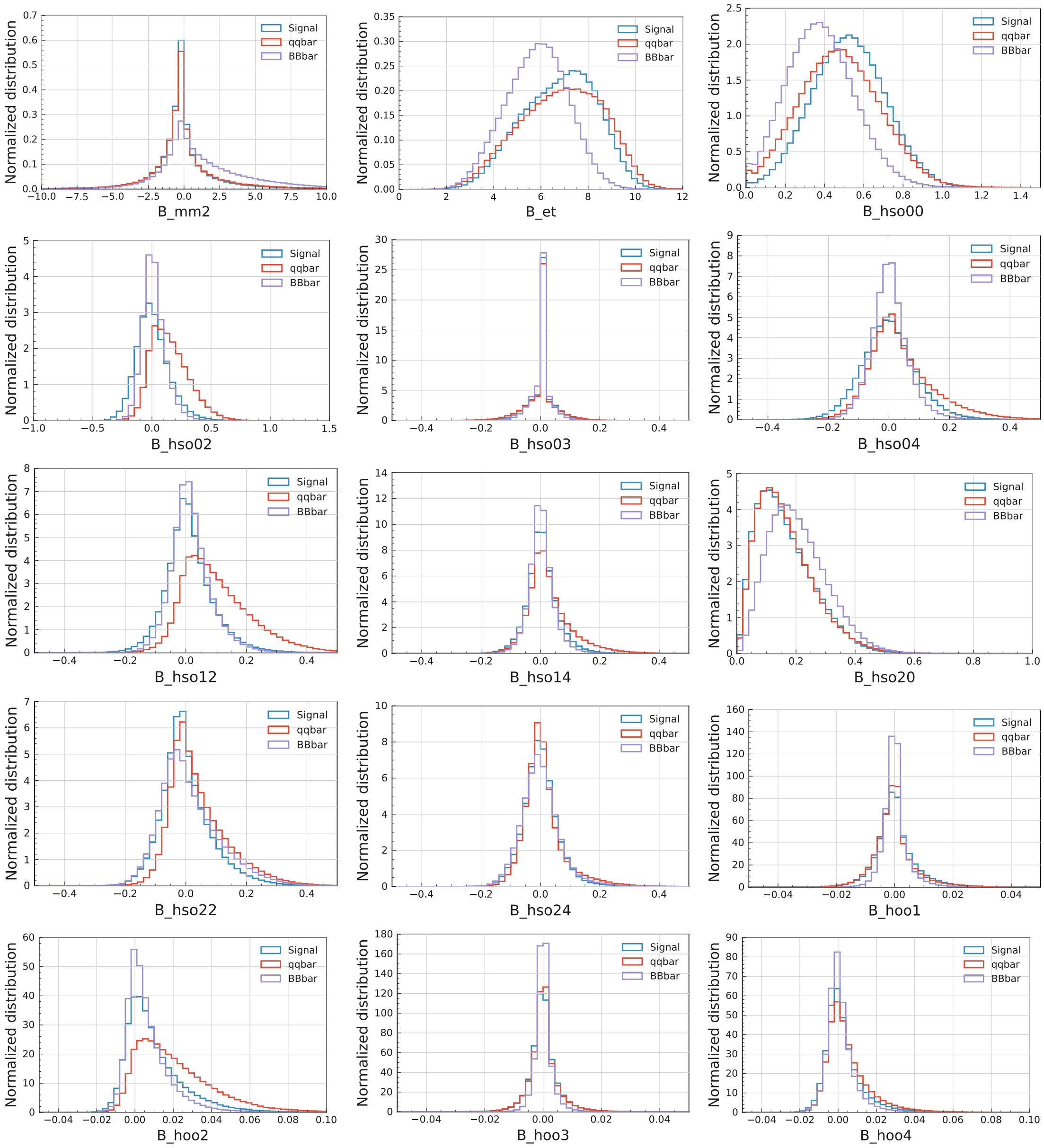


Figure 6.12: Distribution of KSFW variables for signal and background samples.

Table 6.2: Summary of training variables

Variables	Explanation
B_et, B_mm2 B_hso00, B_hso02, B_hso03, B_hso04 B_hso12, B_hso14, B_hso20, B_hso22, B_hso24, B_hoo1, B_hoo2, B_hoo3, B_hoo4,	15 variables of KSFWMoment calculated from primary daughters.
B_cleo1, B_cleo2, B_cleo3, B_cleo4, B_cleo5, B_cleo6, B_cleo7, B_cleo8, B_cleo9	9 CLEO Cone variables
B_thrustBm	Thrust magnitude of B meson.
$\cos\theta_{\mathrm{TBz}}$ (B_cosTBz)	Cosine of angle between the thrust axis of B and the beam axis.
$\cos\theta_{\mathrm{TBTO}}$ (B_cosTBTO)	Cosine of angle between the thrust axis of B and that of ROE.
M^2_{missing} (missingMass2OfEvent)	Missing mass squared.
E^*_{visible} (visibleEnergyOfEventCMS)	Visible energy in the CMS frame.
ΔE (B_deltaE)	$E^*_B - E_{\mathrm{beam}}$.
chiProb$_B$ (B_chiProb)	The χ^2 probability of B vertex fit.
$\cos\theta^*_B$ (B_cosTheta_CMS)	Cosine of θ of B at the CMS frame.
qrGNN (B_qrGNN)	GNN flavor tagger output
chiProb$_{\ell^+\ell^-}$ (jpsi_chiProb)	The χ^2 probability of di-lepton vertex fit.
$\Delta z_{\ell^+\ell^-}$ (jpsi_deltaZll)	Distance between two leptons along the beam axis.
Xs_nDaughters_pi0	number of π^0 (0 or 1) used for reconstruction.

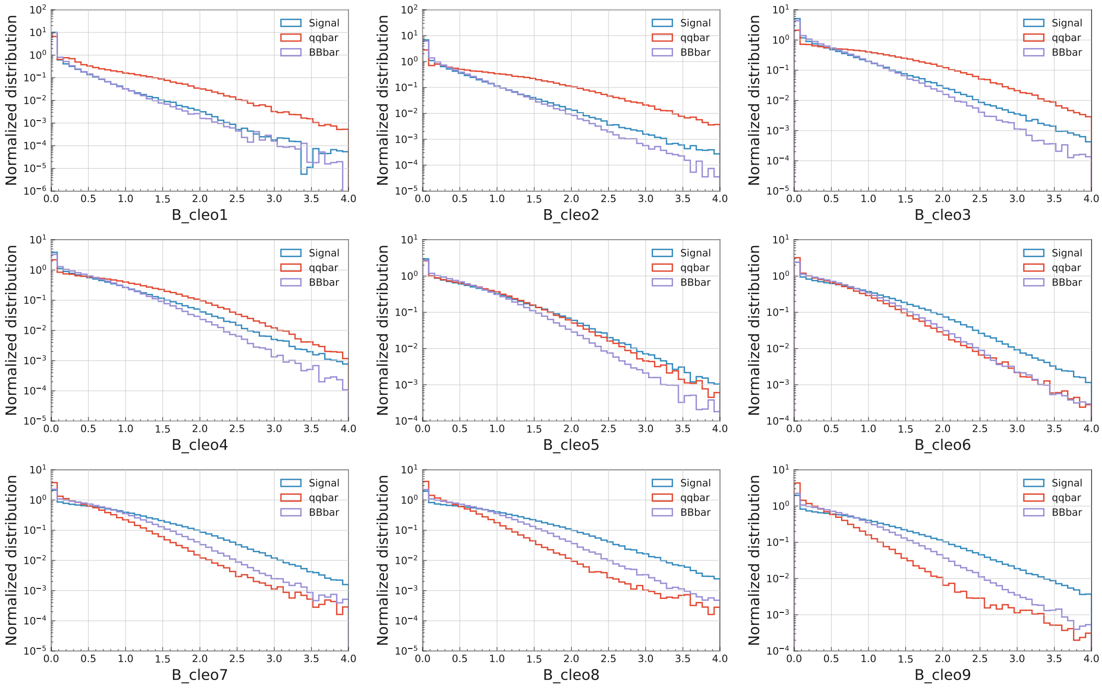


Figure 6.13: Distribution of CLEO variables for signal and background samples.

6.3.2 Data/MC Consistency Check

We have confirmed that the data and MC distributions are consistent for most of the training variables. Plots are shown in Appendix C, except for ΔE and $\Delta z_{\ell^+\ell^-}$, whose corrections are explained below. This comparison is performed with the control channel $B \to X_s J/\psi(\to \ell^+\ell^-)$, following the same event reconstruction and selection process as the signal channel except the charmonium veto. Instead, we require mass window cuts $2.9\,\mathrm{GeV}/c^2 < m(J/\psi) < 3.2\,\mathrm{GeV}/c^2$ and $M_{bc} > 5.27\,\mathrm{GeV}/c^2$ to enhance control channel events.

We calibrate the ΔE distributions using the exclusive $B \to K^{(*)}J/\psi$ decays. This calibration is carried out by performing an unbinned maximal likelihood fit to the ΔE distribution and comparing the difference in shape parameters between data and MC. In this fitting, the signal shape is described using a crystal ball function:

$$CB(\Delta E|\mu,\sigma,\alpha,\beta) = \begin{cases} C\exp\left(-x^2/2\right), & \text{for } x > -\beta \\ CA(B-x)^{-\alpha} & \text{for } x \leq -\beta \end{cases}, \text{ with } x = \frac{\Delta E - \mu}{\sigma}, \tag{6.6}$$

where

$$A = (\alpha/|\beta|)^{\alpha}\exp(-\beta^2/2), B = \alpha/|\beta| - |\beta|$$

and C is the normalization constant. And a first-order Chebyshev polynomial is used to describe the background. This calibration is performed in four different categories: muon channel with or without π^0 and electron channel with or without π^0. Fitting plots to these four categories are shown in Fig. 6.16. The summary of the data/MC discrepancy

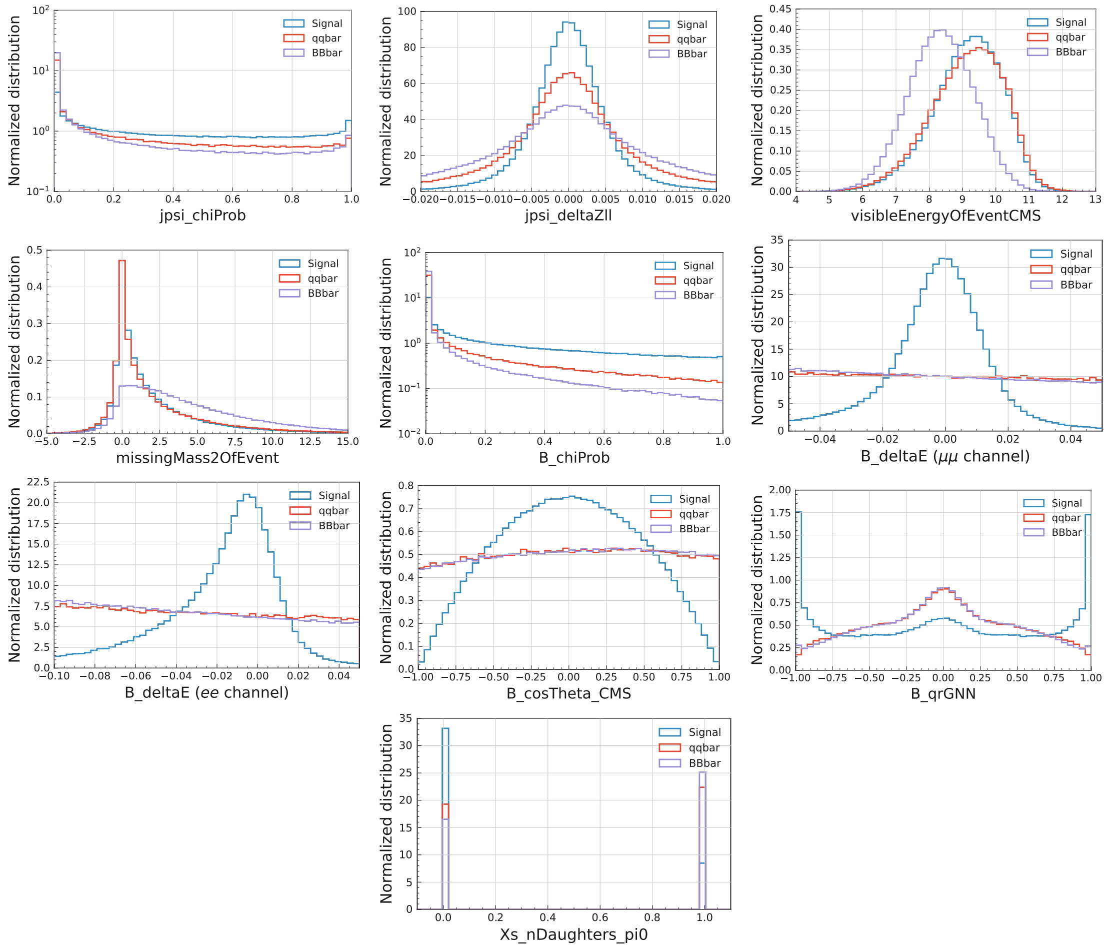


Figure 6.14: Distribution of training variables for signal and background samples.

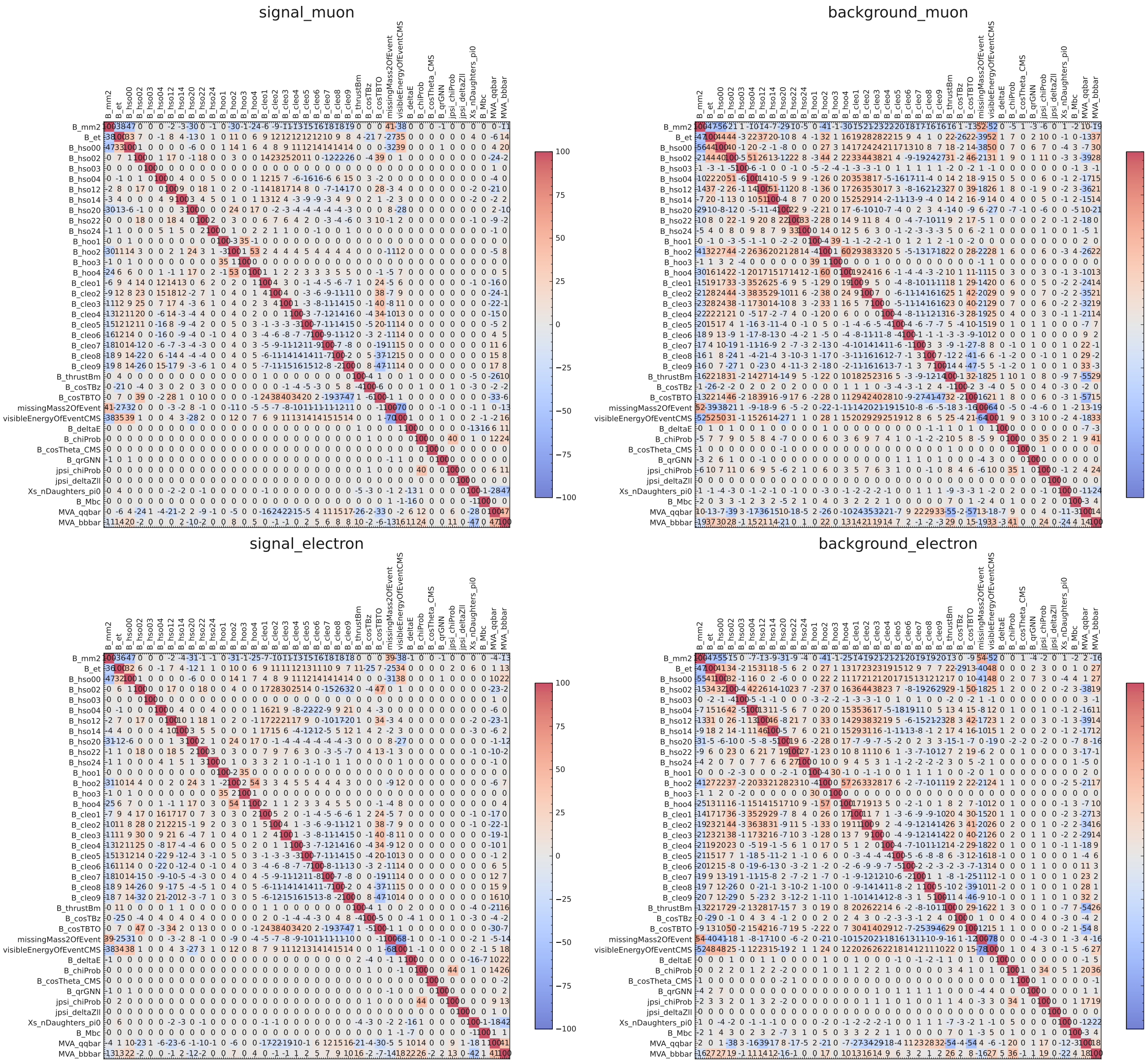


Figure 6.15: Correlations between training variables. The left column is derived in signal MC samples while the right column is derived in generic MC (background) samples. The top row is for muon channel and the bottom row is for electron channel.

is shown in Fig. 6.17. This plot also shows the discrepancy derived from each single channel, and they are in good consistency with the inclusive measurement of four categories. In this result, discrepancies are observed in the center values of electron channels and width in channels without π^0. These discrepancies are corrected by convoluting a smearing Gaussian $\mathcal{N}(0, 3.8)$, $\mathcal{N}(-1.1, 3.8)$, and $\mathcal{N}(-1.7, 0)$ to the ΔE variable of the muon channel without π^0, the electron channel without π^0 and the electron channel with π^0, respectively. The two numbers in parentheses are mean and width in units of MeV, respectively. The width of the smearing Gaussian mentioned above is derived by $\sqrt{\sigma^2_{data} - \sigma^2_{MC}}$.

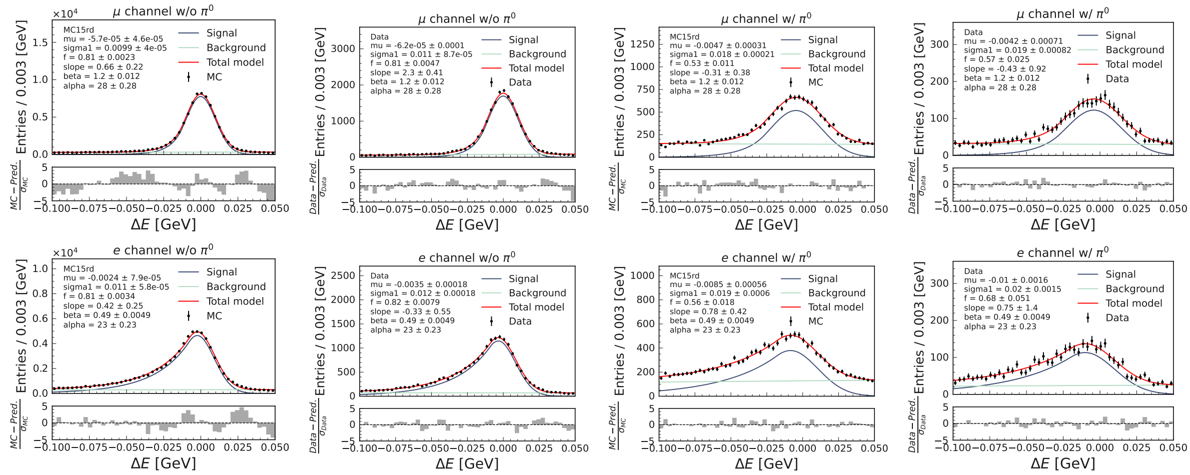


Figure 6.16: Fitting to ΔE distributions in $B \to K^{(*)} J/\psi$ channels in MC and data. The difference in the ΔE shape is used to correct MC samples.

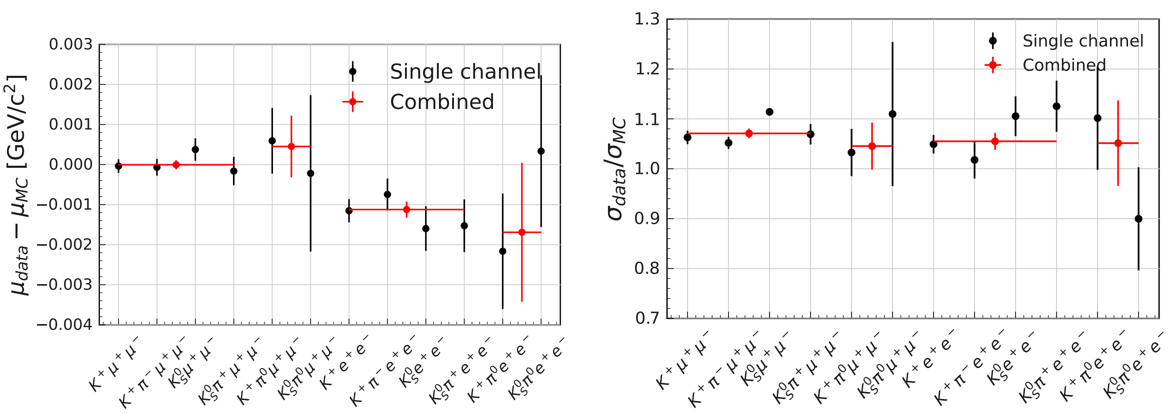


Figure 6.17: Summary of ΔE shape discrepancy between data and MC. Both results derived from fitting to one category inclusively (red) and derived independently from each channel (black) are shown.

The distribution of $\Delta z_{\ell^+\ell^-}$ is calibrated using the $B^+ \to K^{*+}(\to K^+\pi^0)J/\psi(\to \ell^+\ell^-)$ channel. The distribution is fitted to a double Gaussian function with shared center values, and the fitting result is shown in Fig. 6.18. Center values derived from fitting in both data and MC are zero-consistent, with little difference in the width of the narrower Gaussian. This little difference is calibrated by smearing a Gaussian with width $\sqrt{\sigma_{data}^2 - \sigma_{MC}^2} = 0.001$ cm.

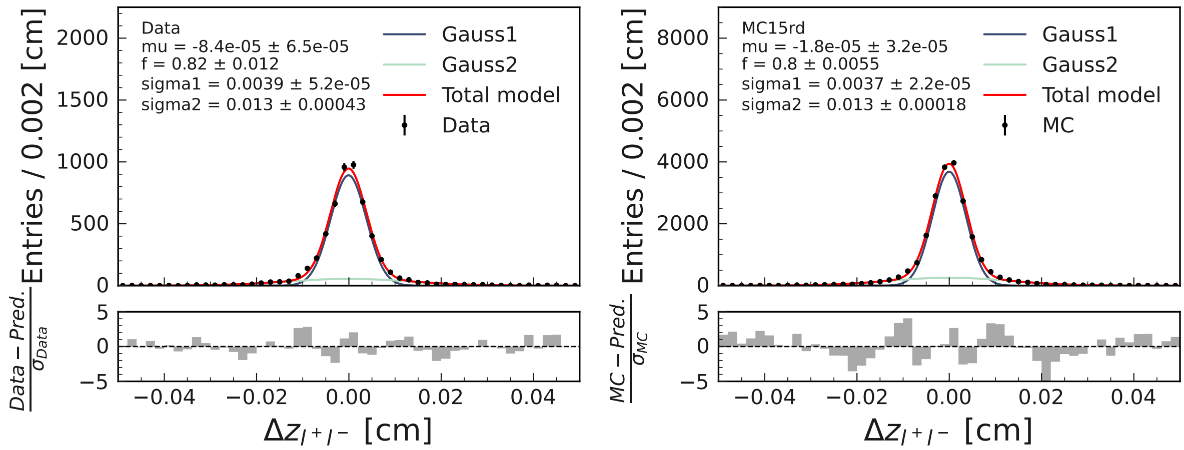


Figure 6.18: Fitting to $\Delta z_{\ell^+\ell^-}$ in data (left) and MC (right) in $B^+ \to K^{*+}(\to K^+\pi^0)J/\psi(\to \ell^+\ell^-)$ channel. The fitted resolution is 39 μm in data and 37 μm in MC.

6.3.3 Training and test samples

MVA classifiers are trained independently in four separate sub-samples, depending on the lepton channel and whether $m(X_s)$ is above $1.1\,\mathrm{GeV}/c^2$ or not. This is because the electron and muon channels suffer differently from bremsstrahlung, which leads to differences in the distribution of ΔE. The exclusive region ($m(X_s) < 1.1\,\mathrm{GeV}/c^2$) and inclusive ($m(X_s) > 1.1\,\mathrm{GeV}/c^2$) region are trained separately due to their significantly different background level. These four samples are denoted muon_K muon_Xs, electron_K, electron_Xs accordingly. In each sub-sample, the classifiers against the $q\bar{q}$ and $B\overline{B}$ background are trained separately as well, denoted as MVA_qqbar and MVA_bbbar, respectively. In total, eight classifiers are trained in this analysis.

The training and test samples of the signal are derived by splitting the simulated signal sample randomly in an equal ratio. For background samples, the full run-dependent sample is used as the test sample, while the full run-independent dataset is used as the training set. The consistency of the variable distribution is confirmed between the run-dependent and run-independent samples. The sizes of each sample are summarized in Tab. 6.3.

6.3.4 Methods and training

Five different machine learning methods are used in this analysis: FBDT [78], Xgboost, Deep Neural Network (DNN), TabNet [79] and Feature-token transformer (FTTransformer) [80]. Hyper parameters for each model are tuned to maximize their performance

Table 6.3: Traing and test sample sizes used in MVA. Numbers in the parentheses are the corresponding expected number of events in run1 dataset after preselection.

Sample		**Training sample size**	**Test sample size**
muon_K	signal	171563 (118.5)	171719 (118.2)
	$q\overline{q}$	34649 (12269.6)	47542 (11431.6)
	$B\overline{B}$	50273 (5827.4)	22520 (5209.8)
muon_Xs	signal	80052 (54.2)	79937 (54.0)
	$q\overline{q}$	76664 (27167.7)	102709 (24750.5)
	$B\overline{B}$	406899 (47079.8)	178349 (41075.3)
electron_K	signal	152436 (114.5)	152454 (114.3)
	$q\overline{q}$	15678 (5656.0)	23122 (5734.0)
	$B\overline{B}$	47009 (5648.8)	21337 (5291.2)
electron_Xs	signal	58395 (63.4)	58205 (63.1)
	$q\overline{q}$	31824 (11469.9)	48277 (11981.7)
	$B\overline{B}$	388117 (46511.1)	174670 (43244.0)

while avoiding overtraining of the model. Because inclusive samples have much larger sample sizes for training, a more complex model is trained on these datasets compared to the exclusive ones.

Table 6.4 shows the hyper parameters for the FBDT and Xgboost models.

Table 6.4: Hyper parameters of FBDT and Xgboost. Parenthesis shows the corresponding name in Xgboost package. Early stopping rounds is only set in Xgboost.

Parameter	K model	Xs model
nTrees	300	400
depth (max depth)	3	4
shrinkage (learning rate)	0.1	0.1
subsample	0.2	0.2
binning	8	9
early stopping rounds [X]	10	10

The DNN model is built with the PyTorch [81] library. The input array is first processed by the batch normalization module, followed by a fully connected linear model. There are three (six) linear layers with 25 (35) nodes in each layer for the exclusive (inclusive) samples. The output of each node is processed by a LeakyReLU activation function before being put into the next layer. In the output of the last layer, there is a sigmoid activation function that is used to give the signal probability. The DNN model was trained with the Adam optimizer [82] with an initial learning rate of 0.001 for 70 epochs. The Mean Squared Error (MSE) loss function is adopted in this training. In this analysis, the model of the last epoch was adopted.

The TabNet model was trained with the parameters shown in Tab. 6.5. Both exclusive and inclusive samples share the same set of parameters.

Table 6.5: Hyper parameters of TabNet.

n_d	n_a	n_steps	gamma	mask type	EPOCH
64	64	10	2.0	sparsemax	100

Table 6.6 shows the hyper parameters of FTTransformer.

Table 6.6: Hyper parameters of FTTransformer.

Sample	d_block	attention_n_heads	EPOCH
K	16	4	50
Xs	32	8	50

No obvious over-training is observed in all five methods mentioned above. As an example, Fig. 6.19 shows the overtraining test for FTTransformer classifiers.

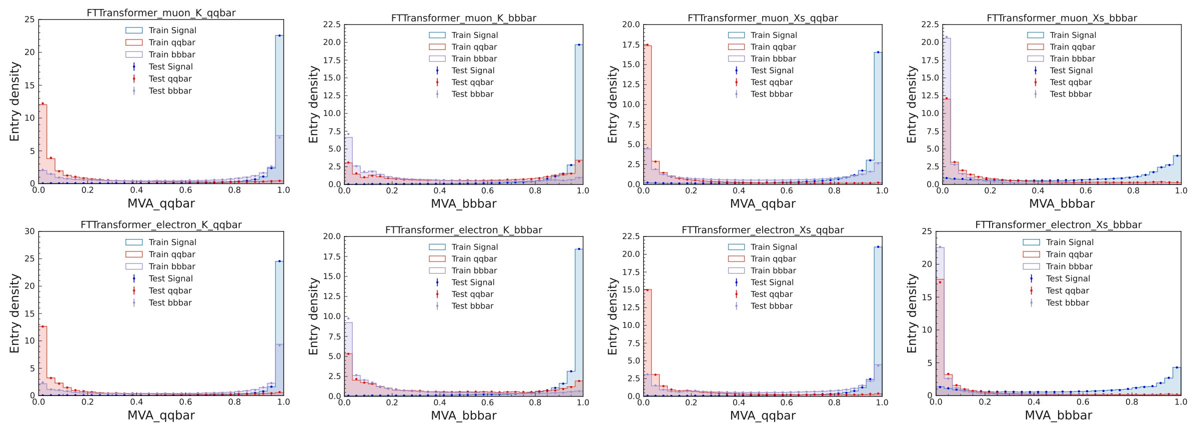


Figure 6.19: Overtraining test of FTTransformer

The best method is determined after selection criteria optimization and best candidate selection to be explained in Sec. 6.3.6.

6.3.5 Optimization

In this analysis, measurements in the $M(X_s) > 1.1\,\mathrm{GeV}/c^2$, $1 < q^2 < 6$ $[(\mathrm{GeV}/c)^2]$ and $q^2 > 14.4$ $[(\mathrm{GeV}/c)^2]$ areas are of the most interest. To optimize the signal significance in these areas, the working point of the MVA classifiers was determined in 12 different regions defined in Tab. 6.7. In addition, the working points of the muon and electron samples are determined separately as well. In total, the working points are determined in 24 (12 different regions multiplied by two different lepton flavors) different sub-samples for signal, $q\overline{q}$ and $B\overline{B}$ samples as shown in Fig. 6.20.

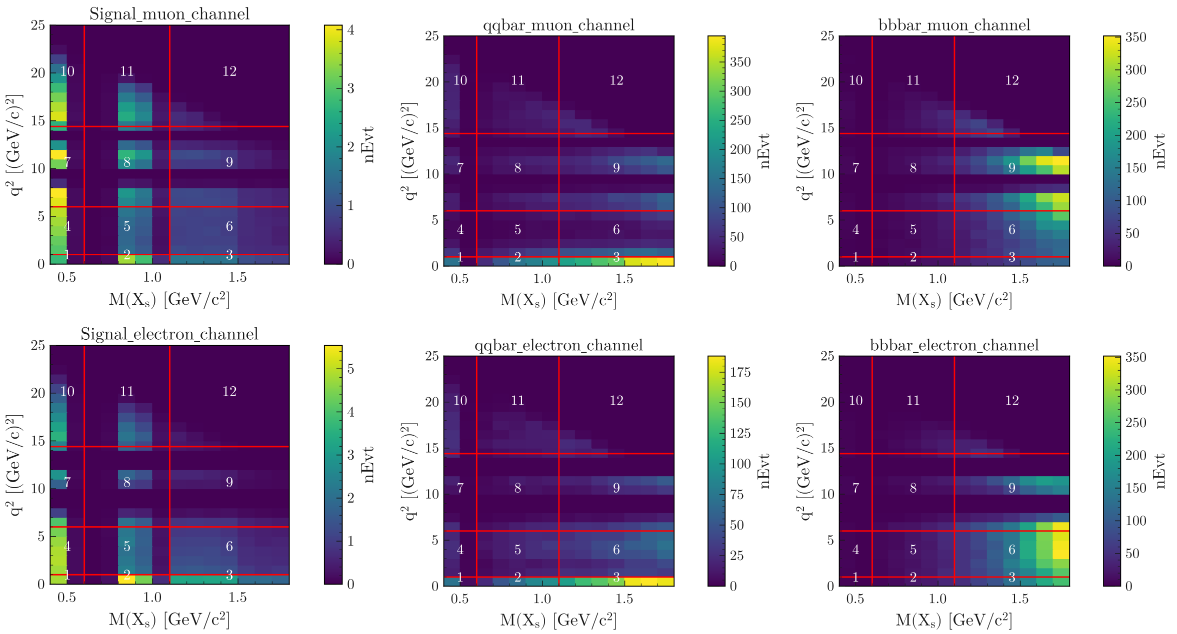


Figure 6.20: Optimization region splitting. The colorbar in the plot shows the expected number of events after pre-selection in signal (left column), $q\overline{q}$ (middle column), and $B\overline{B}$ (right column) samples. The top row is for muon channel and bottom row for electron channel. The white numbers in the plots indicates the region index while the red lines are the boundaries of different regions.

In each sub-sample, the selections in MVA_qqbar and MVA_bbbar are optimized simultaneously to achieve the highest Figure of Merit (FoM) $S/\sqrt{S+B}$ in the region $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$.

6.3.6 Best Candidate Selection (BCS)

After the background suppression by MVA, the average number of candidates per event (referred to as event multiplicity) is about 1.14 for the signal sample, as shown in Fig. 6.21. The best candidate for each event is selected by requiring the greatest output of the MVA_bbbar classifier, which can suppress signal events that are not correctly reconstructed as well.

Figure 6.22 shows the FoM in the whole sample and three $m(X_s)$ bins after BCS for five methods. Given that the $1.1 < m(X_s) < 1.8$ [GeV/c^2] region is of the most interest, FTTransformer that gives the best performance in that region is determined as the method to be used in this analysis. Meanwhile, FTTansformer also has reasonably

Table 6.7: Definition of optimization regions. Illustration of these regions are presented in Fig. 6.20.

Region	Definition
1	$M(X_s) < 0.6$ and $q^2 \leq 1$
2	$0.6 < M(X_s) \leq 1.1$ and $q^2 \leq 1$
3	$1.1 \leq M(X_s)$ and $q^2 \leq 1$
4	$M(X_s) < 0.6$ and $1 < q^2 \leq 6$
5	$0.6 < M(X_s) \leq 1.1$ and $1 < q^2 \leq 6$
6	$1.1 \leq M(X_s)$ and $1 < q^2 \leq 6$
7	$M(X_s) < 0.6$ and $6 < q^2 \leq 14.4$
8	$0.6 < M(X_s) \leq 1.1$ and $6 < q^2 \leq 14.4$
9	$1.1 \leq M(X_s)$ and $6 < q^2 \leq 14.4$
10	$M(X_s) < 0.6$ and $14.4 < q^2$
11	$0.6 < M(X_s) \leq 1.1$ and $14.4 < q^2$
12	$1.1 \leq M(X_s)$ and $14.4 < q^2$

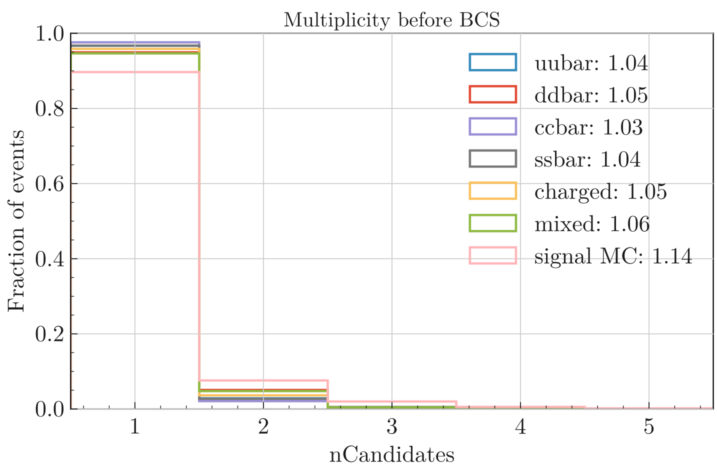


Figure 6.21: Event multiplicity in each MC sample after MVA suppression.

good FoM in other regions.

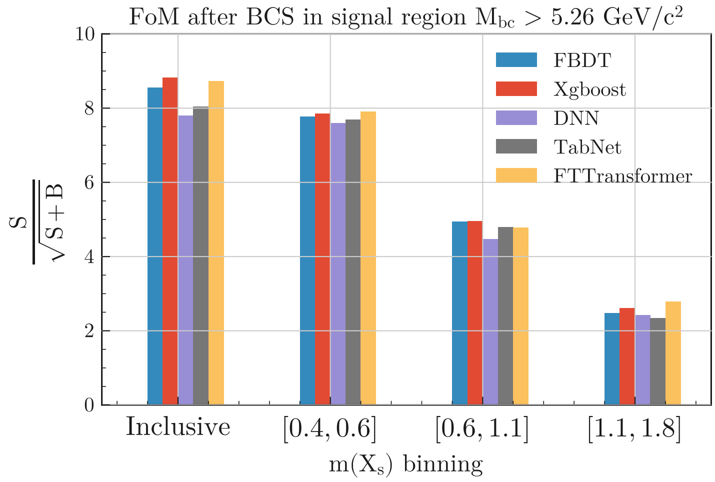


Figure 6.22: FoM after best candidate selection in each $m(X_s)$ region for five methods.

Fig. 6.23 shows the M_{bc} distributions after BCS. While signal peak become significant after BCS, it is obvious that there are peaking backgrounds under the signal peak that need careful estimation. In addition, we present the $m(X_s)$ distribution after BCS in Fig. 6.24.

6.3.7 Variable importance evaluation

In this analysis, the importance of each variable is evaluated using the permutation importance method. We first define the classifier score as the prediction accuracy on the training sample when requiring classifier output greater than 0.8. Then, the values of a certain corresponding variable are shuffled among the training samples, and the reduction of the model score is defined as the variable importance. Each variable is evaluated five times and the average importance and the standard deviation are shown in Fig. 6.25 and Fig. 6.26 for classifiers against $q\bar{q}$ and $B\overline{B}$ backgrounds, respectively.

For $q\bar{q}$ classifiers, B_thrustBm, B_deltaE, B_chiProb and event shape variables are the most important. For classifiers against $B\overline{B}$ background, B_deltaE, B_chiProb and some event shape variables are the most important ones.

6.4 Summary

The reconstruction efficiency after all selections mentioned above is shown in Fig. 6.27.

The cut flow table is shown in Tabs. 6.8 and 6.9 for muon and electron channels. All numbers are scaled to the Belle II run 1 dataset corresponding to a luminosity of 365 fb^{-1}. In this table, the number of signal N_{sig} is defined to be the number of candidates that is correctly reconstructed. For this reason, N_{sig} decreases even during BCS, which does not

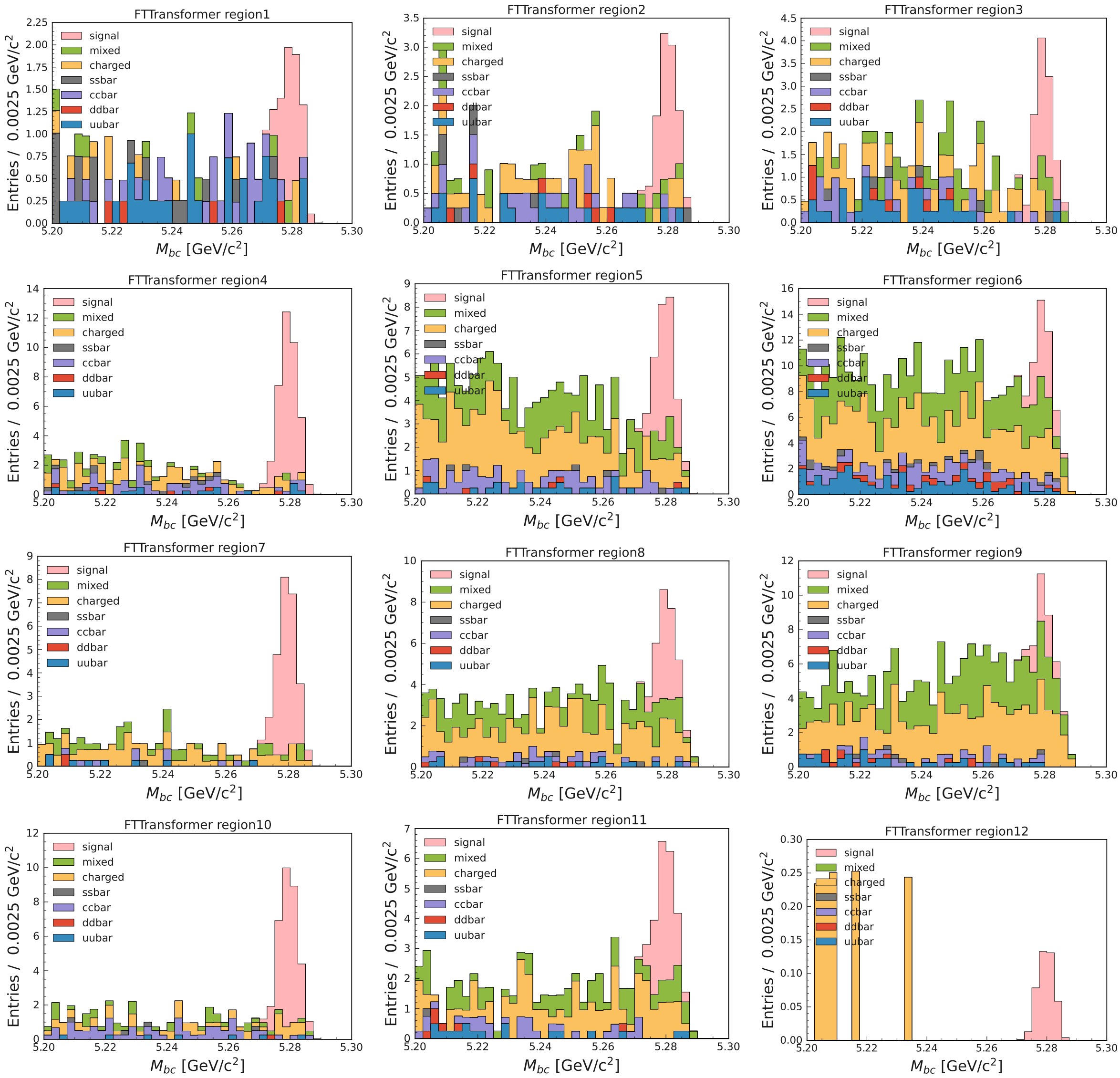


Figure 6.23: M_{bc} of B meson distributions after BCS in each region.

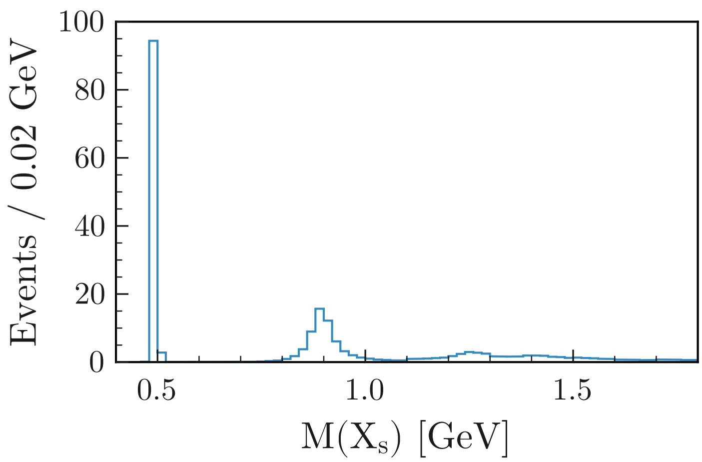


Figure 6.24: $m(X_s)$ distribution of signal MC sample after event selection.

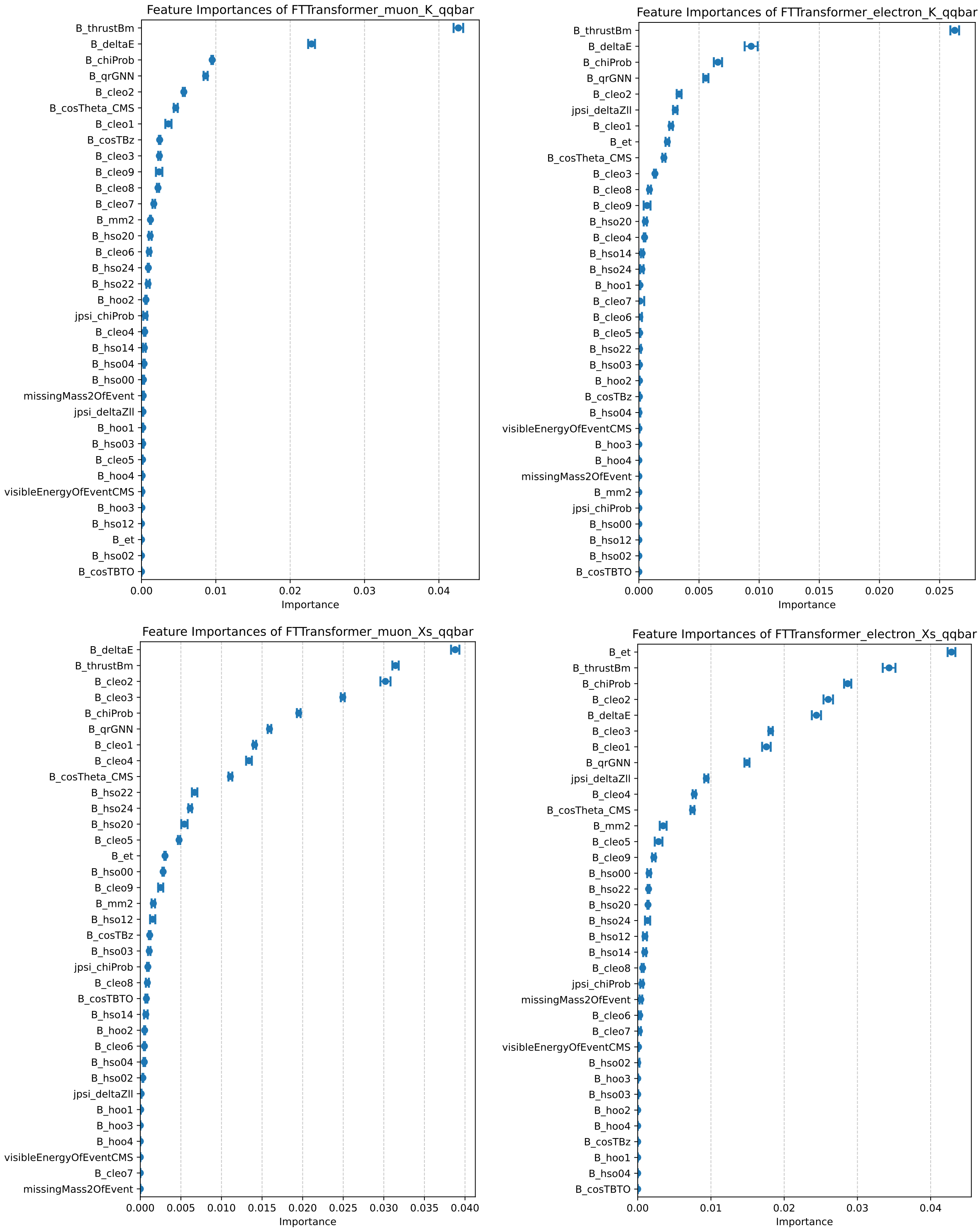


Figure 6.25: Feature importance of classifiers against $q\bar{q}$ backgrounds.

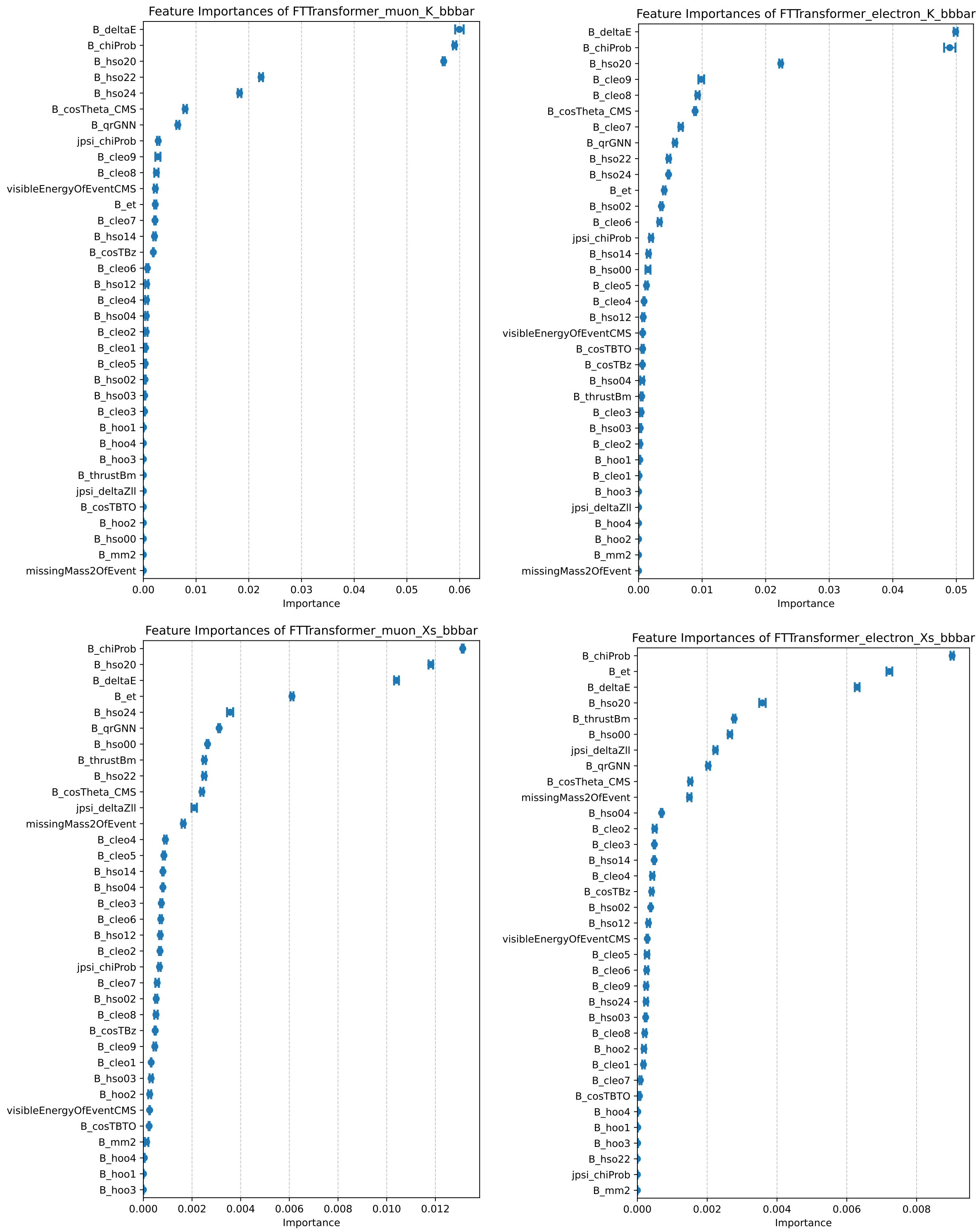


Figure 6.26: Feature importance of classifiers against $B\overline{B}$ backgrounds.

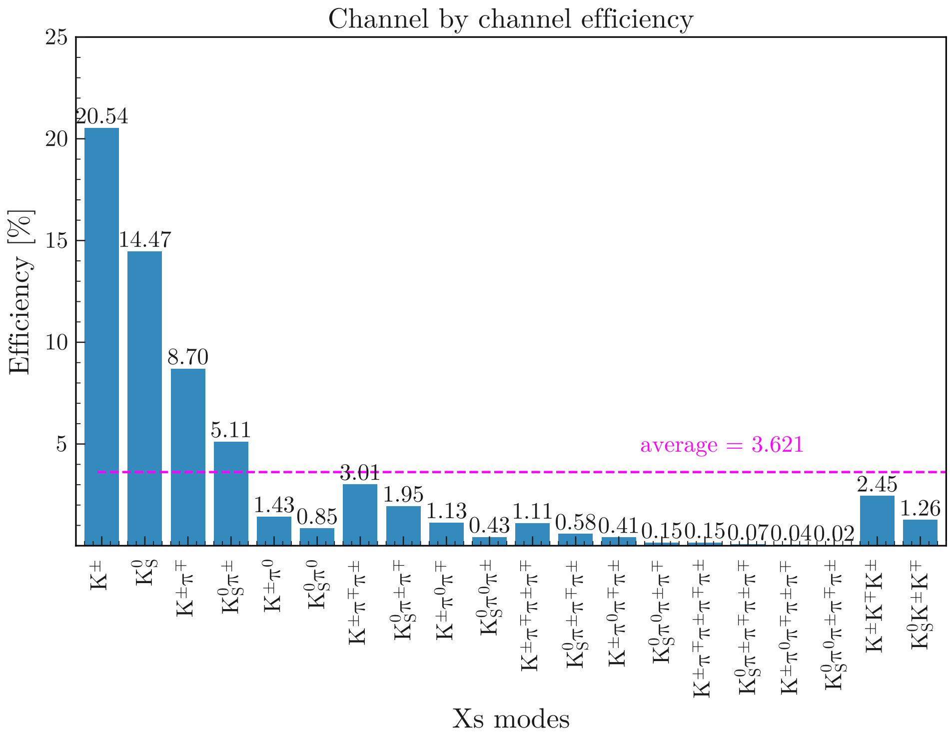


Figure 6.27: Reconstruction efficiency (in %) of each mode after event selection.

change the number of signal event, but could reject correctly reconstructed candidates accidentally.

The number of signal and background, as well as the signal-to-noise ratio and FoM estimated from run-dependent samples, are summarized in Tabs. 6.10 and 6.11 for muon and electron channels, respectively. For q^2 binning, the q^2 reconstructed without any bremsstrahlung correction (labeled as $q^2_{\rm raw}$) is used in order to avoid including any Final State Radiation (FSR) photons, such that we can make a fair comparison with theoretical predictions. For the same reason, the truth variables $m^{\rm true}(X_s)$ and $q^2_{\rm true}$ are calculated without FSR.

To derive the differential branching fractions using Eq. 3.1, signal unfolding matrices (Eq. 3.3) are required. We derived them using signal MC samples and they are shown in Figs. 6.28 and 6.29. The unfolding matrix of $m(X_s)$ binning is almost diagonal due to the good reconstruction of the hadron track momentum and the photon energy of the detector. For the unfolding matrix of q^2, a significant feed-down from higher q^2 bin to lower q^2 bins is observed in the electron channel due to bremsstrahlung.

Table 6.8: Cut flow table for muon channel at 365 fb^{-1}. The N_{sig}, $N_{B\bar{B}}$, and $N_{q\bar{q}}$ at generator level represents the number of event. In the rest of the rows, N_{sig} represents the number of correctly reconstructed B candidates; $N_{B\bar{B}}$ and $N_{q\bar{q}}$ represent the number of remaining B candidates in the $B\overline{B}$ and $q\overline{q}$ generic MC samples, respectively.

	N_{sig}	$N_{B\bar{B}}$	$N_{q\bar{q}}$	Efficiency [%]	$S/\sqrt{S+B}$
Generator	2702	3.83×10^{8}	1.33×10^{9}	100	0.0653
Reconstruction	281	2.14×10^{6}	5.48×10^{5}	10.4	0.172
$\mathrm{chiProb}_B > 0$	275	1.65×10^{6}	3.94×10^{5}	10.2	0.192
$M(X_s) < 1.8$ [GeV/c^2]	260	6.32×10^{5}	1.63×10^{5}	9.61	0.291
ΔE	242	2.33×10^{5}	5.43×10^{4}	8.97	0.452
$M_{\ell\ell} > 0.2$ [GeV/c^2]	242	2.33×10^{5}	5.43×10^{4}	8.97	0.452
Charmonium veto	186	6.53×10^{4}	4.31×10^{4}	6.89	0.565
D veto	174	5.05×10^{4}	3.76×10^{4}	6.45	0.586
MVA	100	679	340	3.72	3.00
BCS	93.5	651	326	3.46	2.86

Table 6.9: Cut flow table for electron channel at 365 fb^{-1}. The N_{sig}, $N_{B\bar{B}}$, and $N_{q\bar{q}}$ at generator level represents the number of event. In the rest of the rows, N_{sig} represents the number of correctly reconstructed B candidates; $N_{B\bar{B}}$ and $N_{q\bar{q}}$ represent the number of remaining B candidates in the $B\overline{B}$ and $q\overline{q}$ generic MC samples, respectively.

	N_{sig}	$N_{B\bar{B}}$	$N_{q\bar{q}}$	Efficiency [%]	$S/\sqrt{S+B}$
Generator	2715	3.83×10^{8}	1.33×10^{9}	100	0.0656
Reconstruction	396	1.91×10^{6}	4.65×10^{5}	14.6	0.257
$\mathrm{chiProb}_B > 0$	386	1.48×10^{6}	2.98×10^{5}	14.2	0.289
$M(X_s) < 1.8$ [GeV/c^2]	356	5.57×10^{5}	1.16×10^{5}	13.1	0.434
ΔE	336	3.1×10^{5}	6×10^{4}	12.4	0.551
$M_{\ell\ell} > 0.2$ [GeV/c^2]	293	3.09×10^{5}	3.16×10^{4}	10.8	0.501
Charmonium veto	191	6.2×10^{4}	2.03×10^{4}	7.02	0.664
D veto	178	4.99×10^{4}	1.8×10^{4}	6.56	0.682
MVA	109	501	173	4.03	3.91
BCS	103	467	171	3.78	3.77

Table 6.10: Expected signal and background numbers in $M_{bc} > 5.26$ GeV/c^2 region for muon channel at $365\,\text{fb}^{-1}$.

	N_{sig}	N_{bkg}	N_{sig}/N_{bkg}	$N_{sig}/\sqrt{N_{sig}+N_{bkg}}$
$0.04 < q^2_{raw} < 1.0$ [(GeV/c)^2]	7.2	40.2	0.18	1.0
$1.0 < q^2_{raw} < 6.0$ [(GeV/c)^2]	29.2	103.7	0.28	2.5
$6.0 < q^2_{raw} < 14.4$ [(GeV/c)^2]	34.1	110.1	0.31	2.8
$14.4 < q^2_{raw} < 25.0$ [(GeV/c)^2]	23.0	21.2	1.1	3.5
$1 < q^2_{raw} < 7.3$ [(GeV/c)^2]	39.3	133.0	0.3	3.0
$9.6 < q^2_{raw} < 13.6$ [(GeV/c)^2]	17.1	47.9	0.36	2.1
$0.4 < m(X_s)_{raw} < 0.6$ [GeV/c^2]	45.0	23.0	2.0	5.5
$0.6 < m(X_s)_{raw} < 1.0$ [GeV/c^2]	25.8	27.5	0.94	3.5
$1.0 < m(X_s)_{raw} < 1.4$ [GeV/c^2]	13.5	69.0	0.2	1.5
$1.4 < m(X_s)_{raw} < 1.8$ [GeV/c^2]	9.2	155.7	0.059	0.72

Table 6.11: Expected signal and background numbers in $M_{bc} > 5.26$ GeV/c^2 region for electron channel at $365\,\text{fb}^{-1}$.

	N_{sig}	N_{bkg}	N_{sig}/N_{bkg}	$N_{sig}/\sqrt{N_{sig}+N_{bkg}}$
$0.04 < q^2_{raw} < 1.0$ [(GeV/c)^2]	18.6	42.0	0.44	2.4
$1.0 < q^2_{raw} < 6.0$ [(GeV/c)^2]	45.4	75.1	0.61	4.1
$6.0 < q^2_{raw} < 14.4$ [(GeV/c)^2]	17.0	55.3	0.31	2.0
$14.4 < q^2_{raw} < 25.0$ [(GeV/c)^2]	21.3	13.6	1.6	3.6
$1 < q^2_{raw} < 7.3$ [(GeV/c)^2]	54.4	113.2	0.48	4.2
$9.6 < q^2_{raw} < 13.6$ [(GeV/c)^2]	7.6	16.9	0.45	1.5
$0.4 < m(X_s)_{raw} < 0.6$ [GeV/c^2]	49.9	27.9	1.8	5.7
$0.6 < m(X_s)_{raw} < 1.0$ [GeV/c^2]	27.9	24.8	1.1	3.8
$1.0 < m(X_s)_{raw} < 1.4$ [GeV/c^2]	14.8	48.4	0.3	1.9
$1.4 < m(X_s)_{raw} < 1.8$ [GeV/c^2]	10.0	85.0	0.12	1.0

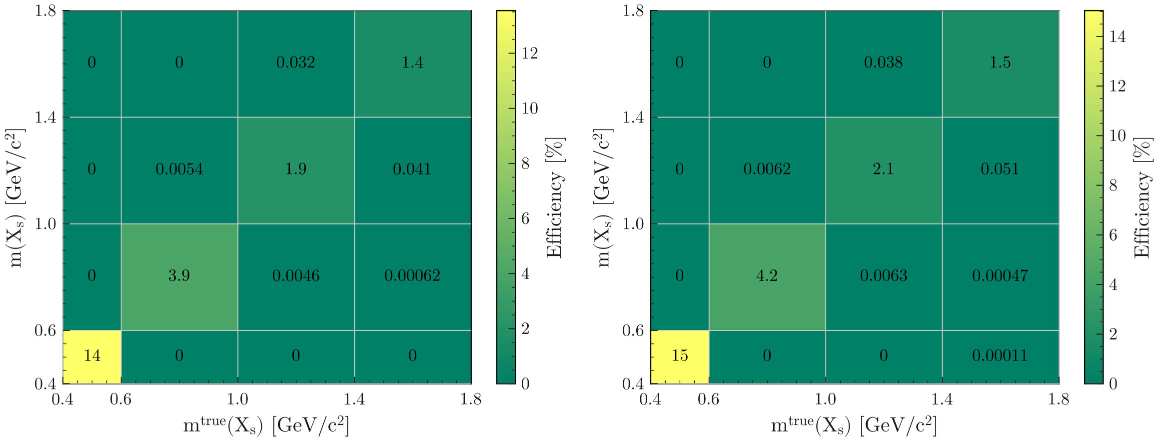


Figure 6.28: Signal unfolding matrix in $m(X_s)$ bins. Horizontal axes represents the true value while the vertical axes is the reconstructed value in signal MC samples. The left plot is for muon channel while the right plot is for the electron channel.

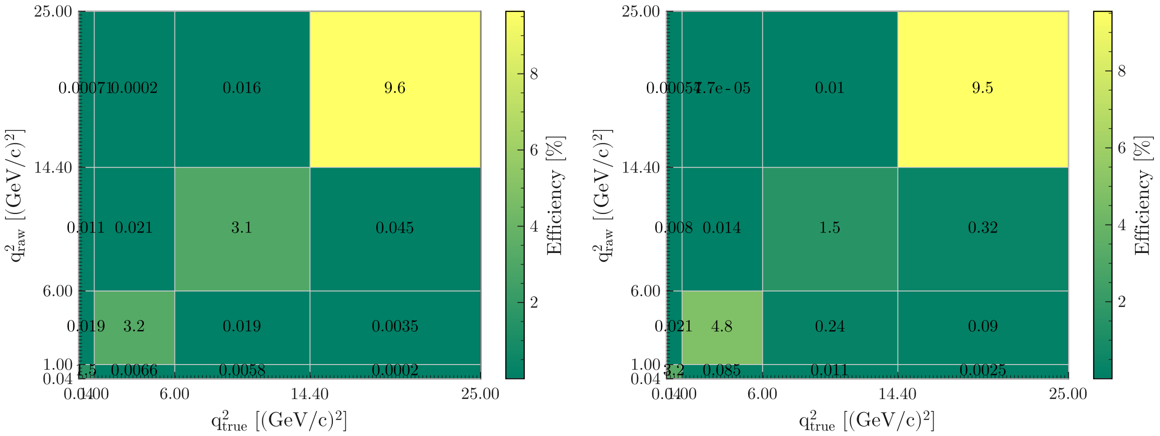


Figure 6.29: Signal unfolding matrix in q^2 bins. Horizontal axes represents the true value while the vertical axes is the reconstructed value without bremsstrahlung correction in signal MC samples. The left plot is for muon channel while the right plot is for the electron channel.

Chapter 7

Peaking Background Estimation

In this analysis, the signal yield uncertainties are extracted by fitting to the M_{bc} spectrum of the B meson. And for this reason, a good understanding of background events that has a Gaussian distribution the same as signal events (called peaking background) is of vital importance. By analyzing the generic MC after selection, we learned that there remain three types of peaking background:

- Double mis-identification background (will be discussed in Sec. 7.1).
- Swap mis-identification background (will be discussed in Sec. 7.2).
- Charmonium background (will be discussed in Sec. 7.3).

Another possible peaking background is from the $b \to d\ell^+\ell^-$ transition. However, it is suppressed by $|V_{td}/V_{ts}|^2 \approx 0.05$ and requiring one kaon in the reconstruction makes the expected yield negligibly small ($O(0.1\%)$) comparing to the signal process.

7.1 Double mis-identification background

Hadronic B decays $B \to Xh^+h^-$ ($h = K, \pi$) can mimic signal components when two hadrons are mis-identified as two leptons. A pure data-driven method is applied to estimate contributions from such a process by analyzing the PID sideband. The general idea behind this is that the number of hadrons surviving the leptonID selections (called in signal region) N_{SR} can be estimated by the number of hadrons not surviving the same selections (called in sideband region) N_{SB}, and they are connected by:

$$N_{SR} = \frac{f_{h^+\to\ell^+}}{\varepsilon_{h^+}} \cdot \frac{f_{h^-\to\ell^-}}{\varepsilon_{h^-}} N_{SB}. \tag{7.1}$$

Here the $\varepsilon_{h^\pm}$ is the fraction of hadrons in the sideband region and $f_{h^\pm\to\ell^\pm}$ is the fraction of hadrons in the signal region (or we call it the fake rate of $h^\pm$ to $\ell^\pm$), which are estimated in a data-driven method using reference channel $D^{*+} \to D^0(\to K^-\pi^+)\pi_s^+$ whose procedure is described in Appendix. G.4 and Ref. [56].

In this analysis, we define the signal region (SR) as muonID > 0.99 for muon channels and $e\mathrm{ID}_{\mathrm{BDT}} > 0.9$ for electron channels, and we define the PID sideband (SB) requiring muonID < 0.99 and $e\mathrm{ID}_{\mathrm{BDT}} < 0.9$ for both muon and electron channels. This is to avoid cross-feed between lepton channels in SB (e.g. signals from electron channels also appear in the muonID < 0.99 region). An illustration of the definition of SR and SB is shown in Fig. 7.1.

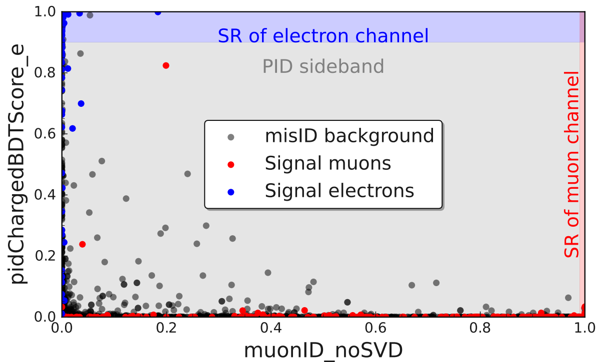


Figure 7.1: Illustration of signal region and PID sideband. Horizontal axis is muonID, and vertical axis is $e\mathrm{ID}_{\mathrm{BDT}}$.

Our target is to derive both the number and the shape of M_{bc} spectrum of double mis-ID events such that their contributions can be subtracted during signal extraction procedure by fitting. To do this, we reconstruct SB events with the same selection criteria as signal events other than PID of leptons. A weight is applied to each candidate of the SB sample:

$$w = \frac{f_{h^+\to\ell^+}}{\varepsilon_{h^+}} \cdot \frac{f_{h^-\to\ell^-}}{\varepsilon_{h^-}} \tag{7.2}$$

Both $f_{h^\pm\to\ell^\pm}$ and $\varepsilon_{h^\pm}$ are functions of the track momenta, polar angle θ, and track charge. Fig. 7.2 shows the M_{bc} spectrum of SB events before (in black histogram) and after (in red histogram) applying the weight in Eq. 7.2. By fitting these distributions after applying the weight with a Gaussian (peaking part) and Argus function (combinatorial part), we derive the number of peaking background is 9.3 ± 0.4 and 0.5 ± 0.1 for muon and electron channels, respectively. Here the uncertainties are dominated by uncertainties of ε_h and $f_{h\to\ell}$.

Table 7.1: Remaining double mis-ID events in SR and estimated double mis-ID events from SB in generic MC sample. Uncertainties of the SR are statistical only, while uncertainties of SB estimation comes from ε_h and $f_{h\to\ell}$.

	Muon channel	Electron channel
Events in signal region	9.8 ± 1.6	0.25 ± 0.25
Events estimated from sideband region	11.2 ± 0.5	0.05 ± 0.02

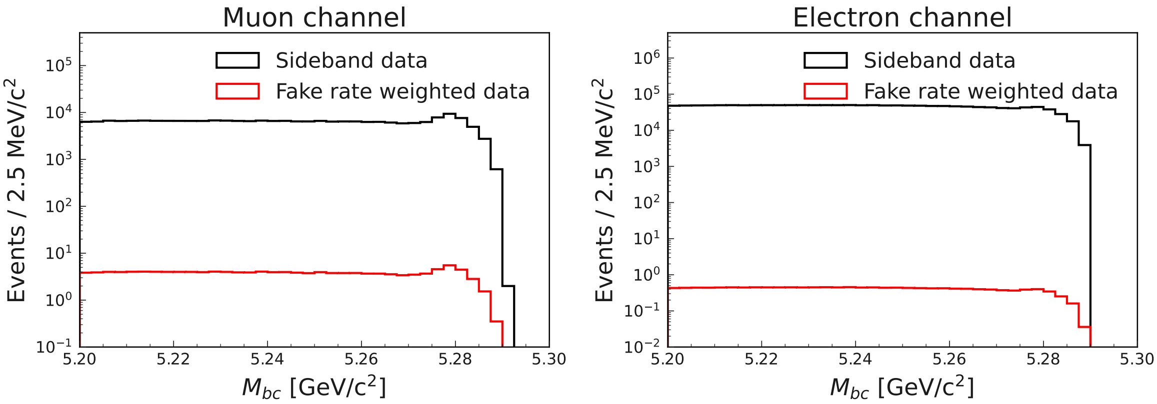


Figure 7.2: M_{bc} distribution of PID SB events before (black) and after (red) applying the weight in Eq. 7.2. The vertical axis is in log scale. The left plot is for muon channel, and the right plot is for electron channel.

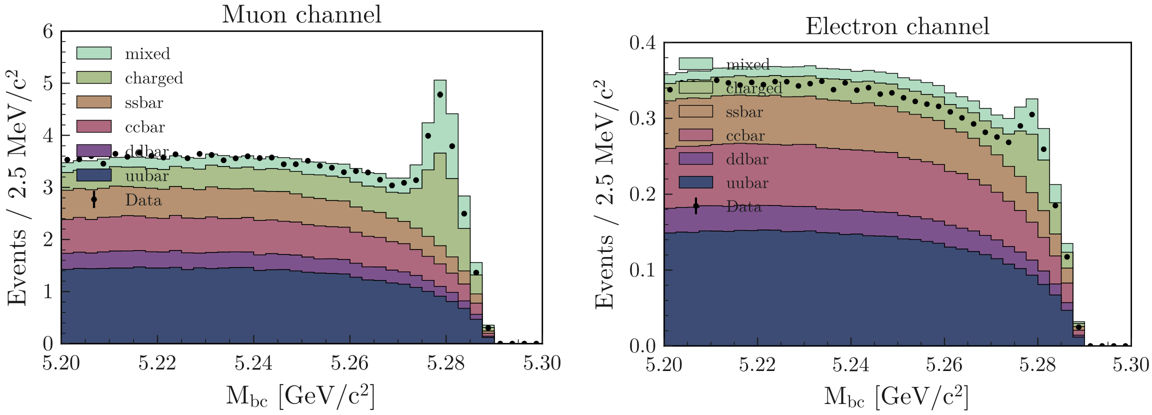


Figure 7.3: Comparison of weighted M_{bc} distribution of PID sideband in data and MC. PID sideband efficiency and fake rates measured in real data is applied to both data and MC samples in this plot. There is a discrepancy in normalization and we interpret it as improper configurations of the branching fractions of the decay of B meson in the generation of generic MC.

We validate this method using simulation samples and the results are shown in Tab. 7.1. For the muon channel, the number of events estimated from SB region is consistent with the number of events in SR region. For electrons, both the number of events estimated from SB region and the number of events in SR region are zero-consistent. Figure 7.3 shows the weighted M_{bc} distribution in PID sideband. We find the shape of the estimated M_{bc} distributions between data and MC are similar to each other, while there are minor discrepancies and we interpret them as improper configurations of the branching fractions of the decay of B meson in the generation of generic MC. Because of this, we use the peaking background shape derived from data directly and this discrepancy has no influence on the measurement.

In addition, we performed several checks to make sure that the peak in the PID sideband is dominated by double mis-ID background. Fig. 7.4 shows that SB candidates, especially the peak at $M_{bc} = 5.28\,\mathrm{GeV}/c^2$, are mostly from mis-identification of hadrons ($K,\ \pi$) to leptons utilizing the MC truth information. It also shows that most of the double misidentification background is from $\pi^+\pi^-$, with the number of events being 8.7 ± 0.4, as expected. $K^{\pm}\pi^{\mp}$ and K^+K^- contributes 1.0 ± 0.1 and 0.20 ± 0.01 events, respectively. We also checked that the remaining signal events in the SB are negligible: 1.4 and 2.8 for the muon and electron channels, respectively.

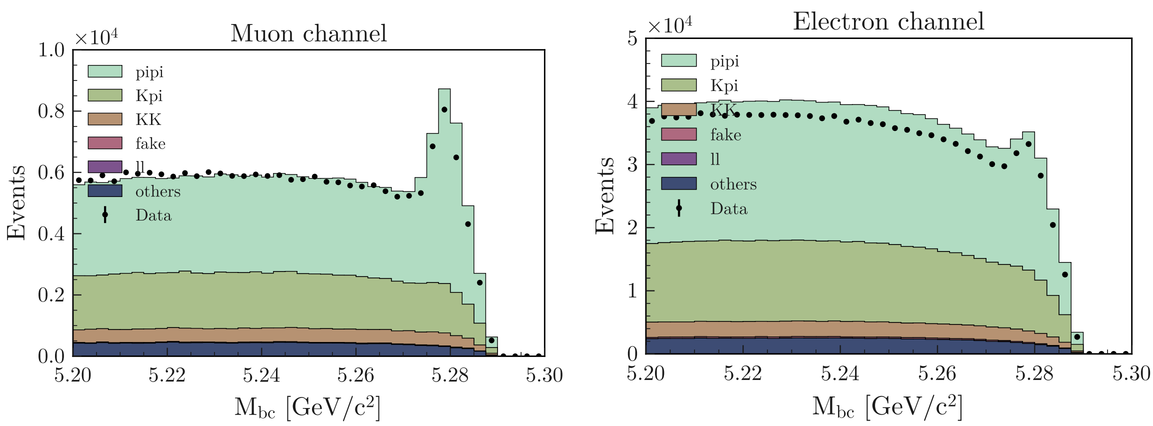


Figure 7.4: Stacks of different misID categories of SB candidates. These plots indicates that the peak at $M_{bc} = 5.28\,\mathrm{GeV}/c^2$ mainly comes from mis-identification of hadrons ($K,\ \pi$) to leptons.

7.2 Swapped mis-identification background

The charmonium process $B \to X_s\psi(\to \ell^+\ell^-)$ can escape from the charmonium veto by misidentifying one of leptons as a hadron, and one of hadrons as a lepton. This kind of background is called swap mis-ID background. Similarly to the double mis-ID background, the swap mis-ID is also estimated using a pure data-driven method. We first reconstruct events inside the charmonium veto, applying exactly the same event selection criteria as we reconstruct the signal process. Then we rerun the reconstruction and selection procedure, swapping one lepton track with one hadron track, adjusting the mass hypothesis accordingly. All relevant kinematic variables including $m(\ell^+\ell^-)$, $\Delta z_{\ell^+\ell^-}$, ΔE, chiProb of di-lepton vertex fitting, and the FTTransformer output are recalculated and candidates that pass the selections are called swapped events. We weight these swapped events by:

$$w = \frac{f_{\ell\to h}}{\varepsilon_\ell} \cdot \frac{f_{h\to\ell}}{\varepsilon_h} \tag{7.3}$$

where $f_{\ell\to h(h\to\ell)}$ is the mis-ID rate from lepton to hadron (hadron to lepton), and $\varepsilon_{\ell(h)}$ is the PID efficiency of lepton (hadron). Similarly, ε_h and $f_{h\to\ell}$ as functions of track momenta, polar angle, and charge are calculated from the $D^{*+} \to D^0(\to K^-\pi^+)\pi^+$ channel as described in Ref. [56]. The ε_ℓ is calculated by data-driven method using the radiative Bhabha $e^+e^- \to \ell^+\ell^-\gamma$, two-photon process $e^+e^- \to e^+e^-\ell^+\ell^-$, and $J/\psi \to \ell^+\ell^-$ processes [56]. $f_{\mu\to h}$ is calculated by the author in the $e^+e^- \to \mu^+\mu^-\gamma$ process using the tag and probe method described in the Appendix. D. $f_{e\to h}$ is derived from an analysis in the two-photon process $e^+e^- \to e^+e^-e^+e^-$ [56]. The M_{bc} distributions of swapped events in data before (in black histogram) and after (in red histogram) applying the weight in Eq. 7.3 are presented in Fig. 7.5.

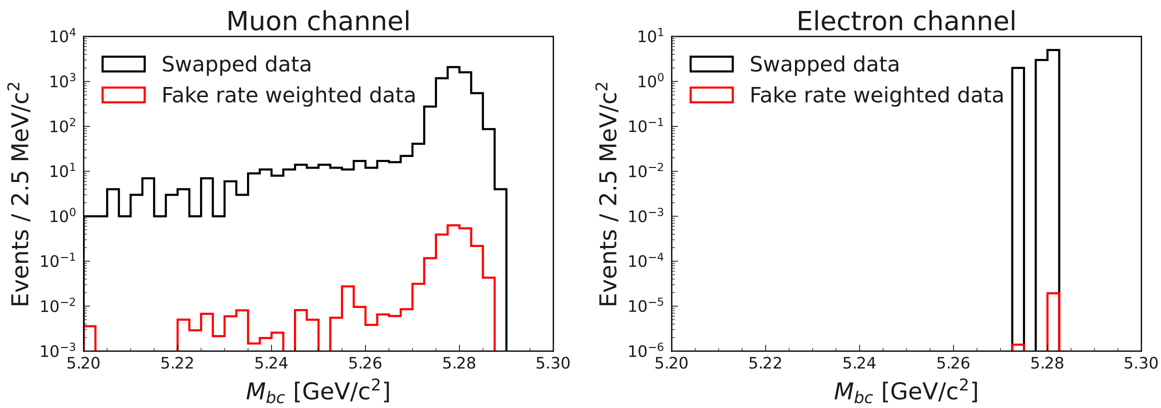


Figure 7.5: M_{bc} distribution of swapped events before (black) and after (red) applying the weight in Eq. 7.3. The vertical axis is in log scale. The left plot is for muon channel, and the right plot is for electron channel.

We validate this method using simulation samples by comparing the remaining swap mis-ID events in the standard event selection procedure and the estimated number from charmonium vetoed samples, with the results shown in Tab. 7.2. Estimation from real data is also shown in the table. The results for the muon channel are in good consistency between the simulation samples. For the electron channel, the contribution is almost negligible, so we do not consider the swap mis-ID in electrons in the fitting procedure. Fig. 7.6 shows the comparison of the weighted distribution M_{bc} in swapped samples between the data and simulation.

Table 7.2: Comparison of swap mis-ID events remained after the standard selection and the estimated number using charmonium events in simulation and data.

	Muon channel	Electron channel
Remaining (Simulation)	2.90 ± 0.80	0 ± 0
Swapping estimation (Simulation)	2.67 ± 0.16	$(8 \pm 5) \times 10^{-5}$
Swapping estimation (Data)	2.29 ± 0.17	$(2 \pm 4) \times 10^{-5}$

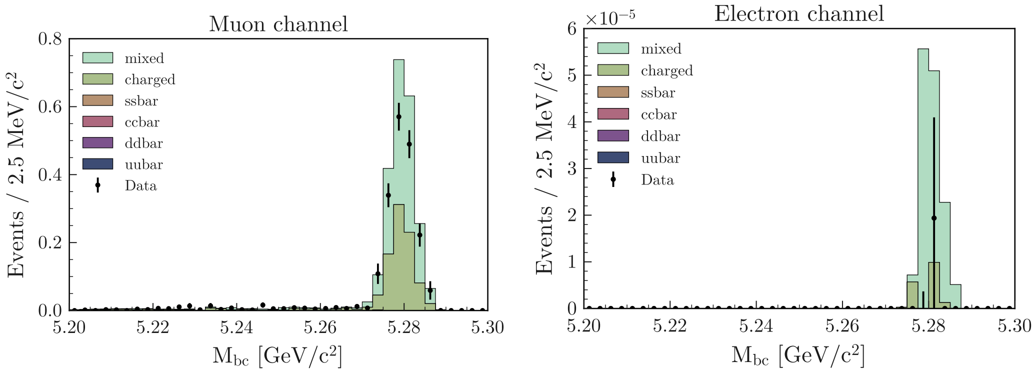


Figure 7.6: Weighted M_{bc} distribution of swap mis-ID samples. Contributions to electron channel is negligible. In this plot, the PID fake rates in data are applied to both data and MC samples.

7.3 Charmonium background

Even if a charmonium veto is applied in the signal reconstruction, some charmonium events can escape from the veto window due to detector resolution and bremsstrahlung in the electron case. In this analysis, the signal shapes of this background are derived from the signal MC of $B \to X_s J/\psi(\to \ell^+\ell^-)$ and $B \to X_s\psi(2S)(\to \ell^+\ell^-)$ processes.

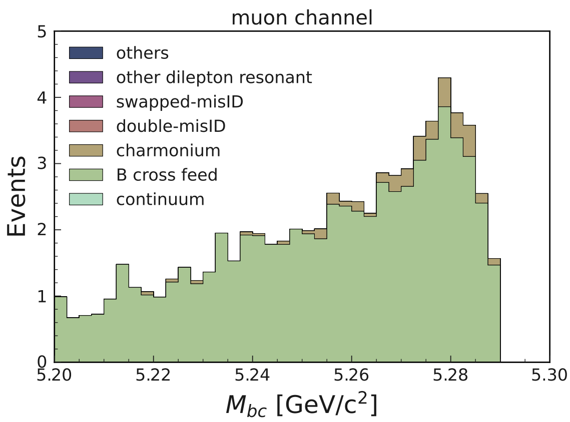
For $B \to X_s J/\psi(\to \ell^+\ell^-)$, a signal MC with a sample size of 20M events is used. This sample is generated by summing each exclusive decay of $B \to K_J J/\psi$ and feed-down from higher charmonium resonances such as $\psi(2S)$ and χ_c are also taken into account, with branching fractions of each individual decay channel taken from PDG. This sample is weighted according to the expected event production with $\mathcal{B}(B \to J/\psi X) = (1.057 \pm 0.042)\%$ measured by the BABAR experiment [83]. For $B \to X_s\psi(2S)(\to \ell^+\ell^-)$, we generated a signal MC sample of one million events. The decay channels considered are listed in Tab. 7.3 with most of the branching fractions taken from PDG [84]. The $B \to \psi(2S)K\pi^0\pi^0$, $B \to \psi(2S)K\pi^+\pi^0$, and $B \to \psi(2S)K_1(1270)$ channels are not available in PDG and their branching fractions are taken from the configuration file of the generic MC assigning 100% uncertainties. In total, the branching fractions of $B^+ \to \psi(2S)X^+$ and $B^0 \to \psi(2S)X^0$ are $(3.18 \pm 0.47) \times 10^{-3}$ and $(2.57 \pm 0.44) \times 10^{-3}$, respectively.

Table 7.3: Branching fractions of $B \to \psi(2S)X$ decays considered in this analysis.

$B \to \psi(2S)X$	$\mathcal{B}(\times 10^4)$ of charged/neutral X
K^+/K^0	$(6.24 \pm 0.21)/(5.8 \pm 0.5)$
K^{*+}/K^{*0}	$(6.7 \pm 1.4)/(5.9 \pm 0.4)$
$K^0\pi^+/K^+\pi^-$	$(1.43 \pm 1.00)/(1.87 \pm 0.5)$
$K^+\pi^0/K^0\pi^0$	$(0.72 \pm 0.50)/(0.93 \pm 0.25)$
$K^+\pi^+\pi^-/K^0\pi^+\pi^-$	$(4.3 \pm 0.5)/(2.81 \pm 0.3)$
$K^+\pi^0\pi^0/K^0\pi^0\pi^0$	$(1 \pm 1)/(1 \pm 1)$
$K^0\pi^+\pi^0/K^+\pi^-\pi^0$	$(1 \pm 1)/(1 \pm 1)$
K_1^+/K_1^0	$(4 \pm 4)/(4 \pm 4)$
π^+/π^0	$(0.244 \pm 0.03)/(0.117 \pm 0.019)$
$\phi K^+/\pi^+\pi^-$	$(0.04 \pm 0.007)/(0.224 \pm 0.035)$
Total	$(31.8 \pm 4.7)/(25.7 \pm 4.4)$

The number of remaining charmonium processes after signal reconstruction and selection are 3.0 ± 0.4 and 5.9 ± 0.6 in muon and electron channels, respectively. Uncertainties from MC statistics and branching fractions are taken into consideration.

Besides the events that are identified as charmonium from MC truth information, cross-feeds by combining a soft pion from the other B meson could also contribute to a small peak in the M_{bc} distributions as shown in Fig. 7.7. To take such events into consideration, we use events to produce the probability density function of charmonium background as long as it has two leptons coming from the same J/ψ or $\psi(2S)$ in the charmonium signal MC.



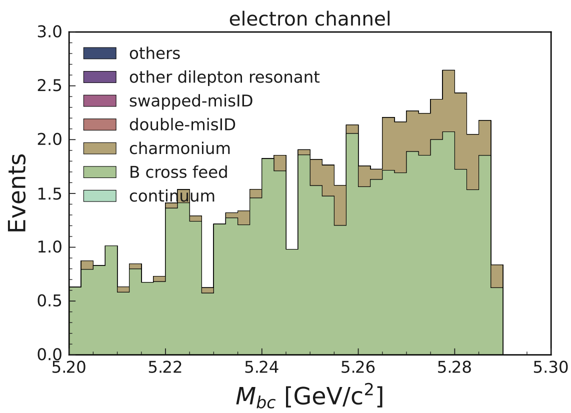


Figure 7.7: Remaining charmonium events after event selection in charmonium signal MC.

Chapter 8

Signal Extraction

8.1 Fitting strategy

As explained in Eq. 3.1, in this analysis, we extract the signal yield uncertainties (statistical uncertainties) in each kinematic bin for the muon and electron channels separately. This is performed using an extended unbinned maximum likelihood fitting method. The likelihood is defined as following:

$$\mathcal{L} = \frac{\exp\left(-\sum_j N_j\right)}{N!} \prod_{i=1}^{N} \left(\sum_j N_j P_j(M_{bc}^i,\ E_{\text{beam}}^i) \right) \tag{8.1}$$

where N_j is the number of events in category j, which runs over six components of signal, self corss-feed (SxF), combinatorial background, and three kinds of peaking backgrounds. $P_j(M_{bc}, E_{\text{beam}})$ is the probability function of category j as a function of M_{bc} and E_{beam}. E_{beam} is defined as half of the center-of-mass energy $\sqrt{s}$, which may differ slightly for data taken at different time periods.

During the fitting procedure, the center values of the yields are blinded to the analyst while the uncertainties are exposed.

8.2 Probability Density Function (PDF)

8.2.1 Signal

The signal component is described with a signal Gaussian:

$$P_{\text{sig}}(M_{bc}|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(M_{bc} - \mu)^2}{2\sigma^2} \right) \tag{8.2}$$

where μ, σ is the center value and standard deviation of Gaussian. Since in some kinematic bins the signal-to-noise ratio is not good enough, we fix (μ, σ) to the fitting result in the $B \to X_s J/\psi$ data as shown in Fig. 8.1. In this fitting, the background is described with an Argus function. The fitting result is summarized in Tab. 8.1. Since these numbers are used universally to fit each kinematic bin, we also confirmed that (μ, σ) is consistent across different kinematic bins in the simulated signal samples, as shown in Fig. 8.2.

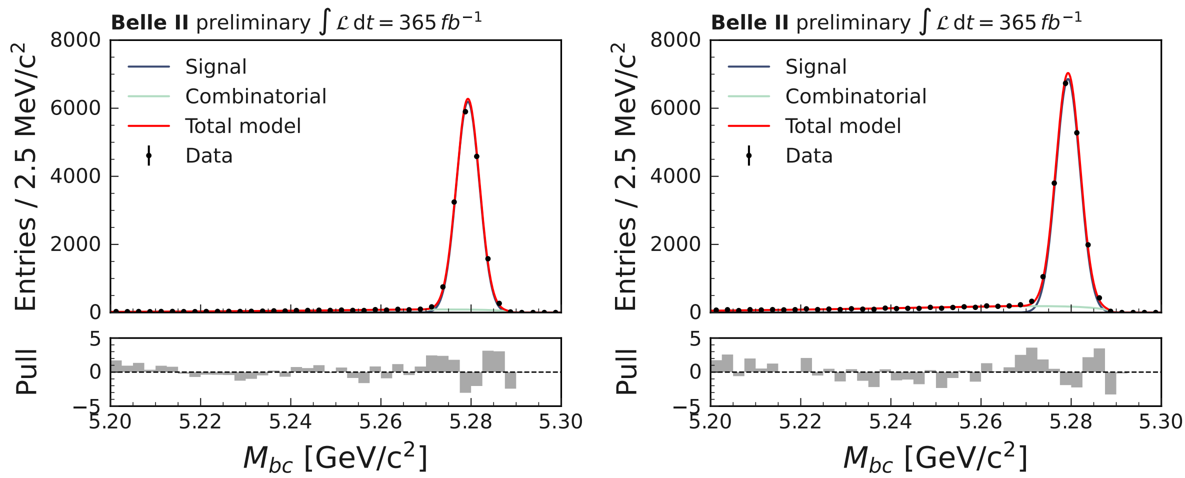


Figure 8.1: Fitting result to $B \to X_s\psi(\to \ell^+\ell^-)$ data. Left for muon channel while right for electron channel.

Table 8.1: Signal shape parameters fitted from $B \to X_s J/\psi$ data samples.

	μ[GeV/c^2]	σ [MeV]
Muon channel	5.27930 ± 0.00002	2.567 ± 0.015
Electron channel	5.27931 ± 0.00002	2.684 ± 0.017

8.2.2 Self cross-feed

Self cross-feed (SxF) events come from partially reconstructed signal events or combining a soft particle from the other side of the event with the signal B meson. For this reason, the yield of SxF events should be proportional to signal events. In this analysis, the shape and ratio to the signal yield of the SxF events are fixed from the signal MC, as shown in Fig. F.1 and Fig. F.2 of Appendix. F, where the shape is smoothened by fitting to a Gaussian and Argus function. Also, it should be noted that by fixing the N_{SxF}/N_{sig} to MC values, we are, in fact, considering SxF as part of the signal templates.

8.2.3 Combinatorial background

The combinatorial background is described with an Argus function:

$$P_{\text{comb}}(M_{bc}, E_{\text{beam}}|\chi) = \frac{\chi^3}{\sqrt{2\pi}\Psi(\chi)} x\sqrt{1-x^2} \exp\left(-\frac{\chi^2}{2}(1-x^2)\right), x = \frac{M_{bc}}{E_{\text{beam}}}, \tag{8.3}$$

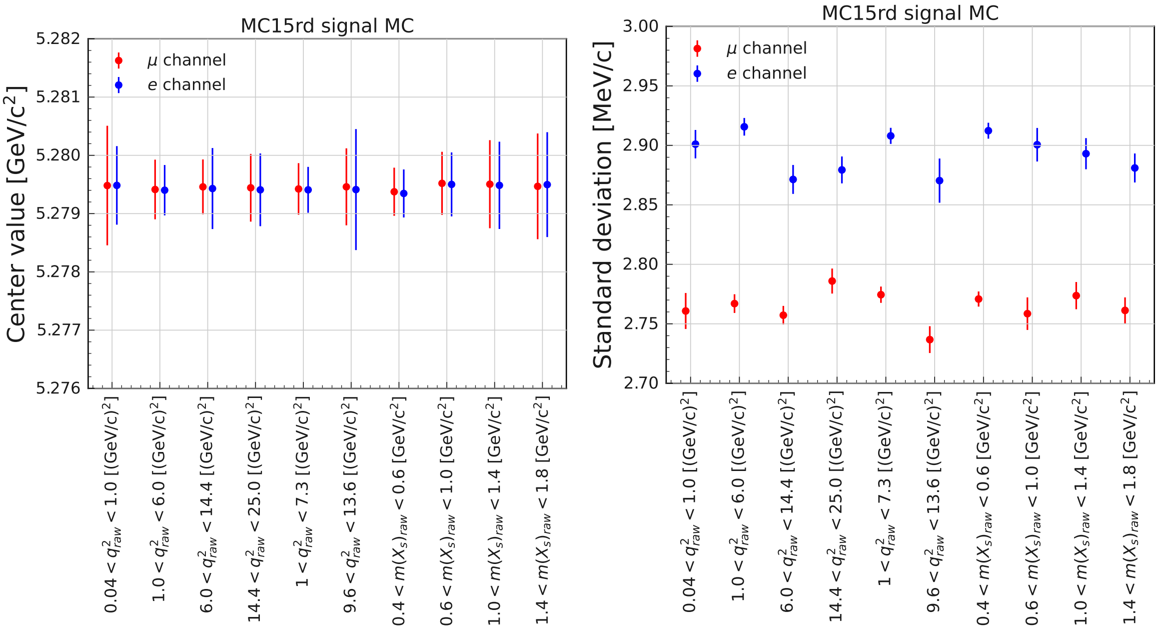


Figure 8.2: Center value and standard deviation of signal M_{bc} distributions in each kinematic bin of $B \to X_s\ell^+\ell^-$ signal MC.

where

$$\Psi(\chi) = \Phi(\chi) - \chi\phi(\chi) - 1/2 \tag{8.4}$$

with Φ and ϕ being the cumulative distribution function and the probability density function of a standard Gaussian, respectively. During fitting, the shape parameter χ is left to be free.

8.2.4 Peaking background

In this analysis, three types of peaking background are considered as explained in Chap. 7. Figures. F.3 and F.4 and Fig. F.5 in Appendix. F show the double mis-ID and the swap mis-ID background templates, respectively. Uncertainties from the PID fake rate and efficiency are reflected in the variation of the PDFs. For the swap mis-ID background, only contributions to muon channels are taken into consideration. The PDFs of charmonium background are shown in Figs. F.6 and F.7 of Appendix. F, with uncertainties from the MC statistics and branch fractions of the decays $B \to \psi X$ taken into account. We smoothen the histogram PDFs by fitting to a Gaussian and Argus function for double mis-ID and charmonium background. For swap mis-ID, it is fitted with a single Gaussian. Both the shape and the yield of three peaking background components are fixed during fitting. Systematic uncertainties of the peaking backgrounds are estimated by varying the PDFs to their upper and lower variations, as shown in figures mentioned above.

8.3 Fitter check

The pull distribution of the fitter output is checked using toy MC samples. The toy MC of each component mentioned above is derived by filling the corresponding sample into a four-dimensional histogram of $(M_{bc}, E_{\mathrm{beam}}, m(X_s)_{\mathrm{raw}}, q^2_{\mathrm{raw}})$ and the number of events in each experiment is determined by a Poisson random number around the expected value. In addition, the M_{bc} of the signal components is replaced by Gaussian random numbers with center value and standard deviation taken from Tab. 8.1. The M_{bc} of the combinatorial background components is replaced by an Argus random number with a shape parameter derived from fitting to the generic MC sample. This is to avoid possible peaking components, such as $B \to K\pi(\ell)\pi(\ell)$ in the generic MC that are not fully excluded by the truth information. In total, 1000 experiments were generated. Example fits to the generated toy samples are shown in Figs. 8.3 and 8.4.

The pull of the fitted yield N_{output} is given by:

$$pull = \frac{N_{output} - N_{input}}{\sigma_{N_{output}}}, \tag{8.5}$$

where N_{input} is the expected signal yield in the corresponding kinematic bin and $\sigma_{N_{output}}$ is the fitting uncertainty of N_{output}. After that, the pull distribution is fitted to a Gaussian and the results are shown in Figs. I.1 and I.2 in Appendix. I. Overall, the pull distributions are in good consistency with standard Gaussian, except for several bins. Systematic uncertainties are assigned to the bins with fitter bias. To understand the source of these biases, we performed another pull test by multiplying the luminosity by a factor of ten and the bias disappeared. This indicates that the fitter bias comes from the low statistics.

We also performed the linearity test using toy MC samples, with an expected signal yield ranging from 50% to 500% of SM predictions. The results are shown in Figs. I.3 and I.4 in Appendix. I. Good linearity is confirmed without significant bias.

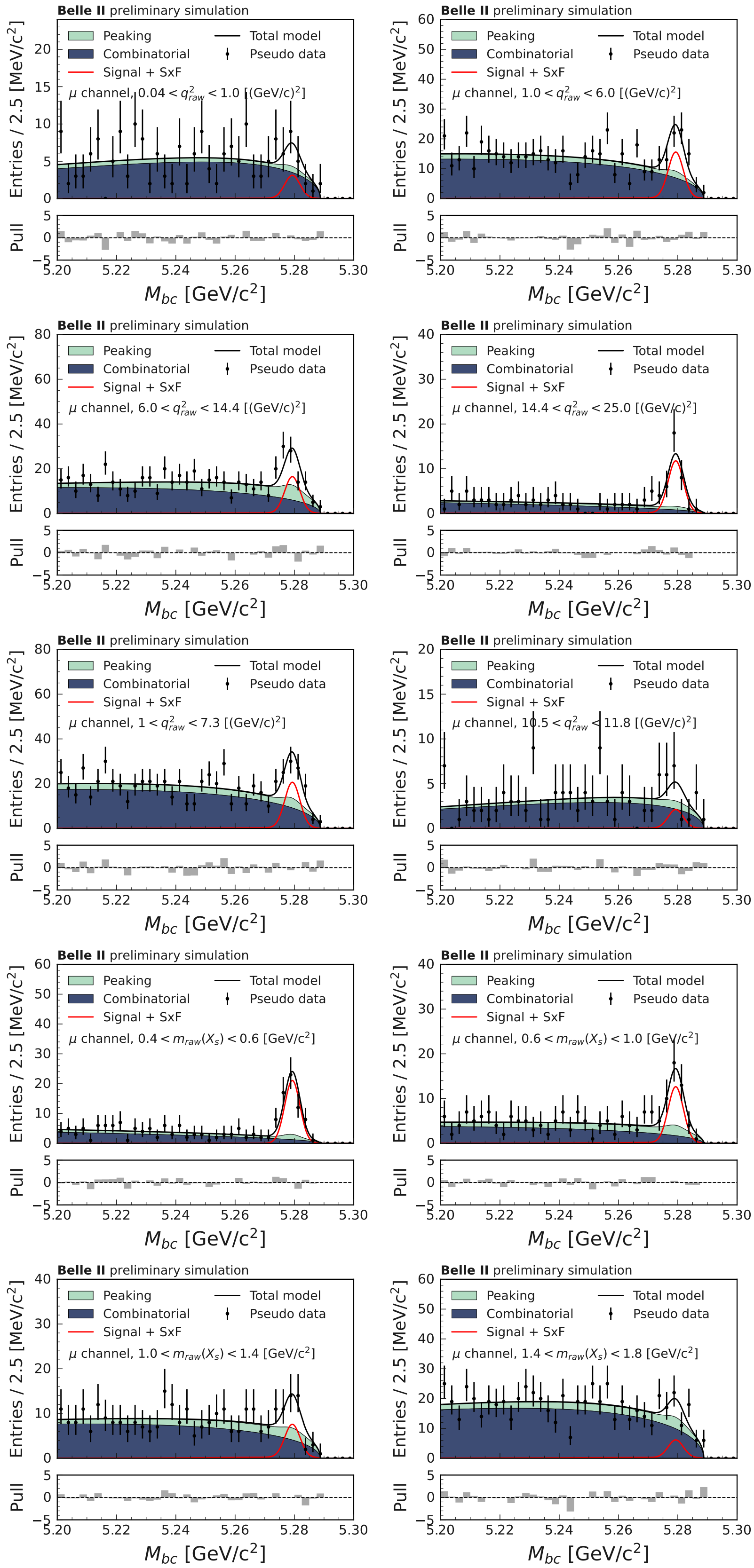


Figure 8.3: Example fitting to toy MC samples in muon channels.

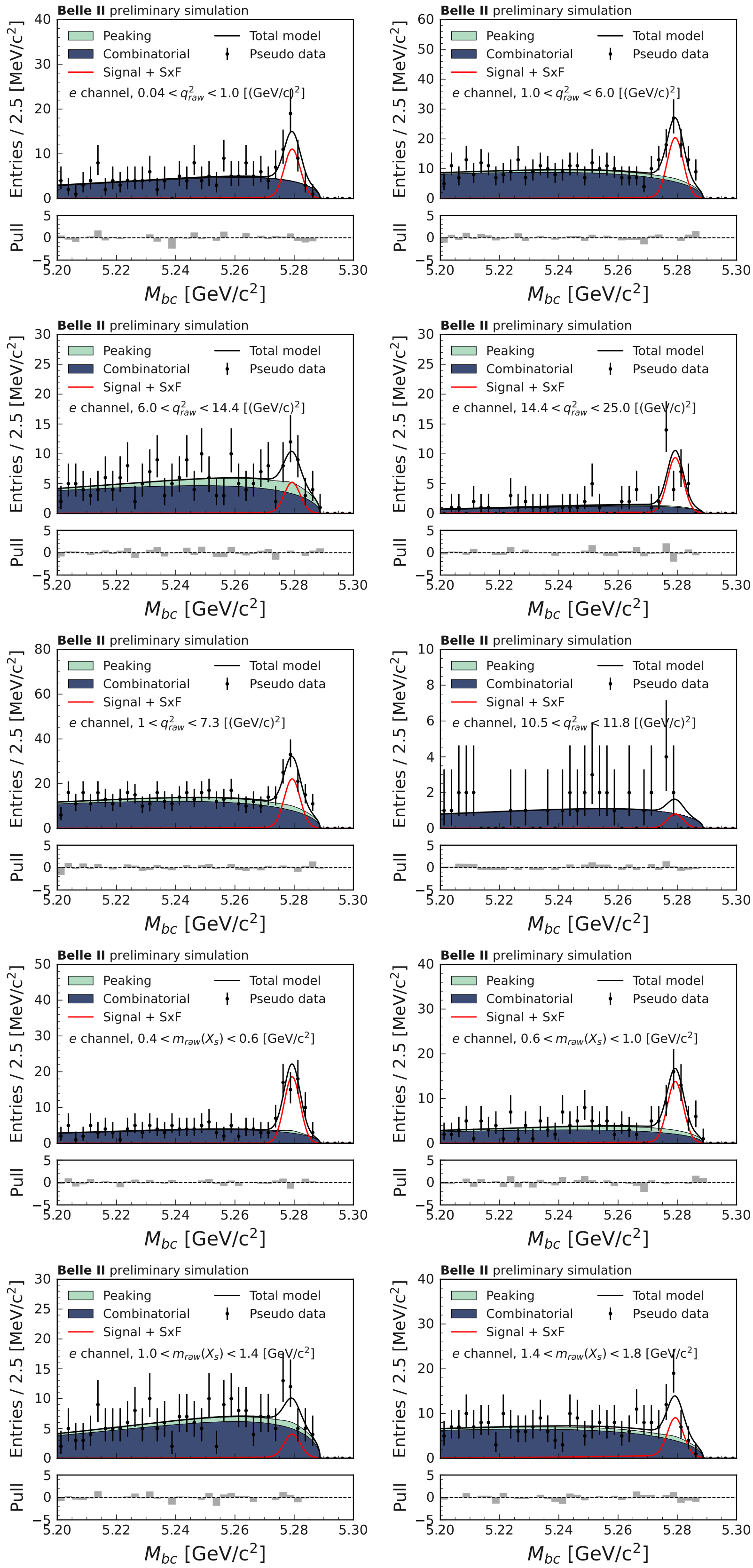


Figure 8.4: Example fitting to toy MC samples in electron channels.

Chapter 9

Systematic Uncertainties

9.1 Signal modeling

Uncertainties related to the modeling of signal MC is the dominant systematic uncertainty in the previous Belle analysis [28]. It can be divided into three different categories: exclusive branching fractions of the $B \to K^{(*)}\ell^+\ell^-$ processes described in Sec. 9.1.1, branching fractions of the newly introduced five high kaon resonances processes $B \to K_J\ell^+\ell^-$ as described in Sec. 9.1.2, X_s line shape described in Sec. 9.1.3 to Sec. 9.1.5, and fragmentation of reconstructed X_s described in Sec. 9.1.6 and finally extrapolation from reconstructed modes to full decay modes of X_s as described in Sec. 9.1.7.

9.1.1 Exclusive branching fractions

The previous Belle analysis adapted the exclusive branching fractions of the theoretical prediction, with huge systematic uncertainties of 30%. Our strategy is to average the latest experiment measurements in order to achieve the best precisions. Table 9.1 shows the latest results of the branching fraction measurement.

We made two assumptions in measurement averaging:

- Lepton flavor universality: $\Gamma(B \to X_s e^+e^-) = \Gamma(B \to X_s\mu^+\mu^-)$ when the photon pole is excluded ($m_{e^+e^-} > 0.2\,\mathrm{GeV}/c^2$).
- Isospin symmetry: $\Gamma(B^+ \to X_s^+\ell^+\ell^-) = \Gamma(B^0 \to X_s^0\ell^+\ell^-)$ with lifetime ratio $\tau_{B^+}/\tau_{B^0} = 1.076$.

The Best Linear Unbiased Estimator (BLUE) method is adapted in the averaging. We use $y_i \pm \sigma_i$ to denote the center value and the standard deviation of measurement i and $E = [\sigma_{i,j}^2]$ to denote the covariance matrix between different measurements. The result is a weighted average of different measurements:

$$\hat{y} = \sum_i w_i y_i, \quad \sum_i w_i = 1 \tag{9.1}$$

Table 9.1: Branching fractions of latest $B \to K^{(*)}\ell^+\ell^-$ measurements from PDG. The Belle measurements marked with (*) are not used in this analysis due to its unclear statement in the $m_{\ell^+\ell^-}$ cut.

Channel	Average BF in PDG ($\times 10^6$)	Measured by
$B^+ \to K^+\mu^+\mu^-$	0.453 ± 0.035	LHCb [1], Belle [22], Babar [85], CDF [86]
$B^0 \to K^0\mu^+\mu^-$	0.339 ± 0.035	LHCb [1], Belle [22], Babar [85], CDF [86]
$B^+ \to K^+e^+e^-$	0.56 ± 0.06	Belle [22], Babar [85]
$B^0 \to K^0e^+e^-$	$0.25^{+0.11}_{-0.09}$	Belle [22], Babar [85]
$B^+ \to K^{*+}\mu^+\mu^-$	0.96 ± 0.10	LHCb [1], Belle [87], Babar [85], CDF [86]
$B^0 \to K^{*0}\mu^+\mu^-$	0.94 ± 0.05	LHCb [2], Belle [87], Babar [85], CDF [86]
$B^+ \to K^{*+}e^+e^-$	$1.55^{+0.40}_{-0.31}$	Belle(*) [87], Babar [85]
$B^0 \to K^{*0}e^+e^-$	$1.03^{+0.19}_{-0.17}$	Belle(*) [87], Babar [85]

where the weight vector $w = [w_1, w_2, ..., w_n]^T$ is chosen to minimize the uncertainty of $\hat{y}$: $\sigma^2_{\hat{y}} = w^T E w$. In case the combination has a large $\chi^2 = [\hat{y} - y_1, \hat{y} - y_2, ..., \hat{y} - y_n]E^{-1}[\hat{y} - y_1, \hat{y} - y_2, ..., \hat{y} - y_n]^T$, a scaling factor $S = \sqrt{\chi^2/(n-1)}$ is applied to $\sigma_{\hat{y}}$ to compensate for possible systematic uncertainties not considered in these measurements.

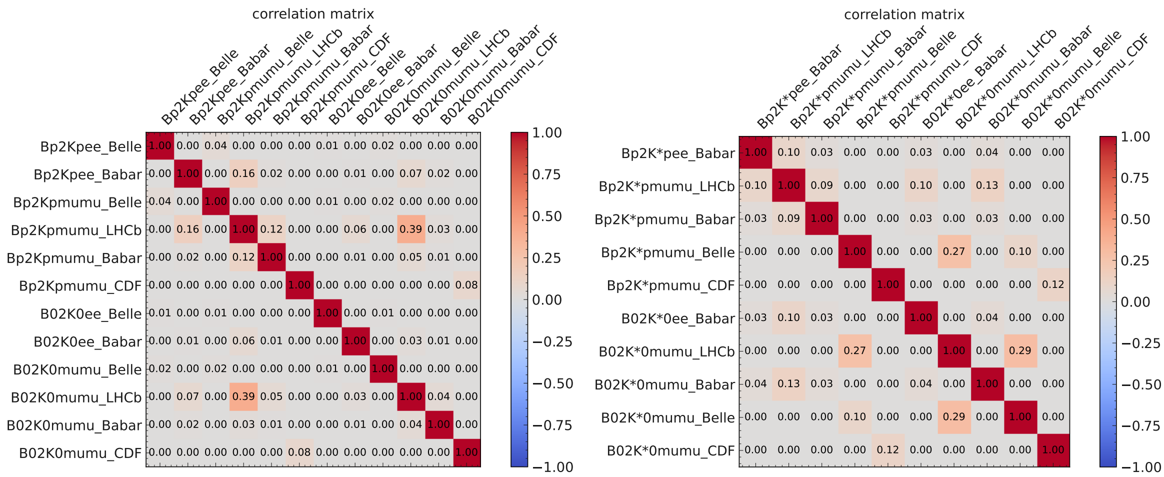


Figure 9.1: Correlation matrix between different exclusive measurements.

In the $B \to K\ell^+\ell^-$ channel, measurements from Belle and *BABAR* experiments [22, 85] are considered independent. For the Belle measurement, systematics from PID, tracking, K^0_S reconstruction, $B\overline{B}$ counting, and $f_{00,\pm}$ are considered fully correlated. For LHCb measurement, the systematics of the branching fractions of $B \to J/\psi\, K$ is considered fully correlated. In addition, the branching fractions of the reference channel are taken from the *BABAR* measurements, so we assume that their systematic uncertainties are 100% correlated. For the Babar measurement, since no detailed breakdown of the systematics is given, we consider all the systematic uncertainties as fully correlated.

For the $B \to K^*(892)\ell^+\ell^-$ channel, results from four papers [1, 2, 87, 85] are taken into

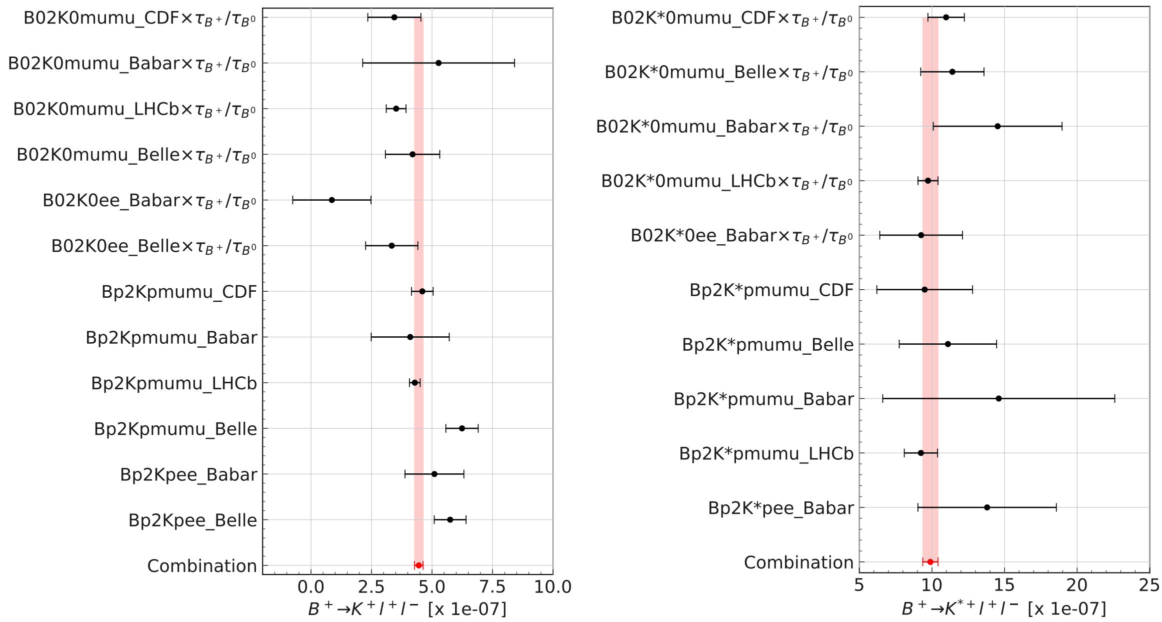


Figure 9.2: Branching fraction of each single measurements and the combination result.

consideration, excluding the $B \to K^*(892)e^+e^-$ measurement from Belle because of its unclear statement of $m_{\ell^+\ell^-}$ range. For the Belle and Babar measurements, all systematic uncertainties are considered fully correlated conservatively, since no detailed breakdown of systematics was given. The LHCb measurement was performed with different control channels, and thus no correlation was assigned between the LHCb measurements. On the other hand, correlations to Belle and BABAR measurements are assigned accordingly depending on the reference channel measurement they adopt.

The correlation matrices between different measurements are shown in Fig. 9.1 and the combination results are

$$\begin{aligned} \mathcal{B}(B^+ \to K^+\ell^+\ell^-) &= (4.45 \pm (0.27 \times S)) \times 10^{-7},\ S = 1.53, \\ \mathcal{B}(B^+ \to K^{*+}\ell^+\ell^-) &= (9.89 \pm 0.52) \times 10^{-7},\ S = 1 \end{aligned} \tag{9.2}$$

with a detailed plot showing each single measurement results and combinations in Fig. 9.2. The uncertainties of the exclusive branching fractions are reduced by a factor of 3–6 compared to the previous Belle analysis.

9.1.2 Kaon resonances

The uncertainties of the contribution of different kaon resonances in the inclusive X_s region are considered by varying their branching fractions shown in Tab. 4.2.

The possible uncertainty from interference between resonances is studied by comparing the efficiency difference with and without interference for the $K_1(1270)$ and $K_1(1400)$ cases, which share the same spin-parity. We perform this study using a toy MC method. First, we derive the reconstruction efficiency as a function of $m(X_s)$ from the signal MC that includes only the contributions of $K_1(1270)$ and $K_1(1400)$. After that, we derive the distributions of $m(X_s)$ with interference in the following way:

$$|\mathcal{A}(K_1(1270)) + \mathcal{A}(K_1(1400))|^2 \tag{9.3}$$

where $\mathcal{A}$ is derived by multiplying the relativistic Breit-Wigner function by the square root of the corresponding $B \to K_J \ell^+\ell^-$ branching fractions. Finally, we derive the efficiency by integrating the reconstruction efficiency over the $m(X_s)$ lineshape with interference. The difference in efficiency with and without interference is measured to be 0.3%, which is negligibly small.

9.1.3 Fermi motion

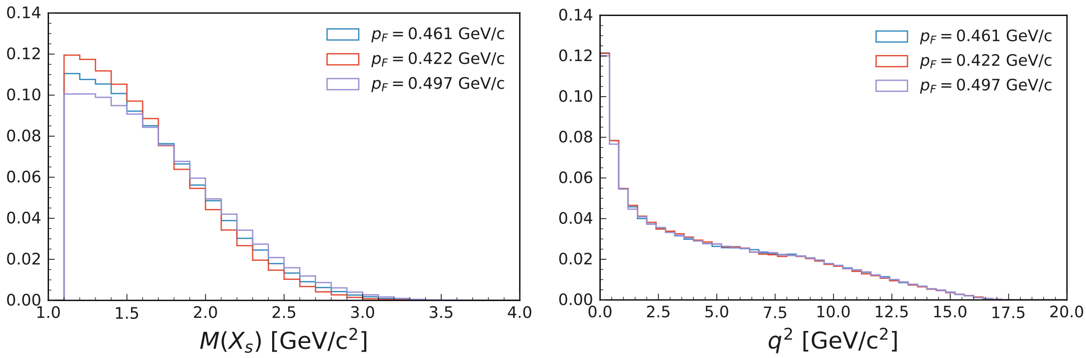


Figure 9.3: Variation of normalized $m(X_s)$ and q^2 distributions as a function of p_F.

The Fermi momentum p_F is verified according to the latest measurements in inclusive channels such as $B \to X_c \ell \nu$ and $B \to X_s \gamma$. HFLAV [33] gave the latest measurement on the average kinematic energy of the b quark in the B meson $\mu_\pi^2 = 0.464 \pm 0.076$ [GeV^2]. According to Ref. [34], the relationship between μ_π^2 and p_F assuming the spectator quark mass $m_{sp} = 0$ is given by:

$$\mu_\pi^2 = 6\left(1 - \frac{2}{\pi}\right) p_F^2 \tag{9.4}$$

And thus we have:

$$p_F = 0.461^{+0.036}_{-0.039}\,\text{GeV}/c \tag{9.5}$$

We verify the p_F parameter in the BTOXSLL generator by one standard deviation. The variation of the $m(X_s)$ and q^2 distributions is shown in Fig. 9.3. The reconstruction efficiency under verified p_F is estimated by re-weighting the MC samples according to Fig. 9.3, and the variation in the reconstruction efficiency is taken as the systematic uncertainties from Fermi momentum.

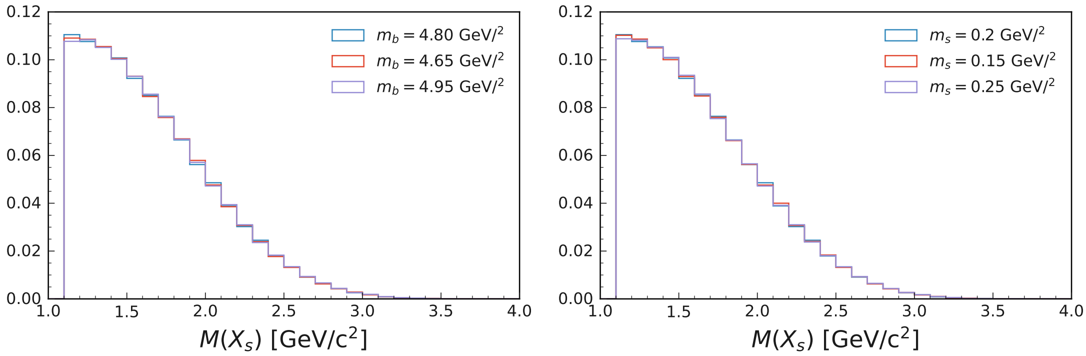


Figure 9.4: $m(X_s)$ distribution variation as a function of quark masses. The left and right plot shows the variation by b quark mass and s quark mass, respectively.

9.1.4 Quark mass

The pole masses of b and s quarks are verified conservatively comparing with PDG results:

$$\begin{aligned} m_b^{\text{pole}} &= 4.8 \pm 0.15\,\text{GeV}/c^2 \\ m_s^{\text{pole}} &= 0.2 \pm 0.05\,\text{GeV}/c^2 \end{aligned} \tag{9.6}$$

Figure. 9.4 shows the $m(X_s)$ distribution variation as a function of quark masses, and similarly, systematic uncertainties are calculated by re-weighting the signal MC samples.

9.1.5 $K^*(892) - X_s$ transition point

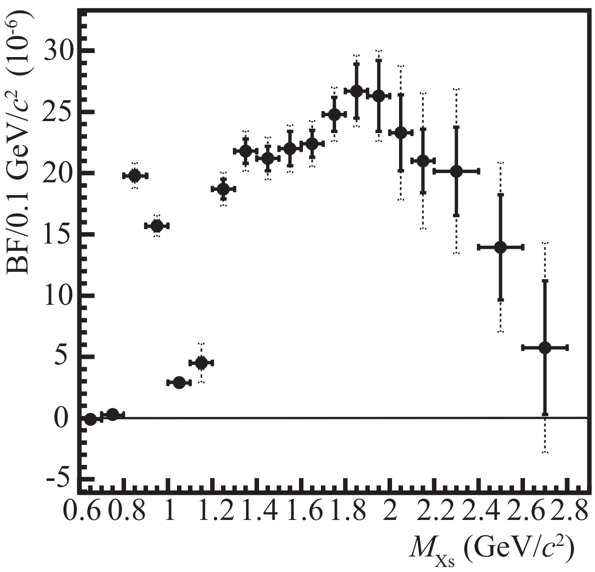


Figure 9.5: Differential branching fraction as a function of $m(X_s)$ in $B \to X_s\gamma$ decay [66]. The solid error bars represent statistical uncertainties and the dashed error bars are the combination of statistical uncertainties and systematic uncertainties.

In $B \to X_s\gamma$ [88, 66] analysis with sum-of-exclusive method, it is confirmed that there is a sharp transition from resonant $K^*(892)$ to inclusive X_s as shown in Fig. 9.5. This transition occurs in the range of $1.05\,\mathrm{GeV}/c^2$ to $1.2\,\mathrm{GeV}/c^2$. For this reason, the $m(X_s)$ cutoff in the BTOXSLL generator is adjusted accordingly from default $1.1\,\mathrm{GeV}/c^2$ to $1.05\,\mathrm{GeV}/c^2$ and $1.2\,\mathrm{GeV}/c^2$.

9.1.6 Fragmentation

In this analysis, the fragmentation of visible modes is determined in a data-driven way. In $B \to X_s\gamma$ [88, 89] analysis, the fractions of each visible mode are measured in three different regions: $1.15 < m(X_s) < 1.5,\ 1.5 < m(X_s) < 2.0,\ 2.0 < m(X_s) < 2.4$ [GeV/c^2]. Since the $B \to X_s\gamma$ analysis covers more modes than this work in the reconstruction of X_s, such as the $2\pi^0$ modes and the η modes, the fractions are normalized accordingly. The results are summarized in Tab. 9.2.

Table 9.2: Fragmentation in $B \to X_s\gamma$ channel. $m(X_s)$ is in the unit of GeV/c^2.

Category	$m(X_s) \in [1.15, 1.5]$	$m(X_s) \in [1.5, 2.0]$	$m(X_s) \in [2.0, 2.4]$
$K1\pi$ w/o π^0	0.112 ± 0.007	0.037 ± 0.004	0.022 ± 0.009
$K1\pi$ w/ π^0	0.062 ± 0.006	0.014 ± 0.004	0.007 ± 0.010
$K2\pi$ w/o π^0	0.244 ± 0.014	0.188 ± 0.013	0.118 ± 0.031
$K2\pi$ w/ π^0	0.467 ± 0.019	0.247 ± 0.018	0.105 ± 0.042
$K3\pi$ w/o π^0	0.005 ± 0.003	0.113 ± 0.011	0.221 ± 0.055
$K3\pi$ w/ π^0	0.105 ± 0.012	0.322 ± 0.024	0.350 ± 0.093
$K4\pi$	0.005 ± 0.004	0.063 ± 0.018	0.135 ± 0.113
$3K$	0.000 ± 0.000	0.016 ± 0.002	0.042 ± 0.012

Another direct measurement on fragmentation is performed in $B \to J/\psi X_s$ channel [90, 37] by measuring the data to MC ratio of the number of events generated in each exclusive channel. In this table, only data to MC up to $K3\pi$ channels are measured. To convert this ratio to the fractions of each visible channel, a signal MC is generated in the Belle configuration and re-weighted. The results are shown in Table. 9.3. Due to the limitation in phase space, the maximal mass of X_s in this channel is around $2.2\,\mathrm{GeV}/c^2$.

The comparison of fractions of each visible mode is shown in Fig. 9.6 in three different $m(X_s)$ intervals. In the intervals below $2.0\,\mathrm{GeV}/c^2$ and $Kn\pi$ ($n \leq 3$) channels, the weighted average of fragmentation from two channels is adopted as the nominal fragmentation result:

$$x_{\mathrm{comb}} = \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2} \tag{9.7}$$

where $x_{1(2)}$ and $\sigma_{1(2)}$ represent the fraction and standard deviation of the $X_s\gamma$ (X_sJ/ψ) channel, respectively. Generally, the fragmentation of two channels has good consistency. In case the 1σ deviation from x_{comb} does not cover the difference in the center value

Table 9.3: Fragmentation in $B \to X_s J/\psi$ channel. $m(X_s)$ is in the unit of GeV/c^2. The $m(X_s) \in [2.0, 2.4]$ data are not used for fragmentation estimation for $B \to X_s \ell^+\ell^-$.

Category	$m(X_s) \in [1.15, 1.5]$	$m(X_s) \in [1.5, 2.0]$	$m(X_s) \in [2.0, 2.4]$
$K1\pi$ w/o π^0	0.125 ± 0.005	0.055 ± 0.002	0.040 ± 0.002
$K1\pi$ w/ π^0	0.074 ± 0.007	0.033 ± 0.003	0.024 ± 0.002
$K2\pi$ w/o π^0	0.295 ± 0.010	0.179 ± 0.006	0.137 ± 0.005
$K2\pi$ w/ π^0	0.297 ± 0.024	0.190 ± 0.015	0.145 ± 0.011
$K3\pi$ w/o π^0	0.047 ± 0.006	0.126 ± 0.015	0.145 ± 0.017
$K3\pi$ w/ π^0	0.145 ± 0.026	0.270 ± 0.048	0.273 ± 0.049
$K4\pi$	0.015 ± 0.000	0.125 ± 0.000	0.214 ± 0.000
$3K$	0.000 ± 0.000	0.021 ± 0.000	0.020 ± 0.000

between x_1 or x_2, the uncertainty extends to $|x_{\text{comb}} - x_{1(2)}|$. An overall scaling factor is applied to x_{comb}s to make sure that their sum equals to one. In the $2.0 < m(X_s) < 2.4$ [GeV/c^2] region and the $K4\pi$, $3K$ channels, the measurement $X_s\gamma$ is adapted directly as the nominal fragmentation result.

The nominal efficiency of the signal MC can be expressed as follows:

$$\varepsilon = \frac{\sum_{mode} N_{mode}^{survived}}{N^{visible}}, \tag{9.8}$$

where $N^{survived}$ represents the number of events remaining in each reconstructed channel while $N^{visible}$ represents the total number of events in all visible channels at the generator level. The efficiency variance $\Delta\varepsilon_k$ from the mode k is defined as:

$$\varepsilon \pm \Delta\varepsilon_k = \frac{\sum_{mode \neq k} N_{mode}^{survived} \pm N_k^{survived} \times \delta_k}{N^{visible}} \tag{9.9}$$

where δ_k denotes the relative uncertainty of fragmentation in Fig. 9.6. The total systematic uncertainty from fragmentation is defined as the root-sum-square of uncertainty from each mode:

$$\Delta\varepsilon = \sqrt{\sum_{mode} (\Delta\varepsilon_{mode})^2} \tag{9.10}$$

9.1.7 Missing modes

We use PYTHIA [62, 35] to extrapolate visible events (the 20 reconstructed modes in Tab. 3.1) to all decay modes of X_s. The X_s modes that are not included in Tab. 3.1 are called missing modes.

In $B \to X_s\gamma$ analysis [88], a set of PYTHIA parameters is tuned to match the fractions of visible events. The PYTHIA parameters sensitive to X_s fragmentation and the values used in the $B \to X_s\gamma$ analysis are listed in Tab. 9.4. We use the signal MC generated under such configurations as the nominal MC.

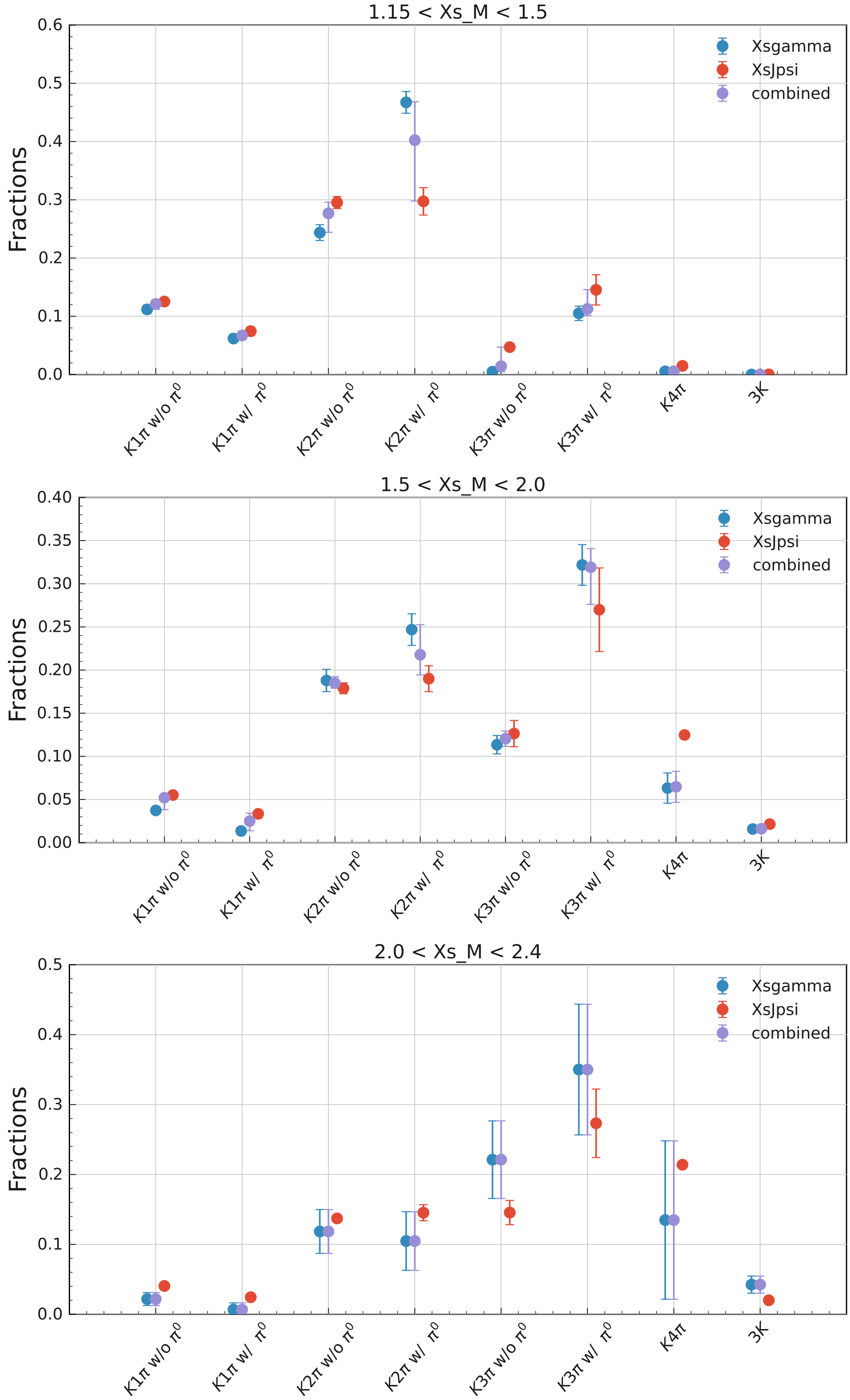


Figure 9.6: Fraction of each visible channel.

Table 9.4: List of effective parameters in the fragmentation of X_s.

PYTHIA paramters	description	Value
PARJ(2)	Suppression of s quark pair production in the field compared with u or d pair production	0.1
PARJ(11)	Probability that a light meson (containing u and d quarks only) has spin 1	0.85
PARJ(12)	Probability that a strange meson has spin 1	0.6
PARJ(15)	Probability that a spin = 1 meson is produced with an orbital angular momentum 1, total spin 0.	0.15
PARJ(25)	Extra suppression factor for η production in fragmentation	0.01

To estimate the uncertainty of this missing mode estimation, we use the seven sets of PYTHIA parameters provided by the other $B \to X_s\gamma$ analysis [66]. These configurations are confirmed to have fair consistency with the data in the fractions of different modes reconstructed. The list of seven configurations is shown in Tab. 9.5 and their fraction of visible modes are shown in the left plot of Fig. 9.7. In each $m(X_s)$ interval, the greatest deviation from the nominal MC is assigned as the systematic uncertainty of missing modes, as shown in the right plot of Fig. 9.7. For conservative estimation, uncertainties in all $m(X_s)$ intervals are considered fully correlated.

Table 9.5: Seven sets of PYTHIA parameters tuned in analysis [66].

Paramters	Config1	Config2	Config3	Config4	Config5	Config6	Config7
PARJ(2)	0.3	0.1	0.1	0.1	0.1	0.1	0.15
PARJ(11)	0.95	0.95	0.95	0.95	0.95	0.95	0.95
PARJ(12)	0.6	0.6	0.6	0.6	0.7	0.7	0.7
PARJ(15)	0.05	0.05	0.30	0.30	0.30	0.40	0.40
PARJ(25)	1.00	1.00	1.00	0.03	0.03	0.03	0.03

The fraction of visible modes used as diag$(f_1, \ldots, f_n)$ term in Eq. 3.1 in each fitting bin (defined in Tab. 3.2) are summarized in Tab. 9.6.

9.2 B meson production

The number of $\Upsilon(4S)$ in the Belle II run 1 dataset is measured as:

$$N_{\Upsilon(4S)} = (387.1 \pm 5.6) \times 10^6. \tag{9.11}$$

A 1.45% relative uncertainty is assigned to the measurement of the branching fraction. In addition, the branching fractions of $\Upsilon(4S)$ decaying to charged and neutral B meson

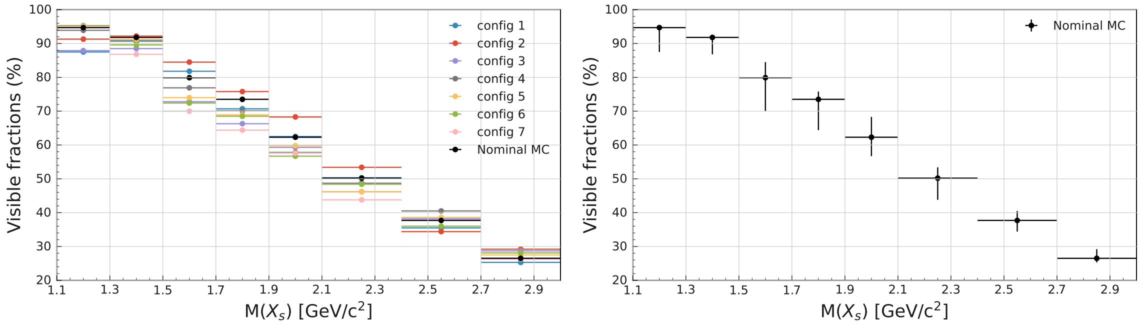


Figure 9.7: Left: visible fractions in nominal MC and seven different PYTHIA configurations. Right: visible fractions in nominal MC and its uncertainty.

Table 9.6: Fractions of visible events in each fitting bin.

q_1^2	q_2^2	q_3^2	q_4^2	q_5^2	q_6^2
$0.76^{+0.02}_{-0.06}$	$0.80^{+0.02}_{-0.05}$	$0.90^{+0.01}_{-0.04}$	$0.98^{+0.00}_{-0.01}$	$0.81^{+0.02}_{-0.05}$	$0.92^{+0.01}_{-0.04}$
$m_1(X_s)$	$m_2(X_s)$	$m_3(X_s)$	$m_4(X_s)$		
1.00	1.00	$0.91^{+0.00}_{-0.06}$	$0.81^{+0.03}_{-0.08}$		

pairs are:

$$f^{\pm} = 0.5113 \pm 0.0108, \ f^{00} = 0.4861 \pm 0.0080. \tag{9.12}$$

The total uncertainty assigned to $N_{B\bar{B}}$ is 2.4%.

9.3 Signal reconstruction

9.3.1 Charged track reconstruction

The systematic uncertainty of track reconstruction efficiency with momenta greater than $0.2\,\mathrm{GeV}/c^2$ is estimated using the $e^+e^- \to \tau^+\tau^-$ events, where one tau lepton decays leptonically ($\tau \to \ell\nu_\tau\nu_\ell$, $\ell = e, \mu$) and the other lepton decays to three charged pions ($\tau \to 3\pi^{\pm}\nu_\tau + n\pi^0$). This event is tagged with three high-quality tracks, and the track reconstruction efficiency is inferred by the ratio of events that can reconstruct the fourth track over the number of tagged events. The tracking efficiency is consistent in the data and the MC with an uncertainty of 0.27% [91], which is assigned as the systematic uncertainty of high-momentum tracking.

For slow pions with momentum smaller than $0.2\,\mathrm{GeV}/c$, their efficiencies are calibrated using the slow pion from D^* decay: $D^* \to D\pi_s$. In this follow-up, a small data to MC discrepancy is observed and the corresponding correction factors are assigned shown in the Tab. 9.7 [92].

Table 9.7: Correction factors for tracking efficiency of slow pions.

Momentum [GeV/c]	Correction factors
$0.05 < p \leq 0.12$	0.947 ± 0.020
$0.12 < p \leq 0.16$	0.985 ± 0.017
$0.16 < p \leq 0.20$	0.983 ± 0.019

9.3.2 Particle identification

The calibration of particle identification, including both leptons and hadrons, is performed by comparing the efficiency difference between the data and MC. The efficiencies of MC can be derived using the MC truth information, and that in the data is derived using various control channels.

For lepton identification, take muon for example, the radiative Bhabha events $e^+e^- \to \mu^+\mu^-\gamma$ is applied using the tag-and-probe method. One first select the events that with a positive charged muon satisfying muonID > 0.9 (the number of left events referred to as N_{tag}), and then applies muonID selection to the negative charged muon (the number of left events referred to as N_{probe}). The lepton identification efficiency of negative charged muon is measured by taking the ratio N_{probe}/N_{tag}. The same procedure can be applied to positive charged muon, and also electrons by using $e^+e^- \to e^+e^-\gamma$ events. A detailed example of evaluating leptonID efficiency in data is available in Appendix. G.4. Other reference channels of two-photon processes $e^+e^- \to e^+e^-\ell^+\ell^-$ and $J/\psi \to \ell^+\ell^-$ are also used to calibrate lepton identification efficiency and their weighted average is adapted as official calibration. The correction factors are applied to MC as a function of track momentum, polar angle, and charge are shown in Fig. H.2 of Appendix. H. On average, correction factors of $(95.0 \pm 0.5)\%$ and $(99.6 \pm 0.5)\%$ are applied to reconstruction efficiencies of each muon and electron particle, respectively.

The charged K, π particles are calibrated mainly using the $D^{*+} \to D^0(\to K^-\pi^+)\pi_s^+$ channel. This is a Cabibbo-favored process and one can utilize the slow pion $\pi_s^\pm$ to tag the flavor of D^0 daughters: the track with the same charge as $\pi_s^\pm$ is pion and the track with opposite charge to $\pi_s^\pm$ is kaon. Hadron identification efficiency is then calibrated by taking the ratio of the number of events before and after applying particleID cut to the corresponding particle. The number of events is derived by fitting to the D^0 spectrum. Correction factors as a function of track momentum, polar angle, and charge are shown in Fig. H.1. On average, correction factors of $(96.1 \pm 0.8)\%$ and $(98.2 \pm 0.2)\%$ are applied to reconstruction efficiencies of each charged kaon and pion particle, respectively.

9.3.3 K_S^0 reconstruction

We calibrate K_S^0 reconstruction efficiency using the D^{*+}-tagged $D^0 \to K_S^0\pi^+\pi^-$ decays in data and MC. This correction includes two parts: 1) reconstruct K_S^0 candidate using two opposite charged pions followed by vertex fitting and 2) K_S^0 candidate selection using Ksselector tool.

For the first part, we correct K_S^0 candidates decay near the IP (flying distance $d <$ 0.5 cm) using the track reconstruction efficiency correction mentioned in Sec. 9.3.1. Under the assumption that the ratio of the produced number of K_S^0 ($\mathbf{n}$) in data (D) and MC (M) is independent from flying distance:

$$\frac{\mathbf{n}_D(p,\cos\theta,d)}{\mathbf{n}_M(p,\cos\theta,d)} = \frac{\mathbf{n}_D(p,\cos\theta,d<0.5\text{ cm})}{\mathbf{n}_M(p,\cos\theta,d<0.5\text{ cm})}, \tag{9.13}$$

one can derive the correction factors for K_S^0 reconstruction efficiency using the K_S^0 yield (n) in each flying distance bin normalized with the yield in $d < 0.5$ cm bin:

$$\begin{aligned}
\frac{\varepsilon_D(p,\cos\theta,d)}{\varepsilon_M(p,\cos\theta,d)} &= \frac{n_D(p,\cos\theta,d)}{n_M(p,\cos\theta,d)} \times \frac{\mathbf{n}_M(p,\cos\theta,d)}{\mathbf{n}_D(p,\cos\theta,d)} \\
&= \frac{n_D(p,\cos\theta,d)}{n_M(p,\cos\theta,d)} \times \frac{\mathbf{n}_M(p,\cos\theta,d<0.5\text{ cm})}{\mathbf{n}_D(p,\cos\theta,d<0.5\text{ cm})} \\
&= \frac{n_D(p,\cos\theta,d)}{n_M(p,\cos\theta,d)} \times \frac{n_M(p,\cos\theta,d<0.5\text{ cm})}{n_D(p,\cos\theta,d<0.5\text{ cm})} \times \frac{\varepsilon_M(p,\cos\theta,d<0.5\text{ cm})}{\varepsilon_D(p,\cos\theta,d<0.5\text{ cm})} \\
&= \frac{n_D(p,\cos\theta,d)}{n_M(p,\cos\theta,d)} \times \frac{n_M(p,\cos\theta,d<0.5\text{ cm})}{n_D(p,\cos\theta,d<0.5\text{ cm})}.
\end{aligned} \tag{9.14}$$

Notice that $n = \varepsilon \cdot \mathbf{n}$ by definition, and we use an underlying assumption that there is no data/MC discrepancy in reconstruction efficiency of K_S^0s decay near the IP: $\varepsilon_D(p,\cos\theta,d < 0.5\text{ cm})/\varepsilon_M(p,\cos\theta,d < 0.5\text{ cm}) = 1$. This assumption is verified using $e^+e^- \to \tau^+\tau^-$ events where K_S^0s come from $\tau^\pm$ decays [93]. In above formulas, p and θ are momentum and polar angle of K_S^0 candidate and yields n are derived by fitting to K_S^0 invariant mass spectrum.

For the second part of Ksselector efficiency (i.e. `V0Selector` > 0.90 and `LambdaVeto` > 0.11), its correction factor is defined as:

$$\frac{\varepsilon'_D(p,\cos\theta,d)}{\varepsilon'_M(p,\cos\theta,d)} = \frac{n'_D(p,\cos\theta,d)}{n'_M(p,\cos\theta,d)} \times \frac{n_M(p,\cos\theta,d)}{n_D(p,\cos\theta,d)}, \tag{9.15}$$

where n' represents the number of remaining K_S^0 after applying the Ksselector.

Multiplying Eq. 9.15 with Eq. 9.14, one could derive the total correction factor for K_S^0 reconstruction and selection:

$$\frac{\varepsilon_D(p,\cos\theta,d)}{\varepsilon_M(p,\cos\theta,d)} \times \frac{\varepsilon'_D(p,\cos\theta,d)}{\varepsilon'_M(p,\cos\theta,d)} = \frac{n'_D(p,\cos\theta,d)}{n'_M(p,\cos\theta,d)} \times \frac{n_M(p,\cos\theta,d<0.5\text{ cm})}{n_D(p,\cos\theta,d<0.5\text{ cm})}. \tag{9.16}$$

Each term on the right hand side of above formula is calculable by fitting to the K_S^0 invariant mass spectrum. A detailed description of above procedure is presented in Appendix. E.

On average, a correction factor of $(97.3 \pm 2.6)\%$ is applied to each K_S^0 candidates in MC sample.

9.3.4 π^0 reconstruction

The correction factor for the reconstruction efficiency of π^0 is estimated using $e^+e^- \to \tau^+\tau^-$ events, where one tau lepton contains only one charged daughter and the other tau decays to $3\pi^\pm\nu$ final state. We denote the number of such events to be $N^{3\times1}$. After that, we reconstruct π^0 candidates from two photon candidates in selected events in both data and MC and count the number of reconstructed π^0 ($N(\pi^0)$) by fitting to its invariant mass spectrum. The correction factor of π^0 reconstruction efficiency is then calculated by the following formula:

$$\frac{\varepsilon_D(p,\cos\theta)}{\varepsilon_M(p,\cos\theta)} = \frac{N_D(\pi^0)}{N_D^{3\times1}} \times \frac{N_M^{3\times1}}{N_M(\pi^0)}, \tag{9.17}$$

where the p, θ are the momentum and polar angle of π^0 candidate.

On average, a correction factor of $(94.5 \pm 1.5)\%$ is applied to each π^0 candidates in MC sample.

9.3.5 Multivariate analysis

The MVA efficiency is corrected using the $B \to X_s J/\psi$ channel. We apply the same selection criteria as signal reconstruction except that we only select events within the charmonium veto. The number of signals before and after MVA is derived by fitting to the M_{bc} spectrum in both data and MC. In the fitting, the signal is described by a single Gaussian and the background is modeled by an Argus function.

We divide the whole sample into three bins: $0.4 < m(X_s) < 0.6$, $0.6 < m(X_s) < 1.1$, and $1.1 < m(X_s) < 1.8$ and the efficiencies are calibrated separately. The correction factors and uncertainties are shown in the left plot of Fig. 9.8. We also have confirmed that the efficiencies inside $0.6 < m(X_s) < 1.1$ and $1.1 < m(X_s) < 1.8$ [GeV/c^2] bins are uniform using a signal MC sample, as indicated in the right plot of Fig. 9.8.

9.4 Signal extraction

In this section, we describe the systematic uncertainties related to signal yields extraction procedure, with center values of signal yields blinded to the analyst.

9.4.1 Fitter bias

The systematic uncertainty of the fitter bias on the signal yields is estimated from the toyMC test results described in Sec. 8.3. The shift of the mean of the pull distribution indicate an under estimation of signal yield in several kinematic bins. The systematic uncertainty due to the fitter bias is given by:

$$\sigma_{\text{fitter bias}} = \mu \cdot \sigma_{\text{stat}} \tag{9.18}$$

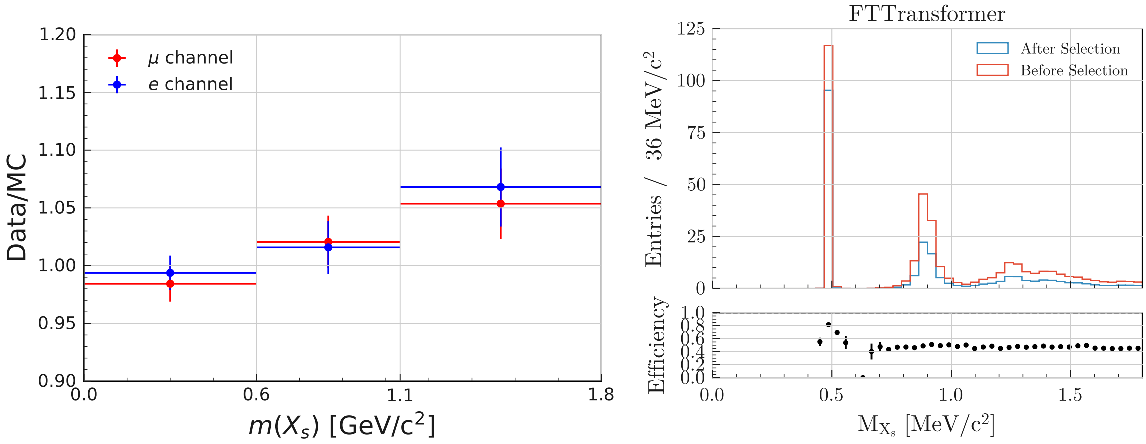


Figure 9.8: Left: Data/MC of MVA selection efficiency. Right: $m(X_s)$ distribution before and after MVA selection in signal MC (upper) and MVA efficiency as a function of $m(X_s)$ (lower).

where μ is the mean of pull distribution and σ_{stat} is the statistical uncertainty of the fitting. A list of bins and the corresponding shifted mean is summarized in Tab. 9.8. Bins not listed in this table are free from fitter bias.

Table 9.8: List of kinematic bins and the corresponding shifted mean of pull distributions. Bins not listed in this table are free from fitter bias.

Bin name	q_1^2	q_4^2	q_6^2	$m_2(X_s)$	$m_4(X_s)$
Shift in muon channel	0.179	0.152	0.115	0.123	–
Shift in electron channel	–	0.123	0.203	0.107	0.104

9.4.2 Signal shape uncertainty

The signal shape is modeled by a Gaussian function with fixed center values and standard deviation as shown in Tab. 8.1. The systematic uncertainties are estimated by varying them by $\pm 1\sigma$.

9.4.3 Self cross-feed

Systematic uncertainty from SxF contribution will be estimated by varying N_{SxF}/N_{sig} ratio by 1σ. Given that the peaking component in SxF could be not well simulated, a 100% uncertainty is assigned to the peaking part of SxF shape for conservative estimation.

9.4.4 Peaking background

The shape of peaking backgrounds are fixed during fitting as shown in Figs. [F.3, F.4], F.5, and [F.6, F.7]. We estimate systematic uncertainties from peaking background by varying the shapes to their upper and lower variations. For double mis-ID and swap mis-ID background, their variations mainly come from PID fake rate and efficiency uncertainties. For charmonium background, the uncertainty is dominated by branching fractions of $B \to X\psi$ decay and MC statistics. We also vary the shape of charmonium background by its fitting uncertainty derived during the fitting to Gaussian and Argus function in Figs. F.6 and F.7.

9.5 Summary of systematic uncertainties

The systematic uncertainties for differential branching fraction, inclusive branching fractions are summarized in Tab. 9.9. Statistical uncertainties are also presented in these tables for comparison. These numbers are derived in real data.

Table 9.9: Breakdown of relative systematic uncertainties (in %) for differential branching fraction and inclusive branching fraction. Statistical uncertainties are also presented for comparison.

	q_1^2	q_2^2	q_3^2	q_4^2	q_5^2	q_6^2	$m_1(X_s)$	$m_2(X_s)$	$m_3(X_s)$	$m_4(X_s)$	$\mathcal{B}_{inc.}$
Signal shape	±0.5	±0.3	±0.1	±0.3	±0.2	±0.5	±0.2	±0.2	±0.4	±0.5	±0.3
Self cross-feed	±1.5	±0.7	±0.9	±1.0	±0.7	±1.0	±0.7	±0.8	±1.8	±1.9	±1.0
Peaking BG	±0.5	±0.4	±2.0	±0.4	±0.7	±2.4	±0.1	±0.4	±4.1	±0.7	±1.0
Fiting bias	±3.6	±0.0	±0.1	±3.2	±0.0	±6.5	±0.0	±1.6	±0.1	±3.6	±1.0
$\mathcal{B}(B \to K\ell^+\ell^-)$	±1.3	±2.9	±2.9	±3.9	±2.9	±2.9	–	–	±0.0	±0.0	±2.6
$\mathcal{B}(B \to K^*\ell^+\ell^-)$	±1.7	±1.4	±1.7	±1.8	±1.4	±2.0	–	–	±1.0	±0.2	±1.6
Kaon resonances	±2.7	±3.4	±3.7	±1.2	±3.4	±4.3	–	–	±3.8	±4.8	±3.2
Fragmentation	${}^{+2.3}_{-3.0}$	${}^{+1.1}_{-1.7}$	${}^{+0.8}_{-1.2}$	${}^{+0.0}_{-0.1}$	${}^{+1.1}_{-1.6}$	${}^{+0.6}_{-0.9}$	–	–	${}^{+3.8}_{-5.5}$	${}^{+4.2}_{-5.9}$	${}^{+1.2}_{-1.7}$
Fermi motion	±0.8	${}^{+0.3}_{-0.5}$	${}^{+0.3}_{-0.2}$	${}^{+0.1}_{-0.0}$	${}^{+0.3}_{-0.5}$	${}^{+0.2}_{-0.1}$	–	–	${}^{+1.7}_{-2.0}$	${}^{+0.7}_{-0.8}$	±0.4
b quark mass	${}^{+0.1}_{-0.0}$	±0.0	±0.0	±0.0	±0.0	±0.0	–	–	±0.1	±0.2	±0.0
s quark mass	${}^{+0.1}_{-0.0}$	±0.0	±0.0	±0.0	±0.0	±0.0	–	–	${}^{+0.4}_{-0.0}$	${}^{+0.0}_{-0.1}$	±0.0
$K^* - X_s$ transition	${}^{+6.0}_{-3.5}$	${}^{+3.4}_{-2.6}$	${}^{+0.8}_{-1.0}$	${}^{+4.6}_{-1.4}$	${}^{+3.1}_{-2.4}$	${}^{+2.2}_{-0.8}$	–	–	${}^{+0.8}_{-6.0}$	±0.1	${}^{+3.1}_{-2.1}$
Missing mode	${}^{+7.4}_{-3.1}$	${}^{+6.6}_{-2.5}$	${}^{+4.7}_{-1.3}$	${}^{+0.8}_{-0.1}$	${}^{+6.5}_{-2.4}$	${}^{+4.2}_{-1.0}$	–	–	${}^{+6.2}_{-0.5}$	${}^{+9.9}_{-3.4}$	${}^{+5.6}_{-2.0}$
Tracking	±1.1	±1.0	±1.0	±0.9	±1.0	±1.0	±0.7	±1.1	±1.2	±1.3	±1.0
K_S^0 reconstruction	±0.7	±0.8	±0.6	±0.5	±0.8	±0.6	±0.7	±0.6	±0.5	±0.5	±0.7
KaonID	±0.8	±0.7	±0.9	±1.5	±0.7	±1.0	±0.6	±1.2	±1.1	±1.5	±0.8
PionID	±0.2	±0.1	±0.1	±0.1	±0.1	±0.1	±0.0	±0.2	±0.3	±0.3	±0.1
LeptonID	±0.7	±0.7	±0.8	±0.6	±0.7	±0.7	±0.7	±0.7	±0.8	±0.7	±0.7
π^0 reconstruction	±0.3	±0.1	±0.1	±0.0	±0.1	±0.1	±0.0	±0.1	±0.3	±0.4	±0.1
MVA	±1.7	±1.3	±1.2	±1.1	±1.3	±1.2	±0.9	±1.1	±2.9	±2.5	±1.4
MC statistics	±0.7	±0.3	±0.4	±0.4	±0.3	±0.8	±0.2	±0.5	±0.8	±0.8	±0.5
B counting	±2.4	±2.4	±2.4	±2.4	±2.4	±2.4	±2.4	±2.4	±2.4	±2.4	±2.4
Total $\sigma_{sys.}$	${}^{+11.7}_{-8.5}$	${}^{+9.4}_{-7.0}$	${}^{+8.0}_{-6.7}$	${}^{+8.0}_{-6.7}$	${}^{+9.3}_{-6.9}$	${}^{+10.6}_{-9.6}$	±3.1	±3.8	${}^{+10.5}_{-11.2}$	${}^{+13.2}_{-10.3}$	${}^{+8.7}_{-6.7}$
Total $\sigma_{stat.}$	±29.8	±17.4	±24.3	±23.6	±15.3	±37.7	±15.9	±13.8	±37.3	±47.5	±12.6

Chapter 10

Conclusion and Prospects

In this chapter, we present the results and prospects to future analysis. Due to constrains from the Belle II collaboration, we only present the uncertainties of all measurements except for the differential branching fraction in $m_1(X_s)$ bin, whose result is fully unblinded. All plots in this chapter presented are made with center values fixed to MC assumption or theoretical predictions while uncertainties from real data analysis.

10.1 Results

We extract signal yields by fitting to the M_{bc} spectrum and uncertainties of derived yields of each lepton flavor in each kinematic bin are presented in Tab. 10.1. Following Eq. 3.1, implementing the signal unfolding matrices (efficiency matrices) in Figs. 6.28 and 6.29 and the extrapolation factors from visible modes to all modes of X_s in Tab. 9.6, we derive the lepton flavor specific branching fractions and derive the lepton flavor averaged branching fractions. Their uncertainties are presented in Tab. 10.2 and Figs. 10.1 and 10.2.

The results in $m_1(X_s)$ bin, which is the $B \to K\ell^+\ell^-$ region, are unblinded for checking the possible bias in this analysis. The derived signal yield is 33.9 ± 7.0 (31.8 ± 7.8) for muon (electron) channels, where the uncertainty is statistical. The branching fractions (in the unit of 10^{-6}) are $0.32 \pm 0.07 \pm 0.01$, $0.27 \pm 0.07 \pm 0.01$, and $0.30 \pm 0.05 \pm 0.01$ for muon, electron, and lepton flavor averaged results, respectively. The first uncertainties are statistical and the second uncertainties are systematic. These branching fraction results are in consistent with the world average values [60][1] within 0.9σ, 1.4σ and 1.7σ for muon, electron, and flavor averaged case, respectively. These deviations are considered

[1] World average (w.a.) values for $\mathcal{B}(B \to K\ell^+\ell^-)$ are:

$$\mathcal{B}^{\text{w.a.}}(B \to K\mu^+\mu^-) = (0.39 \pm 0.02) \times 10^{-6}$$
$$\mathcal{B}^{\text{w.a.}}(B \to Ke^+e^-) = (0.40 \pm 0.06) \times 10^{-6}$$
$$\mathcal{B}^{\text{w.a.}}(B \to K\ell^+\ell^-) = (0.40 \pm 0.03) \times 10^{-6}$$

acceptable within statistical fluctuation.

Table 10.1: Absolute statistical uncertainties of fitted signal yields in each kinematic bin.

Bin	Muon channel	Electron channel
$q^2_{\text{raw}} \in q^2_1$	± 2.9	± 6.9
$q^2_{\text{raw}} \in q^2_2$	± 6.6	± 9.8
$q^2_{\text{raw}} \in q^2_3$	± 9.3	± 5.7
$q^2_{\text{raw}} \in q^2_4$	± 5.9	± 7.2
$q^2_{\text{raw}} \in q^2_5$	± 8.5	± 10.5
$q^2_{\text{raw}} \in q^2_6$	± 4.5	± 4.2
$m_{\text{raw}}(X_s) \in m_1(X_s)$	± 7.0	± 7.8
$m_{\text{raw}}(X_s) \in m_2(X_s)$	± 7.1	± 7.5
$m_{\text{raw}}(X_s) \in m_3(X_s)$	± 5.4	± 6.9
$m_{\text{raw}}(X_s) \in m_4(X_s)$	± 6.5	± 7.4

Table 10.2: Absolute uncertainties of differential branching fractions (in the unit of 10^{-6}) in different kinematic bins. The first and second uncertainties are statistical and systematic, respectively.

Bin name	$\Delta\mathcal{B}(B \to X_s\mu^+\mu^-)$	$\Delta\mathcal{B}(B \to X_s e^+e^-)$	$\Delta\mathcal{B}(B \to X_s\ell^+\ell^-)$
q^2_1	$\pm 0.33 \pm 0.07$	$\pm 0.36^{+0.14}_{-0.10}$	$\pm 0.24^{+0.10}_{-0.07}$
q^2_2	$\pm 0.33^{+0.11}_{-0.08}$	$\pm 0.33^{+0.15}_{-0.11}$	$\pm 0.24^{+0.13}_{-0.09}$
q^2_3	$\pm 0.44^{+0.12}_{-0.11}$	$\pm 0.54^{+0.12}_{-0.10}$	$\pm 0.35^{+0.11}_{-0.10}$
q^2_4	$\pm 0.08^{+0.03}_{-0.02}$	$\pm 0.10 \pm 0.02$	$\pm 0.06 \pm 0.02$
q^2_5	$\pm 0.38^{+0.14}_{-0.11}$	$\pm 0.34^{+0.17}_{-0.13}$	$\pm 0.25^{+0.15}_{-0.11}$
q^2_6	$\pm 0.10 \pm 0.02$	$\pm 0.12 \pm 0.03$	$\pm 0.08 \pm 0.02$
$m_1(X_s)$	$\pm 0.07 \pm 0.01$	$\pm 0.07 \pm 0.01$	$\pm 0.05 \pm 0.01$
$m_2(X_s)$	$\pm 0.23 \pm 0.05$	$\pm 0.23 \pm 0.05$	$\pm 0.16 \pm 0.05$
$m_3(X_s)$	$\pm 0.40 \pm 0.07$	$\pm 0.47^{+0.15}_{-0.17}$	$\pm 0.31 \pm 0.09$
$m_4(X_s)$	$\pm 0.72^{+0.10}_{-0.08}$	$\pm 0.77^{+0.21}_{-0.17}$	$\pm 0.53^{+0.15}_{-0.11}$

By summing the branching fractions of q^2_1 to q^2_4 bins, we derive the lepton flavor specific and lepton flavor averaged inclusive branching fractions in the $m(\ell^+\ell^-) > 0.2\,\text{GeV}/c^2$ region. Their statistical (first term) and systematic (second term) uncertainties are:

$$\begin{aligned} \Delta\mathcal{B}(B \to X_s\mu^+\mu^-) &= (\pm 0.65^{+0.29}_{-0.24}) \times 10^{-6}, \\ \Delta\mathcal{B}(B \to X_s e^+e^-) &= (\pm 0.74^{+0.41}_{-0.31}) \times 10^{-6}, \\ \Delta\mathcal{B}_{inc.}(B \to X_s\ell^+\ell^-) &= (\pm 0.49^{+0.34}_{-0.22}) \times 10^{-6}, \end{aligned} \tag{10.1}$$

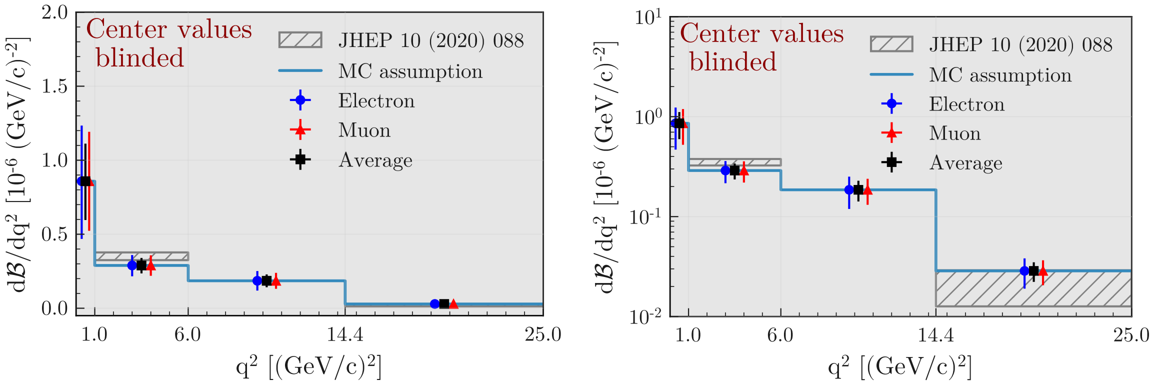


Figure 10.1: Differential branching fraction $d\mathcal{B}/dq^2$ with linear vertical axis on the left and logarithmic scale on the right. The blue circle and red triangle represent results of electron and muon channels, respectively. The black square is the flavor averaged result. The light blue line is the MC assumptions and gray shadowed area represents theoretical prediction [32]. Center values are still blinded due to constrains of Belle II collaboration and they are fixed to MC assumptions. The error bars represent the square root of quadratic sum of statistical and systematic uncertainties.

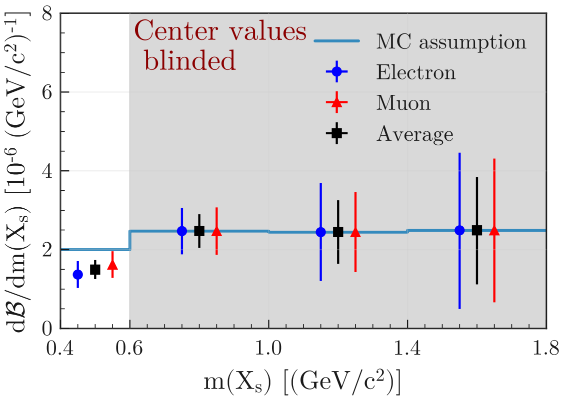


Figure 10.2: Differential branching fraction $d\mathcal{B}/dm(X_s)$. The blue circle and red triangle represent results of electron and muon channels, respectively. The black square is the flavor averaged result. The light blue line is the MC assumptions. Center values except for $m_1(X_s)$ bin are still blinded due to constrains of Belle II collaboration, with blinded values fixed to MC assumption. The error bars represent the square root of quadratic sum of statistical and systematic uncertainties.

10.2 Impact to Physics

Table 10.3: Comparison of sensitivity between this work and previous *BABAR* and Belle analysis. For asymmetric uncertainties, the larger side is taken. Luminosities of analyzed data from each experiment are shown in the parenthesis. Here q_2^2 represents the $1 < q^2 < 6$ [(GeV/c)^2] bin.

Observable		This work (365 fb^{-1})	*BABAR* [30] (428 fb^{-1})	Belle [28] (140 fb^{-1})
$\mathcal{B}_{inc.}$	Stat.	12.6%	10.4%	20.2%
	Sys.	8.7%	9.0%	20.7%
$\mathcal{B}_{q_2^2}$	Stat.	17.4%	25.6%	33.8%
	Sys.	9.4%	15.5%	27.5%

This measurement achieves the best sensitivity in the world. In Tab. 10.3, we summarized the statistical and systematic uncertainties of inclusive branching fraction and differential branching fraction in $1 < q^2 < 6$ [(GeV/c)^2] region from this work, *BABAR* [30] and Belle [28]. For the inclusive branching fraction, comparing to previous *BABAR* measurement, the systematic uncertainty improved from 9.0% to 8.7%, which is not very significant numerically, but we proposed a more reliable data-driven method for estimating fragmentation uncertainty instead of simply varying π^0 and K_S^0 modes by 50%. For the most interesting $1 < q^2 < 6$ (GeV/c)^2 bin, we improved statistical uncertainty significantly from 25.6% (33.8%) to 17.4% comparing to *BABAR* (Belle) even if we have smaller sample luminosity than *BABAR*. This is because of advanced reconstruction and selection algorithms with higher efficiency. Also, there are other reasons such as that *BABAR* has an issue with their muon detector resulting in inefficiency of detecting low-momenta muons and they observe 0-consistent signals in muon channel in this region. The *BABAR* measurement only reconstructs 10 decay modes, which also limits their sensitivity in low-q^2 region. The systematic uncertainty improved from 15.5% (27.5%) to 9.4% comparing to *BABAR* (Belle) thanks to the efforts in reducing systematic uncertainties especially concerning the signal modeling.

Comparison of measured inclusive branching fraction and the branching fraction in $1 < q^2 < 6$ [(GeV/c)^2] region is presented in Fig. 10.3. For the inclusive branching fraction, our result has the best sensitivity with uncertainties comparable with theoretical predictions. For the branching fraction in $1 < q^2 < 6$ [(GeV/c)^2] region, even if our measurement uncertainties are still larger than the theoretical predictions, requiring continuous efforts from the experiment side in the future, we achieved the best sensitivity in the world and we are also better than current world average.

We constrain the new physics parameters $C_{9(10)}^{NP} = C_{9(10)} - C_{9(10)}^{SM}$ using the branching fractions in the cleanest two bins $1 < q^2 < 6$ [(GeV/c)^2] and $14.4 < q^2 < 25$ [(GeV/c)^2]

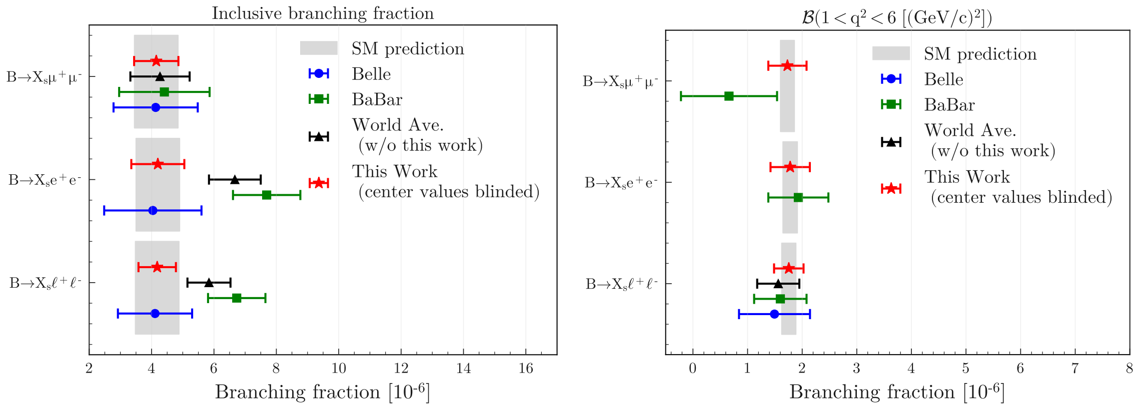


Figure 10.3: Comparison of measured inclusive branching fraction (left) and the branching fraction in $1 < q^2 < 6$ [$(\text{GeV}/c)^2$] region (right). Results from Belle [28] (blue circle), *BABAR* [30] (green square), their average [67] (black triangle) and this work (red star) are presented. The SM prediction for inclusive branching fraction is taken from A. Ali's analysis [68] while the prediction for $1 < q^2 < 6$ [$(\text{GeV}/c)^2$] region is taken from T. Huber et al.'s calculation [32]. Center values of this work are fixed to theoretical predictions.

based on calculations by T. Huber et al. [32]. The two dimensional constrain is shown in Fig. 10.4. In this plot, we show the SM prediction (at $(0, 0)$ by definition), global fit result from Tab. 1.2, and contours of favored region within 1σ by inclusive measurements from Belle [28], *BABAR* [30] and this work (center values fixed to SM prediction). Even though the center values of this work are still blinded, we suppressed the uncertainties significantly and the 1σ area by almost half comparing with previous measurements. The expected 1σ region at Belle II nominal luminosity is also presented and it will be discussed in the next section.

10.3 Possible improvements in future measurements

This analysis is statistically limited with comparable systematic uncertainties. In future measurements, efforts on both sides are of equal importance and they are discussed in this section taking branching fraction at $1 < q^2 < 6$ [$(\text{GeV}/c)^2$] as an example.

10.3.1 Statistical uncertainty

In this work, the statistical uncertainty for muon channel (30%) is larger than that of electron channel (20%) because of muons' relative low particle identification efficiency. At the same time, because of large amount of peaking background introduced by $h \to \mu$ misidentification, it is impossible to lower the muonID selection criteria for high efficiency. As part of the efforts for this dissertation, we developed a new algorithm that improves muon identification performance.

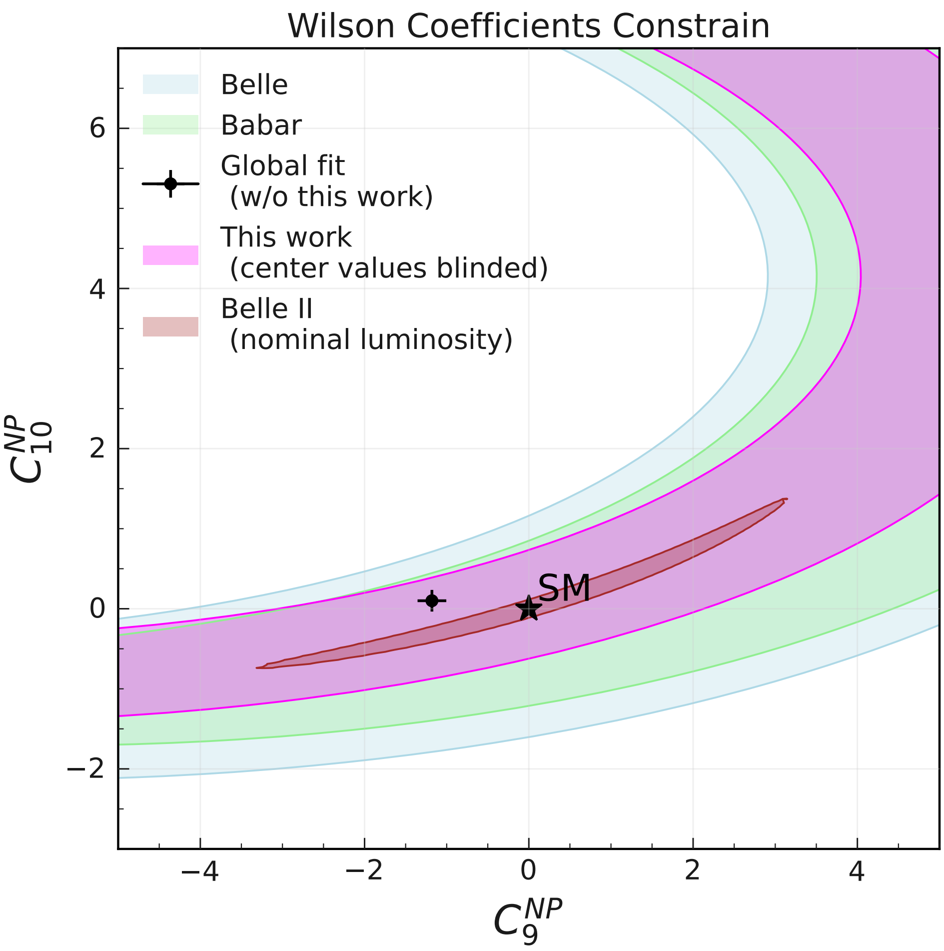


Figure 10.4: 1σ constrains to the new physics parameter C_9^{NP} and C_{10}^{NP} following work of T. Huber et al. [32]. The black star is the SM prediction and the black circle with error bar is the global fit results [5]. The light blue, light green and magenta color bands are constrains from Belle [28], *BABAR* [30] and this work (center values fixed to SM prediction). The brown color band is the expected constrain at the Belle II nominal luminosity ($50\,\mathrm{ab}^{-1}$) assuming SM branching fractions.

Details of this development is described in Appendix. G. As a brief summary, we first discuss the possible room for μ/π identification performance improvement and then present a new method based on Deep Neural Network (DNN). This DNN model utilizes the KLM hit pattern variables as the input and thus can digest the penetration information better than the current algorithm. We test the new method in simulation and find that the pion fake rate (specificity) is reduced from 4.1% to 1.6% at a muon efficiency of 90%. A preliminary validation of this algorithm in a small set of data suggests the pion fake rate reduced from 2.6% to 1.8% while keeping 90% of muon efficiency.

Our simulation sample suggests that this algorithm increases muon identification efficiency from 77% to 92%. If this is validated in data, we can improve the muon identification efficiency for our signal mode with two muons by a factor of 1.4, suppressing current statistical uncertainty from 30% to 25%. Unfortunately, due to limitations from the software and data production of the Belle II collaboration, this algorithm is not implemented for the data analysis presented in this dissertation.

Another method of boosting statistical precision is including larger data samples in analysis. This analyses only uses $365\,\mathrm{fb}^{-1}$ of data collected by Belle II and previous Belle analysis [28] only used $140\,\mathrm{fb}^{-1}$ out of the full data sample of $711\,\mathrm{fb}^{-1}$. By combining the full dataset of Belle and currently collected around $500\,\mathrm{fb}^{-1}$ data from Belle II, the statistical uncertainty can be improved by a factor of 1.6. After 10 years' of operation, Belle II plan to collect $50\,\mathrm{ab}^{-1}$ of data, with which the statistical uncertainty can be suppressed by another factor of 6.4, reducing it from 17.4% to 1.7% of the flavor averaged result.

10.3.2 Systematic uncertainty

The major systematic uncertainty of this analysis is from the extrapolation factor from the 20 visible modes to all decay modes of X_s, followed by uncertainties from the modeling of 20 visible modes. The experimental uncertainties such as B-meson counting and track reconstruction, PID, K_S^0 and π^0 reconstruction, and MVA efficiencies are less significant.

Since the missing mode extrapolation depends on theoretical model, it is irreducible even if Belle II luminosity ($\mathcal{L}$) increases unless future measurements include more modes for measurement or develop better methods of extrapolation.

Systematic uncertainties related to modeling of signal process are dominated by branching fractions of high kaon resonances (3.4%), $K^* - X_s$ transition (3.4%), $\mathcal{B}(B \to K^{(*)}\ell^+\ell^-)$ (2.9%), and fragmentation (1.7%) in the order of magnitude. The branching fractions of high kaon resonances eventually relies on direct measurement from LHCb and in a long time scale, we can assume it decreases following $1/\sqrt{\mathcal{L}}$ as LHCb is also a running experiment. $K^* - X_s$ transition is estimated assuming all X_s is modeled by Pythia in $m(X_s) > 1.1\,\mathrm{GeV}/c^2$ and this is a too conservative estimation as Pythia only contributes to 10% to 30% of our signal modeling. It can be reduced roughly by a constant factor of 3 in the future. $\mathcal{B}(B \to K^{(*)}\ell^+\ell^-)$ relies on LHCb experiment measurements. However, their uncertainties are now dominated by the absolute branching fraction of reference channel $B \to K^{(*)} J/\psi$, and Belle II is the only running experiment that can measure it. Fragmentation uncertainty can be suppressed by direct measurement of different X_s final state so it also decreases as $1/\sqrt{\mathcal{L}}$.

In the future, the experimental uncertainties decreases following $1/\sqrt{\mathcal{L}}$, as most of these uncertainties come from the statistical uncertainties of reference channels.

10.3.3 Summary for discussions on future measurements

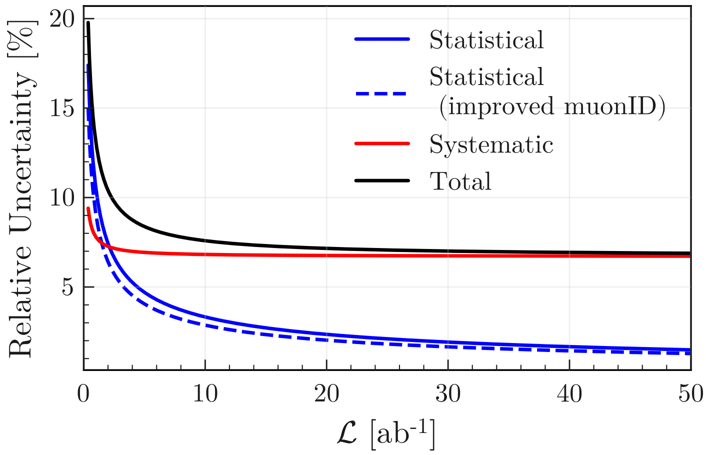


Figure 10.5: Relative uncertainty of $\mathcal{B}(1 < q^2 < 6)$ as a function of Belle II luminosity. The total uncertainty (black) is calculated by the square root of quadratic sum of statistical (blue, solid) and systematic (red) uncertainties. The blue dashed line represents the statistical uncertainty evolution with improved muon identification algorithm.

In conclusion, the uncertainty of branching fraction measurement using methods developed in this dissertation follows:

$$\frac{17.4\%}{\sqrt{\mathcal{L}/365\,\mathrm{fb}^{-1}}}(stat.) \oplus \frac{6.6\%}{\sqrt{\mathcal{L}/365\,\mathrm{fb}^{-1}}}(red.\ sys.) \oplus 6.7\%(irr.\ sys.), \tag{10.2}$$

where the first term is statistical, the second term is reducible systematical uncertainty as luminosity increase, and the third term is the irreducible systematic uncertainty. According to discussions in the last section, the uncertainties from missing mode and $K^* - X_s$ transition terms does not decrease as luminosity increase (thus called irreducible) unless there are developments in the analysis strategy, while the rest of systematic uncertainties are reducible as data luminosity increases. The evolution of relative uncertainties as a function of luminosity is shown in Fig. 10.5. Statistical uncertainty becomes comparable with systematic uncertainty at around $2\,\mathrm{ab}^{-1}$. The total uncertainty reaches the irreducible limit at around $20\,\mathrm{ab}^{-1}$.

The 1σ favored (C_9^{NP}, C_{10}^{NP}) region at Belle II nominal luminosity $(50\,\mathrm{ab}^{-1})$ is also plotted in Fig. 10.4 assuming SM prediction with uncertainties estimated from Eq. 10.2. We can hopefully exclude the global fit result (mainly from exclusive measurements) by 4.5σ under the SM assumption and vise versa.

In the future, as the Belle II experiment collects more data, measurement strategy can be improved in the following aspects:

- including more modes for reconstruction;
- extending $m(X_s)$ cut to values higher than $1.8\,\mathrm{GeV}/c^2$;
- measuring angular decomposition variables (Eq. 1.12);
- measuring branching fraction with denser binning such as splitting the $q^2 \in [1, 6]$ bin into $[1, 3.5]$ and $[3.5, 6]$, where q^2 is in the unit of $(\mathrm{GeV}/c)^2$ [32].

Efforts mentioned above either suppress systematic uncertainties of missing modes or introduce new information in constraining C_9 and C_{10}. Thus, they are promising to enhance the discriminating power of sum-of-exclusive measurements between the SM and global fit to a 5σ level. We leave it as a challenge for people in the future and conclude this dissertation.

Appendix A

Development of $B \to K_J \ell^+ \ell^-$ exclusive decay models in EvtGen

In this appendix, we describe how we implemented the $B \to K_J \ell^+ \ell^-$ exclusive decays into EvtGen framework, where K_J represents a kaon with spin J. In the official release of v02-02-04, which is the adapted by basf2 release-06-01-10 in external library, only the decay for $K_J = K,\ K^*(892)$ can be generated using the BTOSLLBALL generator. In the meantime, decay channels with K_J being higher resonances that are necessary for this analysis. Tab. A.1 shows the decay channels that cannot be generated with current generator in EvtGen.

Table A.1: Missing decay channels in EvtGen. The last two columns are the branching fractions of $K_J \to K\pi$ and $K_J \to K\pi\pi$, respectively. All numbers are taken from PDG.

Decay	J^P	Mass [MeV/c^2]	width [MeV/c^2]	$\mathcal{B}(K\pi)$ [%]	$\mathcal{B}(K\pi\pi)$ [%]
$B \to K^*(1410)\ell^+\ell^-$	1^-	1414 ± 15	232 ± 21	6.6	40
$B \to K_0^*(1430)\ell^+\ell^-$	0^+	1425 ± 50	270 ± 80	93	0
$B \to K_2^*(1430)\ell^+\ell^-$	2^+	1432.4 ± 1.3	109 ± 5	49.9	36.3
$B \to K_1(1270)\ell^+\ell^-$	1^+	1253 ± 7	101 ± 12	0	100
$B \to K_1(1400)\ell^+\ell^-$	1^+	1403 ± 7	174 ± 13	0	100

We start from the free quark decay amplitude in Eq. 1.3 and expand $B \to K_J$ matrix element Eq. 1.4 based on theory papers, and finally express decay amplitude in the following format for all kaons with different spin-parity for easier coding:

$$\mathcal{M}(B \to K_J \ell^+ \ell^-) = \frac{G_F \alpha}{\sqrt{2}\pi} V_{ts}^* V_{tb} m_B [\mathcal{T}_\mu^1 (\bar{\ell}\gamma^\mu \ell) + \mathcal{T}_\mu^2 (\bar{\ell}\gamma^\mu \gamma^5 \ell)], \tag{A.1}$$

absorbing $C_7,\ C_9$ related terms in $\mathcal{T}_\mu^1$ and C_{10} related terms in $\mathcal{T}_\mu^2$.

In the following contents, we use the B-meson rest frame unless otherwise mentioned. We use $p_B,\ p$ and q to denote the 4-momentum of B meson, K meson and di-lepton system. $s = q^2$ denotes the invariant mass of di-lepton system. Hat denotes that the variable is normalized with respect to B mass. For example, $\hat{s} = s/m_B^2$, $\hat{p} = p/m_B$. The metric tensor $g^{\mu\nu} = \mathrm{diag}(+,-,-,-)$ is adapted.

A.1 Scalar Kaon

A.1.1 $J^P = 0^-$

For peudo-scalar kaons, it is already implemented in to the current BTOSLLBALL generator refering to Ref. [6]. For completeness, it is also described in this appendix.

The form factors can be expressed as:

$$\langle K(p)|\bar{s}\gamma_\mu b|B(p_B)\rangle = f_+(s)\left\{(p_B+p)_\mu - \frac{m_B^2 - m_K^2}{s}q_\mu\right\} + f_0(s)\frac{m_B^2 - m_K^2}{s}q_\mu \tag{A.2}$$

noticing that only vector component in quark current is non-zero to keep the same parity on left hand and right hand side. And

$$\langle K(p)|\bar{s}_L\sigma_{\mu\nu}q^\nu b_R|B(p_B)\rangle = \langle K(p)|\bar{s}\sigma_{\mu\nu}q^\nu b|B(p_B)\rangle = i\{(p_B+p)_\mu s - q_\mu(m_B^2 - m_K^2)\}\frac{f_T(s)}{m_B + m_K}. \tag{A.3}$$

And we have:

$$\mathcal{T}_\mu^1 = A(\hat{s})(\hat{p}_\mu + \hat{p}_{B\mu}) + B(\hat{s})\hat{q}_\mu \tag{A.4}$$

$$\mathcal{T}_\mu^2 = C(\hat{s})(\hat{p}_\mu + \hat{p}_{B\mu}) + D(\hat{s})\hat{q}_\mu \tag{A.5}$$

where

$$A(\hat{s}) = C_9^{\mathrm{eff}}(s)f_+(s) + \frac{2\hat{m}_b}{1+\hat{m}_K}C_7^{\mathrm{eff}}f_T(s) \tag{A.6}$$

$$B(\hat{s}) = C_9^{\mathrm{eff}}(s)f_-(s) - \frac{2\hat{m}_b}{\hat{s}}(1-\hat{m}_K)C_7^{\mathrm{eff}}f_T(s) \tag{A.7}$$

$$C(\hat{s}) = C_{10}f_+(\hat{s}) \tag{A.8}$$

$$D(\hat{s}) = C_{10}f_-(\hat{s}) \tag{A.9}$$

A.1.2 $J^P = 0^+$

For scalar kaon case of $K_0^*(1430)$, we refer to Ref. [7].

The form factors can be expressed as:

$$\langle K(p)|\bar{s}\gamma_\mu\gamma_5 b|B(p_B)\rangle = f_+(s)(p_B+p)_\mu + f_-(s)q_\mu. \tag{A.10}$$

Similar to the $J^P = 0^-$ case, only the axial vector current has non-zero contribution due to requirements on parity.

$$\langle K(p)|\bar{s}\sigma_{\mu\nu}\gamma_5 q^\nu b|B(p_B)\rangle = -[2p_\mu q^2 - 2q_\mu(p\cdot q)]\frac{f_T(s)}{m_B + m_K}. \quad (A.11)$$

Using the relation of

$$f_-(s) = \frac{1-\hat{m}_K^2}{\hat{s}}(f_0(s) - f_+(s)), \quad (A.12)$$

one can derive that the expansion of hadron matrix element of $J^P = 0^+$ case is exactly the same as pseudo-scalar case. And hence the expression of $\mathcal{T}_\mu^1$ and $\mathcal{T}_\mu^2$ is also the same.

A.2 Vector Kaon

A.2.1 $J^P = 1^-$

For vector case, it is already implemented in BTOSLLBALL generator for $K^*(892)$ referring to Ref. [6]. For the other vector kaon case $K^*(1410)$, no form factors for this meson is available. Because of this, the form factor for $K^*(892)$ is applied to $K^*(1410)$ directly.

In this section, we give the expression for $\mathcal{T}_1$ and $\mathcal{T}_2$ directly:

$$\mathcal{T}_\mu^1 = A(\hat{s})\epsilon_{\mu\nu\rho\sigma}\epsilon^{*\nu}\hat{p}_B^\rho\hat{p}^\sigma - iB(\hat{s})\epsilon_\mu^* + iC(\hat{s})(\epsilon^*\cdot\hat{p}_B)\hat{p}_\mu + iD(\hat{s})(\epsilon^*\cdot\hat{p}_B)\hat{q}_\mu \quad (A.13)$$

$$\mathcal{T}_\mu^2 = E(\hat{s})\epsilon_{\mu\nu\rho\sigma}\epsilon^{*\nu}\hat{p}_B^\rho\hat{p}^\sigma - iF(\hat{s})\epsilon_\mu^* + iG(\hat{s})(\epsilon^*\cdot\hat{p}_B)\hat{p}_\mu + iH(\hat{s})(\epsilon^*\cdot\hat{p}_B)\hat{q}_\mu \quad (A.14)$$

where $\epsilon_{\mu\nu\rho\sigma}$ (with four indexes) is the Levi-Civita symbol, ϵ_μ (with only one index) is the polarization vector for kaon, and the Lorentz invariant factors are:

$$A(\hat{s}) = \frac{2}{1+\hat{m}_{K_2^*}}C_9^{\text{eff}}(s)V(s) + \frac{4\hat{m}_b}{\hat{s}}C_7^{\text{eff}}T_1(s) \quad (A.15)$$

$$B(\hat{s}) = (1+\hat{m}_{K_2^*})\left[C_9^{\text{eff}}(s)A_1(s) + \frac{2\hat{m}_b}{\hat{s}}(1-\hat{m}_{K_2^*})C_7^{\text{eff}}T_2(s)\right] \quad (A.16)$$

$$C(\hat{s}) = \frac{1}{1-\hat{m}_{K_2^*}^2}\left[(1-\hat{m}_{K_2^*})C_9^{\text{eff}}(s)A_2(s) + 2\hat{m}_b C_7^{\text{eff}}\left(T_3(s) + \frac{1-\hat{m}_{K_2^*}^2}{\hat{s}}T_2(s)\right)\right] \quad (A.17)$$

$$D(\hat{s}) = \frac{1}{\hat{s}}\left\{C_9^{\text{eff}}(s)\left[(1+\hat{m}_{K_2^*})A_1(s) - (1-\hat{m}_{K_2^*})A_2(s) - 2\hat{m}_{K_2^*}A_0(s)\right] - 2\hat{m}_b C_7^{\text{eff}}T_3(s)\right\} \quad (A.18)$$

$$E(\hat{s}) = \frac{2}{1+\hat{m}_{K_2^*}}C_{10}(s)V(s) \quad (A.19)$$

$$F(\hat{s}) = (1+\hat{m}_{K_2^*})C_{10}(s)A_1(s) \quad (A.20)$$

$$G(\hat{s}) = \frac{1}{1+\hat{m}_{K_2^*}}C_{10}(s)A_2(s) \quad (A.21)$$

$$H(\hat{s}) = \frac{1}{\hat{s}}C_{10}(s)\left[(1+\hat{m}_{K_2^*})A_1(s) - (1-\hat{m}_{K_2^*})A_2(s) - 2\hat{m}_{K_2^*}A_0(s)\right] \quad (A.22)$$

A.2.2 $J^P = 1^+$

For axial vector kaons $K_1(1270)$ and $K_1(1400)$, they are mixtures of K_{1A} (1^3P_1) and K_{1B} (1^1P_1) states with mixing angle θ_K:

$$
\begin{aligned}
|K_1(1270)\rangle &= |K_{1A}\rangle \sin\theta_K + |K_{1B}\rangle \cos\theta_K \\
|K_1(1400)\rangle &= |K_{1A}\rangle \cos\theta_K - |K_{1B}\rangle \sin\theta_K.
\end{aligned}
\tag{A.23}
$$

In this analysis, we adapt the convention that $\theta_K = 45°$.

In Ref. [8], only the form factors of K_{1A} and K_{1B} are given. Since each term of matrix element expansion is proportional to the form factors, we can derive the form factors of $K_1(1270)$ and $K_1(1400)$ by mixing the form factors of K_{1A} and K_{1B} using Eq. A.23. The matrix element expansions are expressed as:

$$
\begin{aligned}
\langle K_1(p)|\bar{s}\gamma^\mu(1-\gamma^5)b|B(p_B)\rangle =& i\frac{2A_0(q^2)}{m_B - m_{K_1}}\epsilon^{\mu\nu\rho\sigma}\epsilon^*_\nu p_{B\rho}p_\sigma - 2m_{K_1}V_0(q^2)\frac{\epsilon^*\cdot q}{q^2}q^\mu \\
&- (m_B - m_{K_1})V_1(q^2)\left[\epsilon^{*\mu} - \frac{\epsilon^*\cdot q}{q^2}q^\mu\right] \\
&+ \frac{V_2(q^2)}{m_B - m_{K_1}}(\epsilon^*\cdot q)\left[(p_B+p)^\mu - \frac{m_B^2 - m_{K_1}^2}{q^2}q^\mu\right],
\end{aligned}
\tag{A.24}
$$

$$
\begin{aligned}
\langle K_1(p)|\bar{s}\sigma^{\mu\nu}(1+\gamma_5)q_\nu b|B(p_B)\rangle =& -2T_1(q^2)\epsilon^{\mu\nu\rho\sigma}\epsilon^*_\nu p_{B\rho}p_\sigma \\
&- iT_2(q^2)\left[(m_B^2 - m_{K_1}^2)\epsilon^{*\mu} - (\epsilon^*\cdot q)(p_B+p)^\mu\right] \\
&- iT_3(q^2)(\epsilon^*\cdot q)\left[q^\mu - \frac{q^2}{m_B^2 - m_{K_1}^2}(p_B+p)^\mu\right],
\end{aligned}
\tag{A.25}
$$

We can derive the $\mathcal{T}^1_\mu$ and $\mathcal{T}^2_\mu$ as following.

$$
\mathcal{T}^1_\mu = iA(\hat{s})\epsilon^{\mu\nu\rho\sigma}\epsilon^*_\nu p_{B\rho}p_\sigma - B(\hat{s})\epsilon^*_\mu + C(\hat{s})(\epsilon^*\cdot\hat{q})(\hat{p}_B+\hat{p})_\mu + D(\hat{s})(\epsilon^*\cdot\hat{q})\hat{q}_\mu \tag{A.26}
$$

$$
\mathcal{T}^2_\mu = iE(\hat{s})\epsilon^{\mu\nu\rho\sigma}\epsilon^*_\nu p_{B\rho}p_\sigma - F(\hat{s})\epsilon^*_\mu + G(\hat{s})(\epsilon^*\cdot\hat{q})(\hat{p}_B+\hat{p})_\mu + H(\hat{s})(\epsilon^*\cdot\hat{q})\hat{q}_\mu \tag{A.27}
$$

with

$$
A(\hat{s}) = \frac{2}{1-\hat{m}_{K_1}}C_9^{\text{eff}}(s)A_0(s) + \frac{4\hat{m}_b}{\hat{s}}C_7^{\text{eff}}T_1(s) \tag{A.28}
$$

$$
B(\hat{s}) = (1-\hat{m}_{K_1})\left[C_9^{\text{eff}}(s)V_1(s) + \frac{2\hat{m}_b}{\hat{s}}(1+\hat{m}_{K_1})C_7^{\text{eff}}T_2(s)\right] \tag{A.29}
$$

$$
C(\hat{s}) = \frac{1}{1-\hat{m}_{K_1}^2}\left[(1+\hat{m}_{K_1})C_9^{\text{eff}}(s)V_2(s) + 2\hat{m}_b C_7^{\text{eff}}\left(T_3(s) + \frac{1-\hat{m}_{K_1}^2}{\hat{s}}T_2(s)\right)\right] \tag{A.30}
$$

$$
D(\hat{s}) = \frac{1}{\hat{s}}\left\{C_9^{\text{eff}}(s)\left[(1-\hat{m}_{K_2^*})V_1(s) - (1+\hat{m}_{K_2^*})V_2(s) - 2\hat{m}_{K_2^*}V_0(s)\right] - 2\hat{m}_b C_7^{\text{eff}}T_3(s)\right\} \tag{A.31}
$$

$$E(\hat{s}) = \frac{2}{1-\hat{m}_{K_2^*}} C_{10}(s) A_0(s) \tag{A.32}$$

$$F(\hat{s}) = (1-\hat{m}_{K_2^*}) C_{10}(s) V_1(s) \tag{A.33}$$

$$G(\hat{s}) = \frac{1}{1-\hat{m}_{K_2^*}} C_{10}(s) V_2(s) \tag{A.34}$$

$$H(\hat{s}) = \frac{1}{\hat{s}} C_{10}(s) \left[(1-\hat{m}_{K_2^*})V_1(s) - (1+\hat{m}_{K_2^*})V_2(s) - 2\hat{m}_{K_2^*} V_0(s)\right] \tag{A.35}$$

A.3 Tensor Kaon

For the case of $K_2^*(1430)$, we refer to paper Ref. [9].

For tensor particles, their polarization tensor $\epsilon_{\mu\nu}$ (with two indexes) satisfies $\epsilon_{\mu\nu}p^\nu = 0$. This polarization tensor can be constructed using the spin-1 polarization vector ϵ_μ:

$$\begin{aligned} \epsilon_{\mu\nu}(\pm 2) =& \epsilon_\mu(\pm)\epsilon_\nu(\pm), \\ \epsilon_{\mu\nu}(\pm 1) =& \frac{1}{\sqrt{2}}[\epsilon_\mu(\pm)\epsilon_\nu(0) + \epsilon_\nu(\pm)\epsilon_\mu(0)], \\ \epsilon_{\mu\nu}(0) =& \frac{1}{\sqrt{6}}[\epsilon_\mu(+)\epsilon_\nu(-) + \epsilon_\nu(+)\epsilon_\mu(-)] + \sqrt{\frac{2}{3}}\epsilon_\mu(0)\epsilon_\nu(0). \end{aligned} \tag{A.36}$$

At the rest frame of the kaon, the Lorentz vector ϵ_μ can be expressed as following when adopting the circular polarization representation:

$$\epsilon^\mu(0) = (0,0,0,1)^T, \quad \epsilon^\mu(\pm) = \frac{1}{\sqrt{2}}(0, \mp 1, -i, 0)^T. \tag{A.37}$$

By using this representation, the polarization tensors can be expressed as:

$$\begin{aligned} &\epsilon_{\mu\nu}^{(+2)} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & i & 0 \\ 0 & i & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \epsilon_{\mu\nu}^{(-2)} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & -i & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ &\epsilon_{\mu\nu}^{(+1)} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -i \\ 0 & -1 & -i & 0 \end{pmatrix}, \quad \epsilon_{\mu\nu}^{(-1)} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -i \\ 0 & 1 & -i & 0 \end{pmatrix}, \\ &\epsilon_{\mu\nu}^{(0)} = \frac{1}{\sqrt{6}}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}. \end{aligned} \tag{A.38}$$

In the following calculations, we boost the kaon to the rest frame of B meson and denote B meson and kaon meson momentum to be p_B and p, respectively. Under this frame,

$\epsilon_\mu(0)$ is boosted to $(\beta\gamma, 0, 0, \gamma)$ but $\epsilon_\mu(\pm)$ is left untouched, where $\beta = p_{K_2^*}/E_{K_2^*}$, $\gamma = 1/\sqrt{1-\beta^2}$. We define a new polarization vector which is useful in the decomposition of form factors:

$$\epsilon_{T\mu}(h) = \frac{1}{m_B}\epsilon_{\mu\nu}(h)p_B^\nu. \tag{A.39}$$

Considering the fact that $\epsilon_\mu(0)\cdot p_B/m_B = \beta\gamma$, ϵ_T can be written in a more simple format:

$$\epsilon_{T\mu}(\pm 2) = 0, \quad \epsilon_{T\mu}(\pm 1) = \frac{\beta\gamma}{\sqrt{2}}\epsilon_\mu(\pm), \quad \epsilon_{T\mu}(0) = \sqrt{\frac{2}{3}}\beta\gamma\epsilon_\mu(0) \tag{A.40}$$

Using the ingredients mentioned above, matrix elements of $B \to K_2^*$ transition can be expressed as following:

$$\langle K_2^*(p)|\bar{s}\gamma^\mu b|B(p_B)\rangle = -\frac{2V(q^2)}{m_B + m_{K_2^*}}\epsilon^{\mu\nu\rho\sigma}\epsilon^*_{T\nu}p_{B\rho}p_\sigma, \tag{A.41}$$

$$\begin{aligned}\langle K_2^*(p)|\bar{s}\gamma^\mu\gamma_5 b|B(p_B)\rangle =& 2im_{K_2^*}A_0(q^2)\frac{\epsilon_T^*\cdot q}{q^2}q^\mu + i(m_B + m_{K_2^*})A_1(q^2)\left[\epsilon^*_{T\mu} - \frac{\epsilon_T^*\cdot q}{q^2}q^\mu\right] \\ &- iA_2(q^2)\frac{\epsilon_T^*\cdot q}{m_B + m_{K_2^*}}\left[(p_B+p)^\mu - \frac{m_B^2 - m_{K_2^*}^2}{q^2}q^\mu\right],\end{aligned} \tag{A.42}$$

$$\langle K_2^*(p)|\bar{s}\sigma^{\mu\nu}q_\nu b|B(p_B)\rangle = -2iT_1(q^2)\epsilon^{\mu\nu\rho\sigma}\epsilon^*_{T\nu}p_{B\rho}p_\sigma, \tag{A.43}$$

$$\begin{aligned}\langle K_2^*(p)|\bar{s}\sigma^{\mu\nu}\gamma_5 q_\nu b|B(p_B)\rangle =& T_2(q^2)\left[(m_B^2 - m_{K_2^*}^2)\epsilon^*_{T\mu} - \epsilon_T^*\cdot q(p_B+p)^\mu\right] \\ &+ T_3(q^2)\epsilon_T^*\cdot q\left[q^\mu - \frac{q^2}{m_B^2 - m_{K_2^*}^2}(p_B+p)^\mu\right],\end{aligned} \tag{A.44}$$

It is interesting to notice that the form factors of vector current $V^\mu(= \bar{s}\gamma^\mu b$ or $\bar{s}\sigma^{\mu\nu}q_\nu b)$ are in the format of $\epsilon^{\mu\nu\rho\sigma}\epsilon^*_{T\nu}p_{B\rho}p_\sigma$, which is an axial current. This is because the parity on the left-hand side of has an parity of $P(K_2^*)P(V)P(B) = (+)(-)(-) = (+)$. For the axial vector current $A^\mu(= \bar{s}\gamma^\mu\gamma_5 b$ or $\bar{s}\sigma^{\mu\nu}\gamma_5 q_\nu b)$, the form factors are vectors for the same reason.

We can derive the $\mathcal{T}_\mu^1$ and $\mathcal{T}_\mu^2$ as following.

$$\mathcal{T}_\mu^1 = -A(\hat{s})\epsilon_{\mu\nu\rho\sigma}\epsilon_T^{*\,\nu}\hat{p}_B^\rho\hat{p}^\sigma - iB(\hat{s})\epsilon^*_{T\mu} + iC(\hat{s})(\epsilon_T^*\cdot\hat{q})(\hat{p}_B+\hat{p})_\mu + iD(\hat{s})(\epsilon_T^*\cdot\hat{q})\hat{q}_\mu \tag{A.45}$$

$$\mathcal{T}_\mu^2 = -E(\hat{s})\epsilon_{\mu\nu\rho\sigma}\epsilon_T^{*\,\nu}\hat{p}_B^\rho\hat{p}^\sigma - iF(\hat{s})\epsilon^*_{T\mu} + iG(\hat{s})(\epsilon_T^*\cdot\hat{q})(\hat{p}_B+\hat{p})_\mu + iH(\hat{s})(\epsilon_T^*\cdot\hat{q})\hat{q}_\mu \tag{A.46}$$

with the expression of Lorentz invariant factors $A(\hat{s}),\ B(\hat{s}),\ ,...,\ H(\hat{s})$ being exactly the same expression as the vector kaon case (Eq. A.15–A.22). It is interesting to find that this is in a very similar format as the vector case of Eq. A.13, A.14. This is because the helicity ± 2 of K_2 meson is prohibited in the $B \to K_2 \ell^+ \ell^-$ decay and thus fails to introduce extra degrees of freedom.

The numerical expression of form factors are adapted according to Eq.6, Eq.7 and Table. I of Ref. [9].

In EvtGen, a different polarization tensor representation, denoted as $e_{\mu\nu}(h), h = 0, 1, ..., 4$, is adapted for tensor particles:

$$
\begin{aligned}
e_{\mu\nu}(0) &= \frac{1}{\sqrt{6}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \quad e_{\mu\nu}(1) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
e_{\mu\nu}(2) &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad e_{\mu\nu}(3) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \\
e_{\mu\nu}(4) &= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.
\end{aligned}
\tag{A.47}
$$

And it can be related to the $\epsilon_{\mu\nu}$ representation by:

$$
\begin{aligned}
e_{\mu\nu}(0) &= \epsilon_{\mu\nu}(0), \\
e_{\mu\nu}(1) &= \frac{1}{\sqrt{2}} \left(\epsilon_{\mu\nu}(+2) + \epsilon_{\mu\nu}(-2)\right), \\
e_{\mu\nu}(2) &= \frac{1}{\sqrt{2}i} \left(\epsilon_{\mu\nu}(+2) - \epsilon_{\mu\nu}(-2)\right), \\
e_{\mu\nu}(3) &= -\frac{1}{\sqrt{2}} \left(\epsilon_{\mu\nu}(+1) - \epsilon_{\mu\nu}(-1)\right), \\
e_{\mu\nu}(4) &= -\frac{1}{\sqrt{2}i} \left(\epsilon_{\mu\nu}(+1) + \epsilon_{\mu\nu}(-1)\right).
\end{aligned}
\tag{A.48}
$$

Considering the fact that Decay amplitude $\mathcal{M}$ is an linear function of ϵ_T^* (Eq. A.45, A.46), and that ϵ_T^* is an linear function of $\epsilon_{\mu\nu}$ (Eq. A.39), we can easily convert the amplitude calculated under $\epsilon_{\mu\nu}$ representation to that under $e_{\mu\nu}$ representation using the relations in Eq. A.48.

We use the $\overline{B}^0 \to K_2^*(1430)^0 (\to K^- \pi^+) \mu^+ \mu^-$ decay to examine our model by comparing the generated MC sample with theoretical distributions. We use the modified BTOSLLBALL generator to generate $\overline{B}^0 \to K_2^*(1430)^0 \mu^+ \mu^-$ decay and use TSS model to generate the following $K_2^*(1430)^0 \to K^- \pi^+$ decay. Angular variables in this decay is defined in Fig. 1.4.

The distributions of $d\Gamma/dq^2$, $d\Gamma/d\cos\theta_\ell$, and $d\Gamma/d\cos\theta_K$ in the MC and the theoretical predictions in Eq.10 of Ref. [9] are shown in Fig. A.1. In addition, the forward-backward asymmetry as a function of q^2 is also shown in Fig. A.1, noticing that there is an additional minus sign due to different definitions of $\cos\theta_\ell$. We find the generated distributions are in good consistency with theoretical predictions.

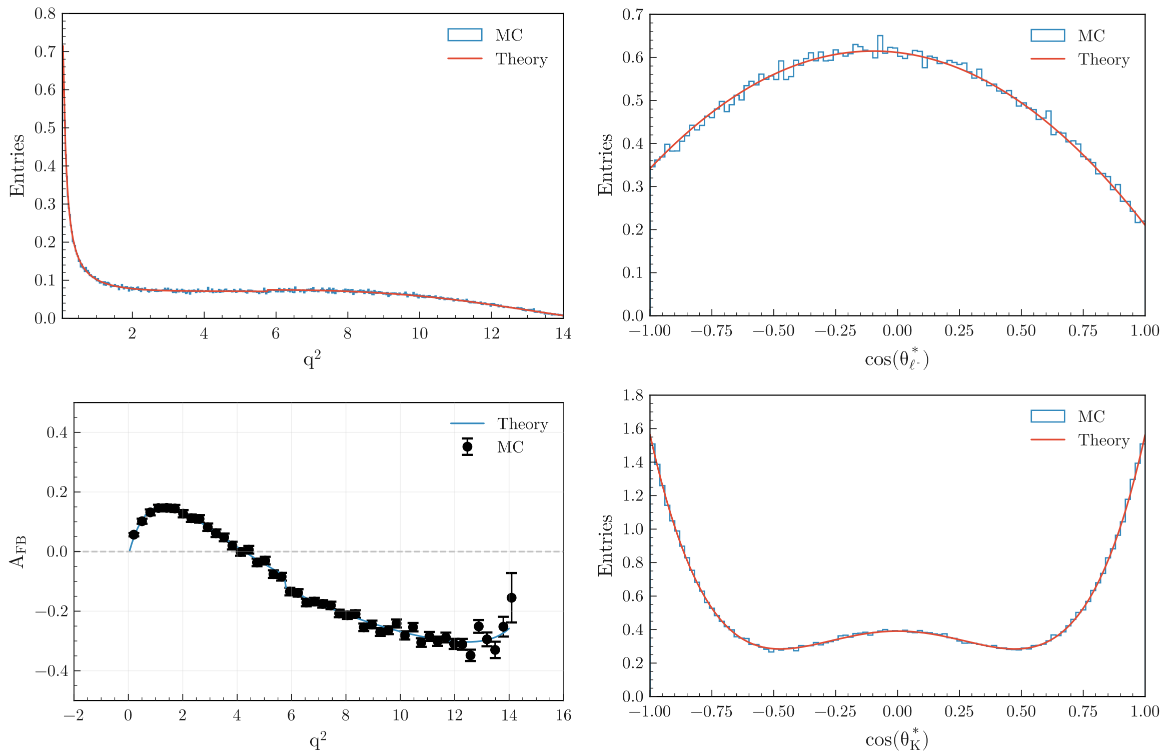


Figure A.1: Comparison of generated MC distributions with theoretical predictions.

We also compared the 2-dimensional distributions of $(\cos\theta_K, \cos\theta_\ell)$ between MC and theory and good consistency is confirmed.

Appendix B

Comparison of reconstruction efficiency between Pythia-based samples and resonance samples in $m(X_s) > 1.1\,\mathrm{GeV}/c^2$ region

In this appendix we compare the reconstruction efficiency between pure Pythia sample and pure resonant sample of each $K\pi$ and $K2\pi$ final state. The result in shown in Fig. B.1 and we find good consistency between pythia and resonance samples.

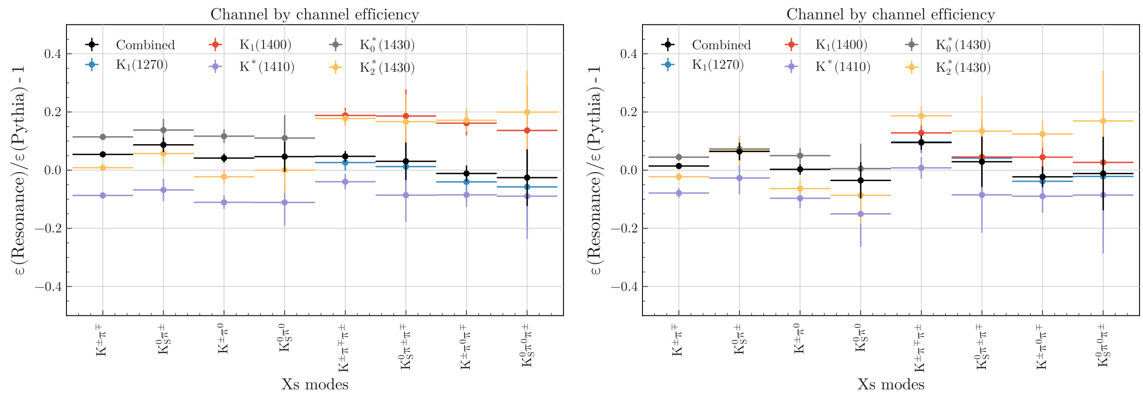


Figure B.1: Efficiency variation of each $K\pi$ and $K2\pi$ final states when produced in pure Pythia sample and combined resonance sample (black). The efficiency variation of each single resonance is also plotted (colored). The left plot is the inclusive result and the right plot is the result in the most interesting $1 < q^2 < 6$ $[(\mathrm{GeV}/c)^2]$ region.

Appendix C

Data/MC consistency check for MVA input variables

This section shows the data/mc of all training variables. Good Data/MC consistency is confirmed in the control channel $B \to J/\psi\, X_s$.

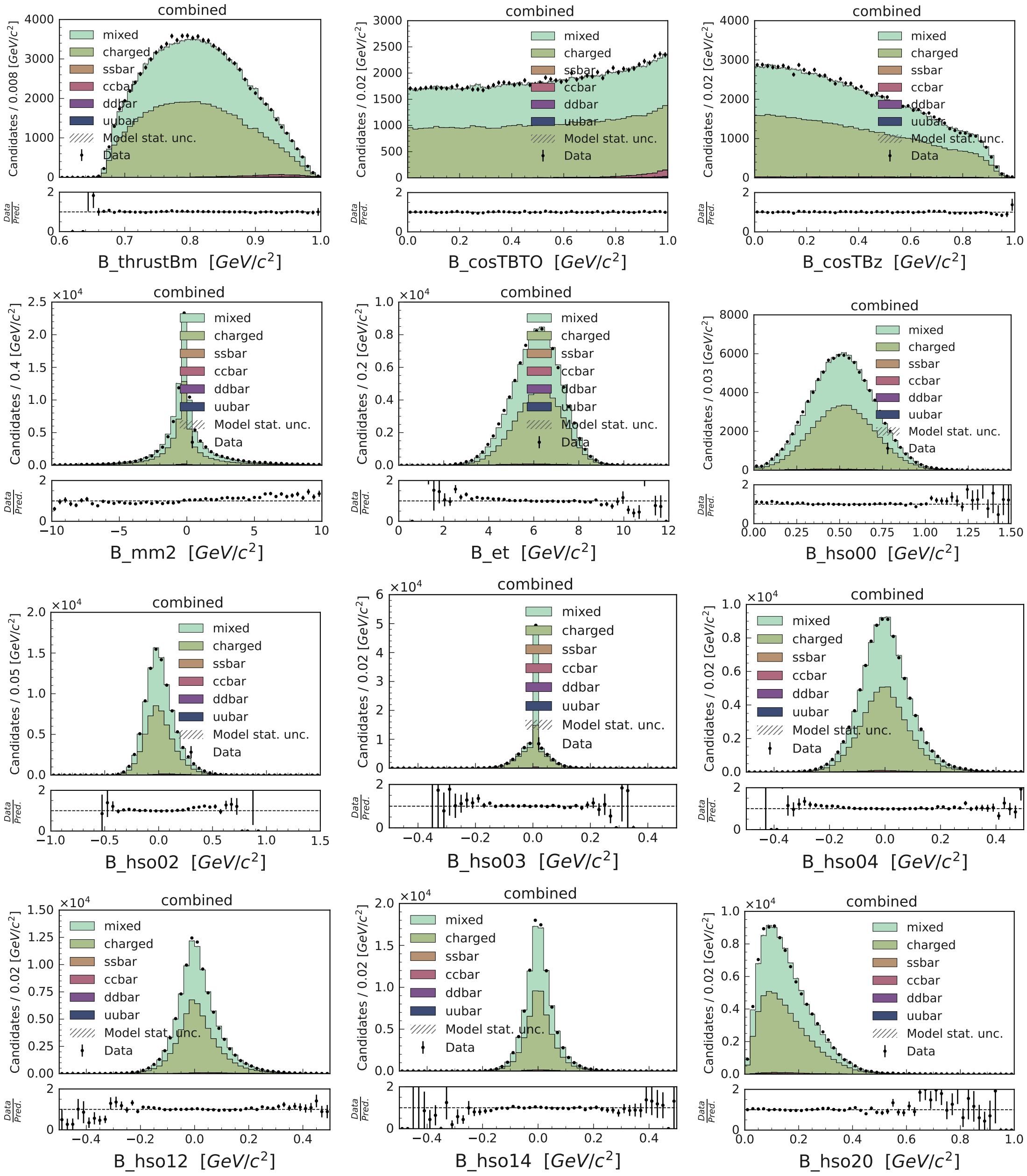


Figure C.1: Data/mc of training variables in $B \to J/\psi\, X_s$ sample.

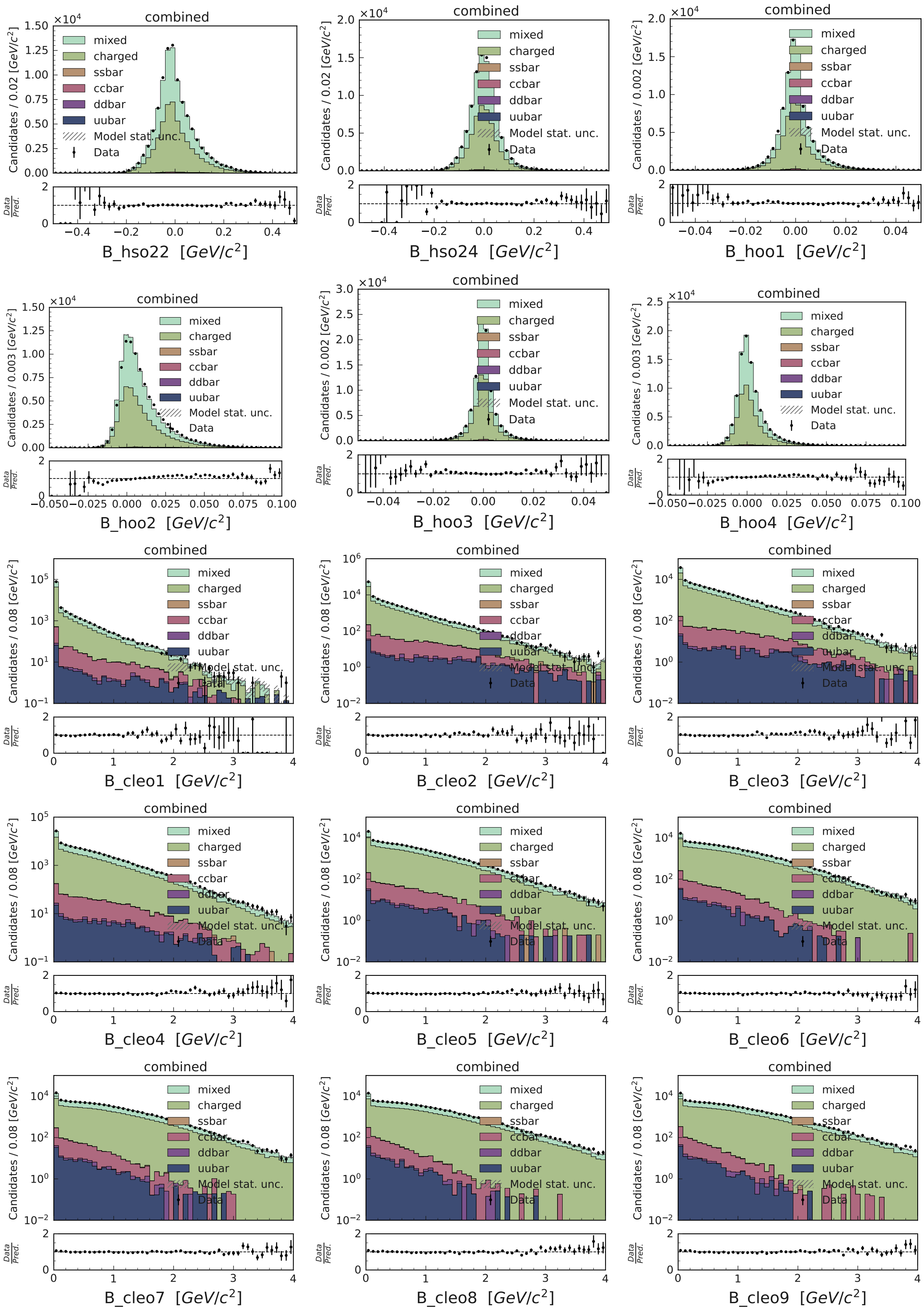


Figure C.2: Data/mc of training variables in $B \to J/\psi\, X_s$ sample.

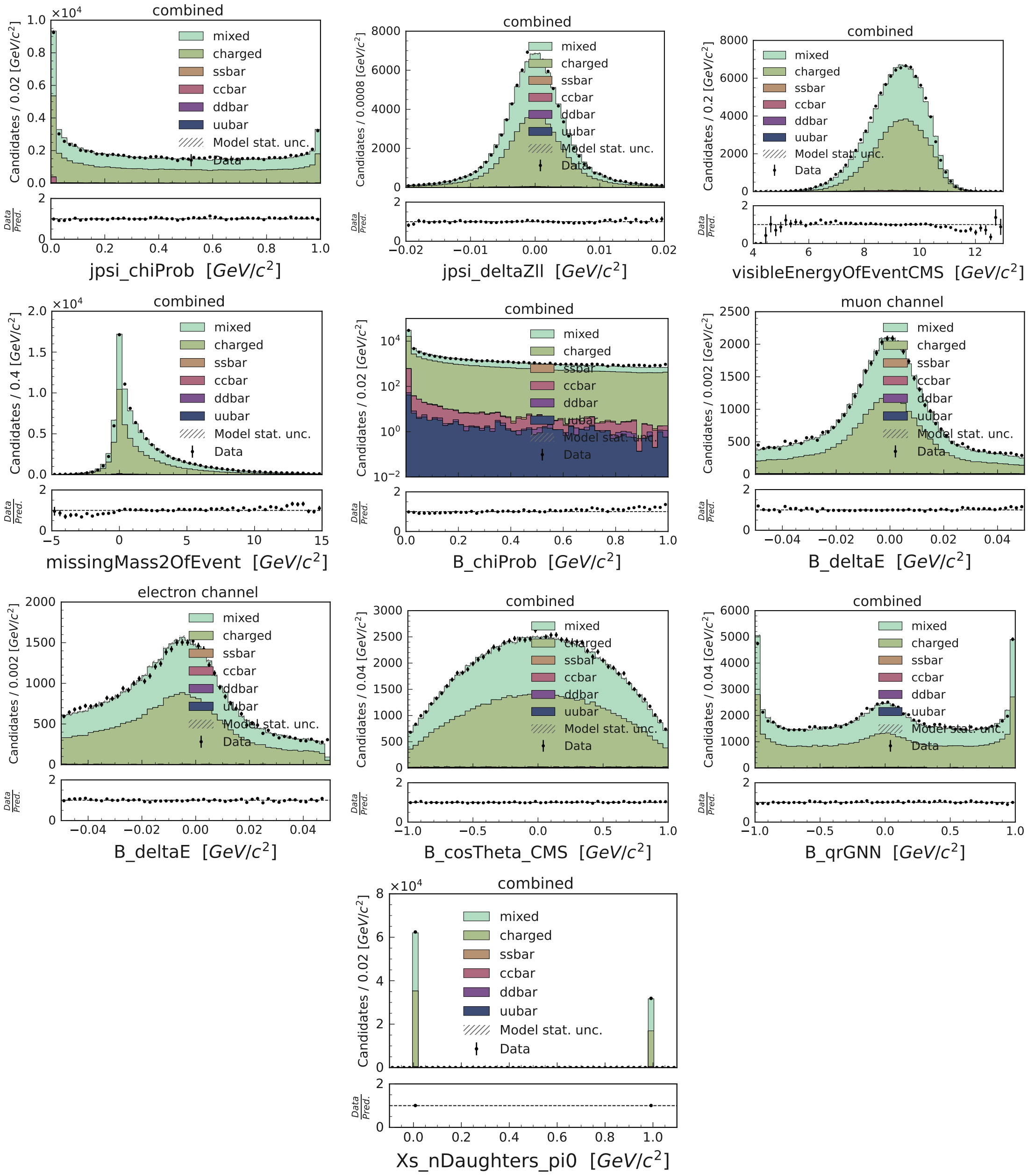


Figure C.3: Data/mc of training variables in $B \to J/\psi\, X_s$ sample.

Appendix D

Estimation of muon to hadron fake rate via $e^+e^- \to \mu^+\mu^-\gamma$ process

The fake rate of muons to hadrons, which is the probability to mis-identify a muon as a hadron, is not given by the PID group for MC15rd and proc13prompt dataset. In this appendix, we explain the method to estimate this fake rate using the $e^+e^- \to \mu^+\mu^-\gamma$ process.

D.1 Dataset

The full run 1 dataset including on-resonance, off-resonance and $\Upsilon(5S)$ scan data are used in this analysis. For MC samples, besides the signal $e^+e^- \to \mu^+\mu^-\gamma$ process, background processes including $e^+e^- \to hh\gamma$ ($h = K, \pi$), $e^+e^- \to \tau^+\tau^-$, $e^+e^- \to \mu^+\mu^-$, $e^+e^- \to e^+e^-(\gamma)$, $e^+e^- \to c\bar{c}$ are used as well. A full list of MC samples as well as the corresponding luminosity are shown in Table D.1.

Table D.1: MC samples and corresponding luminosity.

Samples	$\mu^+\mu^-\gamma$	$hh\gamma$	taupair	$\mu^+\mu^-$	$e^+e^-(\gamma)$	$c\bar{c}$
Luminosity (fb^{-1})	1689.852	422.463	1689.852	422.463	168.985	1689.852

D.2 Selection criteria

The basf2 version of light-2303-iriomote is used to analyze these samples. All events are firstly required to passes the radmumu hlt skim[1]. After that, we require there are only two good tracks in each event, where the good track should satisfy $|dr| < 2$ cm and

[1] `SoftwareTriggerResult(software_trigger_cut&skim&accept_radmumu)`

$|dz| < 5$ cm. In addition, we required the ISR photon to have an energy greater than 1.0 GeV, `clusterNHits` > 1.5, and the polar angle is within CDC acceptance.

D.3 Analysis strategy and results

A tag and probe method is applied to extract the fake rate. We firstly apply muonID_noSVD > 0.9 to one of the two muons in the event, called tag side. The number of remaining events is labeled as N_{tag}. Then we apply the hadronID selections, pionID > 0.4 or kaonID > 0.6, to the other muon, named as probe side. The number of remaining events after selections on both tag and probe side is named as N_{probe}. In MC samples, the number of tag and probe side from signal process can be derived from truth informatino and thus the fake rate is defined as:

$$f_{\mathrm{MC}} = \frac{N_{\mathrm{probe}}}{N_{\mathrm{tag}}}. \tag{D.1}$$

For the data samples, the fake rate is defined by extracting background contributions from N_{tag} and N_{probe}:

$$f_{\mathrm{data}} = \frac{N_{\mathrm{probe}} - N_{\mathrm{probe}}^{\mathrm{bkg}}}{N_{\mathrm{tag}} - N_{\mathrm{tag}}^{\mathrm{bkg}}} = \frac{N_{\mathrm{probe}} - \sum_i n_{\mathrm{probe}}^i \cdot r^i}{N_{\mathrm{tag}} - \sum_i n_{\mathrm{tag}}^i \cdot r^i}, \tag{D.2}$$

where i runs among background processes as indicated in Table D.1, $n_{\mathrm{tag(probe)}}^i$ denotes the number of remaining background events in the corresponding MC sample after tag (probe) selection, and r^i denotes the fake rate correction of hadrons passing the muonID cut on tag side. For background processes from lepton processes like $\mu^+\mu^-$, $e^+e^-(\gamma)$, correction factors are fixed to 1. Usually, contributions from background process is smaller than 0.1%.

The results of MC and data samples are shown in Fig. D.1 and Fig. D.2, respectively.

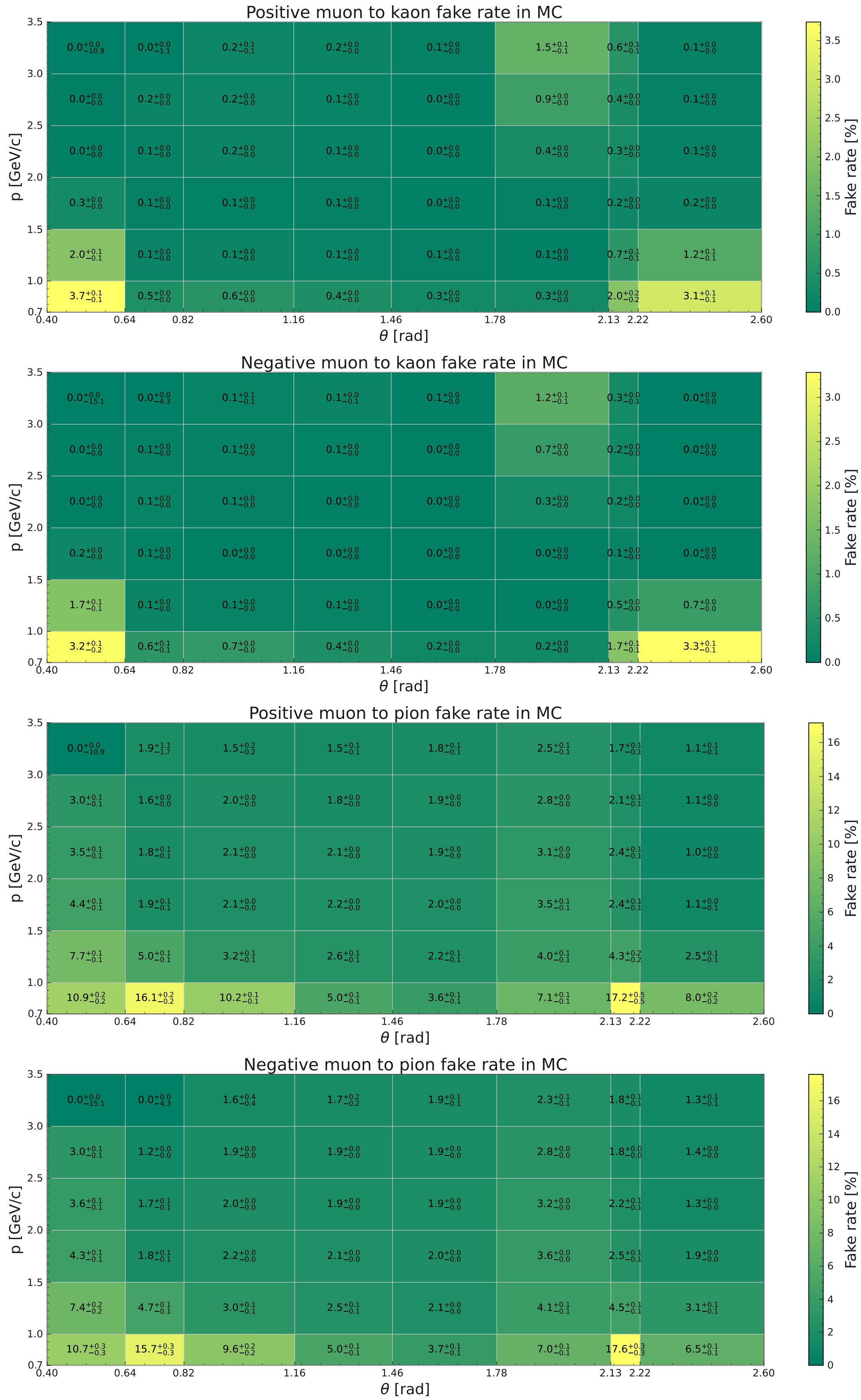


Figure D.1: Estimation of muon to hadron fake rate in MC15rd.

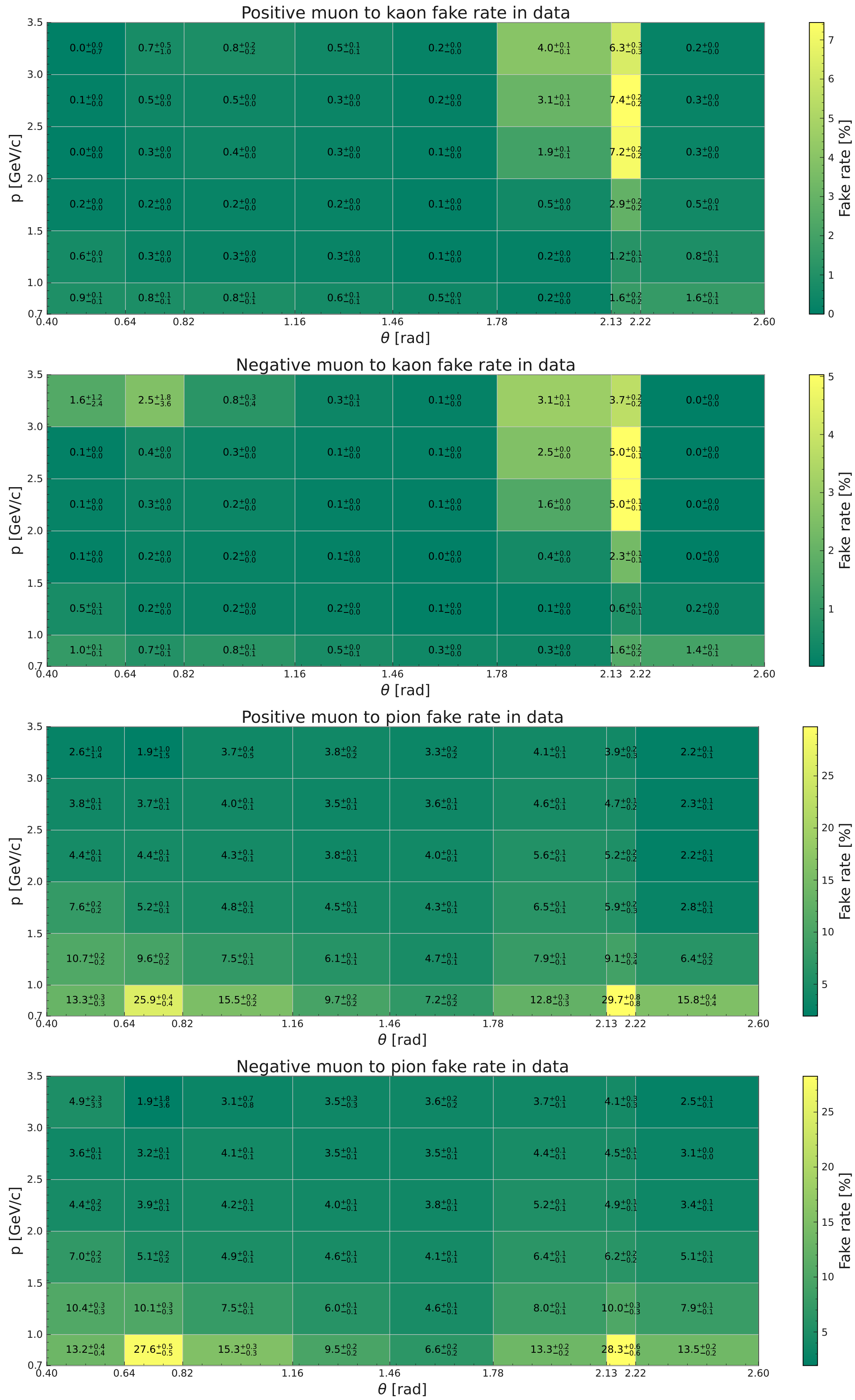


Figure D.2: Estimation of muon to hadron fake rate in data.

Appendix E

Calibration of K_S^0 selector

In this appendix, we explain the method to estimate the data/MC ratio in selection efficiency for the Ksselector tool.

E.1 Dataset

The run 1 dataset including on-resonance and off-resonance data are used in this analysis. For MC samples, the simulated $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$, charged and mixed samples are used.

E.2 Selection criteria

We use the high purity $D^{*+} \to D^0(\to K_S^0\pi^+\pi^-)\pi_s^+$ channel for this calibration. Following the official calibration procedure [93], we applied the following selection criteria in reconstruction of this channel:

- pions $\pi^\pm$: $dr < 2$ cm, $|dz| < 4$ cm, and $p_t > 100\,\mathrm{MeV}/c$.
- slow pion π_s^+: $dr < 2$ cm and $|dz| < 4$ cm
- K_S^0: $0.48 < M < 0.52$ [GeV/c^2] and $|dz| < 4$ cm
- D^0: $1.85 < M < 1.88$ [GeV/c^2]
- D^{*+}: $p^* > 1.6\,\mathrm{GeV}/c$ and $0.144 < \Delta M < 0.147$ [GeV/c^2], where $\Delta M = M(D^{*+}) - M(D^0)$.

E.3 Analysis strategy and results

This calibration is performed to K_S^0 candidates in 93 different kinematic bins whose boundaries are defined in Table. E.1.

As explained in Sec. 9.3.3, the correction factors for K_S^0 reconstruction and selection efficiencies can be derived by Eq. 9.16. The signal yields ($n_{D(M)}^{(\prime)}(p, \cos\theta, d)$) in the formula

Table E.1: For low momentum $p < 1$ GeV/c region, the last three bins of vertex distance are merged to compensate for the lower statistics.

Variable	**Ranges**
Momentum (GeV/c)	$(0, 1)$, $(1, 2)$, $(2, 10)$
$\cos\theta$	$(-1.0, 0.3)$, $(0.3, 0.7)$, $(0.7, 1.0)$
3D vertex distance (cm)	$(0, 0.5)$, $(0.5, 1.4)$, $(1.4, 2.4)$, $(2.4, 3.6)$, $(3.6, 5.2)$, $(5.2, 7.0)$ $(7.0, 9.0)$, $(9.0, 13.)$, $(13., 22.)$, $(22., 35.)$, $(35., 55.)$

are obtained by performing extended unbinned maximal likelihood fit to the K_S^0 invariant mass spectrum. The signal is described by a double-Gaussian function with shared center value, while the background is described by a 1st order Chebyshev polynomial function. Statistical uncertainties from fitting are propagated to the total correction factor.

Two kinds of systematic uncertainties are considered in the official K_S^0 calibration: fitting model and the assumption that Eq. 9.13 holds. Both are also applicable to this analysis and for this reason, we adapt the official systematic uncertainty values in each kinematic bin to our results directly.

Correction factors in each kinematic bin of K_S^0 is shown in Fig. E.1. Both statistical uncertainty and systematic uncertainty are presented in these plots.

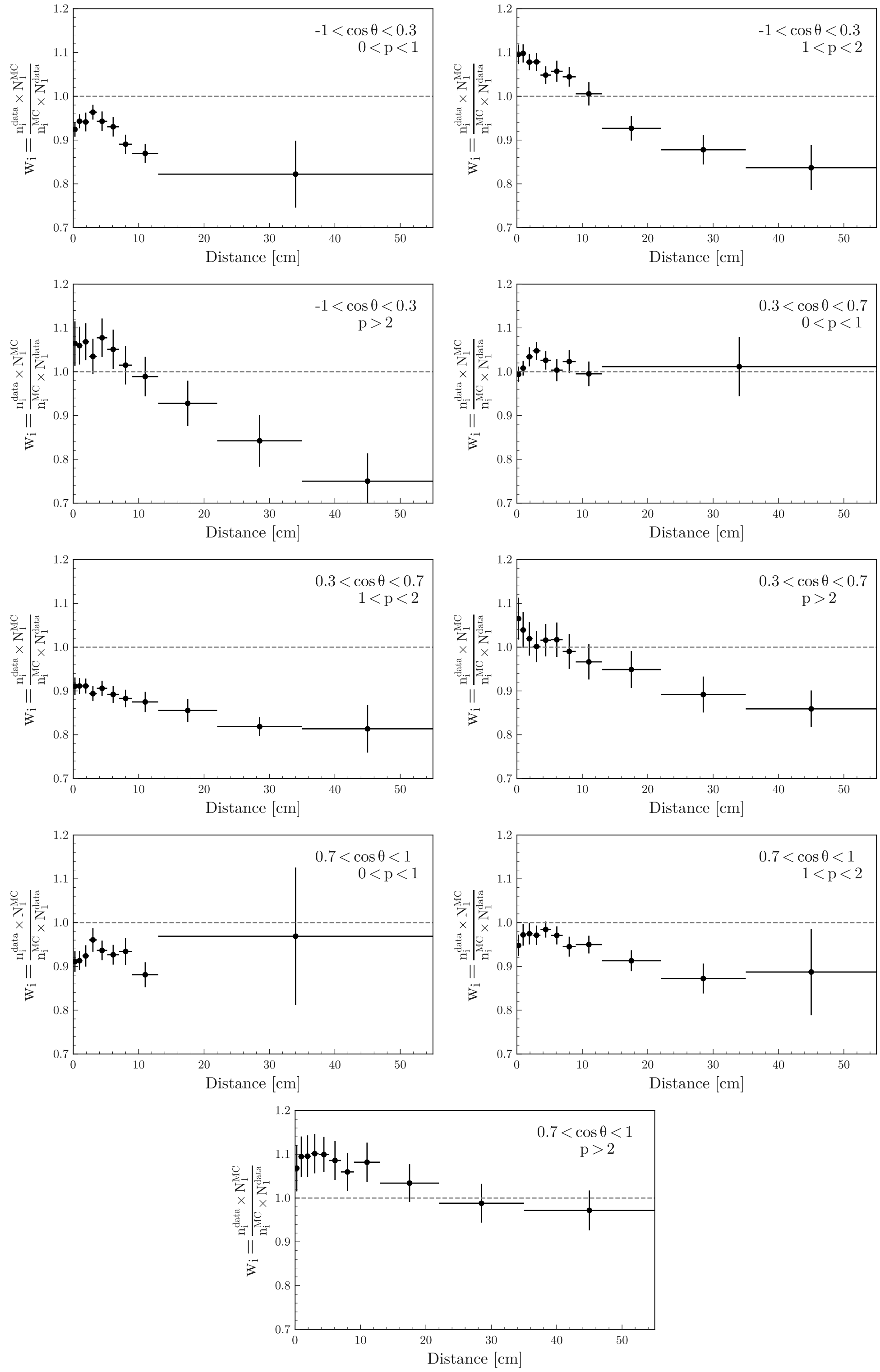


Figure E.1: Weights applied to K^0_s reconstruction. Error bar includes both statistical and systematic uncertainties.

Appendix F

Fitting templates

In this appendix, we present the fitting templates for self cross-feed events and three kinds of peaking background events: double mis-ID, swap mis-ID, and charmonium. These templates are prepared separately for each lepton flavor in each kinematic bins defined in Tab. 3.2.

F.1 Self cross-feed events

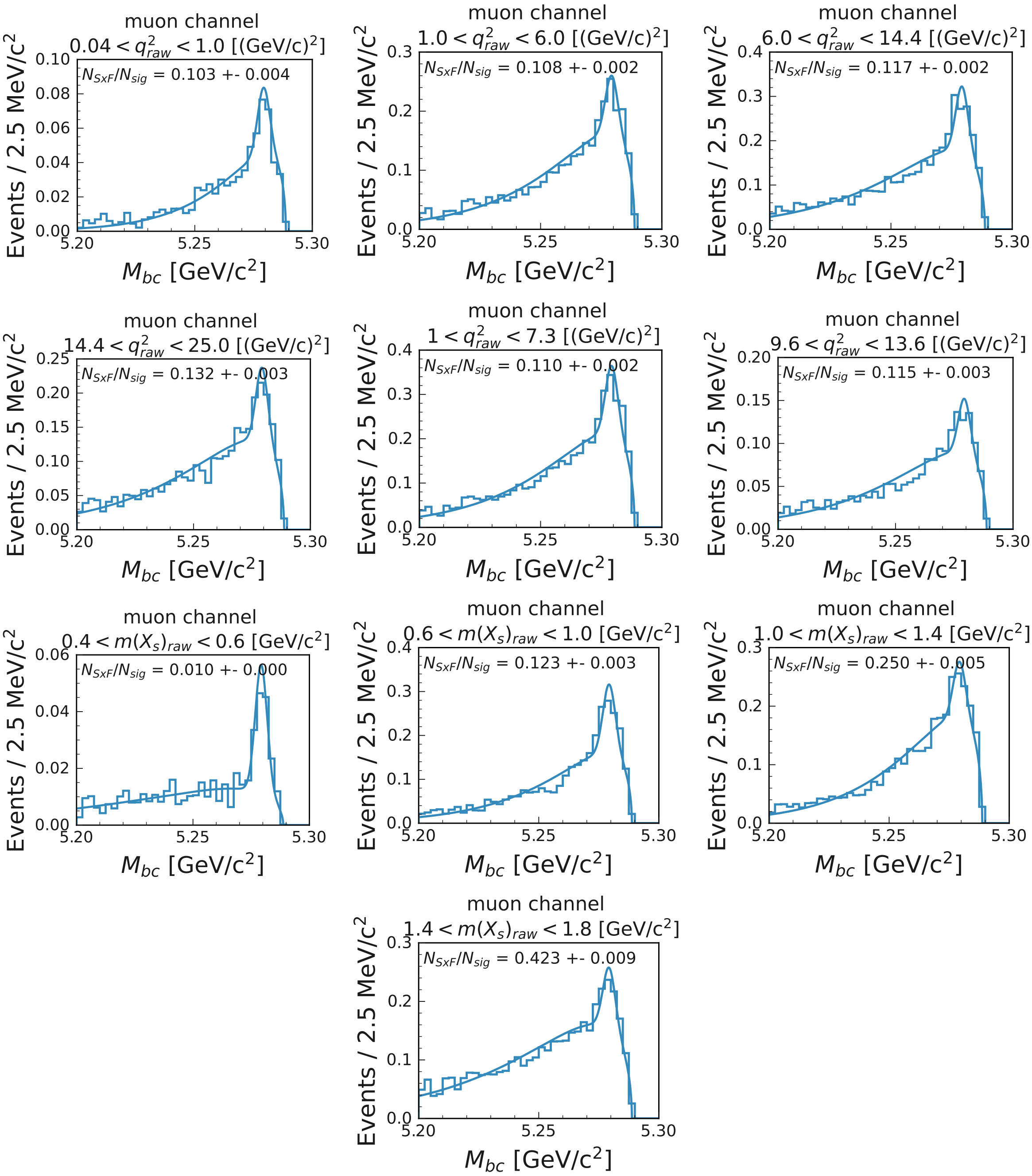


Figure F.1: PDFs for self cross-feed events in muon channel. On the top left of each plot is the ratio between self cross-feed events and signal events derived from signal MC. The histograms are smoothened by fitting to a Gaussian and Argus function as shown by the continuous line overlaid.

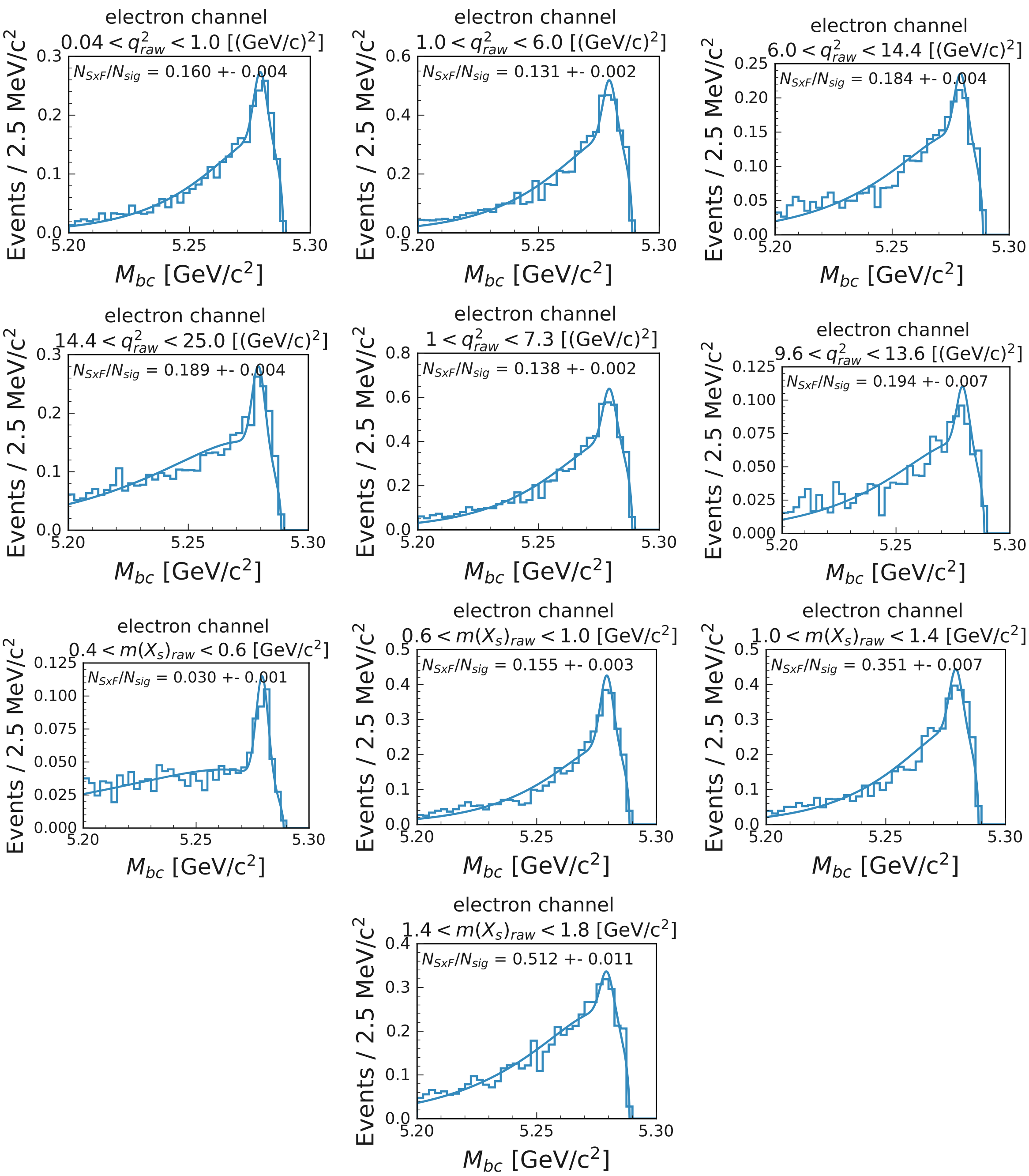


Figure F.2: PDFs for self cross-feed events in electron channel. On the top left of each plot is the ratio between self cross-feed events and signal events derived from signal MC. The histograms are smoothened by fitting to a Gaussian and Argus function as shown by the continuous line overlaid.

F.2 Double mis-ID events

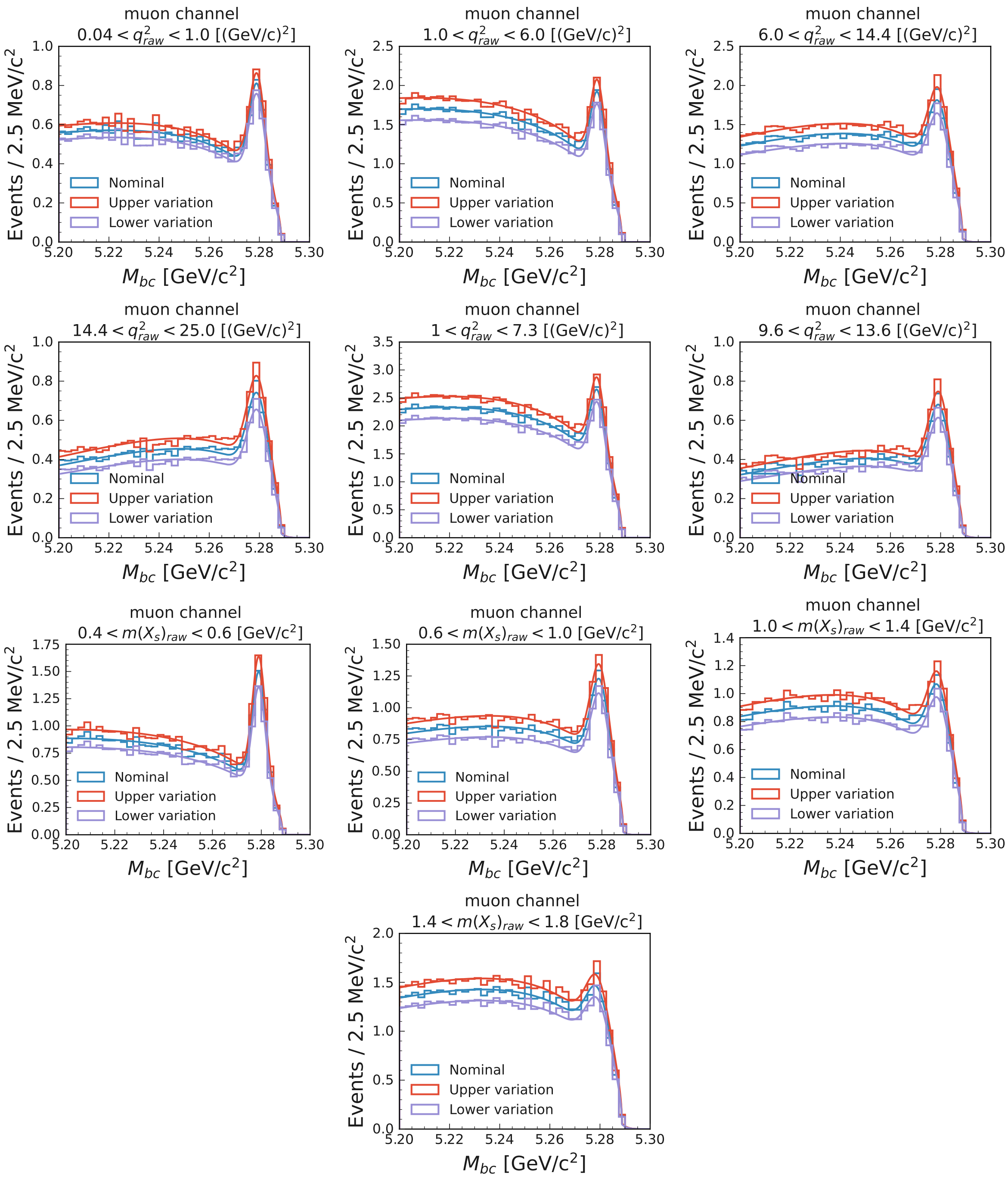


Figure F.3: Fitting templates of double mis-ID backgrounds in muon channel. The histograms are smoothened by fitting to a Gaussian and Argus function as shown by the continuous line overlaid.

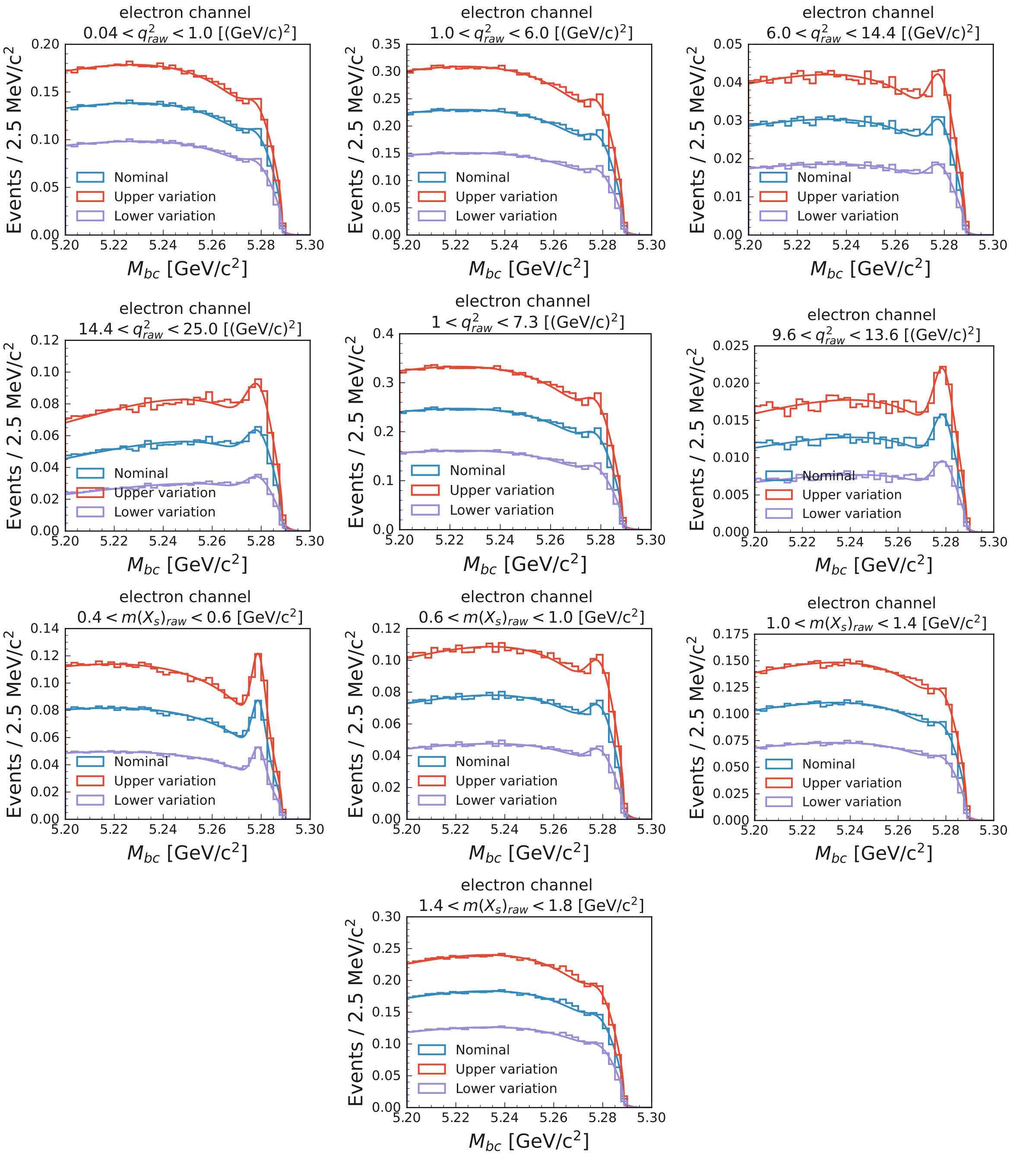


Figure F.4: Fitting templates of double mis-ID backgrounds in electron channel. The histograms are smoothened by fitting to a Gaussian and Argus function as shown by the continuous line overlaid.

F.3 Swap mis-ID events

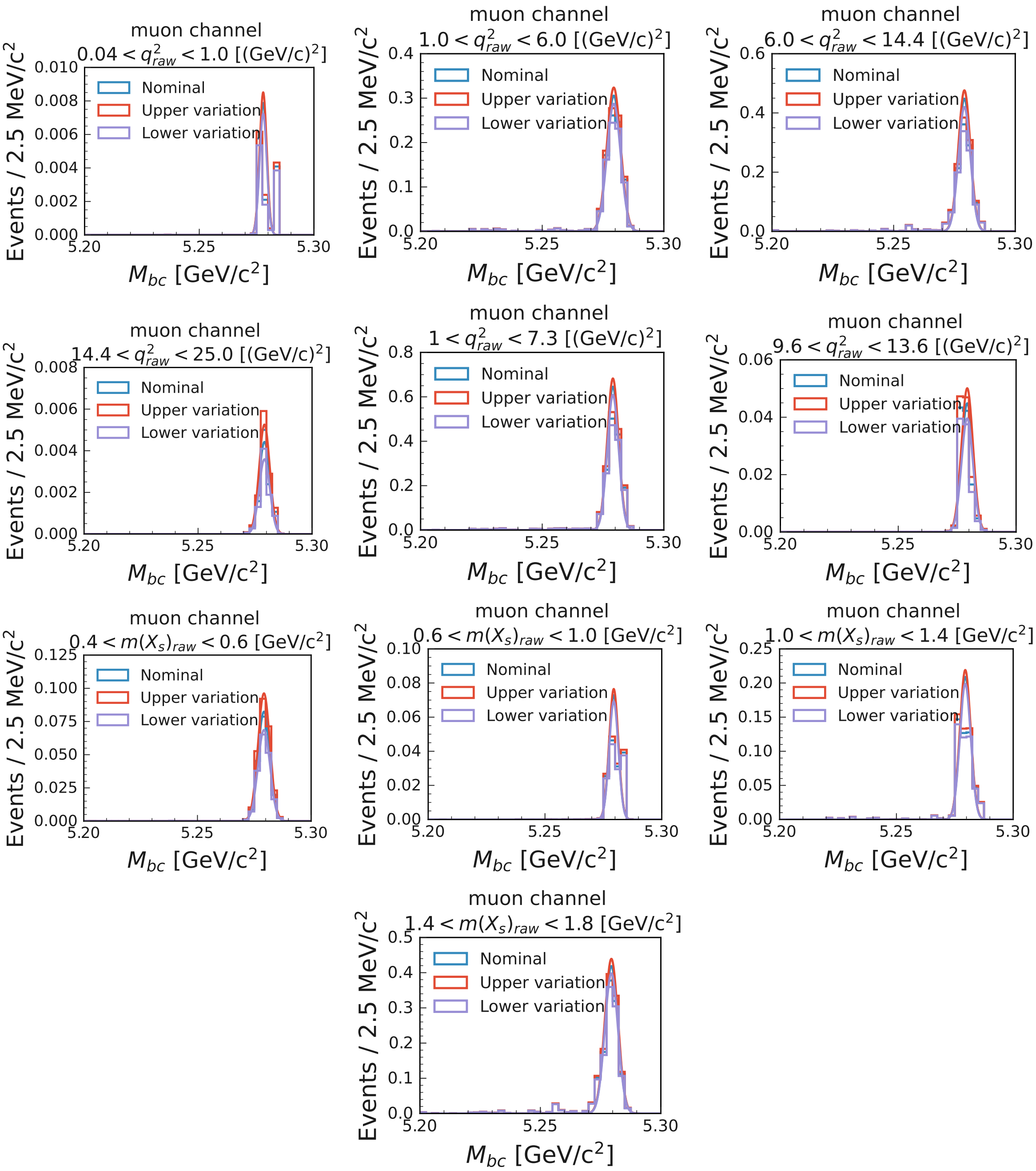


Figure F.5: Fitting templates of swap mis-ID backgrounds. This background is considered only for muon channels. The histograms are smoothened by fitting to a Gaussian function as shown by the continuous line overlaid.

F.4 Charmonium events

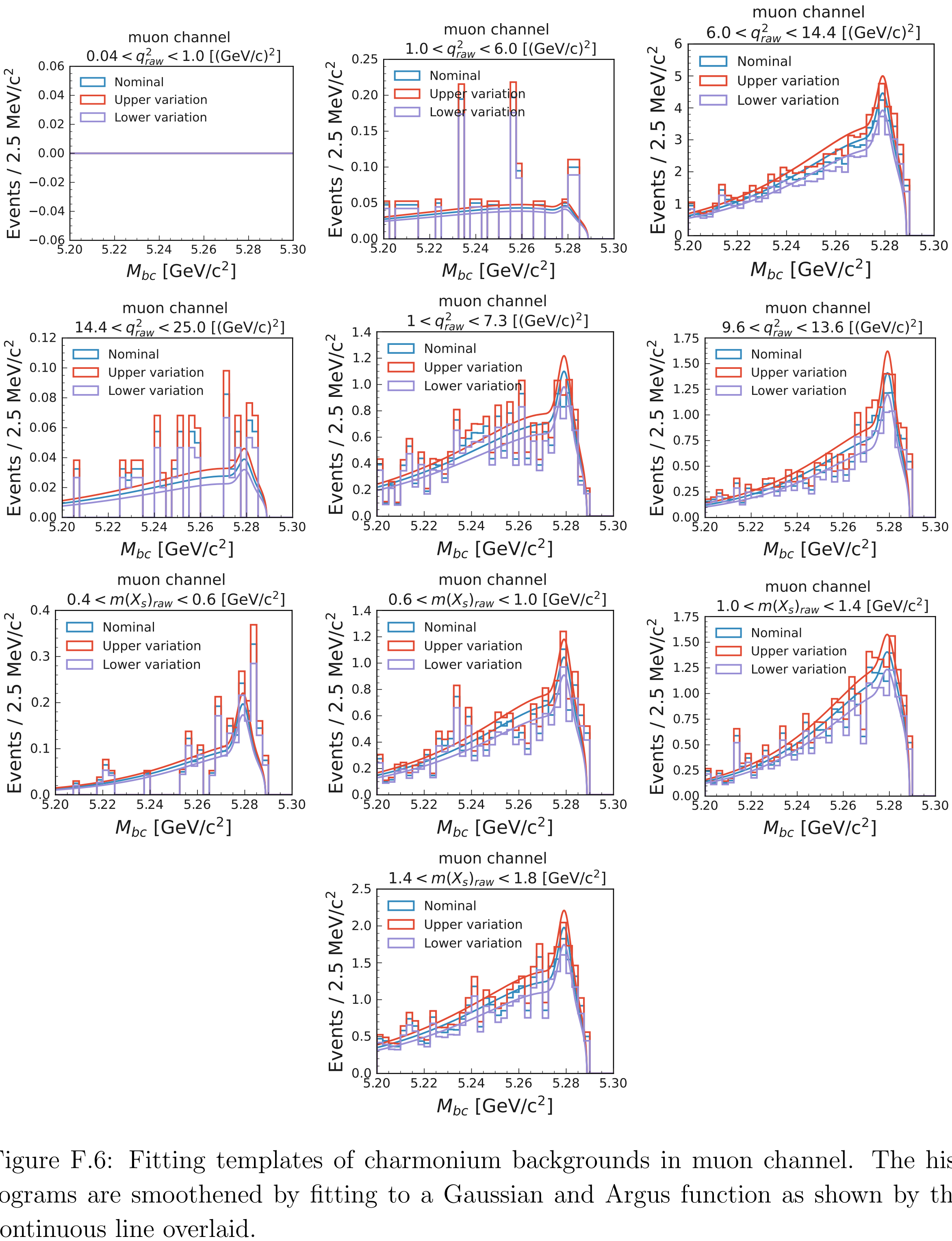


Figure F.6: Fitting templates of charmonium backgrounds in muon channel. The histograms are smoothened by fitting to a Gaussian and Argus function as shown by the continuous line overlaid.

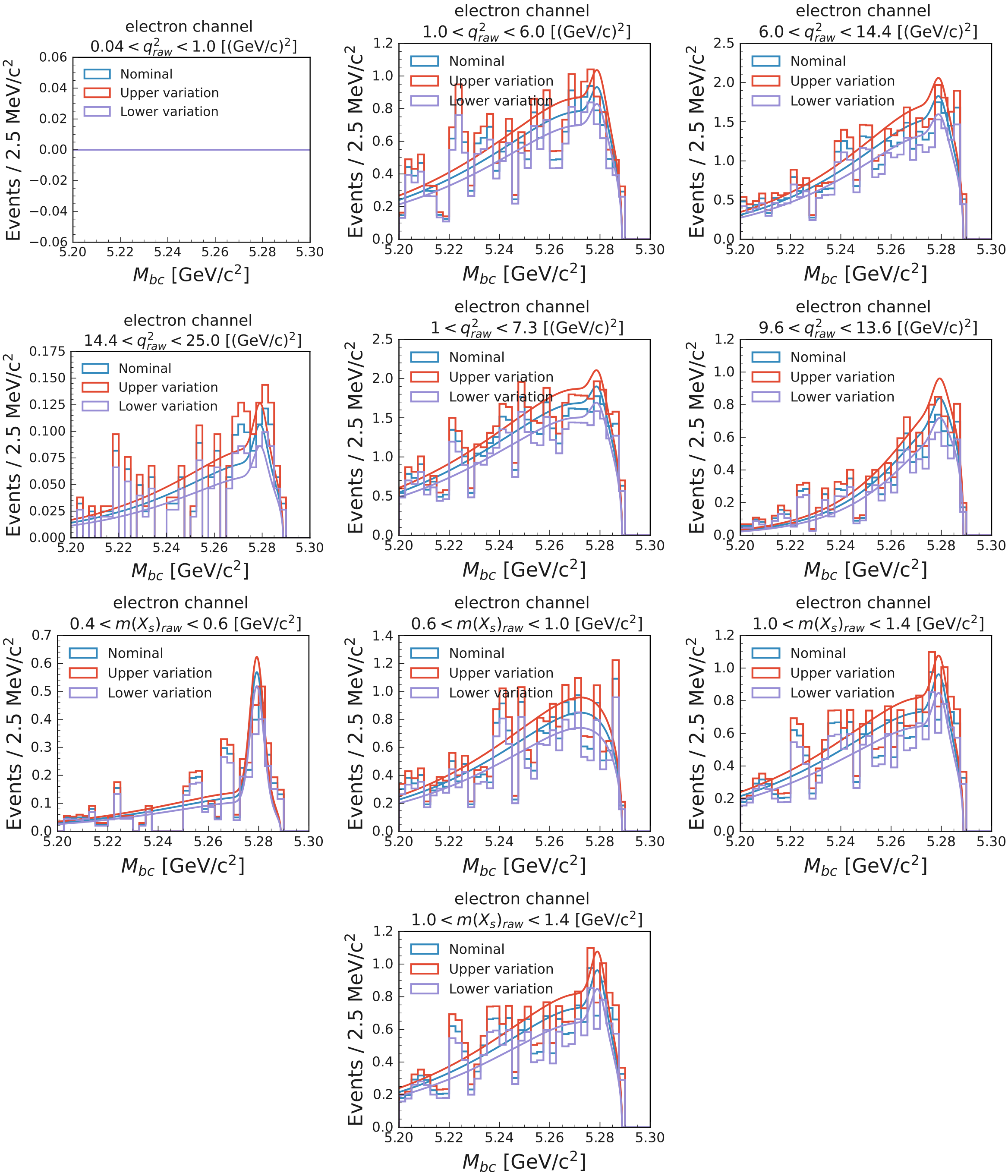


Figure F.7: Fitting templates of charmonium backgrounds in electron channel. The histograms are smoothened by fitting to a Gaussian and Argus function as shown by the continuous line overlaid.

Appendix G

Future Improvements in Muon Identification

In this analysis, the statistical and systematic uncertainties are of similar magnitudes. During the measurement of this channel, we found that there is a large room for muon identification efficiency improvement, with which the statistical uncertainty can be significantly suppressed. In this chapter, we first introduce the current muon identification algorithm and discuss the possible room for performance improvement. Then, we propose a new algorithm based on Deep Neural Network (DNN), and validate its performance both in simulation sample and real data. Finally, we give a breif discussion on its possible influence to our $\mathcal{B}(B \to X_s \ell^+ \ell^-)$ measurement results.

Part of this chapter is published in the journal of Nuclear Instruments and Methods in Physics Research Section A [94].

G.1 Likelihood-based muonID

The traditional muonID algorithm consists of two major steps: (a) track extrapolation and hit association (assuming the muon hypothesis) to estimate the penetration path inside the KLM and (b) likelihood extraction based on the difference between extrapolation and observation.

G.1.1 Track extrapolation and hit association

The track extrapolation is performed by Geant4E [65]. Tracks reconstructed by PXD, SVD, and CDC are extrapolated to KLM with a muon hypothesis, considering characteristic energy loss $\mathrm{d}E/\mathrm{d}x$ and multiple scattering effects to estimate the track momentum and direction. Muons are assumed to not decay or interact through other physics processes. After extrapolating to each KLM layer, the algorithm searches for a single hit associated with the track with a χ^2 method. The χ^2 of each hit on the corresponding layer, which reflects the deviation of extrapolation position to the hit position, is calculated as

$$\chi^2 = \frac{(x_{ext} - x_{hit})^2}{\sigma_{ext}^2 + \sigma_{hit}^2}, \tag{G.1}$$

where $x_{hit(ext)}$ represents the hit (extrapolation) position, σ_{ext} is the extrapolation uncertainty given by Geant4E, and σ_{hit} is the hit position resolution summarized in [45]. This χ^2 is calculated separately along the two directions of the strip sensors. The hit with the smallest sum of χ^2 values in the two directions is selected as the hit associated with the track, if it also satisfies $\chi^2 < (3.5)^2$ in both directions. If the associated hit exists, the extrapolated track properties are adjusted with respect to the hit using Kalman-filter [95]. Extrapolation stops when the particle's energy falls below 2 MeV or the track exits the detector. The number of layers crossed by extrapolation is denoted as N_{ext}.

G.1.2 Likelihood extraction

Binary muonID is defined as the likelihood ratio of the muon and pion hypotheses: $\mathcal{L}_\mu/(\mathcal{L}_\mu + \mathcal{L}_\pi)$. In KLM, the likelihood with hypothesis t (μ or π) is defined as the product of penetration and in-plane likelihoods: $\mathcal{L}_t = \mathcal{L}_t^{\text{pene}} \times \mathcal{L}_t^{\text{in-plane}}$.

The penetration likelihood $\mathcal{L}_t^{\text{pene}} = \prod_{n=1}^{N_{ext}} \mathcal{L}_{t,n}$ is calculated from the likelihoods

$$\mathcal{L}_{t,n} = \begin{cases} P_{t,n} \cdot \varepsilon_n, & \text{with associated hit} \\ 1 - P_{t,n} \cdot \varepsilon_n, & \text{without associated hit} \end{cases} \tag{G.2}$$

of hit pattern in the layers crossed by the extrapolation track. $P_{t,n}$ stands for the probability that track t penetrates to layer n as a function of extrapolated stopping layer N_{ext}. $P_{t,n}$ is measured in the simulation sample in advance. Detector efficiencies ε_n are considered as well, and they are measured in data. In this work, detector efficiencies are assumed to be 100%[1]. An illustration of the penetration likelihood calculation is given in Fig. G.1.

The in-plane likelihood is estimated based on the extrapolation quality described by the sum of χ^2 of all associated hits ($\sum \chi^2$) and the number of degrees of freedom (n_{dof}). The n_{dof} is twice the number of associated hits, because the χ^2 of every hit is calculated once for each of the two directions. For muons, the distribution of $\sum \chi^2/n_{\text{dof}}$ peaks at 1 while for pions, the distribution is wider due to multiple scattering inside KLM, as shown in Fig. G.2.

G.1.3 Discussion

The muonID shows a good performance in μ/π separation. Still, some room for improvement is found, and will be discussed below.

[1] Typically a 96% hit detection efficiency per layer is achieved in operations.

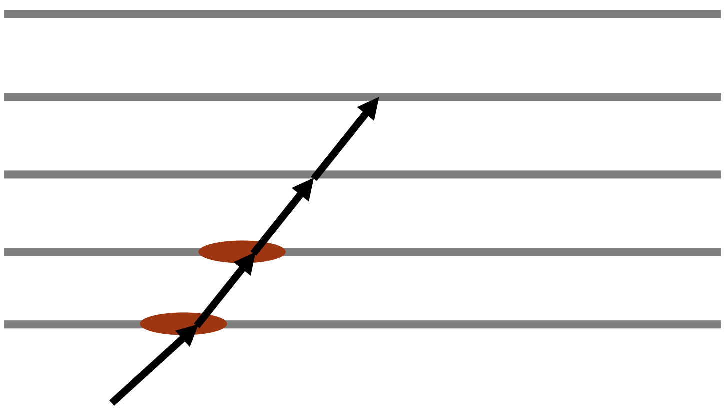


Figure G.1: Illustration of penetration likelihood calculation. The gray horizontal lines represent five KLM layers, with associated hits (brown ellipses) on the first two layers. The track extrapolation represented by the black arrows stops at the fourth layer. In this case, we have $N_{ext} = 4$ and penetration likelihood of hypothesis t given by $\mathcal{L}_t^{\text{pene}} = P_{t,1}\varepsilon_1 P_{t,2}\varepsilon_2(1 - P_{t,3}\varepsilon_3)(1 - P_{t,4}\varepsilon_4)$.

Figure G.3(a) shows the pion rejection rate as a function of the penetration layer (N_{hit}) applying a muonID > 0.9 selection, where N_{hit} is defined as the last layer in which a hit has been detected. The histograms show the probability density of the penetration layer after selection for muon and pion samples. This dataset is a simulation sample with only one track in each event. For illustration, the polar angle is fixed to 90° and only tracks extrapolated to stop at layer 14 of BKLM are selected. From this figure, it is obvious that muonID successfully rejected pions with penetration layers smaller than eight. However, the rejection rate reduces to only around 20% when $N_{hit} > 8$. Meanwhile, we can significantly improve the identification performance by rejecting tracks in the range of $8 \leq N_{hit} \leq 10$, which happens rarely for the muons, but quite frequently for the pions.

To explore the reason why muonID failed to reject pion tracks satisfying $8 \leq N_{hit} \leq 10$, we calculate the penetration likelihood $\mathcal{L}^{\text{pene}}$ as a function of penetration layer (N_{hit}) with muon and pion hypotheses[2]. The result is presented in Fig. G.3(b) and it shows that $\mathcal{L}_\mu^{\text{pene}}$ overwhelms $\mathcal{L}_\pi^{\text{pene}}$ when the track penetration layer is greater than 8, which explains the drastic drop in the rejection rate at layer 8. It suggests that the penetration likelihood used in muonID is not optimally modeled, indicating significant potential for performance improvement. One possible explanation for this mis-modeling is that the penetration likelihood does not consider the correlations between hits in different layers, as it is constructed by simply multiplying the likelihoods assigned to individual layers. This insight motivates us to develop a machine learning-based algorithm capable of incorporating such correlations in this work.

In addition, there is room for improvement by better utilizing the in-plane information. Fig. G.4 shows the distribution of $\sum \chi^2/n_{\text{dof}}$ for tracks satisfying muonID > 0.9 and

[2] In this calculation, we assume extrapolation stops at layer 14 of BKLM, associated hits observed up to layer N_{hit}, and no associated hit observed above layer N_{hit}: $\mathcal{L}_t^{\text{pene}} = \prod_{n=1}^{N_{hit}} \varepsilon_n P_{t,n} \prod_{n=N_{hit}+1}^{14}(1-\varepsilon_n P_{t,n})$.

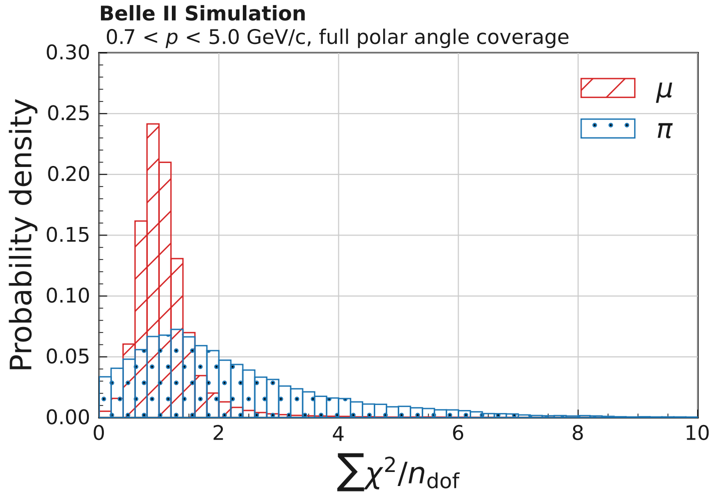


Figure G.2: Distribution of $\sum \chi^2/n_{\rm dof}$ for muon (red, dashed) and pion (blue, dotted).

$N_{ext} - N_{hit} \leq 2$. Still, some remaining pions can be rejected by requiring, for example, $\sum \chi^2/n_{\rm dof} < 2$, at the cost of losing little muon efficiency. However, muonID fails to do so because it relies too heavily on penetration information due to imperfect settings of the scale of penetration and in-plane likelihood. Specifically, $\mathcal{L}_\mu^{\text{in-plane}}/\mathcal{L}_\pi^{\text{in-plane}}$ is at the order of 10^{-1} for tracks with $\sum \chi^2/n_{\rm dof} > 2$, while $\mathcal{L}_\mu^{\rm pene}/\mathcal{L}_\pi^{\rm pene}$ is larger than $10^{-2}/10^{-15} = 10^{13}$ according to Fig. G.3(b). For this reason, the relationship muonID > 0.9 is hardly influenced by the in-plane likelihood.

G.2 Deep Neural Network (DNN) based muon probability

To make better use of penetration and in-plane information, we propose a DNN-based algorithm. The track extrapolation and associated hits information described in Sec. G.1.1 are used in this algorithm. In this section, input variables, network structure and training, as well as the performance evaluation of the new algorithm are described.

G.2.1 Input variables

Five global variables are used as input of the DNN, arranged in the order of $\sum \chi^2$, $n_{\rm dof}$, $N_{ext} - N_{hit}$, N_{ext} and the transverse momentum of the track whose distributions are shown in Fig. G.5(a)–(e). The latter two variables play an important role of indicator since the distributions of the former three variables as well as the hit pattern vary as function of the extrapolation layer and the transverse momentum.

In addition, four hit pattern variables are defined for each KLM layer used as input of the DNN as illustrated in Fig. G.6. Their definitions are explained below.

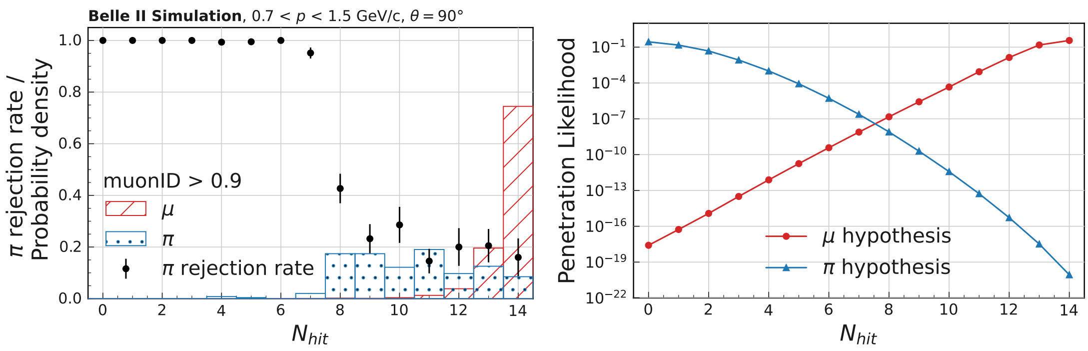


Figure G.3: Left: dots show the pion rejection rate against penetration layer after requiring muonID > 0.9 in samples with $N_{ext} = 14$. Error bars represent statistical uncertainty. The histograms are the penetration layer distribution of remaining muon (red, dashed) and pion (blue, dotted) after selection. Right: penetration likelihood (in logarithmic scale) under muon (red, circle) and pion (blue, triangle) hypothesis as a function of penetration layer assuming extrapolation stops at layer 14 of BKLM.

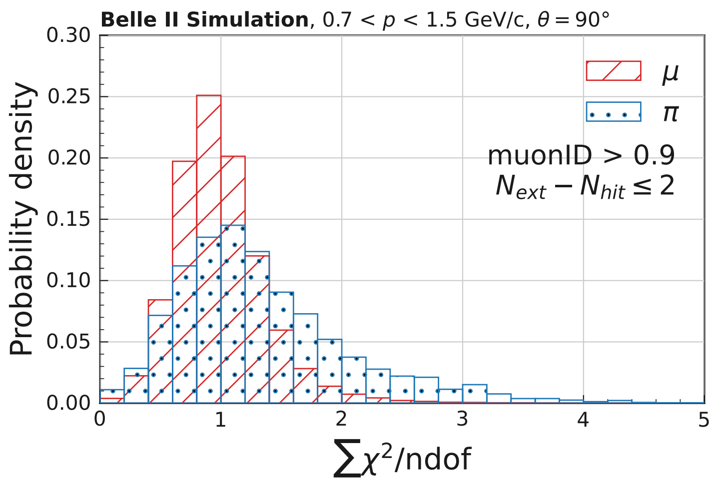


Figure G.4: Distribution of $\sum \chi^2/n_{\mathrm{dof}}$ for muon (red, dashed) and pion (blue, dotted) samples, requiring muonID > 0.9 and $N_{ext} - N_{hit} \leq 2$. This plot indicates that $\sum \chi^2/n_{\mathrm{dof}}$ information is not fully exploited by muonID.

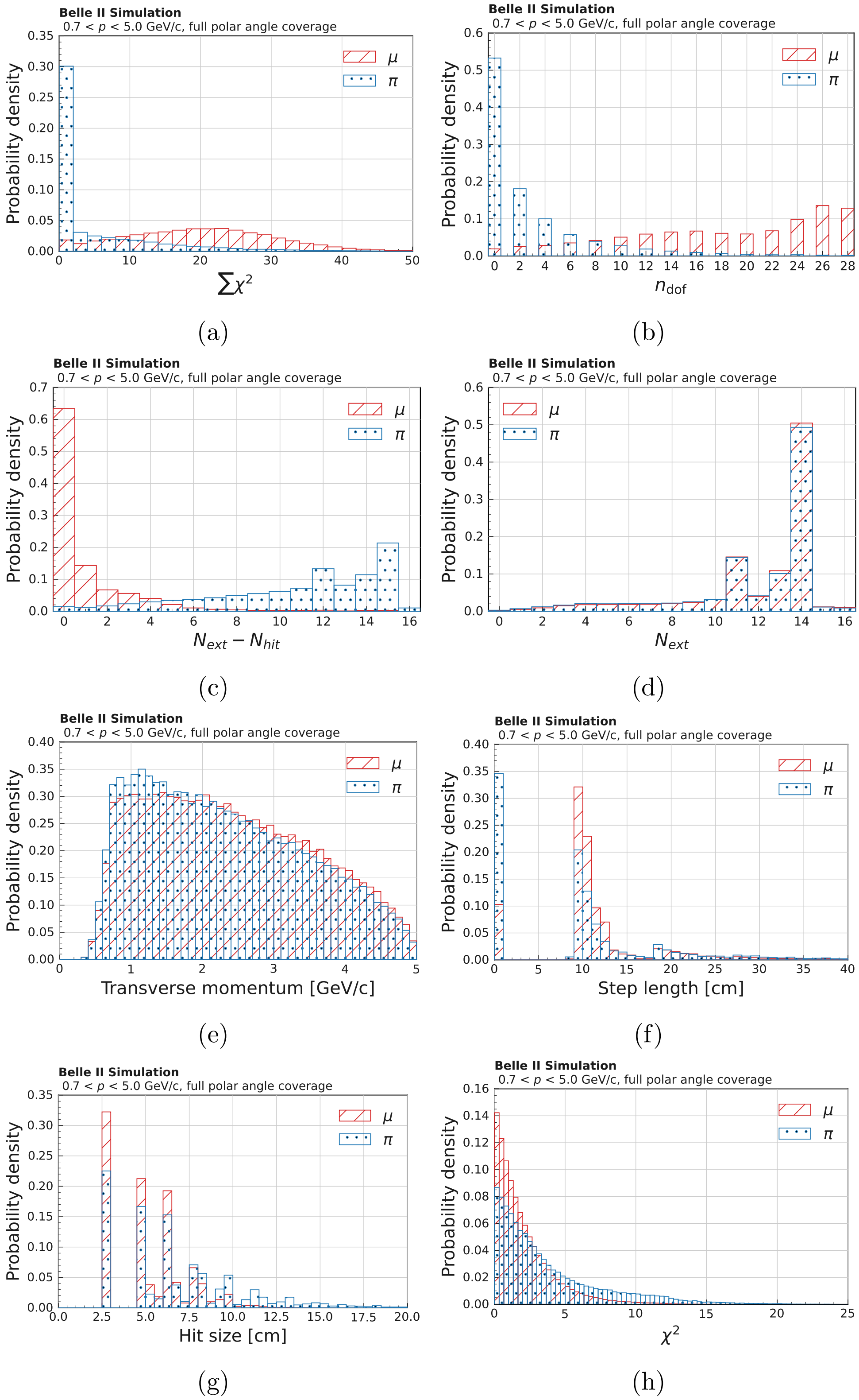


Figure G.5: Distributions of input variables used in DNN in the training sample. Plot (a)–(e) show the distributions of global variables with each entry in the histograms representing one track. Plot (f)–(h) show the distributions of hit pattern variables with each entry in the histograms representing one associated hit. The peak at zero in (f) represents the first associated hits of each track, whose step length are assigned to zero by definition. The peak around 2.5 cm in (g) represents hits with only one strip in both directions. All plots are normalized to unit area.

Step length: defined for each associated hit as the distance to its prior associated hit. If it happens to be the first associated hit of the track, its step length is set to zero.

Hit size: defined for each associated hit as being half of the diagonal length of the rectangular shape of the hit.

χ^2*:* defined in Eq. G.1.

Extrapolation pattern: a binary value indicating whether the extrapolation crossed the corresponding layer or not.

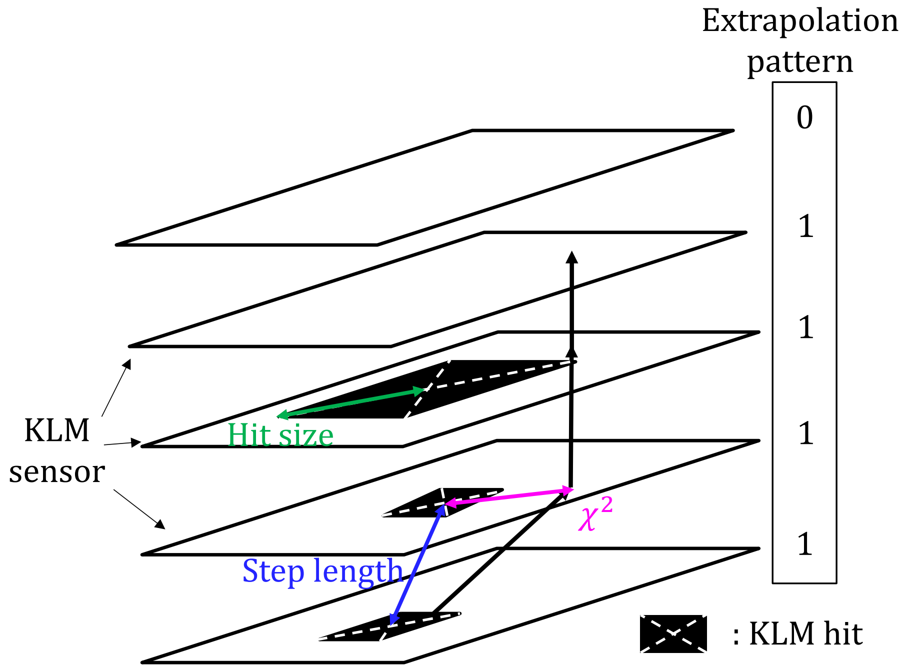


Figure G.6: Illustration of hit pattern variables. The thick black arrows represents the track extrapolation and the extrapolated position at the third layer is adjusted by Kalman-filter. The length of the blue and green arrows represents step length and hit size, respectively. The magenta arrow indicates the χ^2 between hit and extrapolation position. The binary numbers on the right side are the extrapolation pattern of the corresponding layers.

The calculation of muonID penetration likelihood in Eq. G.2 is layer-based, which means that it only reflects the penetration information along the normal direction of the detector layer plane. By introducing the step length into the DNN, the penetration information along the tangent direction (projection of track direction on the sensor plane) is also taken into account. Due to the stronger penetration ability, the total penetration depth (sum of step length) of the muon is greater than that of the pion. And on the other hand, the variation of step length between different layers of pions tends to be larger than that of muons because of strong interaction with detector materials. For the same reason, the hit size and the deviation of extrapolation to the hit position (χ^2) of the pion also tends to be larger than that of the muon, as shown in Fig. G.5(f)–(h).

In total, the input to the DNN model is a 1-dimensional float array with 121 elements. The first five elements are the global variables. The remaining 116 elements are arranged into 29 groups, each group is used to place the hit pattern variables of one layer. The first 15 groups represent the 15 BKLM layers, while the latter 14 groups represent the 14 EKLM layers. In each group, the hit pattern variables are arranged in the order of hit size, step length, χ^2 and extrapolation pattern. If there is no associated hit in the corresponding layer, the hit size, step length, and χ^2 are set to -1.

G.2.2 Network structure and training

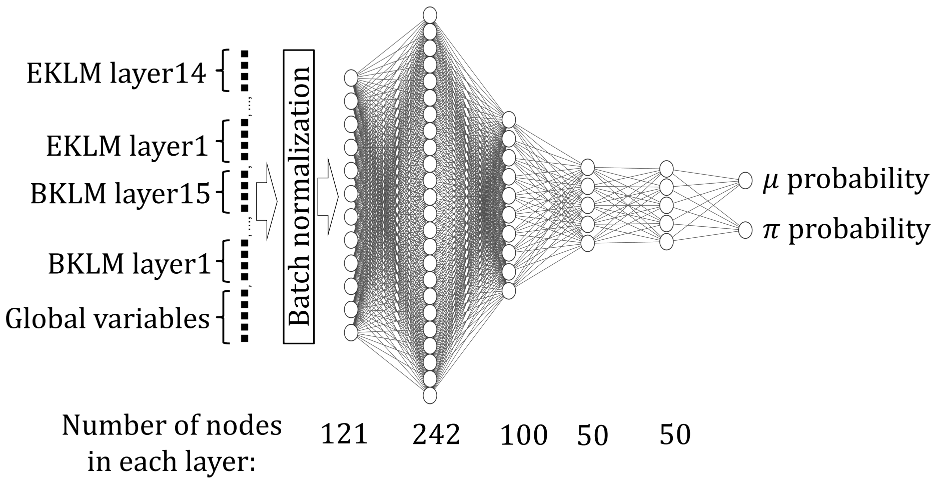


Figure G.7: Structure of DNN.

This neural network is built with the PyTorch [81] library and its structure is shown in Fig. G.7. The input array is first processed by the batch normalization module, followed by a fully connected linear model. There are five linear layers with 121, 242, 100, 50 and 50 nodes, respectively. The output of each node is processed by a LeakyReLU [96] activation function before being input to the next layer. At the output of the last layer there is a softmax activation function used to output the muon probability and the pion probability. In total, there are 64318 trainable parameters in the model.

A simulation sample is generated for training, validation, and test of the model using the Belle II Analysis Software Framework [58, 59]. Each event contains 4 to 16 tracks to simulate different event multiplicity. Each track is randomly generated to be a muon, pion, electron, kaon, proton, or deuteron, with the same probability for each type. The charge of each track is also randomly determined to be positive or negative with equal probability. To improve the robustness against noise hits, simulated beam background [97] at a luminosity about six times higher than the current operation record is overlaid on each event. All tracks are generated with uniform momentum ranging from 0.7 GeV/c to 5.0 GeV/c, cosine of polar angle and azimuthal angle distribution. The polar and azimuthal angular distributions cover the full geometric acceptance of the

KLM. Only the muons and pions in the samples are selected for study using generator information. In addition, pions that decay before KLM are removed from the samples. In total, we generated 559383, 338031 and 153966 tracks for training, validation, and test samples, respectively.

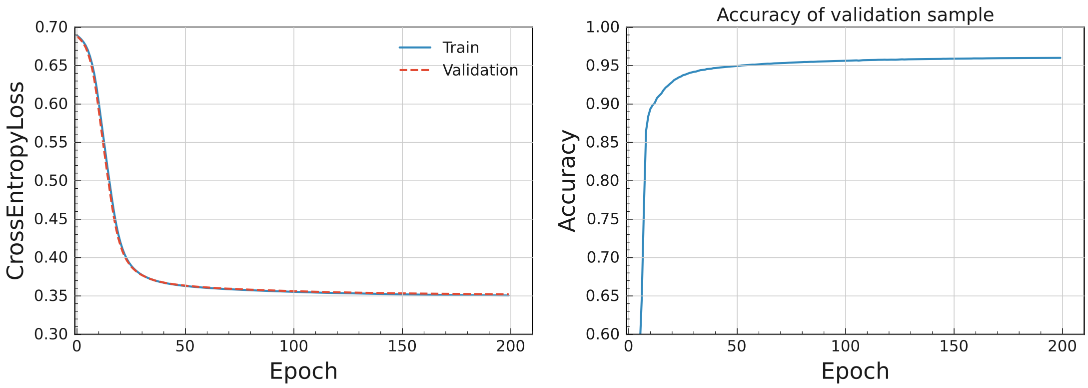


Figure G.8: Left: Loss convergence plot of training (blue, solid) and validation (red, dashed) samples. Right: Accuracy of validation sample as a function of epochs.

In the training, an `Adam` optimizer is adapted with a learning rate of 10^{-5} and the batch size is set to be 10000. `CrossEntropy`, which is suitable for binary classification, is used as the loss to minimize. Training stops when prediction accuracy of the validation sample does not increase for 10 epochs and the epoch with the best accuracy is adapted. The loss convergence plots of training and validation samples, as well as the prediction accuracy in validation sample as a function of epochs are presented in Fig. G.8. The muon efficiency difference at 2% pion fake rate in training, validation and test samples are at the order of $O(0.1\%)$, indicating that no significant over-training is observed. The training is performed on one GeForce RTX 3090 GPU with a training time of around 30 min. Inference latency is 1.81 ± 0.78 ms per $e^+e^- \to \Upsilon(4S) \to B\bar{B}$ event on the KEKCC CPU cluster (AMD EPYC 9534) at KEK.

G.3 Performance in MC Samples

The performance of the model is validated using a Receiver Operation Characteristic (ROC) curve, which plots the true positive rate (μ efficiency) against the false positive rate (π fake rate) as shown in Fig. G.9. The muonID, which is the baseline method, is also plotted for comparison. As demonstrated in the ROC curves, the DNN performs better than muonID. For example, the DNN (muonID) gives a pion fake rate of 1.6% (4.1%) at 90% muon efficiency, or a muon efficiency of 92.2% (76.5%) at a pion fake rate of 2%. To validate the importance of the hit pattern variables, we trained another network using only the five global variables as input. The structure of the network is identical to the default one, except for the batch normalization and the first linear layer, whose number of nodes are adjusted according to input array length. The pion fake rate deteriorates to

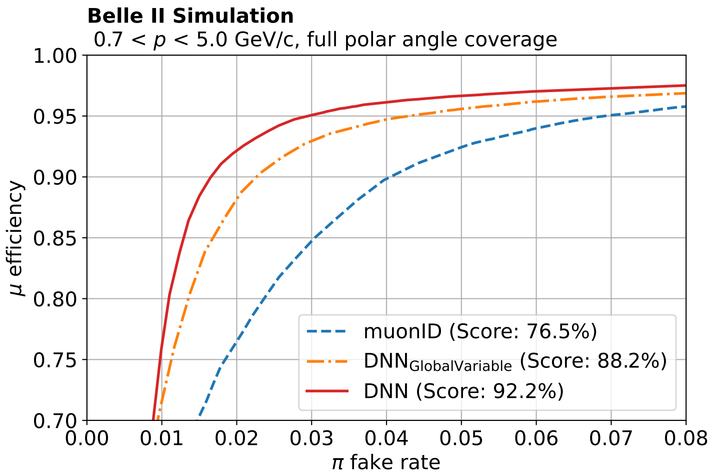


Figure G.9: ROC curve of muonID (blue, dashed), DNN trained with only global variables (orange, dashdot), and the default DNN (red, solid). The score of each model is defined as the muon efficiency at 2% pion fake rate in the test sample.

2.3% at 90% muon efficiency if we only use the five global variables, demonstrating the importance of the hit pattern.

To test the robustness of the model against background hit, we prepared another sample which is generated in the same way as training sample, except that the beam induced background simulated at the Belle II nominal luminosity, which is about three times larger than the training sample. In this sample, the pion fake rate deteriorates to 2.4% at 90% muon efficiency possibly due to the changes of hit patterns under high background environment, but still it performs significantly better than muonID. A re-training of the DNN model can be expected when machine status evolves significantly to achieve a better performance. We also tested the model performance by setting the KLM hit detection efficiency to a uniform 85% and observed slight pion fake rate increased to 2.0%. Finally, we checked the robustness of track extrapolation against inaccuracies of the modeling of the magnetic field inside KLM varying it by 10%. No significant variation of performance is observed.

Figure G.10 shows the pion fake rate of muonID and DNN at each momentum interval, maintaining a uniform muon efficiency of 90%. The pion fake rate is suppressed across the full momentum range, with improvements exceeding 60% in the high momentum range ($p > 2.0$ GeV/c). The improvement is less significant in the low-momentum region, where muons cannot traverse the KLM entirely.

Figure G.11 shows the penetration depth distributions of the pions after selection at the overall muon efficiency 90% using the muonID and DNN method, respectively. Comparing to muonID, DNN successfully rejected about 60% of deeply penetrated pions up to a penetration depth of 125 cm, which aligns well with the detector thickness of around 130 cm for both BKLM and EKLM. This phenomenon may suggest that the DNN has learned a specific pattern: tracks with a penetration depth exceeding 125 cm

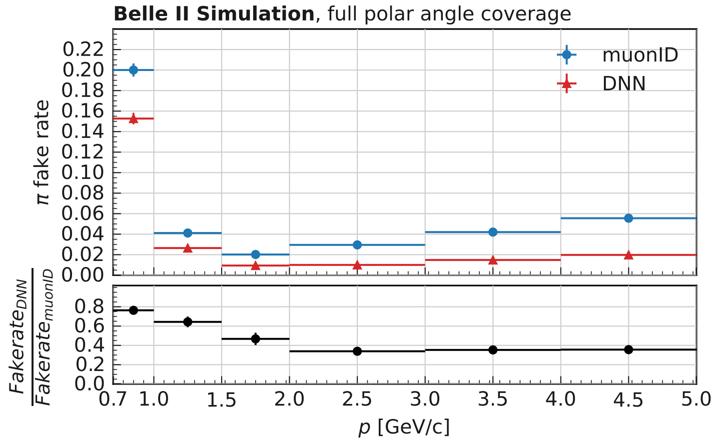


Figure G.10: Upper: pion fake rate of muonID and DNN as a function of track momentum at 90% muon efficiency in each bin of momentum. Lower: pion fake rate ratio of DNN over muonID.

are more likely to escape from the KLM, where μ/π identification based on penetration ability becomes less effective.

G.4 Validation in Data

G.4.1 Dataset and method

In this section, we describe the validation of this new method in real data. We use $e^+e^- \to \mu^+\mu^-\gamma$ channel to evaluate muon efficiency and $D^{*+} \to D^0(\to \pi^+K^-)\pi_s^+$ for pion fake rate estimation.

We apply the following selection criteria to $e^+e^- \to \mu^+\mu^-\gamma$ events:

- Tracks: dr <2 cm; $|dz|$ <5 cm;
- Photon: $E > 1.0\,\text{GeV}$; $-0.8660 < \cos\theta < 0.9563$;
- Event level: number of tracks is 2; $10 < m(\mu^+\mu^-\gamma) < 11$.

And the following requirements are applied for $D^{*+} \to D^0(\to \pi^+K^-)\pi_s^+$ candidates:

- Tracks: dr <2 cm; $|dz|$ <4 cm;
- D^0: $1.75 < m(D^0) < 2.0$ [GeV/c^2];
- D^{*+}: $m(D^{*+}) - m(D^0) < 0.16\,\text{GeV}/c^2$; $p > 1.5\,\text{GeV}/c$ in the center-of-mass frame.

These selection criteria guarantee a sample with > 99% signal purity. In addition, only tracks with information inside KLM (has one on-track hit in KLM or extrapolation reached KLM) are considered in this evaluation. Finally, we get about 17,000 muon tracks and 50,000 pion tracks for evaluation.

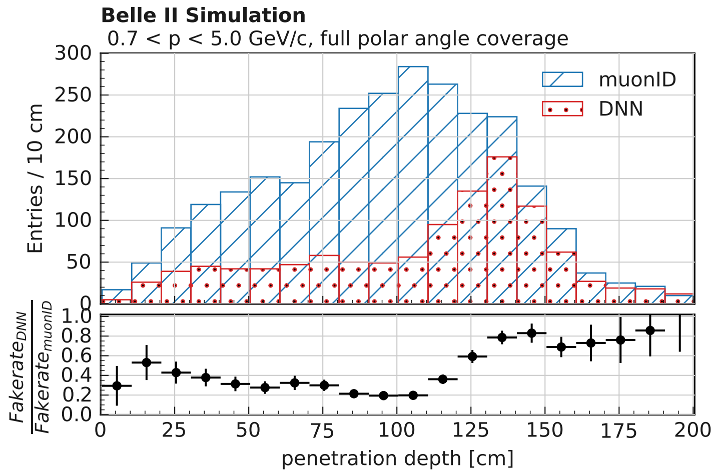


Figure G.11: Upper: distribution of penetration depth of pion after the selection with 90% overall muon efficiency for muonID and DNN method. Lower: pion fake rate ratio of DNN over muonID as a function of penetration depth.

Tag and probe method is used to calculate muon efficiency. muonID>0.9 is applied to the tag side to exclude non-muon events. PID score (muonID and the DNN) is applied to the probe side. By counting the number of events after applying tag side cut (N_{tag}) and after applying cut on both tag and probe side (N_{probe}), the muon efficiency can be estimated by

$$\varepsilon = \frac{N_{probe}}{N_{tag}}. \tag{G.3}$$

For the pion fake rate, the slow pion from D^{*+} decay is used to tag the pion from D^0 decay since they have the same charge. PID score cut is applied to the fast pion from D^0 and count the number of pions by fitting to the mass spectrum of D^{*+} by a double gaussian function. The background is modeled by 1st order Chebyshev polynominal.

Figure G.12 shows the momentum distribution of muon and pion tracks. In the 0.7 GeV/c to 3.0 GeV/c interval, the muon and pion momentum distribution is almost flat. In the evaluation of momentum dependence, the following binning scheme is applied: [0.7, 1.0, 1.5, 2.0, 3.0, 5.0].

Figure G.13 shows the polar angle distribution of muon and pion tracks. In the evaluation of polar angle dependence, the following binning scheme is applied: [0.22, 0.64, 0.82, 1.16, 1.46, 1.78, 2.13, 2.22, 2.71].

G.4.2 Result

We firstly evaluated the muon efficiency and pion fake rate in an integrated phase space 0.7 GeV/$c < p <$ 3.0 GeV/c and $0.64 < \theta < 2.22$ (BKLM). We fixed the muon efficiency to 90% and compare the pion fake rate. The results for positive and negative charge tracks are summarized in Table G.1 and Table G.2, respectively. Fig. G.14 shows the

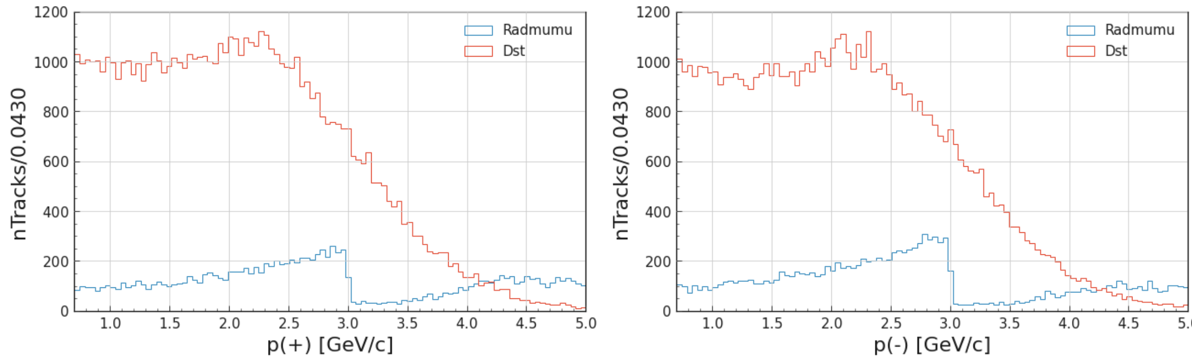


Figure G.12: Distribution of track momentum for plus (left) and minus (right) charge. The blue histogram stands for muons from radiative mumu sample while red for fast pion in the D^{*+} sample.

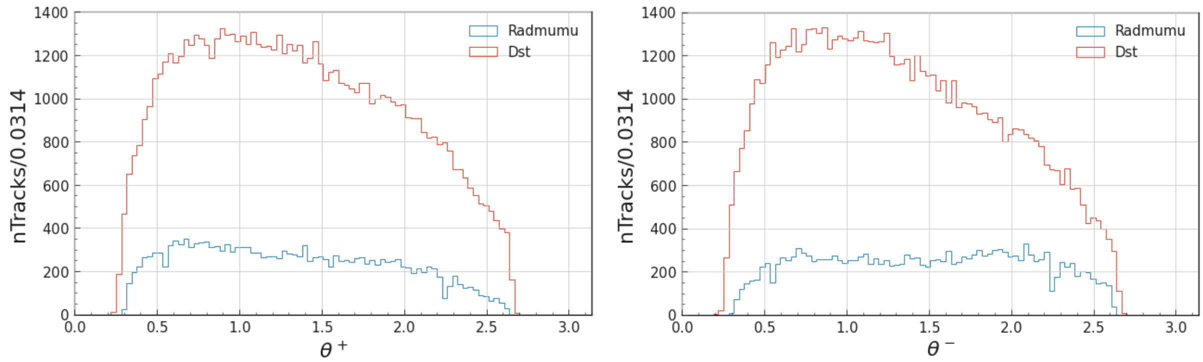


Figure G.13: Distribution of track polar angle for plus (left) and minus (right) charge. The blue histogram stands for muons from radiative mumu sample while red for fast pion in the D^{*+} sample.

fitting results for extracting the number of pions. Considering only the fitting uncertainty, we confirmed the the significance that number of pions is reduced with NN method is 5.7σ and 5.6σ for positive and negative charged tracks, respectively.

Figure G.15 shows the muon efficiency and pion fake rate (scaled by 5) as a function of track momentum in the BKLM ($0.64 < \theta < 2.22$) region. For DNN, uniform PID score larger than 0.9 is applied. Then, the cut for muonID is tuned to have the same muon efficiency as DNN in each kinematic bin. We confirmed pion fake rate is suppressed in full momentum range.

Figure G.16 shows the muon efficiency and pion fake rate (scaled by 5) as a function of track polar angle integrated over track momentum from 0.7 GeV/c to 3.0 GeV/c. For DNN, uniform PID score larger than 0.9 is applied. Then, the cut for muonID is tuned to have the same muon efficiency as DNN in each kinematic bin. Pion fake rate is suppressed in BKLM, backward EKLM and the border of BKLM and EKLM. Performance improvement in the forward end-cap is small which is consistent with training sample.

Table G.1: Working point and performance of muonID and DNN for positive charged tracks.

Working point	Muon efficiency	Pion fake rate
muonID > 0.96875	$(90.0 \pm 0.4)\%$	$(2.8 \pm 0.1)\%$
DNN > 0.9765625	$(90.0 \pm 0.4)\%$	$(1.9 \pm 0.1)\%$

Table G.2: Working point and performance of muonID and DNN for negative charged tracks.

Working point	Muon efficiency	Pion fake rate
muonID > 0.984375	$(90.0 \pm 0.4)\%$	$(2.4 \pm 0.1)\%$
DNN > 0.99268	$(90.0 \pm 0.4)\%$	$(1.8 \pm 0.1)\%$

G.5 Conclusion and prospects

In this chapter, we discuss how the muon identification performance of the Belle II experiment can be improved by better utilizing hit pattern information in the KLM detector. By training a new deep neural network, we reduced the pion fake rate (specificity) from 4.1% to 1.6% at 90% muon efficiency (recall) in the simulation sample. We also tested this new tool in real data, confirming a fake rate reduction by around 30%.

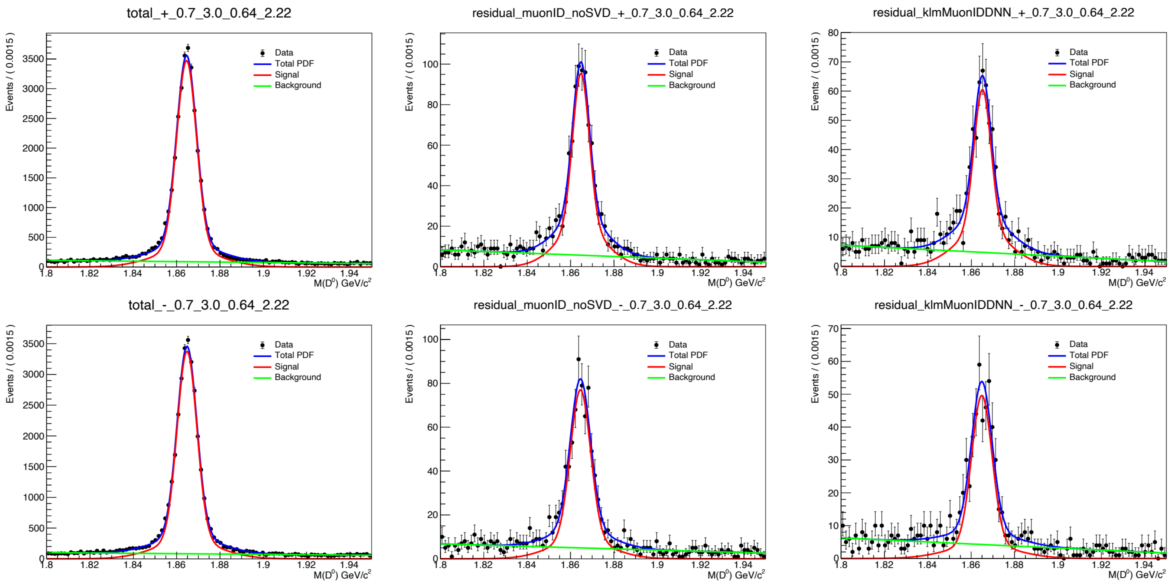


Figure G.14: Fitting result of D^0 mass spectrum for counting the number of pions. The 3 columns from left to right shows the result without any PID cut, after muonID selection, and after DNN selection, respectively. The upper (lower) row is the result for the positive (negative) charged pions.

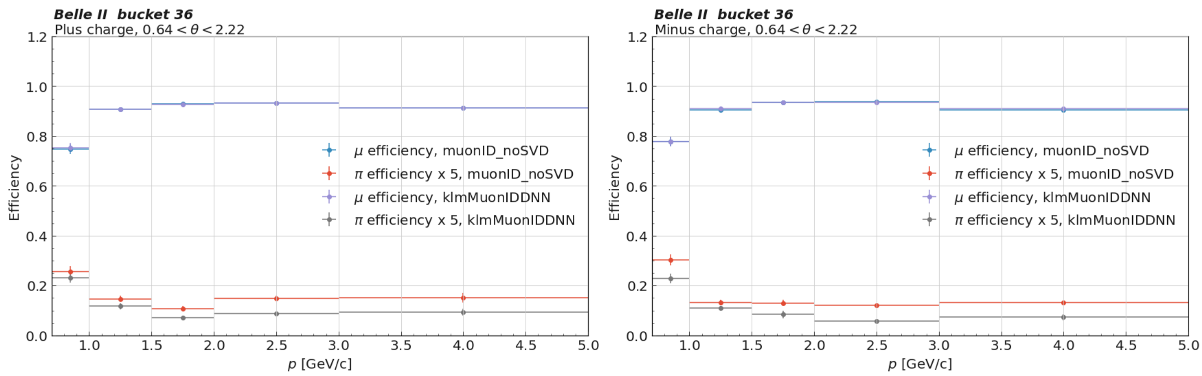


Figure G.15: Muon efficiency and pion fake rate (scaled by 5) as a function of track momentum in BKLM ($0.64 < \theta < 2.22$). The left plot stands for the result for positive charge and the right for negative charge.

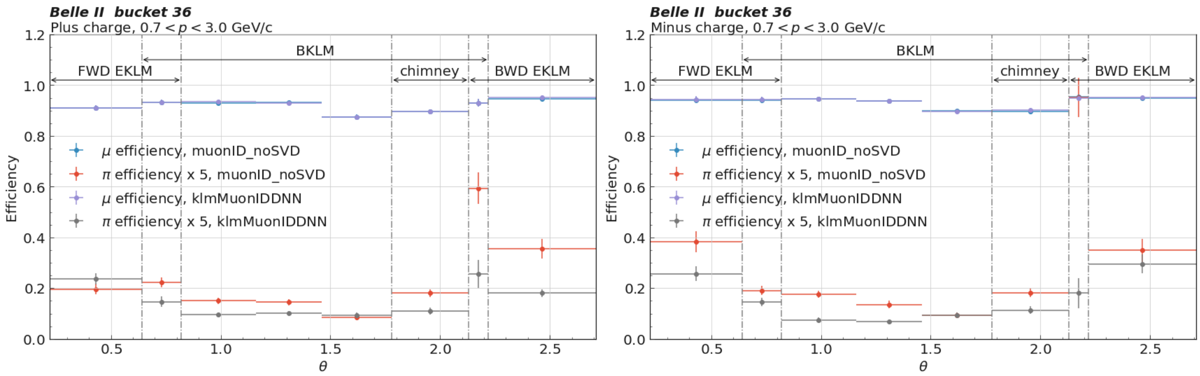


Figure G.16: Muon efficiency and pion fake rate (scaled by 5) as a function of track track polar angle integrated over a momentum range of $0.7\,\mathrm{GeV}/c$ to $3.0\,\mathrm{GeV}/c$. The left plot stands for the result for positive charge and the right for negative charge.

Appendix H

Correction factors of PID efficiency

In this appendix, we present the correction table of PID selection efficiencies. Fig. H.1 is for $K^{\pm}$ and $\pi^{\pm}$ while Fig. H.2 is for $\mu^{\pm}$ and $e^{\pm}$. These tables are generated in three dimensions: track momentum (p), track polar angle ($\cos\theta$), and charge.

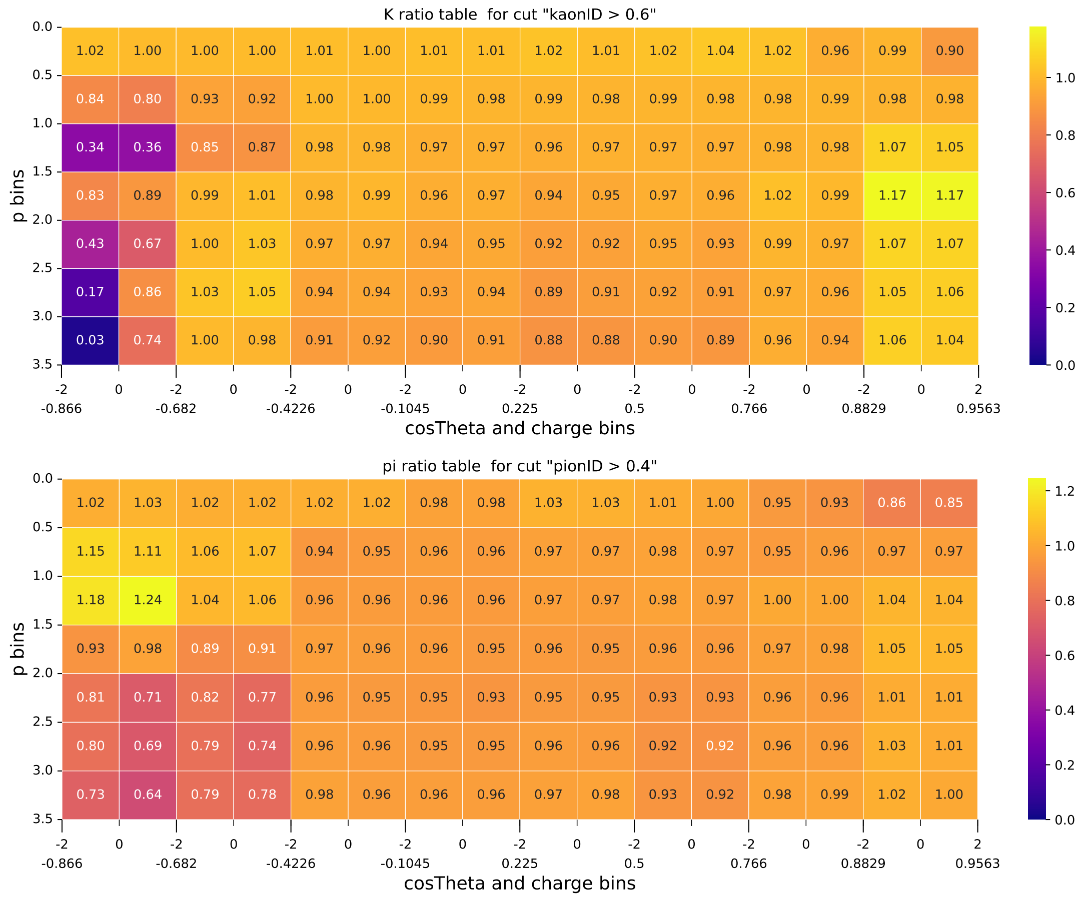


Figure H.1: Correction tables for hadronID efficiency.

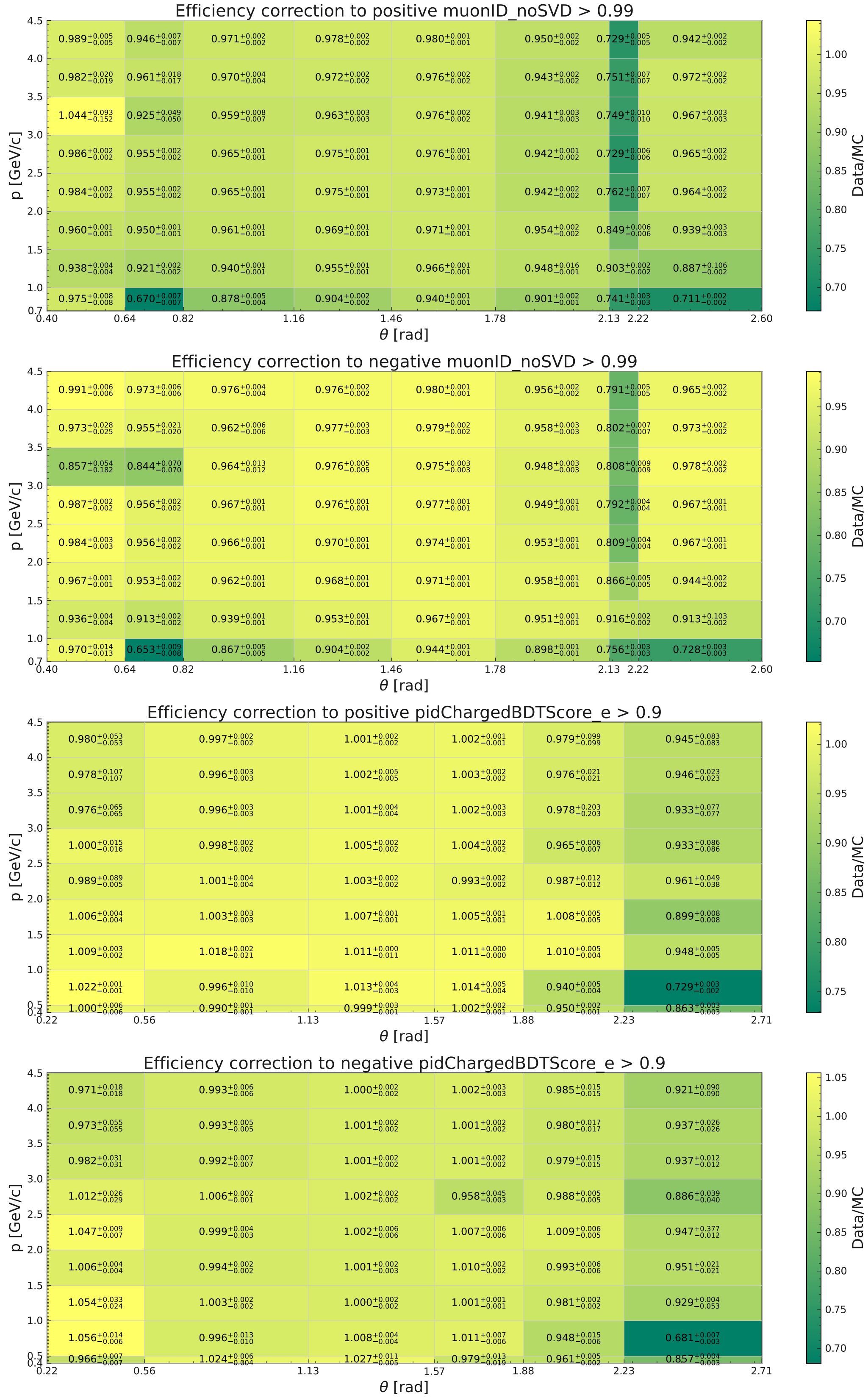


Figure H.2: Correction tables for leptonID efficiency.

Appendix I

Fitter check plots

In this appendix, we present the plots for fitter checks, including pull test (Figs. I.1 and I.2) and linearity test (Figs. I.3 and I.4). These tests results are shown separately for the fitting of each lepton flavor in each kinematic bin.

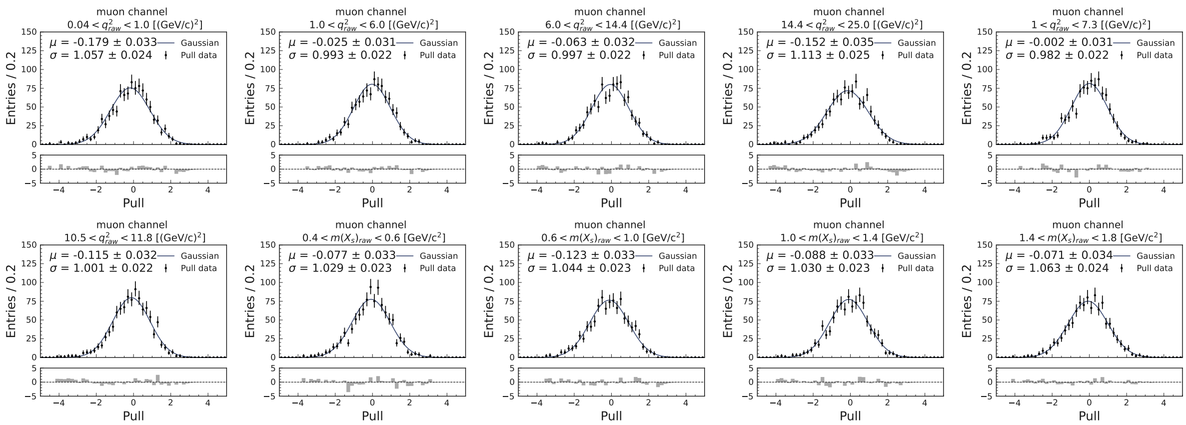


Figure I.1: Pull distributions of signal yield in toy MC study of muon channel.

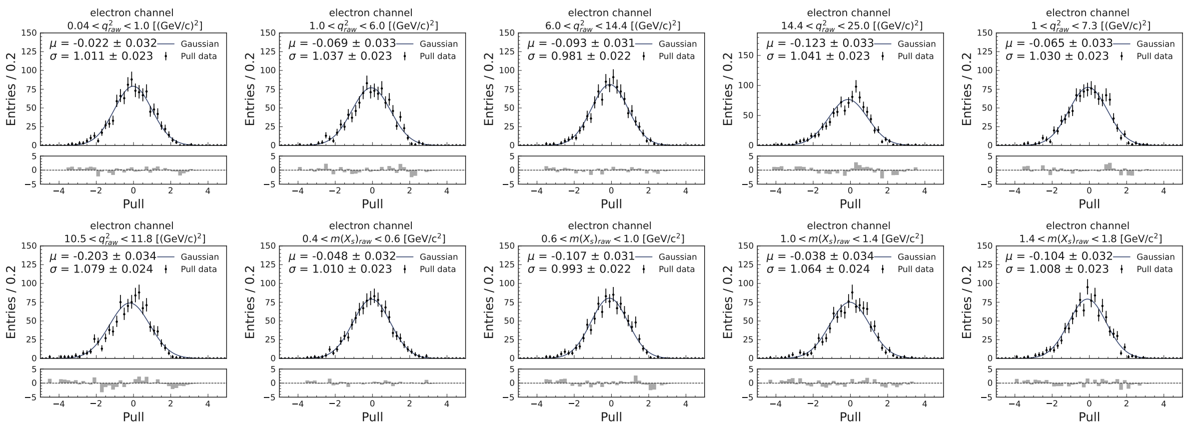


Figure I.2: Pull distributions of signal yield in toy MC study of electron channel.

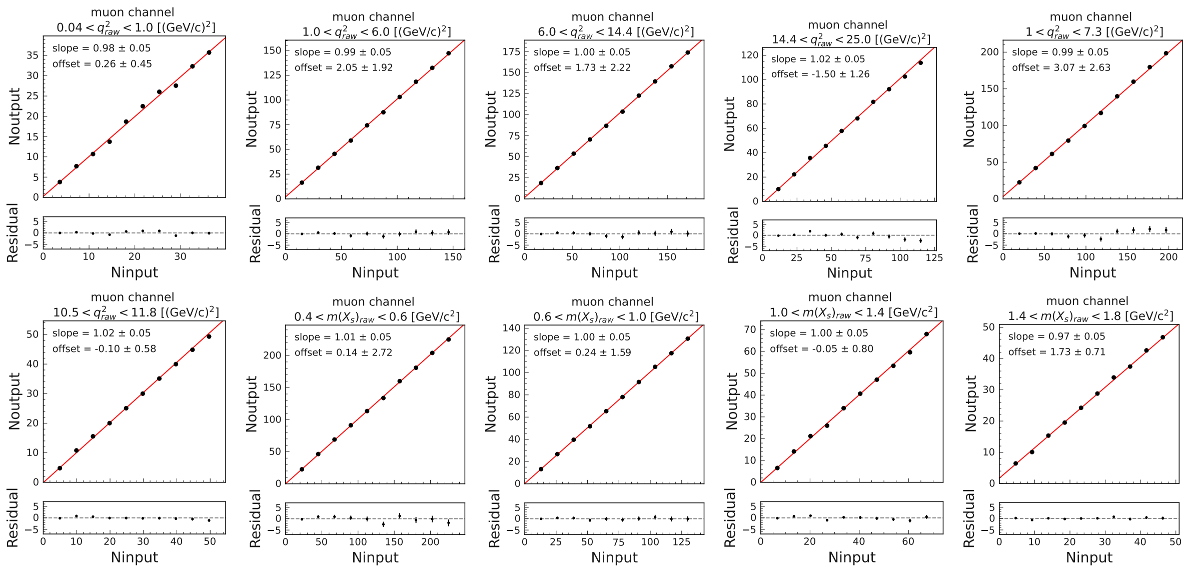


Figure I.3: Linearity test of output and input in muon channel.

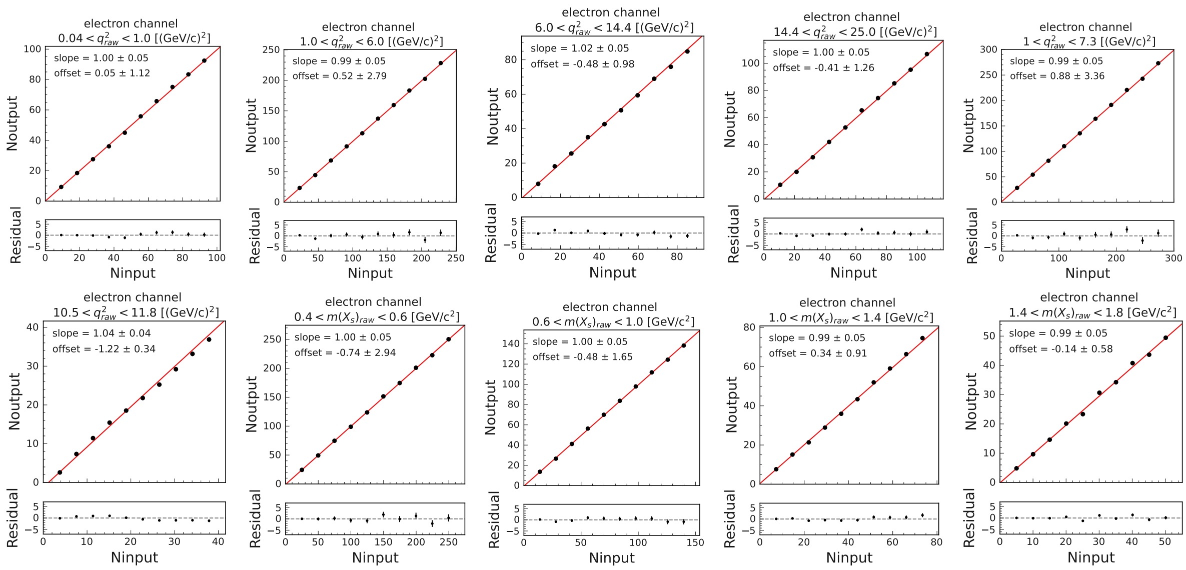


Figure I.4: Linearity test of output and input in electron channel.

Bibliography

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

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

[3] S. Fukae, C. S. Kim, T. Morozumi, and T. Yoshikawa, *A Model independent analysis of the rare B decay $B \to X_s\ell^+\ell^-$*, Phys. Rev. D **59** (1999) 074013, `arXiv:hep-ph/9807254`.

[4] A. J. Buras and M. Münz, *Effective Hamiltonian for $B \to X_s e^+e^-$ beyond leading logarithms in the naive dimensional regularization and Hooft–Veltman schemes*, Physical Review D **52** (1995) 186–195.

[5] B. Capdevila, *Status of the global $b \to s\ell^+\ell^-$ fits*, PoS **FPCP2023** (2023) 010.

[6] A. Ali, P. Ball, L. T. Handoko, and G. Hiller, *A Comparative study of the decays $B \to (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories*, Phys. Rev. D **61** (2000) 074024, `arXiv:hep-ph/9910221`.

[7] Y.-L. Yang *et al.*, *Scrutinizing B^0-meson flavor changing neutral current decay into scalar $K_0^*(1430)$ meson with $b \to s\ell^+\ell^-(\nu\bar{\nu})$ transition*, Eur. Phys. J. C **85** (2025) 64, `arXiv:2410.09363`.

[8] R.-H. Li, C.-D. Lu, and W. Wang, *Transition form factors of B decays into p-wave axial-vector mesons in the perturbative QCD approach*, Phys. Rev. D **79** (2009) 034014, `arXiv:0901.0307`.

[9] R.-H. Li, C.-D. Lu, and W. Wang, *Branching ratios, forward-backward asymmetries and angular distributions of $B \to K_2^* l^+l^-$ in the standard model and new physics scenarios*, Phys. Rev. D **83** (2011) 034034, `arXiv:1012.2129`.

[10] LHCb, R. Aaij *et al.*, *Determination of short- and long-distance contributions in B0→K*0μ+μ- decays*, Phys. Rev. D **109** (2024) 052009, `arXiv:2312.09102`.

[11] CMS, C. collaboration, *Angular analysis of the $B^0 \to K^{*0}(892)\mu^+\mu^-$ decay at $\sqrt{s}$ = 13 TeV*, .

[12] W. Altmannshofer *et al.*, *Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond*, JHEP **01** (2009) 019, `arXiv:0811.1214`.

[13] Belle, S. Wehle *et al.*, *Lepton-Flavor-Dependent Angular Analysis of $B \to K^*\ell^+\ell^-$*, Phys. Rev. Lett. **118** (2017) 111801, `arXiv:1612.05014`.

[14] L. collaboration, *Measurement of CP-Averaged Observables in the $B^0 \to K^{*0}\mu^+\mu^-$ Decay*, Physical Review Letters **125** (2020) .

[15] CMS, A. M. Sirunyan *et al.*, *Angular analysis of the decay $B^+ \to K^*(892)^+\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, JHEP **04** (2021) 124, `arXiv:2010.13968`.

[16] M. Ciuchini *et al.*, *Constraints on lepton universality violation from rare B decays*, Phys. Rev. D **107** (2023) 055036, `arXiv:2212.10516`.

[17] G. Isidori, Z. Polonsky, and A. Tinari, *Charm rescattering in $B^0 \to K^0\bar{\ell}\ell$: an improved analysis*, `arXiv:2507.17824`.

[18] R. Frezzotti *et al.*, *Theoretical framework for lattice QCD computations of $B \to K\ell^+\ell^-$ and $\bar{B}_s \to \ell^+\ell^-\gamma$ decays rates, including contributions from "Charming Penguins"*, `arXiv:2508.03655`.

[19] R. Barbieri, C. W. Murphy, and F. Senia, *B-decay Anomalies in a Composite Leptoquark Model*, Eur. Phys. J. C **77** (2017) 8, `arXiv:1611.04930`.

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

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

[22] BELLE, S. Choudhury *et al.*, *Test of lepton flavor universality and search for lepton flavor violation in $B \to K\ell\ell$ decays*, JHEP **03** (2021) 105, `arXiv:1908.01848`.

[23] LHCb, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, Nature Phys. **18** (2022) 277, `arXiv:2103.11769`, [Addendum: Nature Phys. 19, (2023)].

[24] LHCb, R. Aaij *et al.*, *Test of lepton universality in $b \to s\ell^+\ell^-$ decays*, Phys. Rev. Lett. **131** (2023) 051803, `arXiv:2212.09152`.

[25] LHCb, R. Aaij *et al.*, *Measurement of lepton universality parameters in $B^+ \to K^+\ell^+\ell^-$ and $B^0 \to K^{*0}\ell^+\ell^-$ decays*, Phys. Rev. D **108** (2023) 032002, `arXiv:2212.09153`.

[26] CMS, A. Hayrapetyan *et al.*, *Test of lepton flavor universality in $B^{\pm} \to K^{\pm}\mu^{+}\mu^{-}$ and $B^{\pm} \to K^{\pm}e^{+}e^{-}$ decays in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Rept. Prog. Phys. **87** (2024) 077802, `arXiv:2401.07090`.

[27] Belle, J. Kaneko *et al.*, *Measurement of the electroweak penguin process B —> X(s) l+ l-*, Phys. Rev. Lett. **90** (2003) 021801, `arXiv:hep-ex/0208029`.

[28] Belle, M. Iwasaki *et al.*, *Improved measurement of the electroweak penguin process $B \to X_s l^+ l^-$*, Phys. Rev. D **72** (2005) 092005, `arXiv:hep-ex/0503044`.

[29] BaBar, B. Aubert *et al.*, *Measurement of the $B \to X_s \ell^+ \ell^-$ branching fraction with a sum over exclusive modes*, Phys. Rev. Lett. **93** (2004) 081802, `arXiv:hep-ex/0404006`.

[30] BaBar, J. P. Lees *et al.*, *Measurement of the $B \to X_s l^+ l^-$ branching fraction and search for direct CP violation from a sum of exclusive final states*, Phys. Rev. Lett. **112** (2014) 211802, `arXiv:1312.5364`.

[31] B. Grinstein, M. J. Savage, and M. B. Wise, *B —> X(s) e+ e- in the Six Quark Model*, Nucl. Phys. B **319** (1989) 271.

[32] T. Huber *et al.*, *Phenomenology of inclusive $\overline{B} \to X_s \ell^+ \ell^-$ for the Belle II era*, JHEP **10** (2020) 088, `arXiv:2007.04191`.

[33] HFLAV, Y. S. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, Eur. Phys. J. C **81** (2021) 226, `arXiv:1909.12524`.

[34] C. S. Kim, Y. G. Kim, and K. Y. Lee, *On the motion of b quark inside B meson: The Parton model versus the ACCMM model*, Phys. Rev. D **57** (1998) 4002, `arXiv:hep-ph/9706310`.

[35] T. Sjostrand, *PYTHIA 5.7 and JETSET 7.4: Physics and manual*, `arXiv:hep-ph/9508391`.

[36] K. S. M. Lee, Z. Ligeti, I. W. Stewart, and F. J. Tackmann, *Extracting short distance information from $b \to s l^+ l^-$ effectively*, Phys. Rev. D **75** (2007) 034016, `arXiv:hep-ph/0612156`.

[37] Belle, Y. Sato *et al.*, *Measurement of the lepton forward-backward asymmetry in $B \to X_s \ell^+ \ell^-$ decays with a sum of exclusive modes*, Phys. Rev. D **93** (2016) 032008, `arXiv:1402.7134`, [Addendum: Phys.Rev.D 93, 059901 (2016)].

[38] CMS, LHCb, V. Khachatryan *et al.*, *Observation of the rare $B_s^0 \to \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68, `arXiv:1411.4413`.

[39] LHCb, R. Aaij *et al.*, *Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, `arXiv:1703.05747`.

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

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

[42] W. Altmannshofer, *New Physics Interpretations of the $B \to K^*\mu^+\mu^-$ Anomaly*, in *49th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, pp. 29–34, 2014, `arXiv:1405.5182`.

[43] J. Aebischer, A. J. Buras, M. Cerdà-Sevilla, and F. De Fazio, *Quark-lepton connections in Z' mediated FCNC processes: gauge anomaly cancellations at work*, JHEP **02** (2020) 183, `arXiv:1912.09308`.

[44] Y. Mambrini, *Z-prime and dark matter*, J. Phys. Conf. Ser. **315** (2011) 012014.

[45] Belle II collaboration, T. Abe, *Belle II technical design report*, `arXiv:1011.0352`.

[46] K. Akai, K. Furukawa, and H. Koiso, *SuperKEKB collider*, Nucl. Instrum. Meth. **A907** (2018) 188, `arXiv:1809.01958`.

[47] SuperB, M. Bona *et al.*, *SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report*, `arXiv:0709.0451`.

[48] B. Wang and others. *Operational experience of the belle ii pixel detector*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1032** (2022) 166631.

[49] Belle-II SVD, K. Adamczyk *et al.*, *The design, construction, operation and performance of the Belle II silicon vertex detector*, JINST **17** (2022) P11042, `arXiv:2201.09824`.

[50] T. V. 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 **930** (2019) 132.

[51] D. Kotchetkov *et al.*, *Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector*, Nucl. Instrum. Meth. A **941** (2019) 162342, `arXiv:1804.10782`.

[52] M. a. Yonenaga, *Performance evaluation of the aerogel rich counter for the belle ii spectrometer using early beam collision data*, Progress of Theoretical and Experimental Physics **2020** (2020) .

[53] V. Aulchenko *et al.*, *Development of data acquisition system for belle ii electromagnetic calorimeter*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1030** (2022) 166468.

[54] C. Ketter *et al.*, *Design and commissioning of readout electronics for a KL0 and μ detector at the Belle II experiment*, Nucl. Instrum. Meth. A **1082** (2026) 170893, arXiv:2502.02724.

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

[56] Belle-II, I. Adachi *et al.*, *Charged-hadron identification at Belle II*, Eur. Phys. J. C **85** (2025) 1237, arXiv:2506.04355.

[57] A. Novosel *et al.*, *Identification of light leptons and pions in the electromagnetic calorimeter of Belle II*, Nucl. Instrum. Meth. A **1056** (2023) 168630, arXiv:2301.05074.

[58] Belle II Framework Software Group, T. Kuhr *et al.*, *The Belle II Core Software*, Comput. Softw. Big Sci. **3** (2019) 1, arXiv:1809.04299.

[59] Belle II collaboration, *Belle II Analysis Software Framework (basf2)*, https://doi.org/10.5281/zenodo.5574115.

[60] Particle Data Group, R. L. Workman *et al.*, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[61] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[62] T. Sjöstrand *et al.*, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159, arXiv:1410.3012.

[63] S. Jadach, B. F. L. Ward, and Z. Wąs, *The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions*, Comput. Phys. Commun. **130** (2000) 260, arXiv:hep-ph/9912214.

[64] E. Barberio, B. van Eijk, and Z. Wąs, *PHOTOS: A universal Monte Carlo for QED radiative corrections in decays*, Comput. Phys. Commun. **66** (1991) 115.

[65] GEANT4 collaboration, S. Agostinelli *et al.*, *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[66] Belle, T. Saito *et al.*, *Measurement of the $\bar{B} \to X_s\gamma$ Branching Fraction with a Sum of Exclusive Decays*, Phys. Rev. D **91** (2015) 052004, arXiv:1411.7198.

[67] Heavy Flavor Averaging Group (HFLAV), S. Banerjee *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2023*, `arXiv:2411.18639`.

[68] A. Ali, *Next-to-leading order calculations of the radiative and semileptonic rare B decays in the standard model and comparison with data*, Nucl. Phys. B Proc. Suppl. **117** (2003) 624, `arXiv:hep-ph/0210183`.

[69] A. Ali, E. Lunghi, C. Greub, and G. Hiller, *Improved model independent analysis of semileptonic and radiative rare B decays*, Phys. Rev. D **66** (2002) 034002, `arXiv:hep-ph/0112300`.

[70] LHCb, R. Aaij *et al.*, *Differential branching fraction and angular moments analysis of the decay $B^0 \to K^+\pi^-\mu^+\mu^-$ in the $K^*_{0,2}(1430)^0$ region*, JHEP **12** (2016) 065, `arXiv:1609.04736`.

[71] LHCb, R. Aaij *et al.*, *First observations of the rare decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$*, JHEP **10** (2014) 064, `arXiv:1408.1137`.

[72] Belle-II analysis software Group, J.-F. Krohn *et al.*, *Global decay chain vertex fitting at Belle II*, Nucl. Instrum. Meth. A **976** (2020) 164269, `arXiv:1901.11198`.

[73] R. Sugiura *et al.*, *Measurement of time-dependent CP asymmetries in $B^0 \to K^+K^-K^0_S$ decays with Dalitz-plot analysis*, May 9, 2023. BELLE2-NOTE-PH-2023-021.

[74] R. Okubo *et al.*, *Measurement of Time-Dependent CP Asymmetry in $B^0 \to \rho^+\rho^-$ and Extraction of the CKM angle ϕ_2 using 2019-2022 data*, Feb. 20, 2024. BELLE2-NOTE-PH-2024-011.

[75] Y. Nakazawa *et al.*, *Measurement of branching fraction and longitudinal polarization in $B^0 \to \rho^+\rho^-$ using 2019-2021 data*, May 19, 2022. BELLE2-NOTE-PH-2022-026.

[76] Belle, S. H. Lee *et al.*, *Evidence for $B^0 \to \pi^0\pi^0$*, Phys. Rev. Lett. **91** (2003) 261801, `arXiv:hep-ex/0308040`.

[77] CLEO, D. M. Asner *et al.*, *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53** (1996) 1039, `arXiv:hep-ex/9508004`.

[78] T. Keck, *Fastbdt: A speed-optimized and cache-friendly implementation of stochastic gradient-boosted decision trees for multivariate classification*, 2016.

[79] S. O. Arik and T. Pfister, *Tabnet: Attentive interpretable tabular learning*, 2020.

[80] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, *Revisiting deep learning models for tabular data*, 2023.

[81] A. Paszke *et al.*, in *PyTorch: an imperative style, high-performance deep learning library*, Curran Associates Inc., Red Hook, NY, USA, 2019.

[82] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017.

[83] BaBar, B. Aubert *et al.*, *Study of inclusive production of charmonium mesons in B decay*, Phys. Rev. D **67** (2003) 032002, arXiv:hep-ex/0207097.

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

[85] BaBar, B. Aubert *et al.*, *Direct CP, Lepton Flavor and Isospin Asymmetries in the Decays $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. Lett. **102** (2009) 091803, arXiv:0807.4119.

[86] CDF, T. Aaltonen *et al.*, *Observation of the Baryonic Flavor-Changing Neutral Current Decay $\Lambda_b \to \Lambda\mu^+\mu^-$*, Phys. Rev. Lett. **107** (2011) 201802, arXiv:1107.3753.

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

[88] S. Watanuki *et al.*, *Study of CP and Isospin asymmetries in $b \to s\gamma$ using semi-inclusive reconstruction technique*, 2018. bn1480.

[89] Belle, S. Watanuki *et al.*, *Measurements of isospin asymmetry and difference of direct CP asymmetries in inclusive $B \to X_s\gamma$ decays*, Phys. Rev. D **99** (2019) 032012, arXiv:1807.04236.

[90] Y. Sato *et al.*, *Measurement of Forward-Backward Asymmetry in $B \to X_s\ell^+\ell^-$ Decays with Semi-inclusive Method*, 2013. bn1290.

[91] P. Rados *et al.*, *Tracking Efficiency with taus - Update on MC15rd*,

[92] D. Dorner *et al.*, *Relative tracking efficiency study of charged, low momentum pions*, July 07, 2023. BELLE2-NOTE-PH-2023-035.

[93] G. de Marino *et al.*, *K^0_S efficiency calibration on Run1 data vs MC15*, 2023. BELLE2-NOTE-TE-2023-023.

[94] Z. Wang *et al.*, *Muon identification with Deep Neural Network in the Belle II K-Long and Muon detector*, Nucl. Instrum. Meth. A **1081** (2026) 170814, arXiv:2503.11351.

[95] Belle II Tracking Group, V. Bertacchi *et al.*, *Track finding at Belle II*, Comput. Phys. Commun. **259** (2021) 107610, arXiv:2003.12466.

[96] A. L. Maas *et al.*, in *Rectifier Nonlinearities Improve Neural Network Acoustic Models*, Atlanta, Georgia, USA, 2013.

[97] A. Natochii *et al.*, *Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider*, Nucl. Instrum. Meth. A **1055** (2023) 168550, `arXiv:2302.01566`.