PhD Thesis

Measurement of Branching Fractions of $B \to X_s \ell^+ \ell^-$ Decays at the Belle II experiment

(Belle II 実験における $B \to X_s \ell^+ \ell^-$ 過程の崩壊分岐比の測定)

Yo Sato

Department of Physics
Graduate School of Science
Tohoku University

2020

Abstract

The inclusive $B \to X_s \ell^+ \ell^-$ decays are a great probe to search for physics beyond the standard model (SM) of particle physics. The process is a flavor-changing-neutral-current (FCNC) process which proceeds via loop diagrams in the standard model and thus are strongly suppressed. Since a new heavy particle might be able to enter the loop, the FCNC is sensitive for new physics. Moreover, $B \to X_s \ell^+ \ell^-$ decays provide complementary information with less hadronic uncertainty to the exclusive $B \to K^{(*)} \ell^+ \ell^-$ decays in which tensions from the SM prediction have been observed. Belle II is a unique experiment to explore the process with large statistics to shed light on the anomalies.

We performed the measurements of the branching fractions of $B \to X_s \ell^+ \ell^-$ decay using the data set accumulated by Belle II experiment which corresponds to 37.7 million $B\bar{B}$ pairs. This is the first measurement on $B \to X_s \ell^+ \ell^-$ at Belle II experiment. The obtained results are

$$\mathcal{B}(B \to X_s e^+ e^-) = [4.86^{+2.75}_{-2.42}(\text{stat})^{+1.02}_{-0.92}(\text{syst})] \times 10^{-6} \tag{1}$$

$$\mathcal{B}(B \to X_s \mu^+ \mu^-) < 4.67(5.61) \times 10^{-6} \text{ at } 90\%(95\%) \text{ CL} \tag{2}$$

$$\mathcal{B}(B \to X_s \ell^+ \ell^-) < 5.54(6.30) \times 10^{-6} \text{ at } 90\%(95\%) \text{ CL} \tag{3}$$

Because the statistical significance on $B \to X_s \mu^+ \mu^-$ and $B \to X_s \ell^+ \ell^-$ is less than 2σ, the upper limit on the branching fraction is set for these modes. The branching fraction of $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$ is consistent with previous measurements and the SM prediction. Result of $B \to X_s \ell^+ \ell^-$ is consistent with the world average, Belle measurement and the SM prediction, while the difference from BaBar is at 1.4σ level.

The analysis procedure of $B \to X_s \ell^+ \ell^-$ decays at Belle II experiment well established and we have got ready to lead decisive conclusions regarding the anomalies which are observed in the exclusive $B \to K^{(*)} \ell^+ \ell^-$ decays with upcoming Belle II data.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Standard Model (SM) of elementary particle physics is an outstanding theory to match various experiment results. The SM consists of quark and lepton which compose matter, gauge interaction, and the Higgs mechanism. Higgs boson which was the last piece of SM has been discovered in ATLAS and CMS experiment in 2012 [18] [19]. Despite great successes, the SM is thought to be not perfect theory. There remain some open questions, for example, absence of the dark matter candidates, and a description of gravity is not included in the SM. Therefore, the effort to search for physics beyond the SM from both experimental and theoretical aspects is highly motivated.

Experiment with high energy accelerator is a promising tool for new physics exploration. There are two kinds of experiments in this field, energy frontier experiment and intensity frontier one. In the energy frontier experiment, a direct production of new particles is main target process. Some new physics models such as supersymmetry (SUSY) [20, 21, 22, 23, 24, 25] predict new heavy particles. The highest centre-of-mass energy is around 13 TeV by LHC (Large Hadron Collider) and mass reach at ATLAS and CMS experiments for new heavy particle is $\mathcal{O}(1)$ TeV. Figure 1.1 shows the mass reach for SUSY at ATLAS. Currently, no evidence of new particles is observed, which may suggest new physics candidates exist higher than $\mathcal{O}(1)$ TeV.

While, the intensity frontier experiments perform precise measurements with high statistics data to probe new physics. One of the most promising processes in the intensity frontier experiments is the flavor-changing-neutral-current (FCNC) process, such as a $b \to s\ell^+\ell^-$ transition. The FCNC is suppressed in the SM and proceeds mainly via loop diagrams. New heavy particles might be able to enter the loops or even proceed the FCNC process at tree diagram level, which leads deviations on observables from the SM prediction.

In the analysis of $B \to K^{(*)}\ell^+\ell^-$ decays, which is proceeded by $b \to s\ell^+\ell^-$ at the quark level, some tensions with the SM have been reported by LHCb [3] [26] [27], Belle [4], and ATLAS [28]. The deviation from the SM in an angular observable, so-called P_5', in $B \to K^*\mu^+\mu^-$ decays is at 2.9σ level [3]. In the measurement of the ratio of branching fraction between muon modes and electron modes $R_{K^{(*)}} = \mathcal{B}(B \to K^{(*)}\mu^+\mu^-)/\mathcal{B}(B \to K^{(*)}\mu^+\mu^-)$, the deviation at 2.5σ level has been observed [26] [27]. The mass scale of new physics behind the anomalies is $\mathcal{O}(10)$ TeV which is much higher than the centre-of-mass energy of Belle and BaBar. Furthermore, these measurement may imply a violation of the lepton flavor universality (LFU). In the SM, three charged leptons e, μ and τ are identical except for the Yukawa couplings and thus masses. The LFU violation is a clear evidence of physics beyond the SM and may provide a hint for the fermion generation problem.

Inclusive $B \to X_s\ell^+\ell^-$ decays provide a complementary information to exclusive $B \to K^{(*)}\ell^+\ell^-$. The hadronic uncertainties in inclusive decays are under better control and are largely independent of those in exclusive ones. All existing $B \to X_s\ell^+\ell^-$ measurements are highly statistically limited due to the limited data sample. The Belle II experiment, which is a successor of Belle, is a unique experiment to perform measurements on inclusive $B \to X_s\ell^+\ell^-$ decays. Sensitivity at Belle II with full integrated luminosity is expected to be sufficient to play a decisive role to search for new physics.

In this paper, we report the measurement of the branching fractions of $B \to X_s\ell^+\ell^-$ decays at Belle II experiment. We use a data sample containing 37.7×10^6 $B\bar{B}$ pairs corrected by Belle II with the SuperKEKB accelerator. This is the first measurement on $B \to X_s\ell^+\ell^-$ decays at Belle II and opens a new era in the flavor physics.

ATLAS SUSY Searches* - 95% CL Lower Limits

July 2020

ATLAS Preliminary

$\sqrt{s} = 13$ TeV

Model	Signature		$\int \mathcal{L}\,dt$ [fb^{-1}]	Mass limit		Reference

Inclusive Searches

Model	Signature		$\int \mathcal{L}\,dt$ [fb^{-1}]	Mass limit	Reference
$\tilde{q}\tilde{q}, \tilde{q}\to q\tilde{\chi}_1^0$	$0\ e,\mu$ 2-6 jets mono-jet 1-3 jets	E_T^{miss} E_T^{miss}	139 36.1	$\tilde{q}$ [10× Degen.] 1.9; $\tilde{q}$ [1×, 8× Degen.] 0.43 0.71	ATLAS-CONF-2019-040 1711.03301
$\tilde{g}\tilde{g}, \tilde{g}\to q\bar{q}\tilde{\chi}_1^0$	$0\ e,\mu$ 2-6 jets	E_T^{miss}	139	$\tilde{g}$ 2.35; $\tilde{g}$ Forbidden 1.15-1.95	ATLAS-CONF-2019-040 ATLAS-CONF-2019-040
$\tilde{g}\tilde{g}, \tilde{g}\to q\bar{q}W\tilde{\chi}_1^0$	$1\ e,\mu$ 2-6 jets	E_T^{miss}	139	$\tilde{g}$ 2.2	ATLAS-CONF-2020-047
$\tilde{g}\tilde{g}, \tilde{g}\to q\bar{q}(\ell\ell)\tilde{\chi}_1^0$	$ee,\mu\mu$ 2 jets	E_T^{miss}	36.1	$\tilde{g}$ 1.2	1805.11381
$\tilde{g}\tilde{g}, \tilde{g}\to q\bar{q}WZ\tilde{\chi}_1^0$	$0\ e,\mu$ 7-11 jets SS e,μ 6 jets	E_T^{miss}	139 139	$\tilde{g}$ 1.97; $\tilde{g}$ 1.15	ATLAS-CONF-2020-002 1909.08457
$\tilde{g}\tilde{g}, \tilde{g}\to t\bar{t}\tilde{\chi}_1^0$	$0-1\ e,\mu$ 3 b SS e,μ 6 jets	E_T^{miss}	79.8 139	$\tilde{g}$ 2.25; $\tilde{g}$ 1.25	ATLAS-CONF-2018-041 1909.08457

3rd gen. squarks direct production

Model	Signature		$\int \mathcal{L}\,dt$ [fb^{-1}]	Mass limit	Reference
$\tilde{b}_1\tilde{b}_1, \tilde{b}_1\to b\tilde{\chi}_1^0/t\tilde{\chi}_1^\pm$	Multiple Multiple		36.1 139	$\tilde{b}_1$ Forbidden 0.9; $\tilde{b}_1$ Forbidden 0.74	1708.09266, 1711.03301 1909.08457
$\tilde{b}_1\tilde{b}_1, \tilde{b}_1\to b\tilde{\chi}_2^0\to bh\tilde{\chi}_1^0$	$0\ e,\mu$ 6 b $2\ \tau$ 2 b	E_T^{miss} E_T^{miss}	139 139	$\tilde{b}_1$ Forbidden 0.23-1.35; $\tilde{b}_1$ 0.13-0.85	1908.03122 ATLAS-CONF-2020-031
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to t\tilde{\chi}_1^0$	$0-1\ e,\mu$ ≥ 1 jet	E_T^{miss}	139	$\tilde{t}_1$ 1.25	ATLAS-CONF-2020-003, 2004.14060
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to Wb\tilde{\chi}_1^0$	$1\ e,\mu$ 3 jets/1 b	E_T^{miss}	139	$\tilde{t}_1$ 0.44-0.59	ATLAS-CONF-2019-017
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to \tilde{\tau}_1 bv, \tilde{\tau}_1\to\tau\tilde{G}$	$1\ \tau + 1\ e,\mu,\tau$ 2 jets/1 b	E_T^{miss}	36.1	$\tilde{t}_1$ 1.16	1803.10178
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to c\tilde{\chi}_1^0/\tilde{c}\tilde{c}, \tilde{c}\to c\tilde{\chi}_1^0$	$0\ e,\mu$ 2 c	E_T^{miss}	36.1	$\tilde{c}$ 0.85; $\tilde{t}_1$ 0.46	1805.01649 1805.01649
	$0\ e,\mu$ mono-jet	E_T^{miss}	36.1	$\tilde{t}_1$ 0.43	1711.03301
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to t\tilde{\chi}_2^0, \tilde{\chi}_2^0\to Z/h\tilde{\chi}_1^0$	1-2 e,μ 1-4 b	E_T^{miss}	139	$\tilde{t}_1$ 0.067-1.18	SUSY-2018-09
$\tilde{t}_2\tilde{t}_2, \tilde{t}_2\to \tilde{t}_1 + Z$	3 e,μ 1 b	E_T^{miss}	139	$\tilde{t}_2$ Forbidden 0.86	SUSY-2018-09

EW direct

Model	Signature		$\int \mathcal{L}\,dt$ [fb^{-1}]	Mass limit	Reference
$\tilde{\chi}_1^\pm\tilde{\chi}_2^0$ via WZ	3 e,μ $ee,\mu\mu$ ≥ 1 jet	E_T^{miss} E_T^{miss}	139 139	$\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ 0.64; $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ 0.205	ATLAS-CONF-2020-015 1911.12606
$\tilde{\chi}_1^+\tilde{\chi}_1^-$ via WW	2 e,μ	E_T^{miss}	139	$\tilde{\chi}_1^\pm$ 0.42	1908.08215
$\tilde{\chi}_1^+\tilde{\chi}_2^0$ via Wh	0-1 e,μ 2 $b/2\ \gamma$	E_T^{miss}	139	$\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ Forbidden 0.74	2004.10894, 1909.09226
$\tilde{\chi}_1^+\tilde{\chi}_1^-$ via $\tilde{\ell}_L/\tilde{\nu}$	2 e,μ	E_T^{miss}	139	$\tilde{\chi}_1^\pm$ 1.0	1908.08215
$\tilde{\tau}\tilde{\tau}, \tilde{\tau}\to\tau\tilde{\chi}_1^0$	$2\ \tau$	E_T^{miss}	139	$\tilde{\tau}$ [$\tilde{\tau}_L, \tilde{\tau}_{R,L}$] 0.16-0.3 0.12-0.39	1911.06660
$\tilde{\ell}_{L,R}\tilde{\ell}_{L,R}, \tilde{\ell}\to\ell\tilde{\chi}_1^0$	2 e,μ 0 jets $ee,\mu\mu$ ≥ 1 jet	E_T^{miss} E_T^{miss}	139 139	$\tilde{\ell}$ 0.7; $\tilde{\ell}$ 0.256	1908.08215 1911.12606
$\tilde{H}\tilde{H}, \tilde{H}\to h\tilde{G}/Z\tilde{G}$	0 e,μ ≥ 3 b 4 e,μ 0 jets	E_T^{miss} E_T^{miss}	36.1 139	$\tilde{H}$ 0.13-0.23 0.29-0.88; $\tilde{H}$ 0.55	1806.04030 ATLAS-CONF-2020-040

Long-lived particles

Model	Signature		$\int \mathcal{L}\,dt$ [fb^{-1}]	Mass limit	Reference
Direct $\tilde{\chi}_1^+\tilde{\chi}_1^-$ prod., long-lived $\tilde{\chi}_1^\pm$	Disapp. trk 1 jet	E_T^{miss}	36.1	$\tilde{\chi}_1^\pm$ 0.46; $\tilde{\chi}_1^\pm$ 0.15	1712.02118 ATL-PHYS-PUB-2017-019
Stable $\tilde{g}$ R-hadron	Multiple		36.1	$\tilde{g}$ 2.0	1902.01636, 1808.04095
Metastable $\tilde{g}$ R-hadron, $\tilde{g}\to qq\tilde{\chi}_1^0$	Multiple		36.1	$\tilde{g}$ [$\tau(\tilde{g})$ =10 ns, 0.2 ns] 2.05 2.4	1710.04901, 1808.04095

RPV

Model	Signature		$\int \mathcal{L}\,dt$ [fb^{-1}]	Mass limit	Reference
$\tilde{\chi}_1^+\tilde{\chi}_1^-/\tilde{\chi}_1^0, \tilde{\chi}_1^+\to Z\ell\to\ell\ell\ell$	3 e,μ		139	$\tilde{\chi}_1^\pm/\tilde{\chi}_1^0$ [BR(Zτ)=1, BR(Ze)=1] 0.625 1.05	ATLAS-CONF-2020-009
LFV $pp\to\tilde{\nu}_\tau + X, \tilde{\nu}_\tau\to e\mu/e\tau/\mu\tau$	$e\mu,e\tau,\mu\tau$		3.2	$\tilde{\nu}_\tau$ 1.9	1607.08079
$\tilde{\chi}_1^+\tilde{\chi}_1^-/\tilde{\chi}_2^0\to WW/Z\ell\ell\ell\nu\nu$	4 e,μ 0 jets	E_T^{miss}	36.1	$\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ 0.82 1.33	1804.03602
$\tilde{g}\tilde{g}, \tilde{g}\to qq\tilde{\chi}_1^0, \tilde{\chi}_1^0\to qqq$	4-5 large-R jets Multiple		36.1 36.1	$\tilde{g}$ 1.3 1.9; $\tilde{g}$ 1.05 2.0	1804.03568 ATLAS-CONF-2018-003
$\tilde{t}\tilde{t}, \tilde{t}\to t\tilde{\chi}_1^0, \tilde{\chi}_1^0\to tbs$	Multiple		36.1	$\tilde{t}$ 0.55 1.05	ATLAS-CONF-2018-003
$\tilde{t}\tilde{t}, \tilde{t}\to b\tilde{\chi}_1^\pm, \tilde{\chi}_1^\pm\to bbs$	≥ 4b		139	$\tilde{t}$ Forbidden 0.95	ATLAS-CONF-2020-016
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to bs$	2 jets + 2 b		36.7	$\tilde{t}_1$ [qq, bs] 0.42 0.61	1710.07171
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1\to q\ell$	2 e,μ 2 b 1 μ DV		36.1 136	$\tilde{t}_1$ 0.4-1.45; $\tilde{t}_1$ 1.0 1.6	1710.05544 2003.11956

Mass scale [TeV] — 10^{-1} to 1

Only a selection of the available mass limits on new states or phenomena is shown. Many of the limits are based on simplified models, c.f. refs. for the assumptions made.

FIG 1.1: Mass reach of the ATLAS searches for SUSY.

The outline of this thesis is as follows. The physics motivation of this thesis and measurements on $b \to s\ell^+\ell^-$ are described in Chapter 2. The Belle II experiment and SuperKEKB accelerator are explained in Chapter 3. An overview on the analysis procedure is given in Chapter 4. Event reconstruction procedure is described in Chapter 5. Methods of background suppression and peaking background estimation are described in Chapter 6. A method of the branching fraction extraction is explained in Chapter 7. Sources of systematic uncertainty are listed and estimation methods are decried in Chapter 8. Measurement of the branching fraction on $B \to X_s J/\psi$ control samples is shown for validation of the analysis procedure in Chapter 9. Obtained results of the branching fraction of $B \to X_s\ell^+\ell^-$ and discussion are described in Chapter 10. Finally, we conclude this paper in Chapter 11.

Chapter 2

Physics Motivation

2.1 Overview

The goal of this thesis is to perform the measurement of the branching fraction of $B \to X_s \ell^+ \ell^-$ decays as the first measurement on the $B \to X_s \ell^+ \ell^-$ process at the Belle II experiment.

Radiative penguin $b \to s(d)\gamma$ decay and electroweak penguin $b \to s(d)\ell^+\ell^-$ and $b \to s(d)\nu\bar{\nu}$ decays are FCNC processes. Thanks to the presence of photons and leptons in the final state, the size of non-perturbative QCD corrections can be reduced compared with fully hadronic decay. Figure 2.1 shows the SM Feynman diagrams of these decays. In the SM, t-quark and weak bosons $Z^0, W^\pm$ which are much heavier than b-quark mediate the process. New heavy particles may enter the loop or even proceed via tree diagrams, which induce a deviation on observables with the SM prediction. Thus, these processes are a good probe to search new physics.

The $B \to K\ell^+\ell^-$ decay has been firstly observed by Belle in $b \to s\ell^+\ell^-$ transition [29]. The branching fraction of $b \to s\ell^+\ell^-$ is $\mathcal{O}(10^{-6})$ which is two orders of magnitude smaller than $b \to s\gamma$ because of additional vertex leading α_{em} factor. On the other hand, due to the two leptons, an angular analysis can provide rich information to explore new physics. In this chapter, we introduce the theoretical framework and measurements on $b \to s\ell^+\ell^-$.
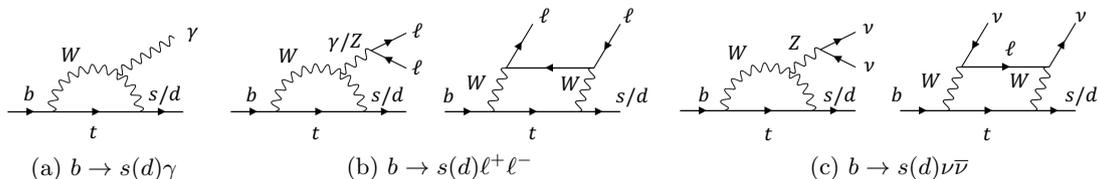
(a) $b \to s(d)\gamma$ (b) $b \to s(d)\ell^+\ell^-$ (c) $b \to s(d)\nu\bar{\nu}$

FIG 2.1: Feynman diagrams of radiative and electroweak penguin decays in the SM. (a) $b \to s(d)\gamma$, (b) $b \to s(d)\ell^+\ell^-$, (c) $b \to s(d)\nu\bar{\nu}$

2.2 Effective Hamiltonian

Flavor changing process is described from wide mass scale from $\Lambda_{\mathrm{QCD}} \sim 400\,\mathrm{MeV}$ over $m_b \sim 4.3\,\mathrm{GeV}$ to $M_W = 80.4\,\mathrm{GeV}$ and $M_t \sim 165\,\mathrm{GeV}$. In scale of m_b and above, the QCD effects can be calculated with perturbation theory, though it is difficulty for the dynamics associated with the energy scale Λ_{QCD}. The effective Hamiltonian is described using the operator product expansion technique to separate these different scales [30] [31] [32]. In this framework, the effective Hamiltonian for $b \to s$ transitions in the SM are given by [33]

$$\mathcal{H}_{\mathrm{eff}}^{\mathrm{SM}} = -\frac{4G_F}{\sqrt{2}} \left[\lambda_t^s \sum_{i=1}^{10} C_i(\mu)\mathcal{O}_i(\mu) + \lambda_u^s \sum_{i=1}^{2} C_i(\mu)(\mathcal{O}_i(\mu) - \mathcal{O}_i^u(\mu)) \right] \tag{2.1}$$

where G_F is the Fermi constant, λ_q^s is a product of CKM matrix elements such as $\lambda_t^s = V_{tb}V_{ts}^*$, and μ is the energy scale at which the calculation is being performed. There is the unitary relation, $\lambda_u^s + \lambda_c^s + \lambda_t^s =$

0. The $\mathcal{O}_i$ are local operators providing the "long-distance" descriptions. The each operator has an associated Wilson coefficient C_i which describes "short-distance" physics with the perturbation theory. The Wilson coefficients C_i are evaluated at a scale μ_W which is of the order of the W-boson mass. The renormalization group equation can be used to evolve the C_i to the scale μ_b which is of the order of the m_b.

Expressions of the current-current ($\mathcal{O}_{1,2}$), photonic dipole ($\mathcal{O}_7$), gluonic dipole ($\mathcal{O}_8$), and the vector and axial-vector of electroweak penguin ($\mathcal{O}_9$ and $\mathcal{O}_{10}$) are described as follows:

$$\mathcal{O}_1 = (\bar{s}_L \gamma_\mu T^a c_L)(\bar{c}_L \gamma^\mu T^a b_L), \tag{2.2}$$

$$\mathcal{O}_2 = (\bar{s}_L \gamma_\mu c_L)(\bar{c}_L \gamma^\mu b_L), \tag{2.3}$$

$$\mathcal{O}_7 = \frac{e}{16\pi^2} m_b (\bar{s}_L \sigma^{\mu\nu} b_R) F_{\mu\nu}, \tag{2.4}$$

$$\mathcal{O}_8 = \frac{e}{16\pi^2} m_b (\bar{s}_L \sigma^{\mu\nu} T^a b_R) G^a_{\mu\nu}, \tag{2.5}$$

$$\mathcal{O}_9 = \frac{e}{16\pi^2} m_b (\bar{s}_L \gamma_\mu b_L) \sum_\ell (\bar{\ell} \gamma^\mu \ell), \tag{2.6}$$

$$\mathcal{O}_{10} = \frac{e}{16\pi^2} m_b (\bar{s}_L \gamma_\mu b_L) \sum_\ell (\bar{\ell} \gamma^\mu \gamma_5 \ell). \tag{2.7}$$

where T^a are the SU(3)$_c$ generators, $F_{\mu\nu}$ and $G^a_{\mu\nu}$ are the photon and gluon field-strength tensors. The subscripts L and R represent the chirality of the quark fields. The operators $\mathcal{O}^u_{1,2}$ are obtained from $\mathcal{O}_{1,2}$ by replacing the c-quark by u-quark fields. The sums run over the quark flavors $q = u, d, s, c, b$. The same set of operators for $b \to d$ processes can be written by replacing s-quark and d-quark fields. The values of the Wilson coefficient $C_{1,2} \sim 1$ at the μ_b scale. In comparison, $C_{3,4,5,6}$ are very small at the scale and hence contributions from the four-quark ($\mathcal{O}_{3,4,5,6}$) operators

$$\mathcal{O}_3 = (\bar{s}_L \gamma_\mu b_L) \sum_q (\bar{q} \gamma^\mu q), \tag{2.8}$$

$$\mathcal{O}_4 = (\bar{s}_L \gamma_\mu T^a b_L) \sum_q (\bar{q} \gamma^\mu T^a q), \tag{2.9}$$

$$\mathcal{O}_5 = (\bar{s}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} b_L) \sum_q (\bar{q} \gamma^{\mu_1} \gamma^{\mu_2} \gamma^{\mu_3} q), \tag{2.10}$$

$$\mathcal{O}_6 = (\bar{s}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} T^a b_L) \sum_q (\bar{q} \gamma^{\mu_1} \gamma^{\mu_2} \gamma^{\mu_3} T^a q). \tag{2.11}$$

can be neglected. In the electroweak penguin decays $b \to s\ell^+\ell^-$, the operators $\mathcal{O}_{7,9,10}$ are the most relevant. Figure 2.2 shows the Feynman diagrams of $b \to s\ell^+\ell^-$ process in the Effective Hamiltonian framework. The values of Wilson coefficients are followings, $C_7(\mu_b) = -0.330, C_9(\mu_b) = 4.069, C_{10}(\mu_b) = -4.231$ [34] [35]. The effects of new physics beyond the SM can appear through modified values of Wilson coefficients and/or through additional operators with different chirality or flavor structure.

(a) $\mathcal{O}_7$ contribution

(b) $\mathcal{O}_{9,10}$ contribution
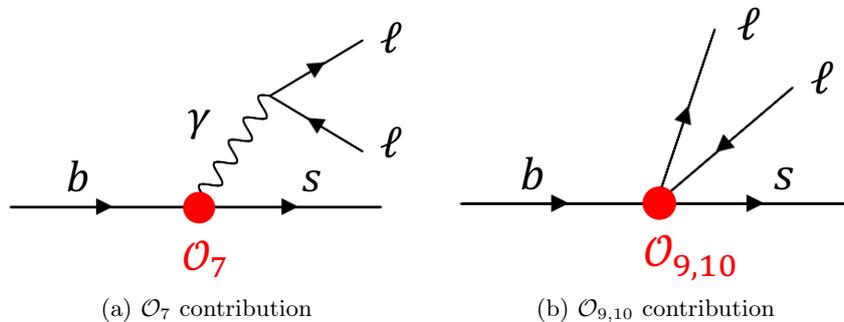
FIG 2.2: Feynman diagrams of $b \to s\ell^+\ell^-$ process in the Effective Hamiltonian framework. (a) $\mathcal{O}_7$ contribution, (b) $\mathcal{O}_{9,10}$ contribution.

2.3 Inclusive $B \to X_s \ell^+ \ell^-$ decays

Inclusive $B \to X_s \ell^+ \ell^-$ decays, where X_s is an inclusive hadronic state including a s-quark, provide information to the $b \to s \ell^+ \ell^-$ process with good theoretical predictions. Compared with exclusive $B \to K^{(*)} \ell^+ \ell^-$ decays, the hadronic uncertainty is under better control. Complementary information to the exclusive decays can be provided to shed light on the anomalies.

There are two main kinematic variables in $B \to X_s \ell^+ \ell^-$, the di-lepton invariant mass-squared $q^2 = M_{\ell^+ \ell^-}$ and the angle between direction of ℓ^+ [ℓ^-] and initial direction of $\bar{B}^0$ or B^- [B^0 or B^+] in the di-lepton centre-of-mass system θ_ℓ. Two observables have been discussed in inclusive $B \to X_s \ell^+ \ell^-$ decays, the q^2 spectrum $d\Gamma/dq^2$ [36] and the forward-backward asymmetry dA_{FB}/dq^2 [37]. They are mainly related to C_7, C_9 and C_{10} and can be derived at lowest order as followings,

$$\frac{d\Gamma}{dq^2} = \Gamma_0 m_b^3 (1-s)^2 \left[(|C_9|^2 + |C_{10}|^2)(1+2s) + \frac{4}{s}|C_7|^2(2+s) + 12\mathrm{Re}(C_7^* C_9) \right], \tag{2.12}$$

$$\frac{dA_{FB}}{dq^2} = \int_{-1}^{1} dz \frac{d^2\Gamma}{dq^2 dz} \mathrm{sign}(z) = -3\Gamma_0 m_b^3 (1-s)^2 s \mathrm{Re}\left(C_{10}\left(C_9 + \frac{2}{s}C_7 \right) \right). \tag{2.13}$$

where $s = q^2/m_b^2$, $z = \cos\theta_\ell$ and

$$\Gamma_0 = \frac{G_F^2}{48\pi^3} \frac{\alpha_{\mathrm{em}}^2}{16\pi^2} |V_{tb}V_{ts}^*|^2. \tag{2.14}$$

An agular decomposition provides a third observable which has different dependency to the Wilson coefficients. The double-differential decay width can be written as following [38].

$$\frac{d^2\Gamma}{dq^2 d\cos\theta_\ell} = \frac{3}{8}[(1+\cos^2\theta_\ell)H_T(q^2) + 2\cos\theta_\ell H_A(q^2) + 2(1-\cos^2\theta_\ell)H_L(q^2)] \tag{2.15}$$

The q^2 spectrum and the forward-backward asymmetry can be derived from the functions $H_i(q^2)$,

$$\frac{d\Gamma}{dq^2} = H_T(q^2) + H_L(q^2), \tag{2.16}$$

$$\frac{dA_{FB}}{dq^2} = \frac{3}{4}H_A(q^2). \tag{2.17}$$

The functions $H_i(q^2)$ are

$$H_T(q^2) = \Gamma_0 m_b^3 \cdot 2s(1-s)^2 \left[\left(|C_9 + \frac{2}{s}C_7|^2 + |C_{10}|^2 \right) \right], \tag{2.18}$$

$$H_L(q^2) = \Gamma_0 m_b^3 \cdot (1-s)^2 \left[(|C_9 + 2C_7|^2 + |C_{10}|^2 \right], \tag{2.19}$$

$$H_A(q^2) = -\frac{9}{4}\Gamma_0 m_b^3 (1-s)^2 s \mathrm{Re}\left(C_{10}\left(C_9 + \frac{2}{s}C_7 \right) \right). \tag{2.20}$$

Additional information on the Wilson coefficients can be obtained from the lepton flavor universality test observables,

$$R_{X_s}[q_0^2, q_1^2] = \frac{\int_{q_0^2}^{q_1^2} dq^2 \frac{d\Gamma(B \to X_s \mu^+ \mu^-)}{dq^2}}{\int_{q_0^2}^{q_1^2} dq^2 \frac{d\Gamma(B \to X_s e^+ e^-)}{dq^2}}. \tag{2.21}$$

Prediction of R_{X_s} in the SM is unity with high precision due to the lepton flavor universality. Taking interference BSM effect with the SM into account, R_{X_s} can be approximated by [39]

$$R_{X_s} \simeq 1 + (\Delta_+ + \Delta_-)/2, \tag{2.22}$$

$$\Delta_\pm = \frac{2}{|C_9^{\mathrm{SM}}|^2 + |C_{10}^{\mathrm{SM}}|^2} \left[\sum_{i=9,10} \mathrm{Re}\left(C_i^{\mathrm{SM}}(C_i^{\mathrm{NP}\mu} + C_i^{'\mu}) \right) + \sum_{i=9,10} \mathrm{Re}\left(C_i^{\mathrm{SM}}(C_i^{\mathrm{NP}e} + C_i^{'e}) \right) \right] \tag{2.23}$$

C' is the Wilson coefficient of the chirality flipped operator which is zero in the SM. The labels SM and NP denote the SM and new-physics contributions, respectively ($C^{\mathrm{NP}} = C - C^{\mathrm{SM}}$).

Measurements of $B \to X_s\ell^+\ell^-$

Belle and BaBar experiments have already performed measurements on $B \to X_s\ell^+\ell^-$ [40] [17] [2] [41] [16]. So far, all measurements are highly statistically limited due to the small branching fraction and limited data sample.

The branching fraction has been measured in Belle using 152×10^6 $B\bar{B}$ pairs [17] and in BaBar using 471×10^6 $B\bar{B}$ pairs [16]. The measurements and world average by Heavy Flavor Averaging Group (HFLAV) [1] is summarized in TABLE 2.1. The SM prediction on the branching fraction is shown in TABLE 2.2 [42] [35]. All measurements are consistent with the SM predictions. While measurement of R_{X_s} has not been performed in previous experiments, the measurements of the branching fraction at BaBar correspond to $R_{X_s} = 0.57^{+0.18}_{-0.17}$ assuming the systematic uncertainties are negligible in the R_{X_s} due to taking ratio. The difference from the unity is at 2.4σ level. This is also consistent with the trend of $R_{K^{(*)}}$ anomalies which are discussed in the following section. Belle II will perform measurements using large statistics leading smaller statistical uncertainty.

TABLE 2.1: Measurements of the branching fraction on $B \to X_s\ell^+\ell^-$ decays.

Experiment	$\mathcal{B}(B \to X_s e^+e^-)$ $[10^{-6}]$	$\mathcal{B}(B \to X_s\mu^+\mu^-)$ $[10^{-6}]$	$\mathcal{B}(B \to X_s\ell^+\ell^-)$ $[10^{-6}]$
Belle ($M_{\ell^+\ell^-} > 0.2\,\mathrm{GeV}$)	$4.05 \pm 1.30^{+0.87}_{-0.83}$	$4.13 \pm 1.05^{+0.85}_{-0.81}$	$4.11 \pm 0.83^{+0.85}_{-0.81}$
BaBar ($M^2_{\ell^+\ell^-} > 0.1\,\mathrm{GeV}^2$)	$7.69^{+0.82+0.71}_{-0.77-0.60}$	$4.41^{+1.31+0.63}_{-1.17-0.50}$	$6.73^{+0.70+0.60}_{-0.64-0.56}$
World Average (HFLAV)	6.67 ± 0.82	$4.27^{+0.98}_{-0.91}$	5.84 ± 0.69

TABLE 2.2: The SM prediction of the branching fraction on $B \to X_s\ell^+\ell^-$ decays.

$\mathcal{B}(B \to X_s e^+e^-)$ $[10^{-6}]$	$\mathcal{B}(B \to X_s\mu^+\mu^-)$ $[10^{-6}]$	$\mathcal{B}(B \to X_s\ell^+\ell^-)$ $[10^{-6}]$
6.89 ± 1.01		
4.2 ± 0.7 ($M_{\ell^+\ell^-} > 0.2\,\mathrm{GeV}$)	4.15 ± 0.70	4.18 ± 0.70

The forward-backward asymmetry is firstly measured in Belle using 772×10^6 $B\bar{B}$ pairs [2] as a function of q^2. Figure 2.3 shows the measurement of the forward-backward asymmetry and the SM prediction. The measurement is consistent with the SM prediction.

To shed further light on the anomalies which are observed in exclusive modes, precise study on inclusive $B \to X_s\ell^+\ell^-$ is important. The $B \to X_s\ell^+\ell^-$ decays are reconstructed with a sum-of-exclusive method, in which X_s is reconstructed from a lot of exclusive final states. High reconstruction efficiency for stable particles are essential for multiplicity final states. The B-factory experiment at a electron-positron collider is suitable for the analysis on $B \to X_s\ell^+\ell^-$ thanks to the low background environment. Belle II is a unique experiment to achieve the measurements on $B \to X_s\ell^+\ell^-$ with large statistics. Thanks to the large statistics, angular decomposition measurement, which has not performed due to small statistics, will be performed in Belle II.

Also a fully-inclusive reconstruction method of $B \to X_s\ell^+\ell^-$ decays, which has been used in $B \to X_s\gamma$ decays, is being explored with dedicated simulation studies. Only di-lepton $\ell^+\ell^-$ is reconstructed explicitly and X_s state is obtained as the recoil in this method. Efficient background suppression is key to the success.

2.4 Exclusive $B \to K^{(*)}\ell^+\ell^-$ decays

In the exclusive $B \to K^{(*)}\ell^+\ell^-$ decays, q^2 and θ_ℓ are also important kinematic observables. In addition to them, two helicity angles θ_K and ϕ can be used in $B \to K^*(\to K\pi)\ell^+\ell^-$ decays. The θ_K is the angle between the direction of K [$\bar{K}$] and that of B [$\bar{B}$] in the rest frame of K^* [$\bar{K}^*$] and ϕ is the angle between the plane defined by the di-lepton pair and the plane defined by the $K\pi$ in the B-meson rest frame.
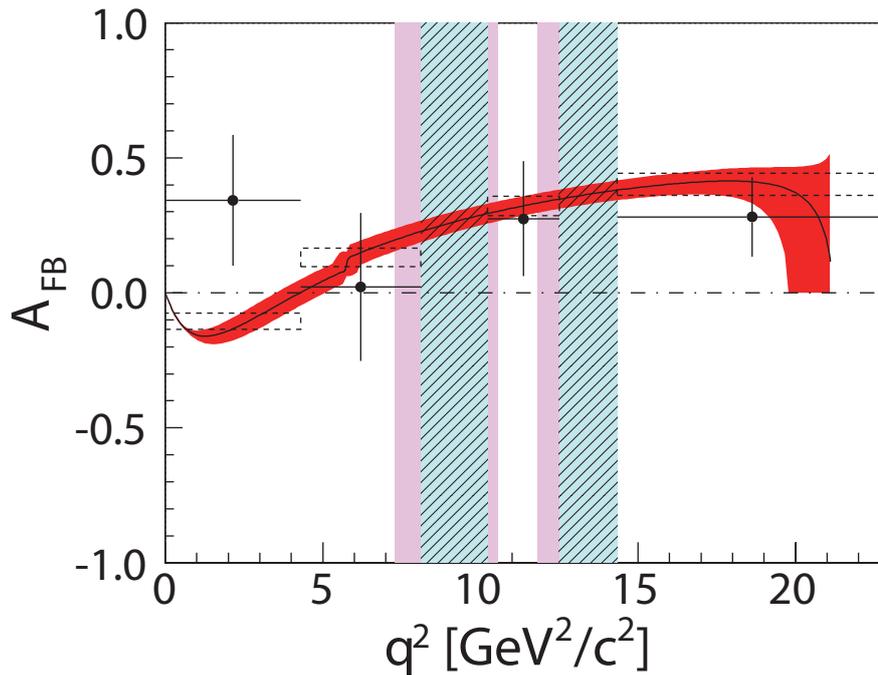
FIG 2.3: Measurement of the forward-backward asymmetry in Belle [2]. Black dots with error bars represent the measurement and curve (black) with the band (red) and dashed boxes (black) shows the SM prediction. The backgrounds form $J/\psi(\to \ell^+\ell^-)$ and $\psi(2S)(\to \ell^+\ell^-)$ events have been vetoed by rejected events in the teal hatched regions. For the electron channel, the pink shaded regions are added to the veto regions due to the large bremsstrahlung effect.

Following the definition given in [43], the CP-averaged angular distribution can be written

$$
\frac{1}{d(\Gamma + \overline{\Gamma}/dq^2)} \frac{d^4(\Gamma + \overline{\Gamma})}{dq^2 d\cos\theta_\ell d\cos\theta_K d\phi} = \frac{9}{32\pi} \left[\frac{3}{4}(1 - F_L)\sin^2\theta_K + F_L \cos^2\theta_K \right.
$$
$$
+ \frac{1}{4}(1 - F_L)\sin^2\theta_K \cos 2\theta_\ell
$$
$$
- F_L \cos^2\theta_K \cos 2\theta_\ell + S_3 \sin^2\theta_K \sin^2\theta_\ell \cos 2\phi
$$
$$
+ S_4 \sin 2\theta_K \sin 2\theta_\ell \cos\phi + S_5 \sin 2\theta_K \sin\theta_\ell \cos\phi
$$
$$
+ \frac{4}{3} A_{FB} \sin^2\theta_K \cos\theta_\ell + S_7 \sin^2\theta_K \sin\theta_\ell \sin\phi
$$
$$
\left. + S_8 \sin 2\theta_K \sin 2\theta_\ell \sin\phi + S_9 \sin^2\theta_K \sin^2\theta_\ell \sin 2\phi \right]. \quad (2.24)
$$

where F_L is the fraction of the longitudinal polarication of the K^* meson and S_i are CP-averaged observables [43]. To cancel the $B \to K^*$ form-factor uncertainties, other notation of observables are also defined.

$$
P'_{4,5,6,8} = \frac{S_{4,5,6,8}}{\sqrt{F_L(1 - F_L)}}. \quad (2.25)
$$

Each observable has independent dependence to the Wilson coefficient and can be compared with the SM prediction. The exclusive $B \to K^*\ell^+\ell^-$ provides rich information to set constraints on the Wilson coefficients from these observables.

In analogy to R_{X_s}, the lepton flavor universality observables $R_{K^{(*)}}$ have additional information to the Wilson coefficient.

$$
R_{K^{(*)}}[q_0^2, q_1^2] = \frac{\int_{q_0^2}^{q_1^2} dq^2 \frac{d\Gamma(B\to K^{(*)}\mu^+\mu^-)}{dq^2}}{\int_{q_0^2}^{q_1^2} dq^2 \frac{d\Gamma(B\to K^{(*)}e^+e^-)}{dq^2}}. \quad (2.26)
$$

Even though they are also estimated to be equal to unity in the SM, they have other chirality dependency due to the polarization. In lowest order, they can be written using $\Delta_\pm$ 2.23

$$R_K \simeq 1 + \Delta_+, \tag{2.27}$$
$$R_{K^*} \simeq 1 + \Delta_+ - p(\Delta_+ - \Delta_-). \tag{2.28}$$

where $p \simeq 0.86$ is the so-called polarization fraction of the K^*. To solve the chirality structures of lepton flavor universality violating new-physics, the correlation of R_K, R_{K^*} and R_{X_s} is important. Furthermore, the double ratios of these variables provide theoretically clean information [39].

Measurements of $B \to K^{(*)} \ell^+ \ell^-$

Exclusive $B \to K^{(*)} \ell^+ \ell^-$ decays are reconstructed from long-lived particles which can be directly reconstructed and intermediate particles such as K^*, K_S^0, π^0 which are formed from the stable particles. The B-factory experiments, Belle and BaBar, have advantages on the reconstruction efficiency on electron modes thanks to the high resolution electromagnetic calorimeter. Also the neutral particle such as γ, π^0 and K_S^0 can be detected efficiently. Experiments with hadron collier, such as LHCb, CMS, ATLAS, and CDF have large statistics of B meson and high muon reconstruction efficiency, and thus an advantage on the muon modes.

The branching fraction of $B \to K^{(*)} \ell^+ \ell^-$ has been measured in BaBar [44] [45] and Belle [46]. The CDF experiment has also performed measurement on $B \to K^{(*)} \mu^+ \mu^-$ [47]. The world average by HFLAV [1] on the branching fraction of $B \to K^{(*)} \ell^+ \ell^-$ is shown in TABLE 2.3. TABLE 2.4 shows the SM prediction on the branching fraction [42] [35]. The uncertainties on the predictions are large compared with those of $B \to X_s \ell^+ \ell^-$ shown in TABLE 2.2 due to the irreducible form factor uncertainty.

TABLE 2.3: Measurements of the branching fraction on $B \to K^{(*)} \ell^+ \ell^-$ decays averaged by HFLAV [1].

Mode	$e^+ e^-$ $[10^{-6}]$	$\mu^+ \mu^-$ $[10^{-6}]$	$\ell^+ \ell^-$ $[10^{-6}]$
$B \to K \ell^+ \ell^-$	0.44 ± 0.06	0.44 ± 0.04	0.48 ± 0.04
$B \to K^* \ell^+ \ell^-$	$1.19^{+0.17}_{-0.16}$	1.06 ± 0.09	1.05 ± 0.10

TABLE 2.4: The SM prediction of the branching fraction on $B \to K^{(*)} \ell^+ \ell^-$ decays.

Mode	$e^+ e^-$ $[10^{-6}]$	$\mu^+ \mu^-$ $[10^{-6}]$
$B \to K \ell^+ \ell^-$	0.35 ± 0.12	0.35 ± 0.12
$B \to K^* \ell^+ \ell^-$	1.58 ± 0.49	1.19 ± 0.39

In the angular measurements and the lepton flavor universality tests, anomalies have been observed on the decays by many experiments. As shown in FIG 2.4, LHCb experiment has reported a discrepancy in the P_5' variables with $B^0 \to K^{*0} \mu^+ \mu^-$ decay at 2.5σ and 2.9σ level in the $4.0 < q^2 < 6.0\,\text{GeV}^2$ and $6.0 < q^2 < 8.0\,\text{GeV}^2$ regions, respectively [3]. Belle experiment has performed the angular analysis using $B \to K^* e^+ e^-$ and $B \to K^* \mu^+ \mu^-$ separately [4]. In the observable P_5', a tension by 2.6σ in $q^2 \in [4, 8]\,\text{GeV}^2$ is observed in the muon mode. In the same region, the electron mode deviates by 1.3σ. Measurement of the lepton-flavor-violating observables $Q_5 = P_5'^{\mu} - P_5'^{e}$ has been performed in Belle for the first time. Figure 2.5 shows the measurement Q_5 which is compared with the SM prediction and a new-physics model given in [5].

LHCb measurements of $R_K = 0.846^{+0.060+0.016}_{-0.054-0.014}$ in $q^2 \in [1.1, 6.0]\,\text{GeV}^2$ [26] and $R_{K^*} = 0.69^{+0.11}_{-0.07} \pm 0.05$ in $q^2 \in [1.1, 6.0]\,\text{GeV}^2$ [27] deviates by 2.5σ and 2.5σ from the unity. Measurement on BaBar [45] and Belle [46] are compatible with the SM prediction and the LHCb results. Figure 2.6 summarizes measurements on R_K and R_{K^*}.
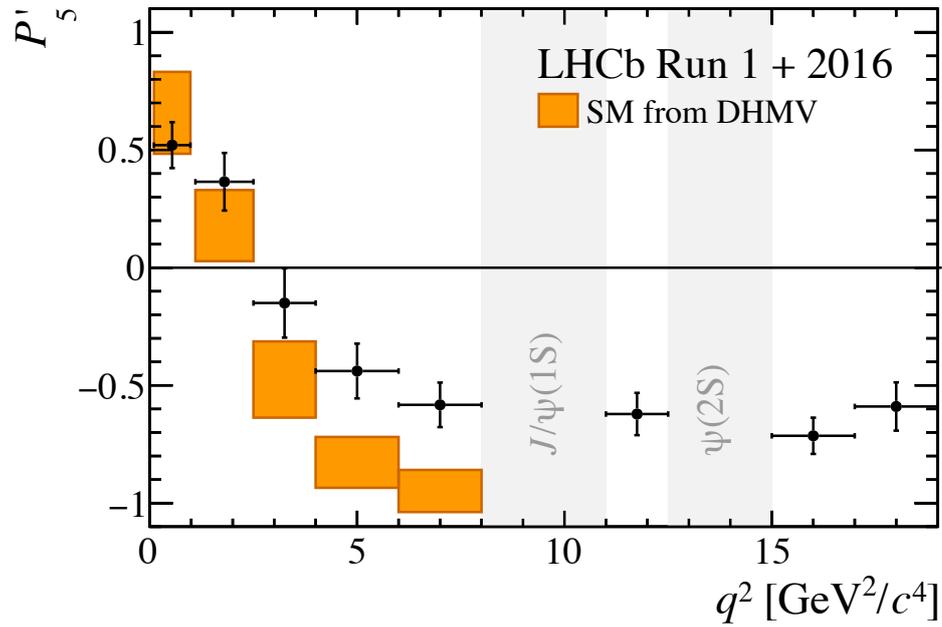
FIG 2.4: Measurement of the CP-averaged observable P_5' in LHCb [3]. Black dots with error bars represent the measurement and orange boxes show the SM prediction.
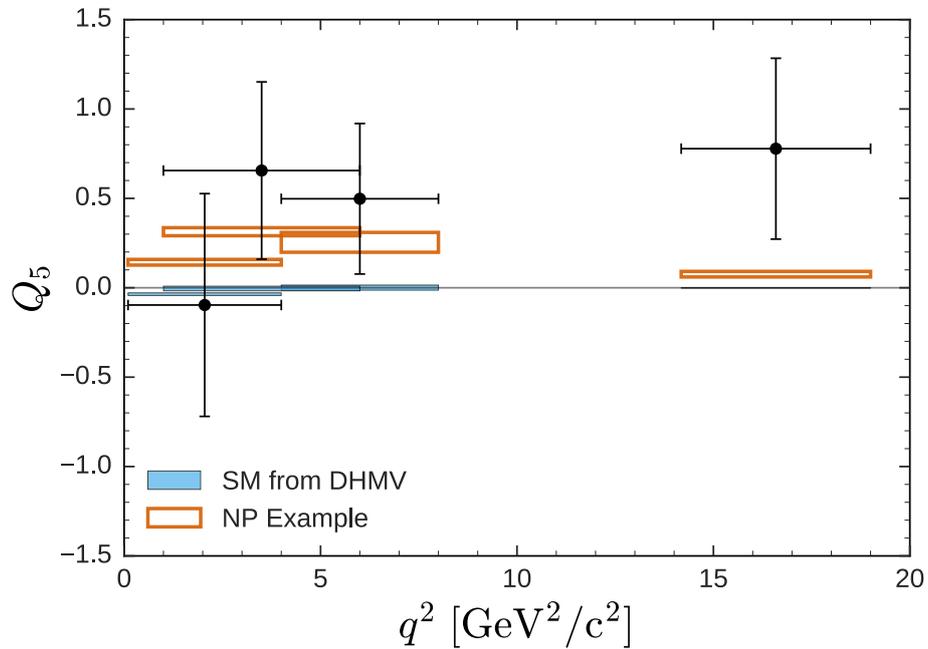
FIG 2.5: Measurement of the lepton-flavor-violating observable Q_5 in Belle [4]. Black dots with error bars represent the measurement. Blue boxes show the SM prediction and orange boxed show a prediction in a new-physics model [5]
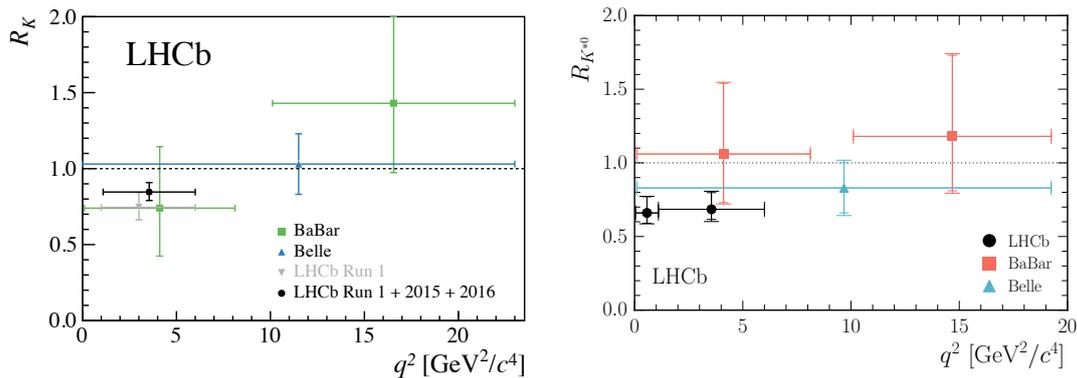
FIG 2.6: Measurement of the lepton-flavor-universality test observables R_K (Left) and R_{K^*} (Right).

2.5   Constraints on the Wilson coefficients

A model-independent global fit based on the Effective Hamiltonian framework is powerful way to search for new physics. In addition to the measurements which are discussed in previous sections, other measurements on $B_s \rightarrow \mu^+\mu^-$ [48] [49] [50] and $B_s \rightarrow \phi\mu^+\mu^-$ [51] are combined to set constraints on the Wilson coefficients. Constraints on the Wilson coefficients are investigated by several groups [6] [52]. Figure 2.7 shows contours in 2D planes of the Wilson coefficients given by [6]. The difference between the SM prediction and best fit values is at more than 5σ level for these three cases. Smaller C_9 than the SM prediction only in the muon mode is favored from the fits, which may imply the lepton flavor universality violation.
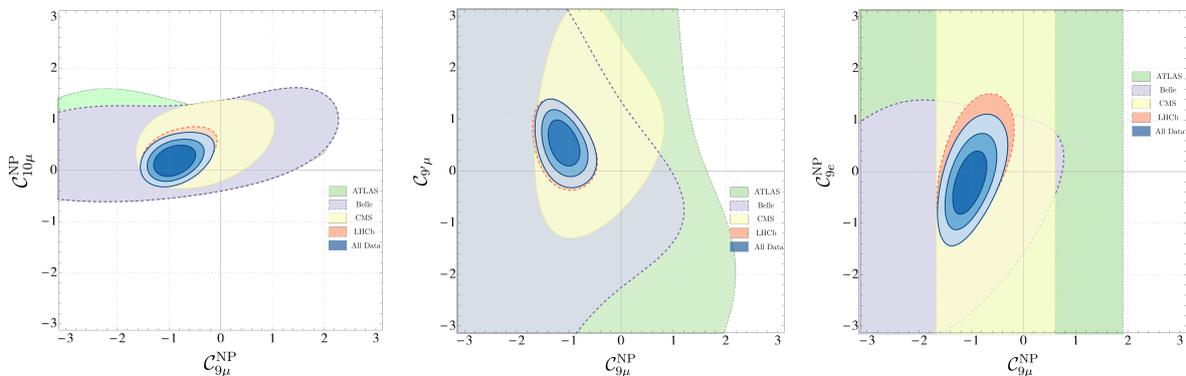
FIG 2.7: Constraints on the Wilson coefficients in $(C_9^{\mathrm{NP}\mu}, C_{10}^{\mathrm{NP}\mu})$, $(C_9^{\mathrm{NP}\mu}, C_9^{'\mu})$, and $(C_9^{\mathrm{NP}\mu}, C_9^{\mathrm{NP}e})$ from left to right [6]. NP denotes new-physics contribution, $C^{\mathrm{NP}} = C - C^{\mathrm{SM}}$. The origin is the SM prediction. The contours of each experiment correspond to the 3σ constraint and blue contours of All Data shows $1, 2, 3\sigma$ constraints.

2.6   Interplay of inclusive and exclusive $b \rightarrow s\ell^+\ell^-$

An impact from the Belle II measurements on inclusive $B \rightarrow X_s\ell^+\ell^-$ decays with an integrated luminosity of $50\,\mathrm{ab}^{-1}$ to the anomalies which have been observed on exclusive $B \rightarrow K^{(*)}\ell^+\ell^-$ is discussed in this section. The sensitivities for inclusive $B \rightarrow X_s\ell^+\ell^-$ observables are estimated in [7]. From the branching fraction and forward-backward asymmetry, the constraints on the Wilson coefficients C_9 and C_{10} are derived. Contours shown in FIG 2.8 are the Belle II sensitivities for each significance. For example, $C_9^{\mathrm{NP}}(\equiv C_9 - C_9^{\mathrm{NP}}) = -1$ can be excluded at 5σ level if the true values are at the SM prediction.

Figure 2.8 also shows red contours which are global fit results mainly from analyses on exclusive decays [8]. If the true values are at the global fit result, the inclusive analyses will exclude the SM with 6σ level. Moreover, the hadronic uncertainty of inclusive transition is independent with that of exclusive decay. Thus, a cross-check of analyses on exclusive decays can be performed. Inclusive $B \to X_s \ell^+ \ell^-$ measurements at Belle II play a critical role to search for new physics.
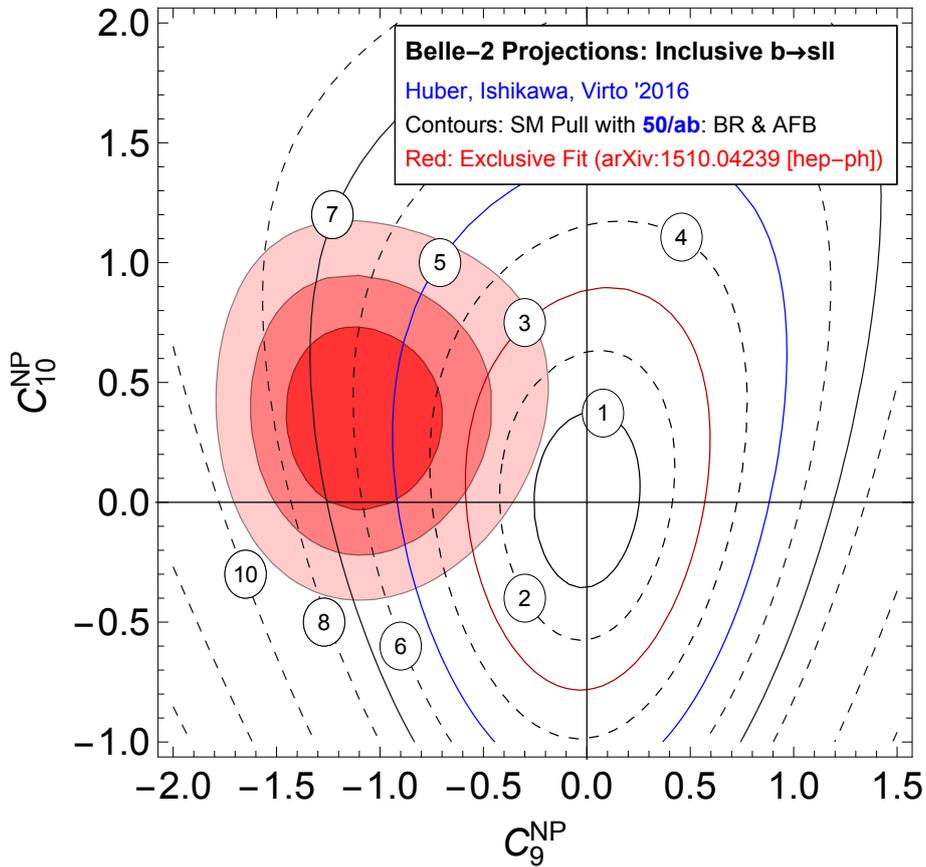
FIG 2.8: Expected sensitivity on the Wilson coefficients in $(C_9^{\mathrm{NP}}, C_{10}^{\mathrm{NP}})$ from inclusive $B \to X_s \ell^+ \ell^-$ measurements at Belle II [7]. NP denotes new-physics contribution, $C^{\mathrm{NP}} = C - C^{\mathrm{SM}}$. Black, red and blue lines show 1σ, 3σ, and 5σ level constraints, respectively. Red contours show the current global fits results at level of $1, 2, 3\sigma$ [8].

Chapter 3

Belle II experiment

The Belle II experiment at the SuperKEKB accelerator is a B factory facility at KEK, Tsukuba Japan. The SuperKEKB is designed to collide electron and positron at the centre-of-mass energy of Upsilon resonances. A main physics goal is to search for new physics beyond the standard model with high precision measurements of B meson decay. Most of the data is taken at the Upsilon 4S resonance which decays to B meson pair without fragmentation particles.

The design luminosity of the SuperKEKB is $8 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$, about 40 times larger than KEKB. The Belle II experiment aims to collect data corresponding to integrated luminosity of $50\,\mathrm{ab^{-1}}$, a factor of 50 more than the Belle experiments which is its predecessor. The Belle II detector is upgraded from the Belle detector to take data with a 40 times higher event rates.

3.1 SuperKEKB accelerator

The SuperKEKB accelerator is a two-ring energy-asymmetric electron-positron collider. Figure 3.1 shows a schematic view of the SuperKEKB accelerator. Electrons and positrons are accelerated by a linear accelerator (Linac) up to 7.007 GeV and 4.000 GeV, respectively. The ring to storage electrons is called the high energy ring (HER) and the other ring to storage positrons is called the low energy ring (LER). The two rings are crossed at the interaction point (IP) where the Belle II detector is placed. The centre-of-mass energy is $\sqrt{s} = \sqrt{4E_{e^-}E_{e^+}} = 10.58\,\mathrm{GeV}$.

The beam size at the IP is reduced by a factor of 20, from 1 μm to 50 nm, from the KEKB design parameter. This is known as a nano-beam scheme which is invented for the Italian super B factory project. The beam current increases by a factor of 2 as well. Combining these two, the target luminosity of the SuperKEKB which is 40 times larger than that of the KEKB will be achieved. On the other hand, the beam energy asymmetry is reduced from 8 GeV (electrons) and 3.5 GeV (positrons) to 7 GeV and 4 GeV to avoid the beam losses due to the Touschek scattering [53]. This leads to lower magnitude of the boost which is essential to study time dependent CP asymmetry by measuring the spatial separation of B meson decays.

3.2 Belle II detector

The Belle II detector (FIG 3.2) [9] is a complex 4π detector which consists of following sub-detectors.

- Vertex detector (VXD)

- Central Drift Chamber (CDC)

- Time of Propagation (TOP) counter

- Aerogel Ring Imaging Cherenkov (ARICH)

- Electromagnetic Calorimeter (ECL)

- K_L^0-Muon detector (KLM)
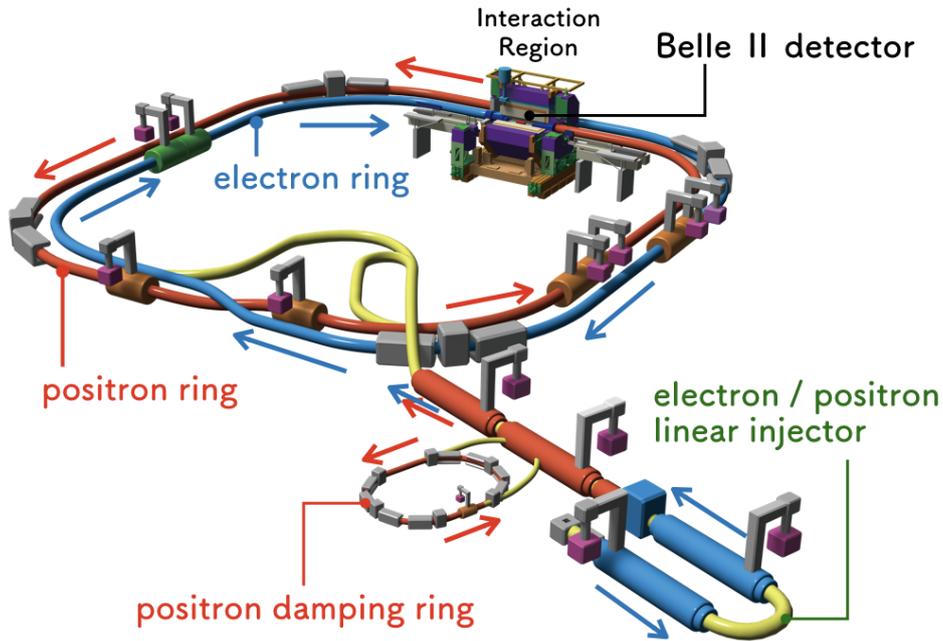
FIG 3.1: Schematic view of the SuperKEKB accelerator.

The detector is designed forward-backward asymmetrically so that boosted particles due to the asymmetric beam energy can be reconstructed effectively. The superconducting solenoid magnet is placed between ECL and KLM and creates a uniform magnetic field of 1.5T.

B meson and other heavy particles immediately decay to stable particles. To study properties of B meson, the Belle II detector is required to measure energy and momentum of these particles, detect the vertex position, and identify the type of particle. The Belle II detector is upgraded to perform in one order higher background condition than the predecessor. The data acquisition (DAQ) systems are also modified to take data with 40 times larger event rates.

3.2.1   VXD

The Vertex detector (VXD) is placed at the innermost of the Belle II detector. The VXD is a six layers system to measure the decay vertices. It consists of two inner layers of the pixel detector (PXD) and four outer layers of the silicon vertex detector (SVD).

The PXD is composed of two layers of pixelated sensors with DEPFET (DEPleted Field Effect Transistor) which allows to make detector thin (50 μm). This helps to reduce the multiple scattering thanks to low material budget. The radii of two layers are 14 mm and 22 mm, which are smaller than the Belle vertex detector which was made of silicon strip. The schematic view of PXD is shown in FIG 3.3. While the vertex resolution is expected to be improved, the background rate increases considerably. The pixel detector is newly introduced to sustain higher hit rate.

The SVD is composed of four layers of double-sided silicon strip detector. The radii of four layers are 38 mm, 80 mm, 115 mm, and 140 mm. The longitudinal section of SVD is shown in FIG 3.4. In comparison, the outermost vertex detector of the Belle detector is placed at a radius of 88 mm. With the large vertex detector, the significant improvement in the reconstruction efficiency of $K_s^0 \rightarrow \pi^+\pi^-$ is expected.

3.2.2   CDC

The central drift chamber (CDC) is the main tracking device of the Belle II detector. The radius of the CDC is extended compared to the Belle detector from 880 mm to 1130 mm because of thinner barrel

FIG 3.2: Top view of the Belle II detector.

FIG 3.3: CAD rendering of PXD.
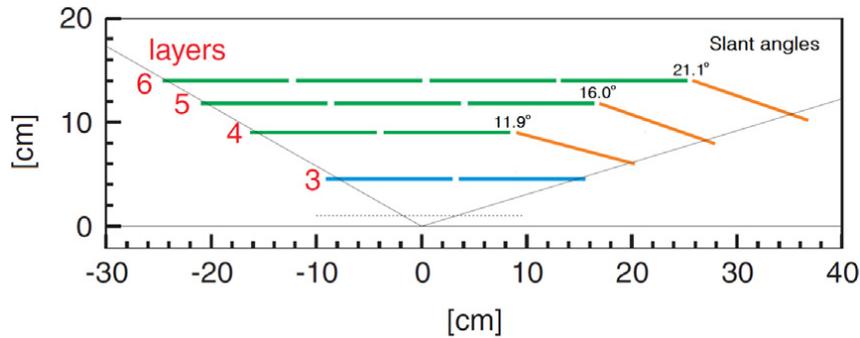
FIG 3.4: Schematic view of SVD longitudinal section. [9]

particle identification detector (TOP). Figure 3.5 shows the comparison of the wire configuration between Belle and Belle II. The CDC is comprised of 14336 sense wires arranged in 56 layers. There are two kinds of layers: the axial layer which is composed of wires aligned with the solenoidal magnetic field and the stereo layer in which wires are skewed with respect to the axial layer. As a result of two kinds of layers, it is possible to reconstruct a 3D helix track. The momentum of a track is measured from the helix parameter. The 50 % helium-50 % ethane gas mixture is filled.
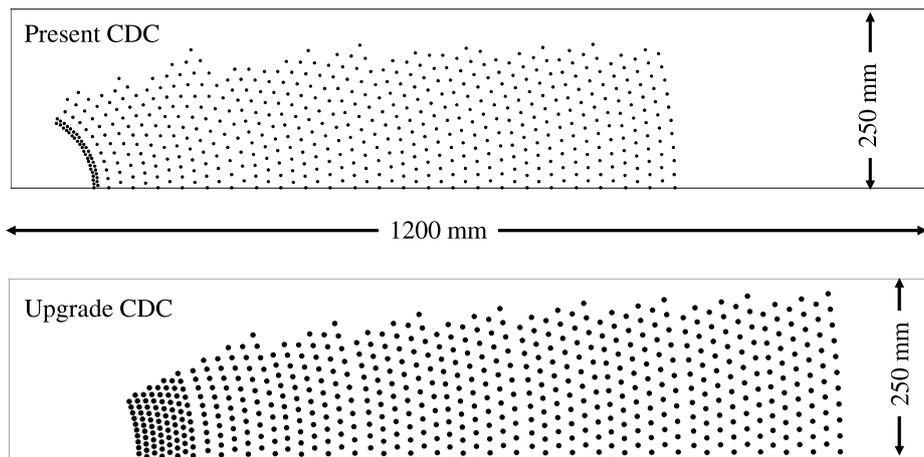
FIG 3.5: Wire configuration of Belle CDC (upper) and Belle II CDC (lower). [9]

The CDC contributes the particle identification by measuring the characteristic energy loss of charged particles. Figure 3.6 shows the energy loss dE/dx at the CDC as a function of track momentum. Two dimensional distributions are measured in data and the black solid line shows the predictions with the simulation. As shown in FIG 3.6, the CDC provides the discrimination information especially for the low-momentum particles. The CDC also provides the reliable information to the trigger system.

3.2.3 TOP

The time of propagation (TOP) counter is used for the particle identification in the barrel region. The TOP is comprised of 16 modules surrounding the CDC. Each module is made of 2.5 m long quartz bar, a prism, a focusing mirror, and the photon detector MCP-PMT. Figure 3.7 shows the schematic view of a TOP module. The TOP measure the Cherenkov photons to distinguish particle types. The Cherenkov angle, θ_C, can be expressed by

$$\cos \theta_c = \frac{1}{n\beta}. \tag{3.1}$$
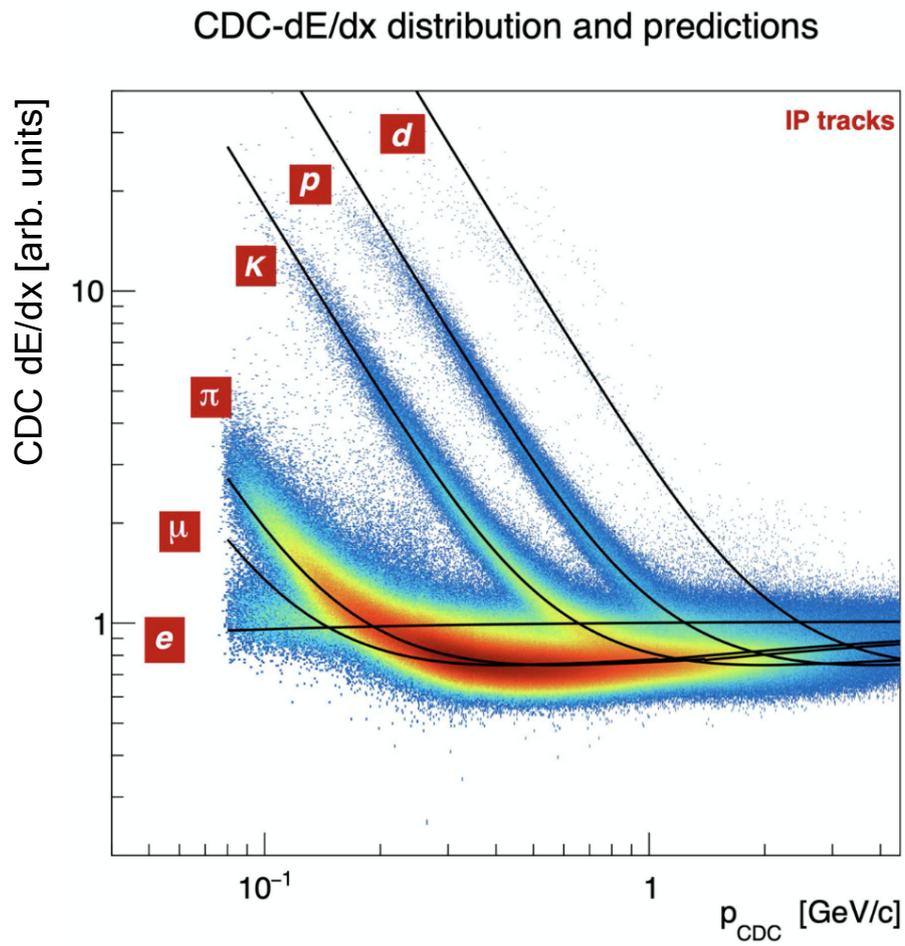
FIG 3.6: Energy loss dE/dx at CDC as a function of track momentum. Two dimensional distributions are measured in data and the black solid line shows the predictions with the simulation

where n is a refractive index and β is the velocity of particle ($\beta = p/E$). Combining momentum and arrival position measured by the CDC, the mass of the particles can be determined.
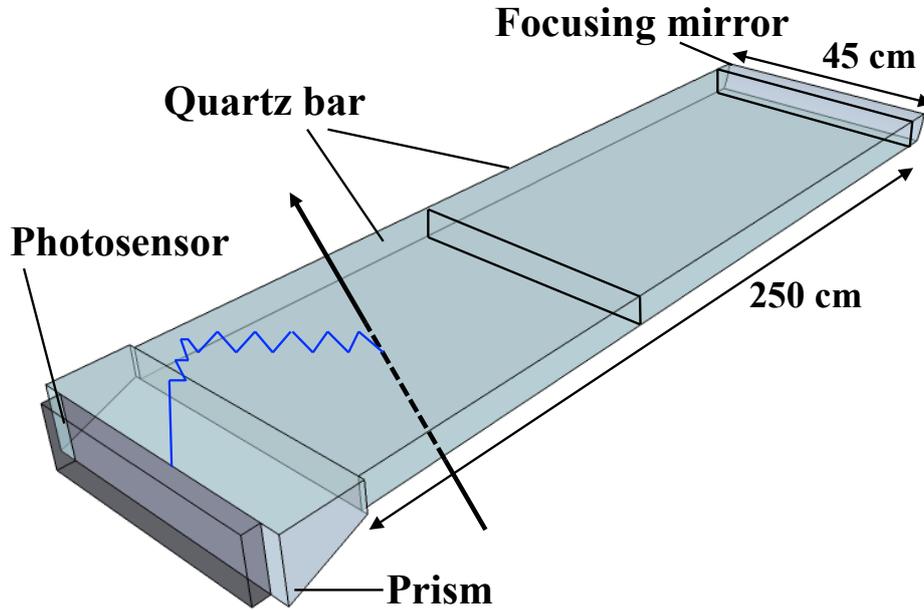
FIG 3.7: Schematic view of a TOP module. [10]

The impact position and propagation time of Cherenkov photons which are reflected in the quartz bar are measured at the MCP-PMT. Relation between impact position and propagation time is depend on the arrival position and the Cherenkov angle θ_C. Figure 3.8 shows the relation between the hit time and the position of MCP-PMT. Kaons were tagged via the decay chain $D^{*+} \to D^0\pi^+_{slow}, D^0 \to K^-\pi^+$ where the charge of π_{slow} determines which D^0 daughter is the kaon. The probability density functions (PDFs) for the pion and kaon hypothesis are graphed in left side and right side, respectively. As shown in FIG 3.8, the kaon hypothesis fits much better than that of pion.

3.2.4   ARICH

In the forward endcap region, the particle identification is provided by the proximity focusing aerogel ring imaging Cherenkov (ARICH) counter. It consists of two layers of aerogel as radiator and hybrid avalanche photon detector (HAPD). An expansion volume of 20 cm between the aerogel tile and the HAPD enlarges the Cherenkov photons rings. The thicker aerogel radiator provides the larger number of detected photons, but the resolution of Cherenkov ring becomes worse due to the uncertainty of the emission point. To prevent the degradation, two 2 cm layers of aerogel with different refractive indexes (n=1.045 upstream and n = 1.055 downstream) are used. The produced rings are overlapped on the detection surface. Figure 3.9 shows the schematic principle of ARICH detector.

The Cherenkov angle with a refractive index of 1.5 is shown in FIG 3.10 as a function of track momentum. The difference of the Cherenkov angle between pion and kaon at $p = 4\,\text{GeV}/c$ is 23 mrad ($\theta_C = 308, 285\,\text{mrad}$ for pion and kaon). Figure 3.11 shows the distribution of the reconstructed Cherenkov angles for $e^+e^- \to \mu^+\mu^-$ sample with muon momentum selection $6.4\,\text{GeV}/c < p < 7.0\,\text{GeV}/c$, and for the $K_S^0 \to \pi^+\pi^-$ sample with pion momentum selection $1.0\,\text{GeV}/c < p < 1.1\,\text{GeV}/c$. The position of the large peak is consistent with the expectation shown in FIG 3.10. The peak at $\theta_C \simeq 0.02$ is made by the Cherenkov photons emitted at the HAPD glass. The width of the large peak is evaluated by using these control events. The evaluated width at $p = 4\,\text{GeV}/c$ is about 13 mrad which enable to distinguish between pion and kaon.
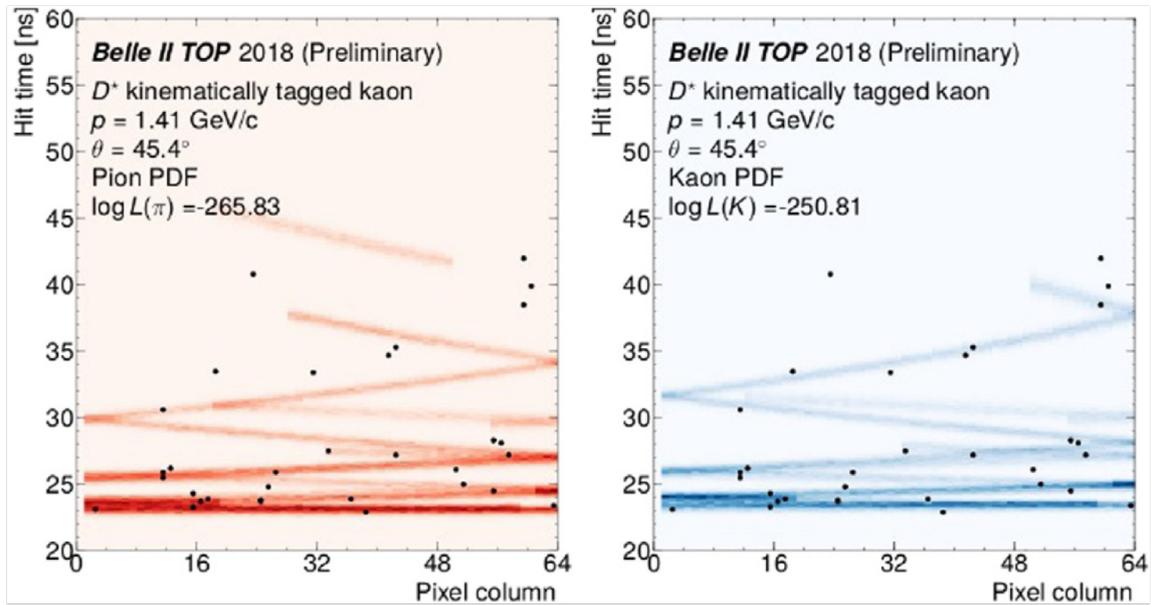
FIG 3.8: Relation between the hit time and the position of MCP-PMT. Black points shows the data of a kaon. The left side is the pion PDF and the right side is the kaon PDF. [11]
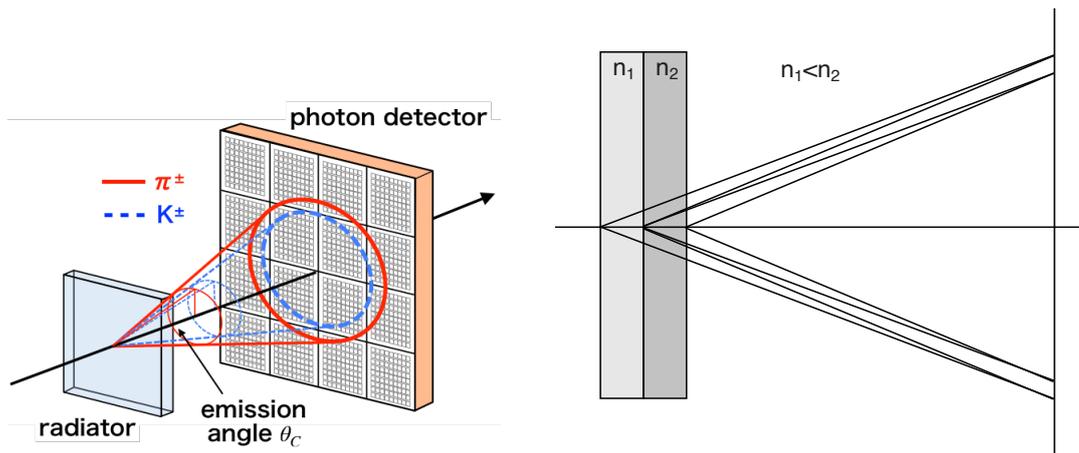
FIG 3.9: Left : The principle of the particle identification of the ARICH. Right : Principle of operation of the proximity focusing with non-homogeneous aerogel radiator.
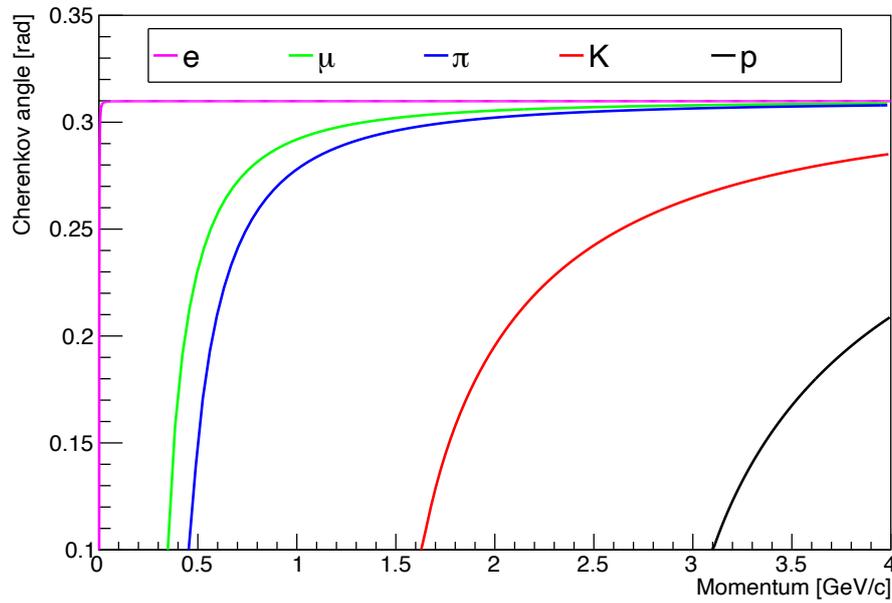
FIG 3.10: The Cherenkov angle as a function of track momentum with refractive index $n = 1.5$. [12]
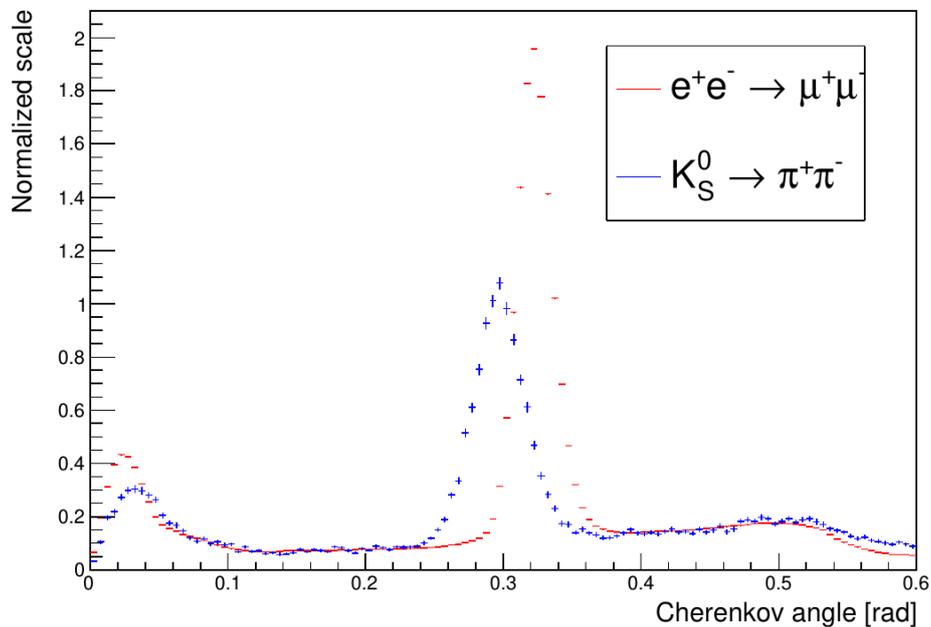
FIG 3.11: Distribution of the Cherenkov angle for $e^+e^- \rightarrow \mu^+\mu^-$ sample with muon momentum selection $6.4\,\mathrm{GeV}/c < p < 7.0\,\mathrm{GeV}/c$ in red, and for the $K_S^0 \rightarrow \pi^+\pi^-$ sample with pion momentum selection $1.0\,\mathrm{GeV}/c < p < 1.1\,\mathrm{GeV}/c$ in blue. [13]

3.2.5   ECL

The electromagnetic calorimeter (ECL) is used to measure the energy and angle of photons. It also generates a photon trigger. The Belle II ECL exploits the same crystals of Belle calorimeter with an upgrade of the readout electronics to cope with the SuperKEKB luminosity. The ECL is composed of 8736 CsI(Tl) crystals assembled in a projective geometry covering about 90 % of the solid angle in the centre-of-mass-system. The crystals have a 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. At the external bases, two photodiodes are glued with a 1 mm plexiglass plate to detect the scintillation light.

In the Belle experiment, the energy resolution was $\sigma/E = 4\%$ at 100 MeV, 1.6% at 8 GeV, and the angler resolution was 13 mrad (3 mrad) at low (high) energies; π^0 mass resolution was 4.5 MeV. In Belle II, a very similar performance is expected in absence of backgrounds. However, due to the relatively long decay time of scintillation in CSI(Tl) (1 μs), the overlapping of pulses are considerably increased in the SuperKEKB luminosity. The scintillator photo-sensors are equipped with wave-form-sampling read-out electronics to mitigate the pile-up noise.

The ECL also contribute to identify electron. Electrons loose most of own energy in the ECL due to the electromagnetic shower. Thus, the ratio between the energy deposited at the ECL and track momentum, E/p, tends to be unity. Figure 3.12 shows the E/p distributions for a variety of momentum ranges. As shown in FIG 3.12, the E/p is an good discriminator for electron in high momentum region. The PDF for each particle hypothesis is constructed from the E/p distributions.
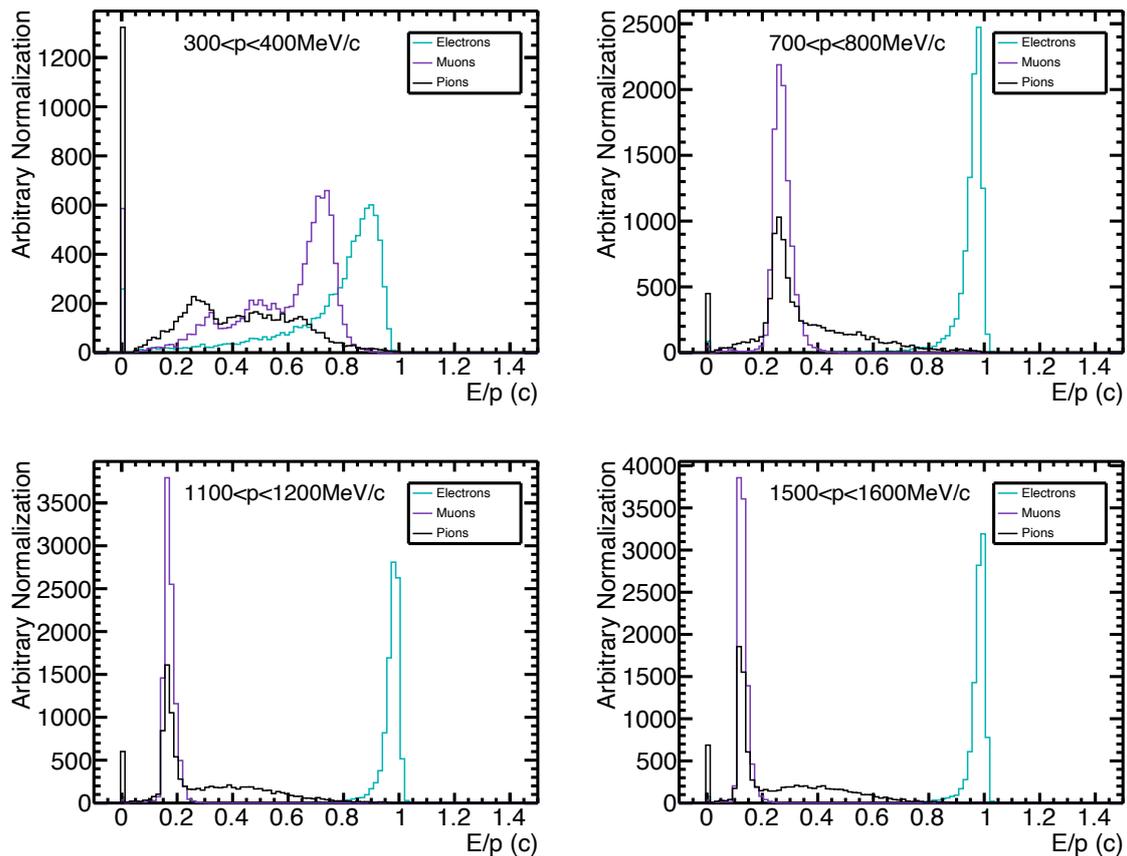
FIG 3.12: The E/p distributions for a variety of momentum ranges. Turquoise line is for electron, purple line is for muon and black line is for pion. The peak at 0 means that a track does not much any ECL cluster. [7]

3.2.6   KLM

The K_L^0 and muon detector (KLM) consists of alternating 4.7 cm thick iron plates and active material detectors located outside the superconducting solenoid. In the barrel region there are 15 layers, while in the forward endcap 14 layers are placed and in the backward endcap there are 12 layers. The iron plates serve as the magnetic flux return and provides 3.9 interaction lengths (λ) in total, in addition to the 0.8 λ of the ECL, in which KL can shower hadronically. Most of barrel layers is equipped with glass electrode Resistive Plate Chambers (RPC). Because RPCs have the long dead time during the recovery of the electric field after a discharge, the inner two barrel layers and the endcaps employ layers of scintillator strips with silicon photomultiplier (SiPM).

   The muon identification mostly relied on the KLM, since the mass of muon and that of pion is very close ($m_\mu = 105.7\,\mathrm{MeV}/c^2, m_{\pi\pm} = 139.6\,\mathrm{MeV}/c^2$). Muons often penetrate the KLM detector, while most of pion captured within the KLM. A few percent of pions possibly penetrate the KLM without the inelastic hadronic scattering ($e^{-3.9} = 0.02$) and might be identified as muons. The identification of muon is performed by extrapolating tracks reconstructed by CDC and comparing with the actual measured hit in the KLM.

3.2.7   Trigger

The trigger system of the Belle II experiment plays an important role to identity and record physics events. TABLE 3.1 shows the cross section and rate of various physics events at the target luminosity of the SuperKEKB. Bhabha scattering and $e^+e^- \to \gamma\gamma$ events which are useful for calibration of detectors and luminosity measurement have large cross-section and event rate. These events will be reduced by a factor of 100 with the pre-scale. The total expected event rate is about 15 kHz, while the limitation of the data acquisition system (DAQ) is 30 kHz. Therefore, the trigger system is required to suppress beam-induced backgrounds up to same level of physics events.

TABLE 3.1: Production cross-section and event rate.

Process	Cross section [nb]	Event rate [Hz]	Pre-scaled event rate [Hz]
$B\bar{B}$	1.2	960	960
Continuum $q\bar{q}$	2.8	2200	2200
$e^+e^- \to \mu^+\mu^-$	0.8	640	640
$e^+e^- \to \tau^+\tau^-$	0.8	640	640
Bhabha	44	35000	350
$e^+e^- \to \gamma\gamma$	2.4	1900	19
Two-photon (e.g. $e^+e^- \to e^+e^-e^+e^-$)	12	~10000	~10000
Total	67	~50000	~15000

   The Belle II trigger is comprised of the hardware based Level 1 (L1) trigger and the software based High Level Trigger (HLT).

L1 trigger

The L1 trigger combines and analyzes signals from sub-detectors. The trigger system is implemented with the Field Programmable Gate Array (FPGA) which provides configurable logics. The track trigger based on the CDC hits and the calorimeter trigger based the ECL clusters are the main information of the L1 trigger. These signals are merged with TOP and KLM hit information by the global reconstruction logic (GRL). Then, the output of GRL is sent to the global decision logic (GDL), which makes decision whether an event should be taken. The L1 trigger output rate is required to satisfy the DAQ read-in rate, 30 kHz. Figure 3.13 shows the schematic view of the Belle II L1 trigger system.

   The track trigger processes 2D and 3D tracking algorithms. At the Belle experiment, only the 2D track information was reconstructed. The 3D tracking gives the vertex position along the beam direction. This is used to eliminate the beam-induced backgrounds in which tracks do not originate from the interaction point.
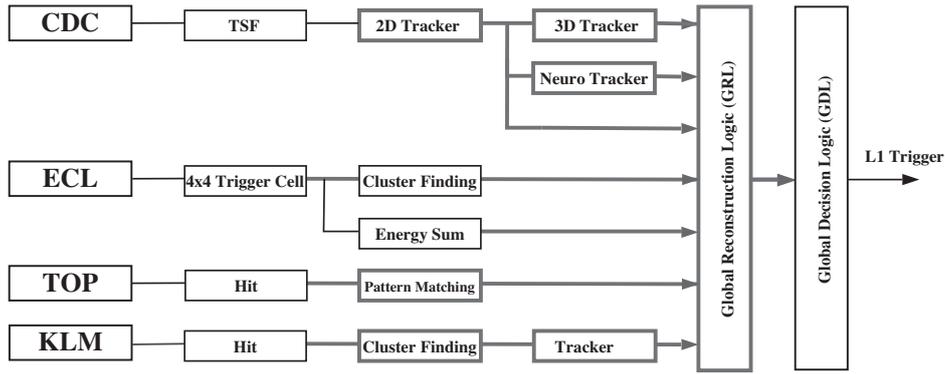
FIG 3.13: Schematic view of L1 trigger system. CDC find a track using small segment information, which is called Track Segment Finder (TSF), and reconstruct 2D track. In addition, 3D track information which is constructed from 2D track and beam along information. Neural network based track finder is going to be implemented. ECL provides both each cluster information and total energy deposit in ECL to GRL. TOP and KLM information helps to identify particle type at the trigger level.

The calorimeter trigger provides the total energy information and the isolated counting information. They characterize the Bhabha scattering and $e^+e^- \to \gamma\gamma$ which have large energy deposit in the ECL and the back-to-back event topology to apply the pre-scale.

The GRL combines the sub-detector information. For example, a track found in CDC and a cluster of ECL are matched. Combining a track of CDC and KLM hits can be used to identify muon events.

The hadronic events such as $B\bar{B}$ and continuum have high multiplicity of charged tracks. They are trigger by requiring that there are at least three tracks in CDC. For the low-multiplicity events, dedicated trigger menus are developed. One photon trigger, for example, is equipped for the dark matter search with the initial state radiation photon.

HLT

The HLT suppresses event rates from 30 kHz which is L1 trigger rate to 10 kHz for offline storage. The HLT uses full detector information except for the PXD and reconstructs events. The events are categorized into each physics process. Currently, all events are kept in the HLT since the total event rates are acceptable. Sophisticated software for the HLT is being developed for the higher luminosity.

3.3   Status of the Belle II experiment

The commissioning of SuperKEKB accelerator is proceeded in three phases.

- Phase 1 : A background study is performed without collisions. The dedicated detector system is installed at the IP instead of the Belle II detector.

- Phase 2 : Main purpose of the Phase 2 is the calibration of accelerator and the Belle II detector response study. The VXD detector which is the most affected by the beam background is partially installed. The first collision at Phase 2 is achieved in April 26th 2018 and the collected data set by July 17th is 472 pb^{-1}. The dark-sector analysis using these data is performed and first Belle II papers are published.

- Phase 3 : Phase 3 is the data taking phase with the full Belle II detector system. Currently only inner layers and 2 outer ladders of PXD are installed due to its production delay. The data taking resumed March 2019. On 15th June 2019, SuperKEKB achieved the world's highest luminosity of $2.22 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$. The accumulated data with the Belle II detector by the summer shutdown 2020 is 74.10 fb^{-1} (FIG 3.14).
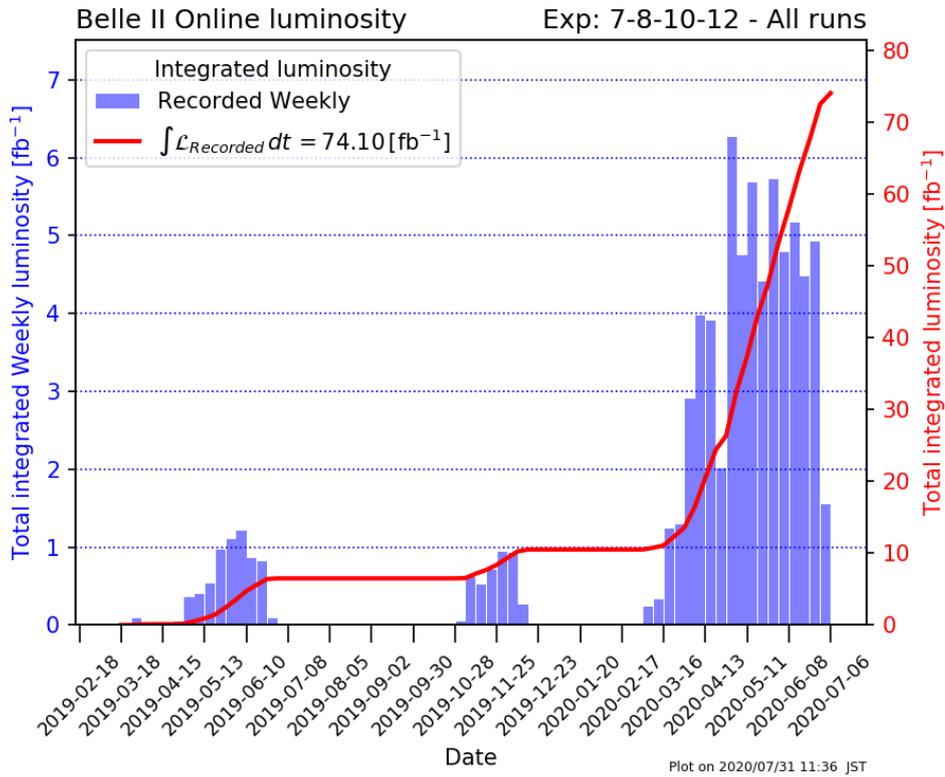
FIG 3.14: Luminosity integration in Phase 3 until 2020 summer.

3.4 Particle Identification (PID)

The identification of charged particle types is crucial for physics analysis at the Belle II experiment. In the analysis of $B \to X_s \ell^+ \ell^-$, especially, the electron and the muon identifications are important to reject backgrounds and to suppress systematic uncertainty. In order to identify particle type, a likelihood $\mathcal{L}^{det}(\vec{x}|i)$ for each particle hypothesis (i) is constructed at each sub-detector from a given set of observables, $\vec{x}$. The dominant observables in each detector are as follows:

- Energy loss dE/dx measurement in CDC

- Ratio between energy measured in ECL and track momentum E/p

- Position and time of Cherenkov photons measured in TOP counter

- Cherenkov angle measured in ARICH from the ring image

- Difference between KLM hits position and the extrapolated track

Assuming each sub-detector information is independent, a global likelihood for each particle hypothesis $\mathcal{L}(\vec{x}|i)$ is defined by:

$$\mathcal{L}(\vec{x}|i) = \prod_{det}^{CDC,TOP,ARICH,ECL,KLM} \mathcal{L}^{det}(\vec{x}|i). \tag{3.2}$$

The likelihood ratio PID_i of all particle hypotheses is used as an evaluation indicator in this analysis:

$$PID_i = \frac{\mathcal{L}_i}{\sum_j^{e,\mu,\pi,K,p,d} \mathcal{L}_j}. \tag{3.3}$$

Figure 3.15 shows the PID distributions for electron, muon, pion and kaon. There are several peaks, for example $PID = 1/6 (= 0.167)$, due to lack of information to distinguish particle type.

(a) PID_e

(b) PID_μ

(c) PID_π
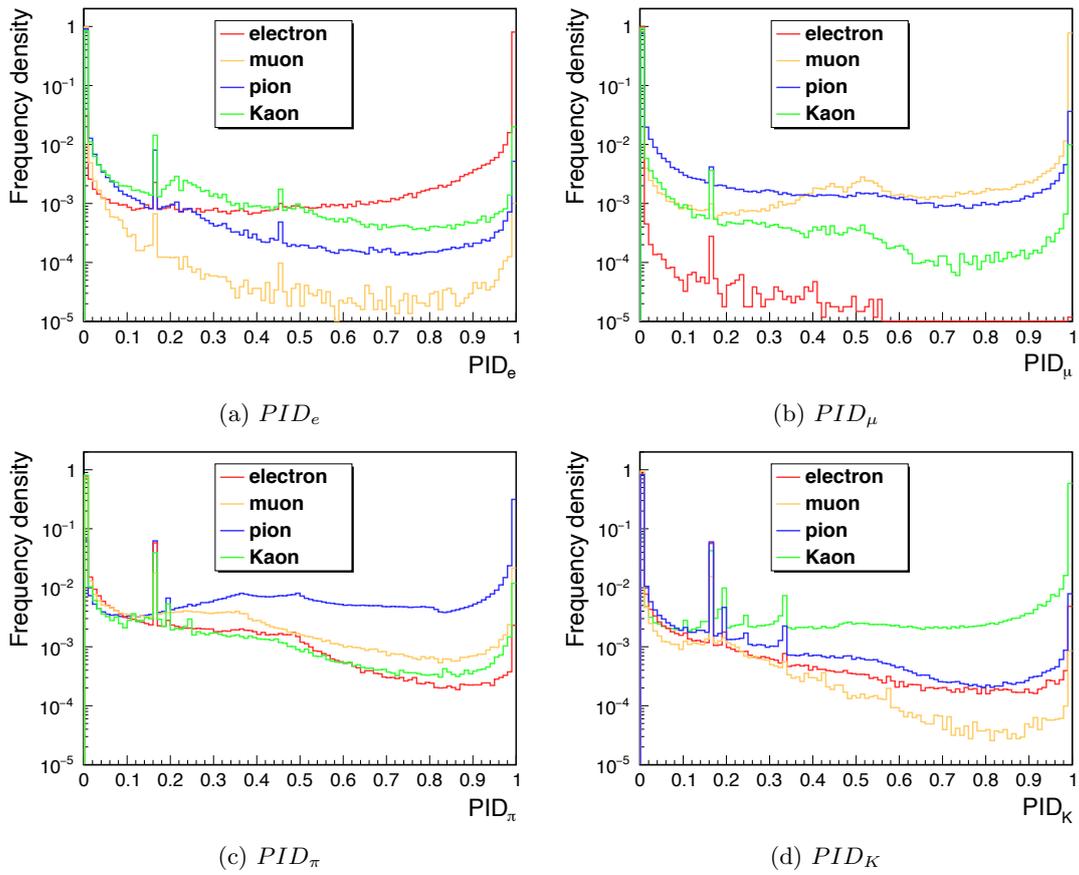
(d) PID_K

FIG 3.15: PID distributions. (a) PID_e, (b) PID_μ, (c) PID_π, (d) PID_K. Red line shows electron, orange line shows muon, blue line shows pion, and green line shows kaon.

Alternative indicator can be constructed from the binary likelihood ratio, for example,

$$\mathcal{R}_{K/\pi} = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi}. \tag{3.4}$$

The PID efficiency and mis-identification (mis-ID) rate are evaluated by using several calibration modes, in which the type of charged particle can be known from event features without biases on the PID performance. The PID efficiency for i-type particle is defined as

$$\text{efficiency} = \frac{\text{Number of } i\text{-type particle identified as } i\text{-type}}{\text{Number of } i\text{-type particle}}. \tag{3.5}$$

and the mis-ID rate from i-type to j-type is defined as

$$\text{mis-ID rate} = \frac{\text{Number of } i\text{-type particle identified as } j\text{-type}}{\text{Number of } i\text{-type particle}}. \tag{3.6}$$

The lepton identification efficiency and mis-ID rate from lepton to hadron are evaluated by combining several methods. The main method is the two-photon analysis, $e^+e^- \to e^+e^-\gamma\gamma^* \to e^+e^-\ell^+\ell^-$. The e^+e^- in the final state enter the beam pipe region and only $\ell^+\ell^-$ go into the detector volume. In this analysis, one of the lepton ($\ell^\pm$) is used to probe the efficiency and the other lepton ($\ell^\mp$) is used to suppress the background by the lepton identification requirement in advance. Since the two-photon events have very large cross-section, especially in low-momentum region, the statistical uncertainty on the identification efficiency can be reduced. Further detail of the two-photon analysis is discussed in Appendix B. The high momentum region is covered by analyses on $J/\psi \to \ell^+\ell^-$, $e^+e^- \to e^+e^-(\gamma)$, and $e^+e^- \to \mu^+\mu^-(\gamma)$ [14].

The hadron identification efficiency and mis-ID rate from hadron to lepton are investigated with $D^{*+} \to D^0(\to K^-\pi^+)\pi^+_{slow}$. The D^0 meson is reconstructed from two oppositely charged tracks. The charged slow pion π_{slow} is used to tag the flavor of D^0, thus to used identify the kaon and pion [15]. Moreover, the mis-ID rate is studied using $K^0_S \to \pi^+\pi^-$ and $e^+e^- \to \tau^+\tau^-$.

The evaluated lepton identification efficiency and mis-ID rate from hadron to lepton for $PID_e > 0.9$ and $PID_\mu > 0.9$ are shown in Figure 3.16. The typical identification efficiency is $\sim 94\%$ for electron and $\sim 90\%$ for muon, while the mis-ID rate of pion is $\sim 2\%$ for electron and $\sim 4\%$ for muon.

Figure 3.17 shows hadron identification performance as a function of threshold on the binary likelihood ratio $\mathcal{R}_{K/\pi}$. The identification efficiency of kaon at $PID_K > 0.6$ is about 80% and that of pion at $PID_\pi > 0.4$ is about 84% in the barrel region.
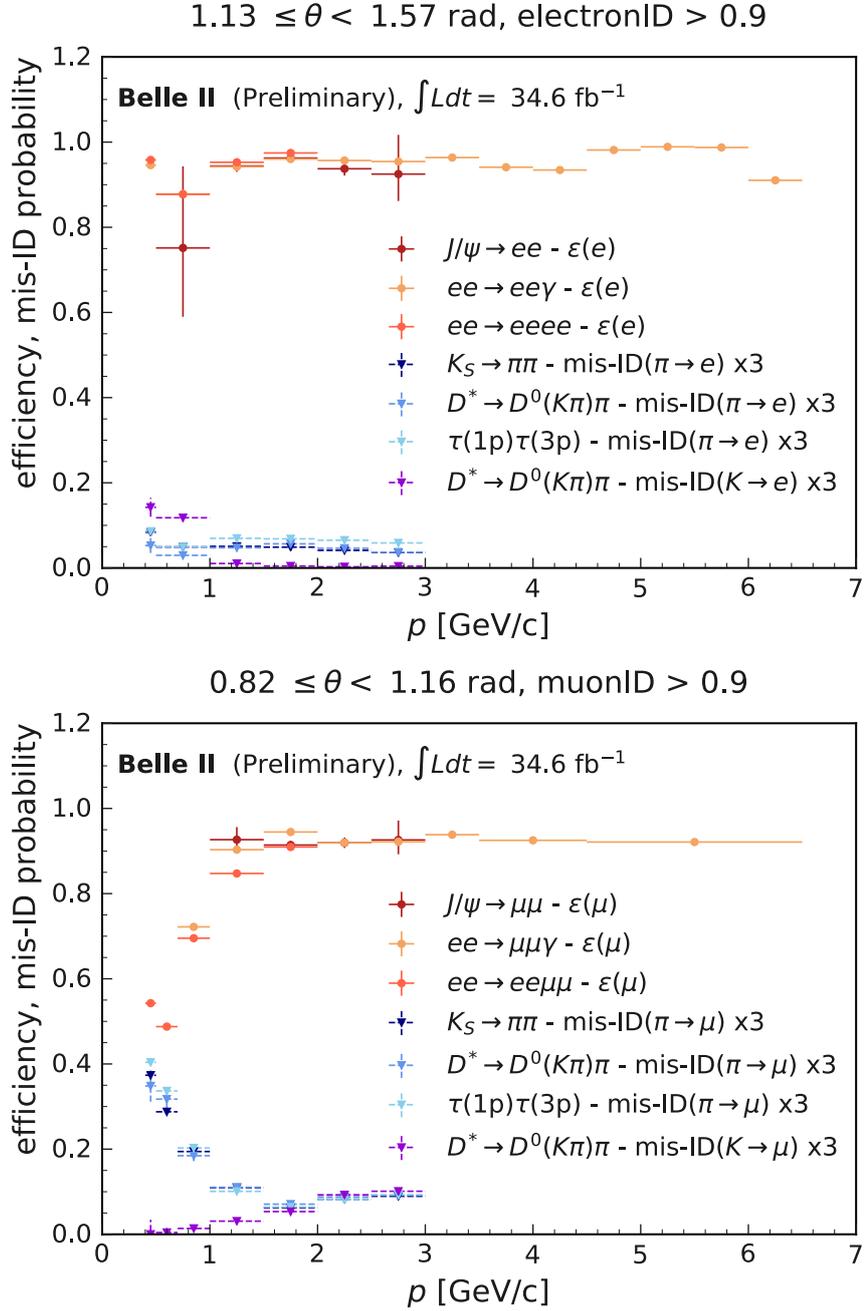
FIG 3.16: Lepton identification efficiency and mis-ID rate from hadron to lepton. Top: Electron identification at $PID_e > 0.9$, Bottom: Muon identification at $PID_\mu > 0.9$. Note that the mis-ID rate has been inflated by a factor 3 for illustration purposes. [14]
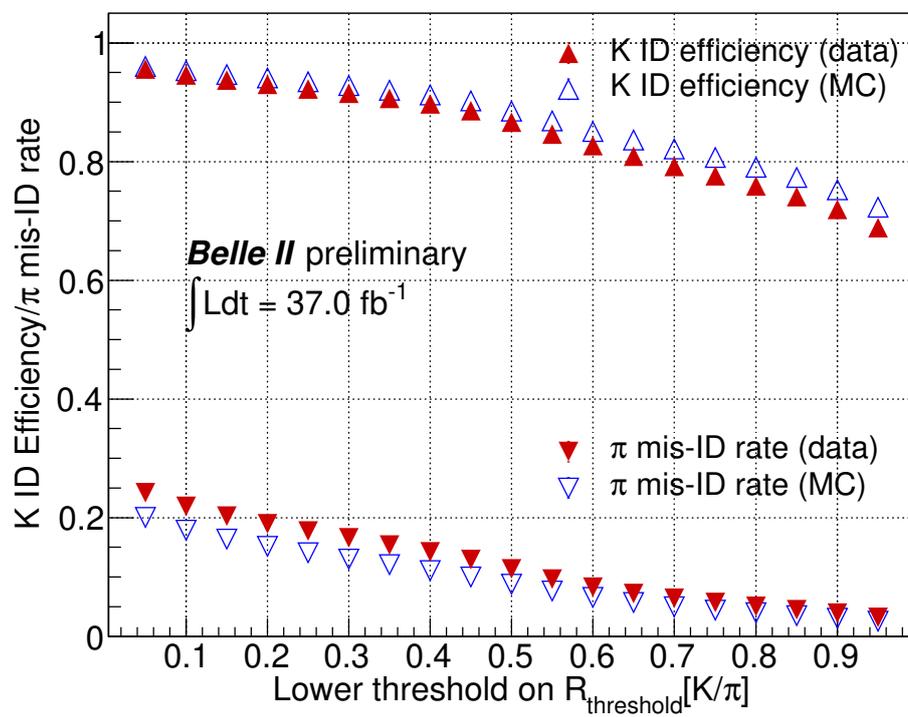
FIG 3.17: Hadron identification performance as a function of $\mathcal{R}_{K/\pi}$. Kaon identification efficiency is shown with triangle and mis-ID rate from pion to kaon is inverted triangle. Red is data and blue is MC. [15]

Chapter 4

Analysis overview

4.1 Analysis strategy

This study is performed as a blind analysis : the analysis procedure is optimized using Monte-Carlo (MC) simulation samples, rather than looking at a signal region. Especially, the reconstruction efficiency of $B \to X_s \ell^+ \ell^-$ is estimated by simulations to measure the branching fraction. A difference of event distribution characteristics between data and MC is corrected with data-driven study and taken into account as systematic uncertainty.

In this analysis, the inclusive $B \to X_s \ell^+ \ell^-$ decays are reconstructed with the sum-of-exclusive approach. The hadronic system X_s is reconstructed from $Kn\pi$ ($0 \leq n \leq 4$) and $3K$ final states allowing for at most one π^0 and one K_S^0. TABLE 4.1 shows the decay mode of X_s.

TABLE 4.1: Decay modes of X_s which are reconstructed in the analysis.

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

4.2 Data sample

This analysis is based on data accumulated on the $\Upsilon(4S)$ resonance in the Phase3 of Belle II. The amount of data corresponds to an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$ which contains 37.7×10^6 $B\overline{B}$ pairs.

4.3 Monte-Carlo simulation sample

We have performed simulation study with large statistics Monte-Carlo (MC) samples to evaluate the reconstruction efficiency of $B \to X_s \ell^+ \ell^-$ and to estimate major background contributions. The MC samples are produced with an event generation tool EvtGen [54]. The hadronization process is implemented with PYTHIA [55] [56] and radiative correction is calculated by PHOTOS [57]. Detector simulation with Belle II Phase3 configuration is performed with GEANT4 [58].

4.3.1   Signal MC sample

Signal MC samples are used to optimize the analysis procedure with the full detector simulation. Especially, the estimation of the reconstruction efficiency is crucial to calculate the actual number of signal events from reconstructed signal yields.

Signal MC sample of $B \to X_s \ell^+ \ell^-$ is produced separately from three components, $B \to K \ell^+ \ell^-$, $B \to K^* \ell^+ \ell^-$, and non-resonant $B \to X_s \ell^+ \ell^-$. Samples of $B \to K \ell^+ \ell^-$ and $B \to K^* \ell^+ \ell^-$ are produced using an EvtGen model of BTOSLLBALL [59], which utilizes a Light Cone Sum Rule approach to estimate the $B \to K^{(*)}$ form factors. Those of non-resonant $B \to X_s \ell^+ \ell^-$ rely on the BTOXSLL model, in which the di-lepton mass spectrum is generated according to [60] and then the two lepton momentum is generated according to [61]. Mass of the non-resonant X_s is required to be larger than $1.1\,\mathrm{GeV}/c^2$. Expressions for the Wilson coefficients and power corrections are taken from [35] and the detailed formulae are taken from [62] and [63]. In the hadronization of non-resonant X_s which is performed by the PYTHIA [55] [56], the partons (s-quark and u-/d-quark) are turned into two to ten hadrons which can be excited states such as K^* and then they are distributed according the allowed phase space.

Since the reconstruction efficiency depends on the multiplicity and momentum of final state particles, the fragmentation of non-resonant X_s should be corrected by using real data and uncertainties of the measurement of the fragmentation should be included in the systematic uncertainty. Since parameters in the EvtGen model can also change the momentum of final state particles, the effect on the reconstruction efficiency from these variation of these parameters. Moreover, the value of the transition point, $m_{X_s} > 1.1\,\mathrm{GeV}/c^2$, which is same with the previous [2] is arbitrary and should be taken into account as systematic uncertainty. It will be described in Section 8.5.

These three components are mixed according to the SM predictions of branching fractions [42] [35] which are shown in TABLE 2.2 and TABLE 2.4. Due to the photon pole contribution, electron modes have larger branching fractions than muon modes for K^* and non-resonant X_s. We have used the SM prediction of $B \to (K, K^*, X_s)\mu^+\mu^-$ on muon modes as well as electron modes applying a cut of $M_{e^+e^-} > 0.2\,\mathrm{GeV}$. The branching fraction of K and K^* modes are assumed to be 0.35×10^{-6} and 1.19×10^{-6}. That of non-resonant X_s is assumed to be $2.61(= 4.15 - 0.35 - 1.19) \times 10^{-6}$. The uncertainty on the branching fraction has to be also taken into account as systematic uncertainty. The detail will be described in Section 8.5.

According to the MC samples, the fraction of the X_s covered by this study (TABLE 4.1) is 63.0%. If the fraction of states containing K_L^0 is taken to be equal to that containing K_S^0, the missing fraction is 16.9%. Figure 4.1 shows the M_{X_s} distribution of the signal MC samples.

4.3.2   Background MC sample

To establish background suppression for $B \to X_s \ell^+ \ell^-$, two kinds of hadronic MC samples are produced. One is the continuum events, $e^+ e^- \to q\bar{q}(q = u, d, s, c)$, and the other is $B\bar{B}$ ($B^0\bar{B}^0$ and B^+B^-) events. Decay process of these samples are simulated generically according to the recent measurements results. The number of background MC samples corresponds to an integrated luminosity of $5\,\mathrm{ab}^{-1}$.
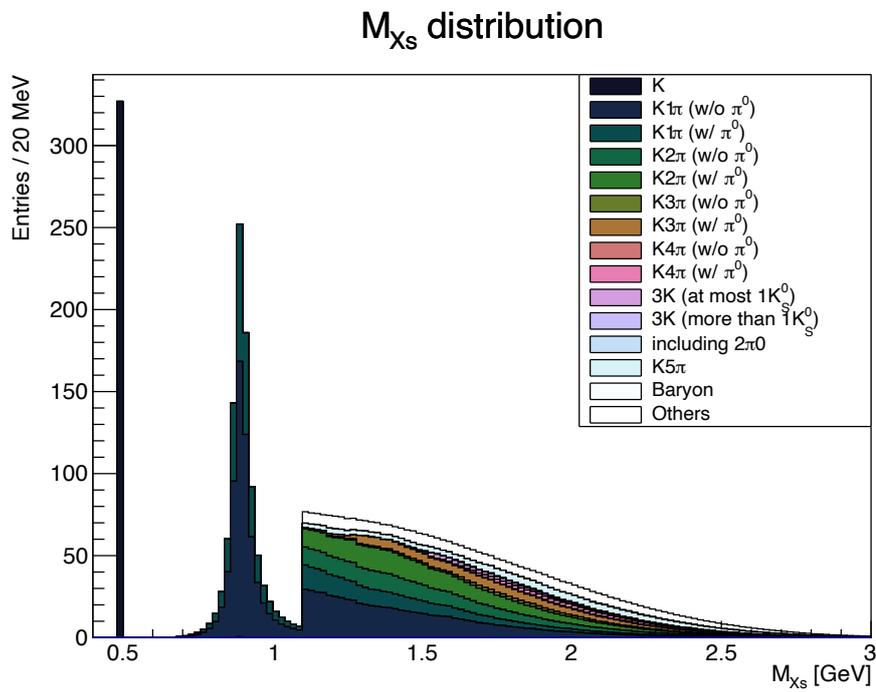
FIG 4.1: M_{X_s} distribution for generated signal MC samples. The histograms are scaled to the 200 fb^{-1}. Decay modes of X_s are separated by color code of histogram.

Chapter 5

Reconstruction

5.1   Particle Selection

Charged particles are selected from tracks (reconstructed with CDC, SVD and PXD) originating from the interaction point (IP) by requiring $dr < 0.5\,\mathrm{cm}$ and $|dz| < 2.0\,\mathrm{cm}$, where dr and dz is distance between a track and IP in the plane to the beam axis and along the beam axis, respectively. Type of the charged particles is identified using the particle identification information obtained by the sub detector systems.

Electron candidates are required to satisfy

- $p > 0.4\,\mathrm{GeV}/c$

- $PID_e > 0.9$

where p is the momentum and PID_e is the likelihood ratio defined in Section 3.4. The momentum selection, $p > 0.4$, is required so that the track can reach the ECL. Due to the light mass, electrons often emit bremsstrahlung photons and loose energy. The energy is recovered by adding four momenta of photons within 0.05 rad cone around the electron to the electron's four momenta.

Muon candidates are required to satisfy

- $p > 0.7\,\mathrm{GeV}/c$

- $PID_\mu > 0.9$

The momentum selection, $p > 0.7\,\mathrm{GeV}/c$ is required so that the track can reach the KLM.

Kaon and pion candidates are selected by requiring $PID_K > 0.6$ and $PID_\pi > 0.4$, respectively. Additional selection on the number of hits in CDC, `nCDCHits`, is required to ensure the CDC dE/dx infromation: `nCDCHits`> 20.

The K_S^0 candidates are reconstructed from two oppositely charged tracks requiring a mass selection, $0.3 < M_{\pi^+\pi^-} < 0.7\,\mathrm{GeV}/c^2$. Kinematics of these tracks are calculated by assuming the pion mass. Neither particle identification nor impact parameter cuts are applied on these tracks. Since the four momenta of tracks are calculated at the closest position to the IP, the sum of the four momenta of daughters shifts from the true four momenta of K_S^0. To correct the difference, the kinematics of daughters are recalculated at the vertex position of K_S^0, which is called the vertex fit. After the vertex fit, the following criteria are applied on K_S^0 candidates: $0.4876 < M_{K_S^0} < 0.5076\,\mathrm{GeV}/c^2$ and `significanceOfDistance`> 50, where $M_{K_S^0}$ is the mass of K_S^0 calculated after the vertex fit and `significanceOfDistance` is the significance of distance from the vertex to IP. The variable `significanceOfDistance` is applied to reject many candidates that arise from random pion tracks originating from the interaction region. Figure 5.1 shows the reconstructed invariant mass $M_{K_S^0}$ of $K_S^0 \to \pi^+\pi^-$ in the MC samples.

The π^0 candidates are formed by combining two photons which are reconstructed from ECL clusters inside the CDC acceptance ($17° < \theta_\gamma < 150°$). In addition, the γ candidates are required to satisfy the following criteria to reduce background photon issued by beam backgrounds: `clusterNHits>1.5` and [`clusterReg == 1` and $E > 0.080\,\mathrm{GeV}$] or [`clusterReg == 2` and $E > 0.030\,\mathrm{GeV}$] or [`clusterReg == 3` and $E > 0.060\,\mathrm{GeV}$], where `clusterReg` is the ECL region of a cluster, 1:forward, 2:barrel, 3:backward. The energy threshold is optimized for each region of ECL and `clusterNHits` is sum of weights of all
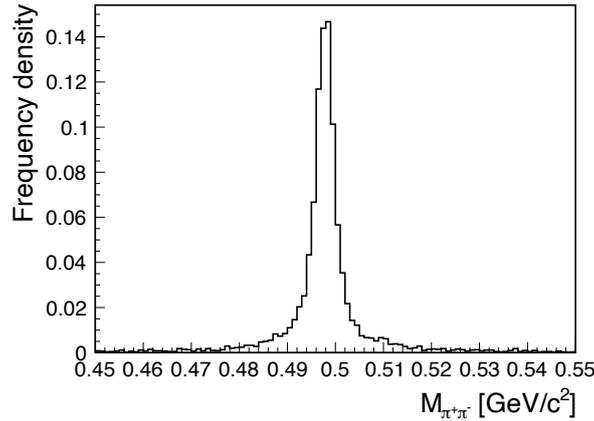
FIG 5.1: $M_{K_S^0}$ distribution of $K_S^0 \to \pi^+\pi^-$ in the MC samples.

crystals in an ECL cluster. For non-overlapping clusters, `clusterNHits` is equal to the number of crystals in the cluster. This, however, can be a non-integer value, when energy splitting among nearby clusters. Finally, π^0 candidates are required to satisfy the following selection, $0.120 < M_{\gamma\gamma} < 0.145\,\text{GeV}/c^2, -1.5 < \Delta\phi_{\gamma\gamma} < 1.5\,\text{rad}, \alpha_{\gamma\gamma} < 1.4\,\text{rad}$ where $\Delta\phi_{\gamma\gamma}$ is difference of ϕ between two gammas and $\alpha_{\gamma\gamma}$ is the angle between two gammas. In the low momentum region, there are a lot of fake π^0 candidates which degrades the signal-to-background ratio. Momentum requirement on π^0 candidates is applied to be more than $0.4\,\text{GeV}/c$. Figure 5.2 shows the reconstructed invariant mass $M_{\gamma\gamma}$ of $\pi^0 \to \gamma\gamma$ in the MC samples.
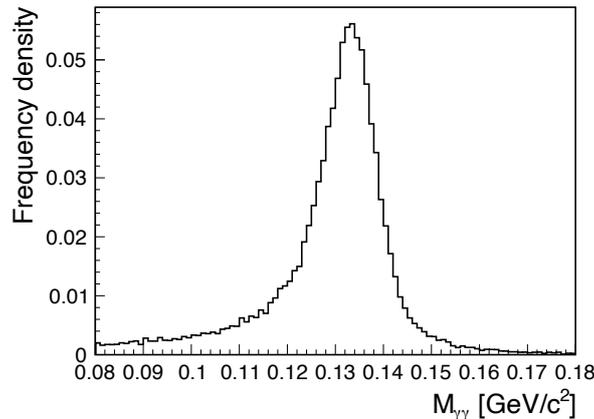
FIG 5.2: $M_{\gamma\gamma}$ distribution of $\pi^0 \to \gamma\gamma$ in the MC samples.

The particle selection criteria are summarized in Table 5.1.

5.2 X_s Reconstruction

The hadronic system X_s is reconstruced with sum-of-exclusive approach. The twenty final states listed in TABLE 4.1 are adopted as X_s candidates. To suppress combinatorial backgrounds, the invariant mass of X_s is required to be smaller than $2.2\,\text{GeV}/c^2$: $M_{X_s} < 2.2\,\text{GeV}/c^2$. Figure 5.3 shows the reconstructed invariant mass M_{X_s} in the signal MC samples.

5.3 B Reconstruction

The B meson produced as a $B\overline{B}$ pair is reconstructed by combining X_s and an electron pair or a muon pair. Two independent kinematic variables, which are the beam constraint mass M_{bc} and the energy

TABLE 5.1: Summary of the particle selection criteria.

Particle	Selection Criteria
Charged tracks	$dr < 0.5\,\mathrm{cm}, |dz| < 2.0\,\mathrm{cm}$
$e^{\pm}$	$p > 0.4\,\mathrm{GeV}/c, PID_e > 0.9,$ Recovering the bremsstrahlung radiation by adding photons which are within 0.05 rad cone along the electron direction.
$\mu^{\pm}$	$p > 0.7\,\mathrm{GeV}/c, PID_\mu > 0.9$
$K^{\pm}$	$PID_K > 0.6, \texttt{nCDCHits} > 20$
$\pi^{\pm}$	$PID_\pi > 0.4, \texttt{nCDCHits} > 20$
K^0_S	Reconstructed from two oppositely charged tracks with the vertex fit. $\texttt{significanceOfDistance} > 50, 0.4876 < M < 0.5076\,\mathrm{GeV}/c^2$
γ	$17° < \theta_\gamma < 150°, \texttt{clusterNHits} > 1.5, \texttt{clusterNHits>1.5}$ and $[\texttt{clusterReg} == 1$ and $E > 0.080\,\mathrm{GeV}]$ or $[\texttt{clusterReg} == 2$ and $E > 0.030\,\mathrm{GeV}]$ or $[\texttt{clusterReg} == 3$ and $E > 0.060\,\mathrm{GeV}].$
π^0	Reconstructed from two γ's. $0.120 < M_{\gamma\gamma} < 0.145\,\mathrm{GeV}/c^2, -1.5 < \Delta\phi_{\gamma\gamma} < 1.5\,\mathrm{rad}, \alpha_{\gamma\gamma} < 1.4\,\mathrm{rad}, p_{\pi^0} > 0.4\,\mathrm{GeV}/c$
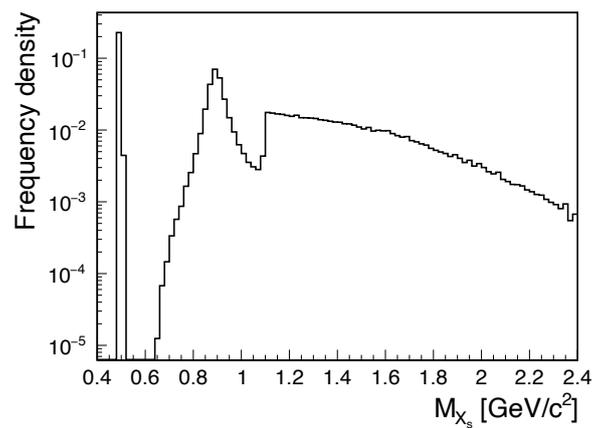
FIG 5.3: M_{X_s} distribution of signal MC samples.

difference ΔE, are calculated in the CM system. The definitions of these variables are

$$M_{\mathrm{bc}} \equiv \sqrt{E_{\mathrm{beam}}^{*2} - |\vec{p}_B^*|^2} \tag{5.1}$$

$$\Delta E \equiv E_B^* - E_{\mathrm{beam}}^* \tag{5.2}$$

where E_{beam}^* is the beam energy in the CM system and $(\vec{p}_B, E_B)$ is the reconstructed four-momentum of B meson in the CM system. Rather than using the B invariant mass, these variables are more effective for background separation, because the energy resolution of the initial e^+e^- is more precise than that of the reconstructed B meson. Figure 5.4 shows distributions of M_{bc} and ΔE of signal MC samples.

B candidates are required to satisfy $5.2 < M_{\mathrm{bc}} < 5.3\,\mathrm{GeV}/c^2$ and $|\Delta E| < 0.15\,\mathrm{GeV}$. M_{bc} is used to extract signal yields with fitting and thus the requirement is very loose. On the other hand ΔE is used to suppress backgrounds.

(a) M_{bc} (b) ΔE
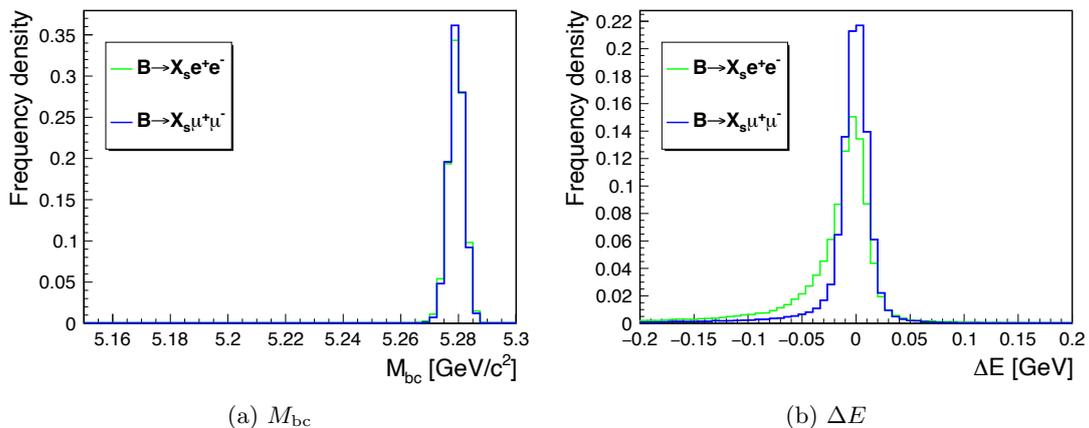
FIG 5.4: M_{bc} (left) and ΔE (right) distributions of the signal MC samples. Green line is $B \to X_s e^+ e^-$ and blue line is $B \to X_s \mu^+ \mu^-$.

Chapter 6

Background study

6.1   Background sources

There are three dominant sources of background. The first is due to $B \to J/\psi \,(\psi(2S)) X_s$ with $J/\psi \,(\psi(2S)) \to \ell^+\ell^-$. The second is associated with $c\bar{c}$ continuum events in which both charm quarks decay semi-leptonically. The third arises from $B\overline{B}$ events where two leptons are generated by semi-leptonic B or D decays.

The $B \to J/\psi \,(\psi(2S)) X_s$ backgrounds are eliminated by imposing vetoes on the invariant mass of the lepton pair. Because the final state and the kinematics of the backgrounds resemble those of the signal, the vetoed events are used as control samples. The selection criteria are described in Section 6.2.2.

The semi-leptonic backgrounds have significant amount of missing energy due to missing neutrinos and large distance between two leptons because they originate from different decay products of B or D. Because of these kinematic features, these backgrounds can be effectively suppressed. Furthermore, the $c\bar{c}$ semi-leptonic backgrounds can be easily identified with event shape variables.

6.2   Background suppression

The number of background events is thousands of times larger than of signal in the analysis of $B \to X_s \ell^+\ell^-$ decays. Thus, the effective background suppression is crucial for this study. At first, the scope of the research is determined and some backgrounds are suppressed in the pre-selection process. Then, the peaking backgrounds which can make a peak at $M_{\mathrm{bc}} = 5.28\,\mathrm{GeV}/c^2$ is eliminated by applying vetos. Final, we use the multivariate analysis technique with FastBDT [64] to suppress non-peaking backgrounds efficiently.

6.2.1   Pre-selection

Firstly, we define the scope of the research by applying a loose pre-selection. To suppress the backgrounds from photon conversion, $\gamma \to e^+e^-$, and Dalitz decay of π^0, $\pi^0 \to e^+e^-\gamma$, electron mode candidates are required to have large di-electron mass, $M_{e^+e^-} > 0.2\,\mathrm{GeV}/c^2$. There are also signal contribution which are sensitive to the Wilson coefficient C_7. The dedicated study in this region with $B \to K^*e^+e^-$ remains for future work. Moreover, the mass of X_s is required to be smaller than $2.0\,\mathrm{GeV}/c^2$. These conditions are same as the Belle experiment [17] [2] which is good to compare with theoretical predictions.

Some non-peaking backgrounds are suppressed here to make the FastBDT performance more efficient. To estimate the distance between each track in a candidate of B meson, the vertex fit is applied on the reconstructed B meson. Since daughters of K_s^0 is displaced from B meson vertex, these particles are eliminated from the fit. Then the χ^2 probability of the vertex fit is calculated $\mathtt{chiProb}_B$, which will be used as an input variable of FastBDT. In the pre-selection, we just require that the vertex fit is converged so that the probability is calculable. The ΔE defined as Equation 5.2 is one of the best variables to distinguish between signal and backgrounds. A loose selection of ΔE rejects large amount of backgrounds while it keeps most of signal. Since the electron may loose energy due to the radiation, the

selection on the electron modes is asymmetric, $-0.10 < \Delta E < 0.05\,\mathrm{GeV}(X_s e^+ e^-)$, while the selection on the muon modes are symmetric, $-0.05 < \Delta E < 0.05\,\mathrm{GeV}(X_s \mu^+ \mu^-)$.

The selection criteria in the pre-selection are summarized in TABLE 6.1 and FIGs 6.1 - 6.3 show the distributions of ΔE, M_{X_s}, and $M_{e^+ e^-}$.

TABLE 6.1: Criteria of the pre-selection

Energy difference	$-0.10 < \Delta E < 0.05\,\mathrm{GeV}(X_s e^+ e^-)$
	$-0.05 < \Delta E < 0.05\,\mathrm{GeV}(X_s \mu^+ \mu^-)$
Condition of the vertex fit of B meson	Converged
X_s mass	$M_{X_s} < 2.0\,\mathrm{GeV}/c^2$
γ conversion and π^0 Dalitz decay veto	$M_{e^+ e^-} > 0.2\,\mathrm{GeV}/c^2$
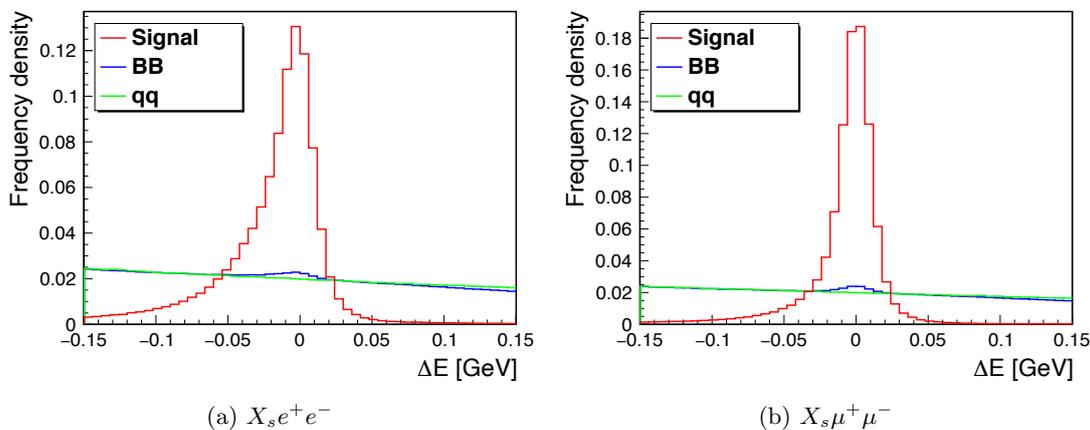
(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$

FIG 6.1: ΔE distributions of $B \to X_s \ell^+ \ell^-$. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events.

6.2.2 Charmonium veto

The charmonium events are the most dominant peaking backgrounds in this study, since they have same final states when $J/\psi\,(\psi(2S))$ decays to two leptons. It is very difficult to estimate the amount of these events from data, it could be a dominant source of systematics uncertainty unless they can be eliminated certainly. To veto the $B \to J/\psi\,(\psi(2S))X_s$ backgrounds, signal candidates that satisfy the following tight criteria are rejected.

$$-0.40\,\mathrm{GeV}/c^2 < M_{e^+(\gamma)e^-(\gamma)} - M_{J/\psi} < 0.15\,\mathrm{GeV}/c^2 \tag{6.1}$$

$$-0.25\,\mathrm{GeV}/c^2 < M_{e^+(\gamma)e^-(\gamma)} - M_{\psi(2S)} < 0.10\,\mathrm{GeV}/c^2 \tag{6.2}$$

$$-0.25\,\mathrm{GeV}/c^2 < M_{\mu^+\mu^-} - M_{J/\psi} < 0.10\,\mathrm{GeV}/c^2 \tag{6.3}$$

$$-0.15\,\mathrm{GeV}/c^2 < M_{\mu^+\mu^-} - M_{\psi(2S)} < 0.10\,\mathrm{GeV}/c^2 \tag{6.4}$$

where $M_{J/\psi}$ is the nominal invariant mass of J/ψ, $M_{J/\psi} = 3.096916\,\mathrm{GeV}/c^2$, and $M_{\psi(2S)}$ is that of $\psi(2S)$, $M_{\psi(2S)} = 3.68609\,\mathrm{GeV}/c^2$. Figure 6.4 shows the $M_{e^+(\gamma)e^-(\gamma)}$ and $M_{\mu^+\mu^-}$ distributions. Because the backgrounds can pass the selection if photons are added wrongly in the bremsstrahlung recovery process, the invariant mass of the electron pair is calculated with and without the photons. Thus, there are four combinations at most, $e^+e^-, e^+\gamma e^-, e^+e^-\gamma, e^+\gamma e^-\gamma$. The amount of remaining backgrounds are estimated in Section 6.3.3.

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$
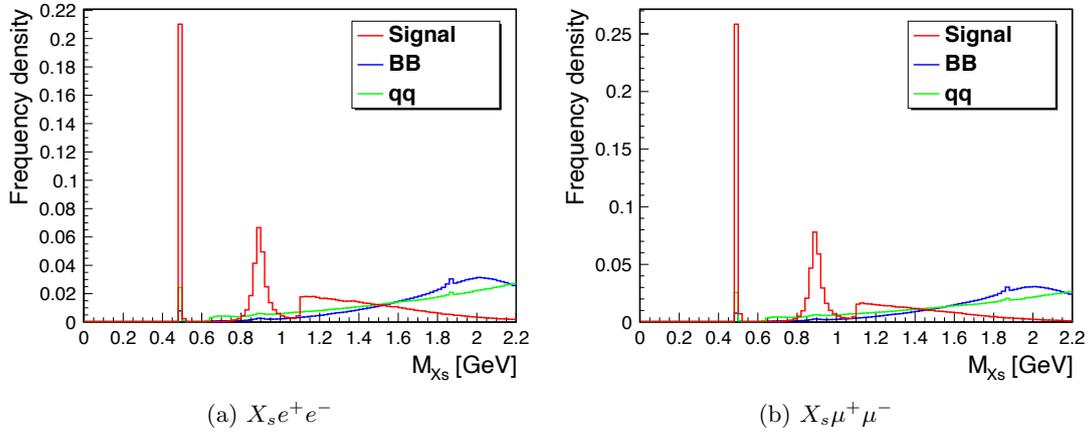
FIG 6.2: M_{X_s} distributions of $B \to X_s \ell^+ \ell^-$. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events.

(a) $X_s e^+ e^-$
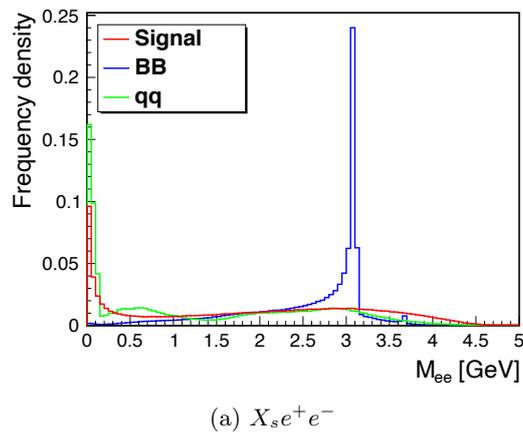
FIG 6.3: $M_{e^+ e^-}$ distributions of $B \to X_s e^+ e^-$. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events. The $q\bar{q}$ continuum background has a peak at $M_{e^+ e^-} = 0$ due to photon conversion and Dalitz decay of π^0.

(a) $M_{e^+e^-}$

(b) $M_{e^+\gamma e^-}$

(c) $M_{e^+e^-\gamma}$

(d) $M_{e^+\gamma e^-\gamma}$

(e) $M_{\mu^+\mu^-}$
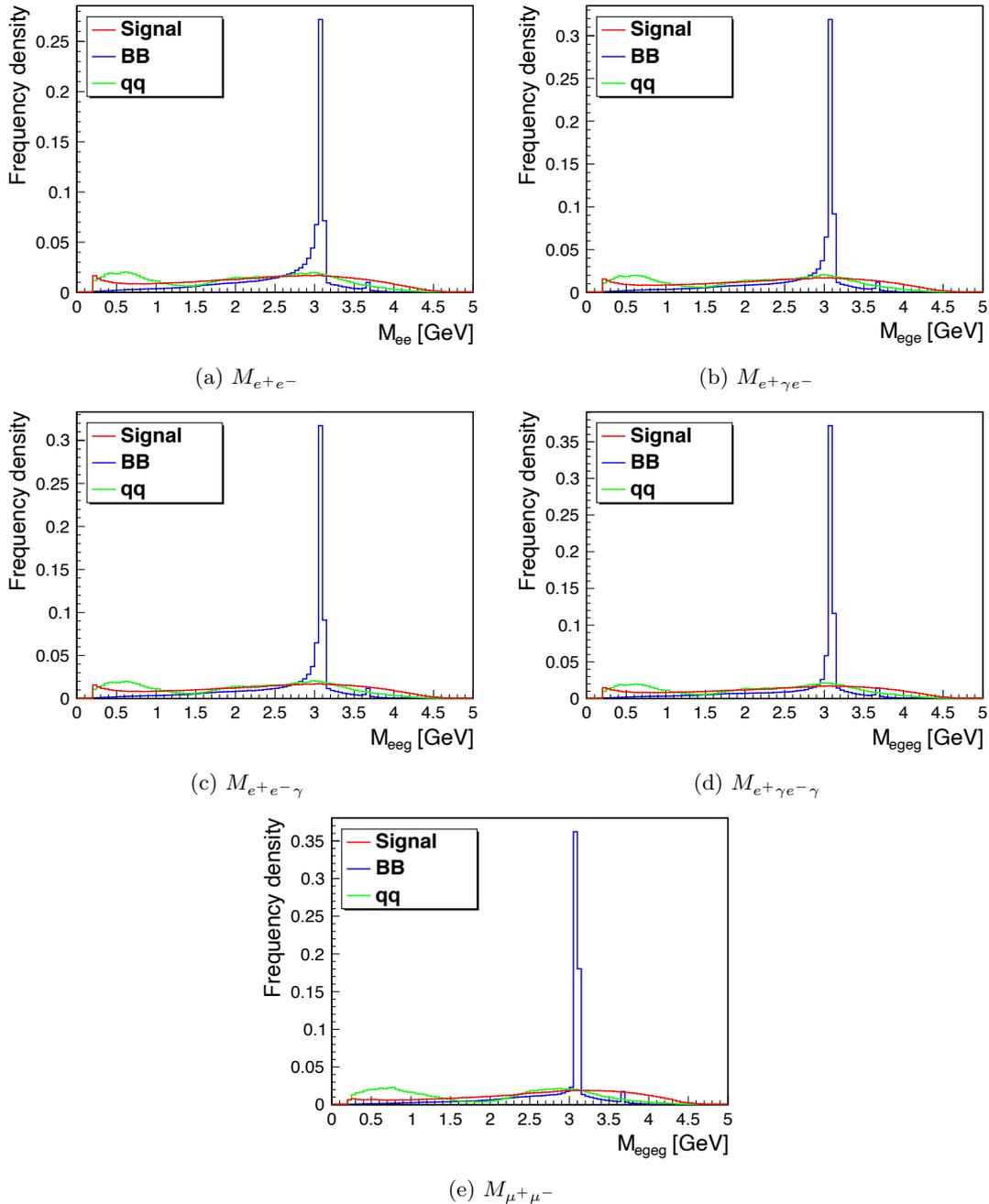
FIG 6.4: $M_{e^+(\gamma)e^-(\gamma)}$ and $M_{\mu^+\mu^-}$ distributions. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events. The $B\bar{B}$ background has peaks at $M_{\ell^+\ell^-} = 3.0969\,\text{GeV}/c^2$ due to J/ψ and at $M_{\ell^+\ell^-} = 3.6861\,\text{GeV}/c^2$ due to $\psi(2S)$.

6.2.3 D veto

The events of $B \to DX$ can be contaminated due to miss identification of particle types. Some events make a peak on the M_{bc} and ΔE distributions, for example, mis-identified $B \to D(\to K\pi)\pi$ as $B \to K\mu\mu$. The amount of these peaking backgrounds can be estimated from data (Section 6.3.1). However, the mis-ID rate is still high in the early stage of Belle II and thus the number of peaking backgrounds are large which induce large systematic uncertainty. We eliminate the peaking backgrounds due to D meson by applying a veto on the invariant mass.

The invariant masses of all hadronic or semi-leptonic combinations of B daughters whose charge is 0 or ± 1 are calculated. The leptonic combination of $\ell^+\ell^-$ is not examined to avoid a bias on the $q^2 (= M_{\ell^+\ell^-})$. If the most D-like mass $M_{\mathrm{most\ D-like}}$ among them is close to nominal D mass, the candidate is vetoed. The veto condition is as follows.

$$1.85 \,\mathrm{GeV}/c^2 < M_{\mathrm{most\ D-like}} < 1.89 \,\mathrm{GeV}/c^2 \tag{6.5}$$

Figure 6.5 show the $M_{\mathrm{most\ D-like}}$ distributions of the events which are identified as the peaking backgrounds using the MC truth information. The veto is applied to the following modes which have significant backgrounds due to D events : $K\ell^+\ell^-, K\pi\ell^+\ell^-, K\pi\pi\ell^+\ell^-, K\pi\pi\pi\ell^+\ell^-$. The remaining peaking backgrounds can be estimated from data to get the signal yields correctly. The detail is described in Section 6.3.1.
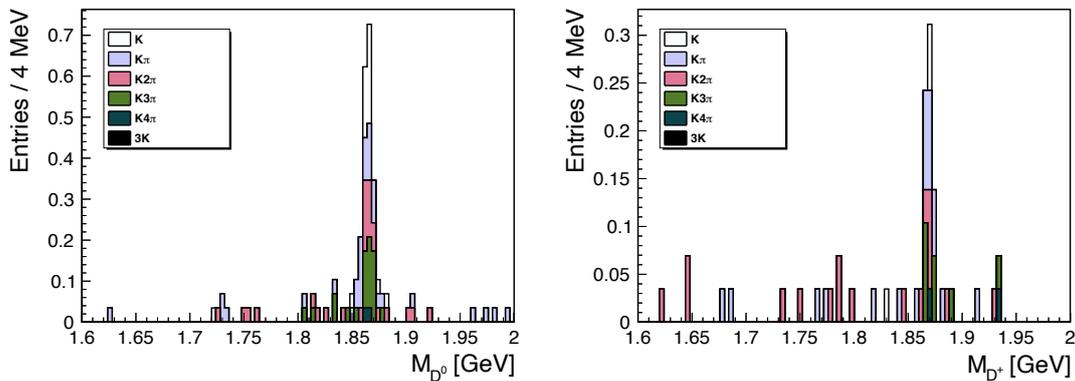
(a) $M_{\mathrm{mostD-like}}$ distribution for neutral combinations in the $X_s e^+ e^-$ mode.

(b) $M_{\mathrm{mostD-like}}$ distribution for charged combinations in the $X_s e^+ e^-$ mode.
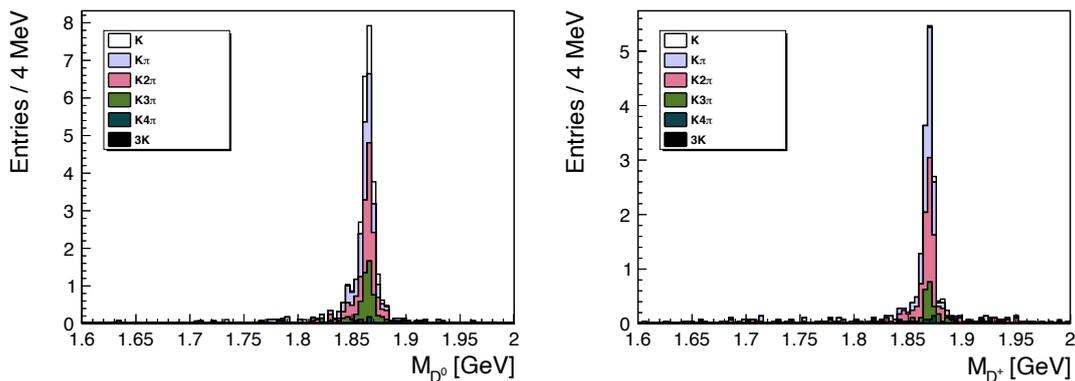
(c) $M_{\mathrm{mostD-like}}$ distribution for neutral combinations in the $X_s \mu^+ \mu^-$ mode.

(d) $M_{\mathrm{mostD-like}}$ distribution for charged combinations in the $X_s \mu^+ \mu^-$ mode.

FIG 6.5: $M_{\mathrm{mostD-like}}$ distributions of events which are identified as the double mis-ID background using the MC truth information. Decay modes of X_s are separated by color. The histograms are scaled for an integrated luminosity of $34.6 \,\mathrm{fb}^{-1}$.

6.2.4   FastBDT

Figure 6.6 shows the M_{bc} distributions after the D veto. The signal events are hardly visible due to the large backgrounds. To suppress the remaining backgrounds, FastBDT [64] is adopted in this study. FastBDT is a speed-optimized and cache-friendly implementation for the stochastic gradient-boosted decision tree (SGBDT) algorithm [65] extensively used by the Belle II experiment.

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$

(c) $X_s \ell^+ \ell^-$
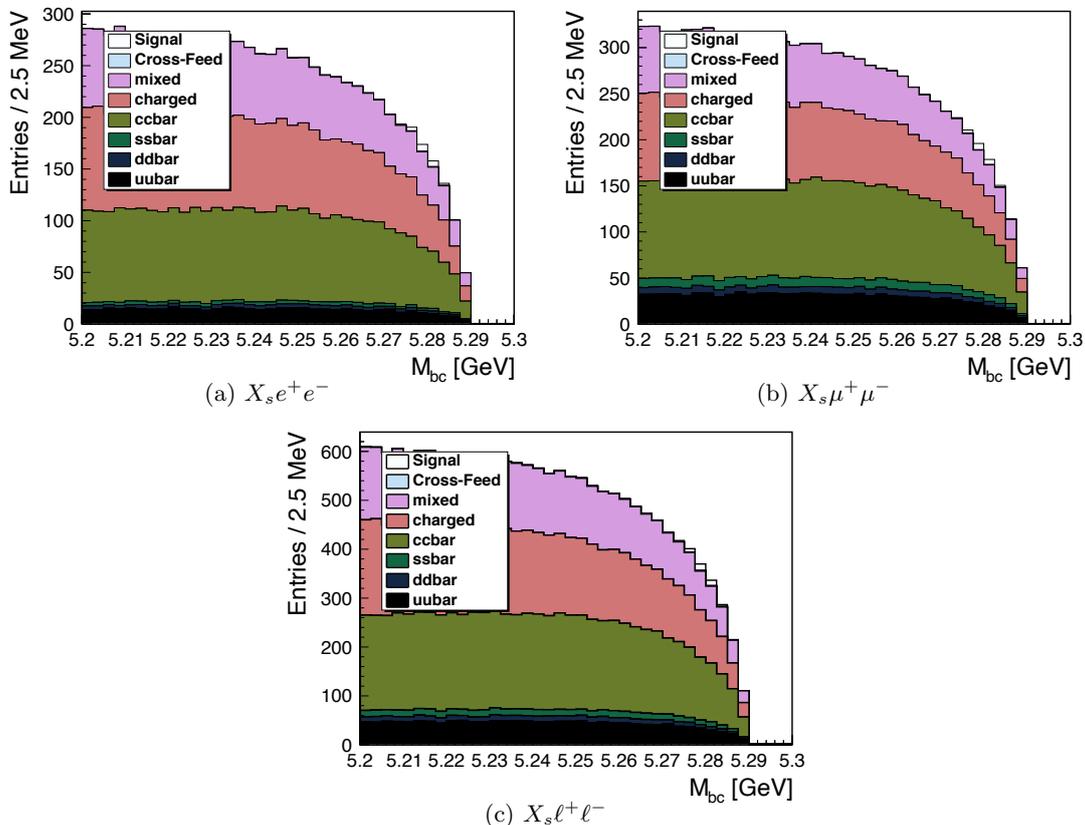
FIG 6.6: M_{bc} distributions before applying the FastBDT. MC samples are identified by color code. White histogram shows signal events. The histograms are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

Here we introduce the FastBDT implementation and the SGBDT algorithm. The base idea of the SGBDT is the decision tree (DT) which is a tree-like decision model. Combination of variables and threshold for each node is trained in DT algorithm. The gradient boosted decision tree (GBDT) is modified from the DT with gradient boosting technique. Instead of single deep DT, many shallow DTs are combined to give a decision. An advantage of GBDT compared with DT is the robustness against to the over-training. The output is given by the sum of weights of many terminate nodes instead of the single value of a terminate node. In the SGBDT, a bunch of training samples are randomly chosen for each training cycle. This feature helps to further avoid the over-training. FastBDT is an implementation of the SGBDT and optimize the pre-processing of data samples and memory access pattern. FastBDT is one order magnitude faster than popular implementations such as TMVA [66]. Figure 6.7 shows the schematic view of the application of FastBDT.

In this study, FastBDT is used to suppress continuum backgrounds and remaining non-peaking $B\bar{B}$ backgrounds. FastBDT classifiers are trained with MC samples using 36 kinematical variables as input variables. To take into account the difference of kinematics between electron modes and muon modes as well as high M_{X_s} region and low M_{X_s} region, events are divided into four categories, namely: $[e^+ e^-,\ M_{X_s} < 1.1\,\mathrm{GeV}/c^2]$, $[e^+ e^-,\ M_{X_s} > 1.1\,\mathrm{GeV}/c^2]$, $[\mu^+ \mu^-,\ M_{X_s} < 1.1\,\mathrm{GeV}/c^2]$, and $[\mu^+ \mu^-,\ M_{X_s} > 1.1\,\mathrm{GeV}/c^2]$. Then, two classifiers are trained for each events classes, one is for $B\bar{B}$ backgrounds and the other is for continuum backgrounds, since they have very different kinematics and it is difficult to optimize single classifier for them. In total, 8 classifiers are trained and each event has two FastBDT

$$\vec{X} = \{x, y, z, ...\}$$
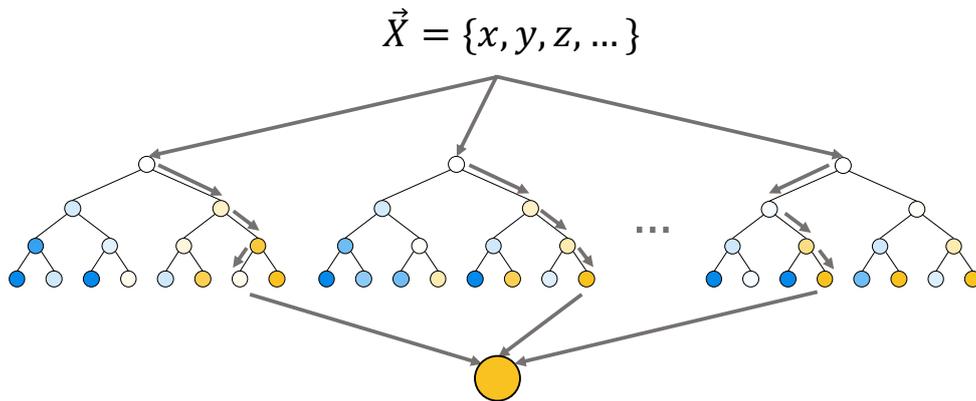
FIG 6.7: Schematic view of the application of FastBDT. The depth of trees is three in this case.

output values. In each classifier, the number of trees are 200 and depth of a tree is 3. The training samples are provided from mixture of equal parts of signal and background. The number of samples are at least 80 thousands which is enough to avoid the over training.

Here we explain why there are difference on the kinematical variables between the $B\overline{B}$ events and the continuum events. Since the $\Upsilon(4S)$ mass, $M_{\Upsilon(4S)} = 10.579 \,\mathrm{GeV}/c^2$, is barely above the $B\overline{B}$ production threshold ($M_B = 5.279$), both B meson are produced almost at rest in the centre-of-mass (CM) frame, as a result, the B decay products are distributed isotropically. By contrast, the continuum $q\overline{q}$ events have back-to-back two jet topology because the light quark q is produced with large momentum. The difference of event topology is quantified in variables which is described in later of this section.

Moreover, the semi-leptonic $B\overline{B}$ backgrounds can be characterized with neutrino information and distance of particles. Since the semi-leptonic decay of B or D always include a neutrino in the decay products, missing energy and momentum are larger than those of signal events. In addition, both B meson and D meson have long life time, as a result the decay vertex is displaced from each other.

Now we introduce the kinematical variables to take into account these features.

- KSFW variables [67]
  KSFW [67] is variables to describe the event kinematical shape which is extended from the Fox-Wolfram moments [68]. The phase-space distribution of momentum and energy flow is parameterized based on the spherical harmonics resulting in the Fox-Wolfram moments. The ℓ-th Fox-Wolfram moment H_ℓ is defined as following expression,

$$H_\ell = \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_\ell(\cos\theta_{ij}). \tag{6.6}$$

where P_ℓ is the ℓ-th Legendre polynomial, $\vec{p}_{i(j)}$ is the momentum of i-th (j-th) particle, and θ_{ij} is angle between i-th and j-th particles. The sum is over the particles in the final state.

The SFW (Super Fox-Wolfram) is extended from the Fox-Wolfram momenta by separating the particles from the signal and those from other side B meson. The H_l is divided into three variables,

$$H_\ell^{SS} = \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_\ell(\cos\theta_{ij}), \tag{6.7}$$

$$H_\ell^{SO} = \sum_{j,k} |\vec{p}_j||\vec{p}_k| P_\ell(\cos\theta_{jk}), \tag{6.8}$$

$$H_\ell^{OO} = \sum_{k,l} |\vec{p}_k||\vec{p}_l| P_\ell(\cos\theta_{kl}), \tag{6.9}$$

$$H_\ell = H_\ell^{SS} + H_\ell^{SO} + H_\ell^{OO}. \tag{6.10}$$

The i and j runs over particles in the signal side final state while k and l runs over the other side particles. Thanks to separating the signal side and the other side, H_ℓ^{SO} and H_ℓ^{OO} have less correlation with signal side variables, such as M_{bc}.

Moreover, the KSFW (Kakuno Super Fox-Wolfram) moments are defined from SFW considering the charge of particles and missing mass of the event. The KSFW moments can be categorized into three, R_ℓ^{SO}, R_ℓ^{OO}, and P_t.

R_ℓ^{SO} is based on H_ℓ^{SO} adding the charge information.

$$R_\ell^{SO} = \frac{\alpha_\ell^c H_\ell^{SO,c} + \alpha_\ell^n H_\ell^{SO,n} + \alpha_\ell^m H_\ell^{SO,m}}{E_{\text{beam}}^* - \Delta E}. \tag{6.11}$$

c, n, and m denotes charged particles, neutral particle, and missing particle, respectively. α's are free parameters and H_ℓ^i is defined as follows,

$$H_\ell^{SO,c} = \begin{cases} \sum_i \sum_j |p_j| P_\ell(\cos\theta_{ij}) & (\ell : \text{even}) \\ \sum_i \sum_j |p_j| Q_i Q_j P_\ell(\cos\theta_{ij}) & (\ell : \text{odd}) \end{cases} \tag{6.12}$$

$$H_\ell^{SO,n} = \begin{cases} \sum_i \sum_j |p_j| P_\ell(\cos\theta_{ij}) & (\ell : \text{even}) \\ 0 & (\ell : \text{odd}) \end{cases} \tag{6.13}$$

$$H_\ell^{SO,m} = \begin{cases} \sum_i |p_M| P_\ell(\cos\theta_{iM}) & (\ell : \text{even}) \\ 0 & (\ell : \text{odd}) \end{cases} \tag{6.14}$$

i runs over particles in the signal final states and j runs over remaining charged particles for $H_\ell^{SO,c}$ and neutral particles for $H_\ell^{SO,n}$. $\vec{p}_M$ is the missing momentum, Q_i is charge of i-th particle and θ_{iM} is the angle between $\vec{p}_i$ and $\vec{p}_M$.

R_ℓ^{OO} is based on H_ℓ^{OO} and defined as follows.

$$R_\ell^{OO} = \frac{\beta_\ell H_\ell^{'OO}}{(E_{\text{beam}}^* - \Delta E)^2}. \tag{6.15}$$

and $H_\ell^{'OO}$ is

$$H_\ell^{'OO} = \begin{cases} \sum_j \sum_k |p_j||p_k| P_\ell(\cos\theta_{jk}) & (\ell : \text{even}) \\ \sum_j \sum_k |p_j||p_k| Q_j Q_k P_\ell(\cos\theta_{jk}) & (\ell : \text{odd}) \end{cases} \tag{6.16}$$

In this thesis, we use the normalized KSFW moments which are defined in the following equation.

$$h_{x,\ell}^{so} = \frac{H_\ell^{SO,x}}{E_{\text{beam}}^* - \Delta E}, \tag{6.17}$$

$$h_\ell^{oo} = \frac{H_\ell^{'OO}}{(E_{\text{beam}}^* - \Delta E)^2}. \tag{6.18}$$

with $x = c, n, m$. Even thought R^{SS} and h^{ss} can be defined in similar manner, they tends to have large correlation with signal side variables and thus they are omitted.

$P_t = \sum_{n=1}^{N_t} |p_{t,n}|$ is the scaler sum of the transverse momentum of all the particle from both signal side and other side.

In addition to 17 KSFW moments, the missing mass squared, $mm2$, is calculated as complementary information.

$$mm^2 = \begin{cases} (E_{\Upsilon(4S)} - \sum_{n=1}^{N_t} E_n)^2 - \sum_{n=1}^{N_t} |p_n|^2 & (E_{\Upsilon(4S)} - \sum_{n=1}^{N_t} E_n) > 0 \\ -(E_{\Upsilon(4S)} - \sum_{n=1}^{N_t} E_n)^2 - \sum_{n=1}^{N_t} |p_n|^2 & (E_{\Upsilon(4S)} - \sum_{n=1}^{N_t} E_n) < 0 \end{cases} \tag{6.19}$$

where E_n is energy of n-th particle and $E_{\Upsilon(4S)}$ is the energy of $\Upsilon(4S)$ corresponding to the centre-of-mass energy. In total, these 18 variables are denoted as KSFW variables in this paper. The KSFW variables are powerful to distinguish signal from not only the continuum $q\bar{q}$ backgrounds but also the semi-leptonic $B\bar{B}$ backgrounds thanks to taking into account the missing components.

- Thrust variables.
  The thrust axis $\vec{n}_T$ for N momenta $\vec{p}_i (i = 1, ..., N)$ is defined as the unit vector which make the thrust T maximum,

$$T = \frac{\sum_i^N |\vec{n}_T \cdot \vec{p}_i|}{\sum_i^N |\vec{p}_i|}. \tag{6.20}$$

  Since the continuum $q\bar{q}$ events have back-to-back topology, thrust of the continuum events are typically larger than that of $B\bar{B}$ events. In this analysis, three thrust related variables are used; thrust of rest-of-events (ROE) other than reconstructed B meson side particle $\text{Thrust}_{\text{ROE}}$, cosine of angle between thrust axis of reconstructed B meson side particles and that of ROE $\cos\theta_{\text{TBTO}}$, and cosine of angle between thrust angle of reconstructed B meson and the beam axis $\cos\theta_{TBz}$. These variables help to suppress the continuum background avoiding a bias on the signal variables.

- CLEO Cone variables [69]
  CLEO Cone variables are extended from thrust variables and investigated by CLEO collaboration [69]. CLEO Cone is calculated from sum of momentum of particles within angular sectors around the thrust axis in intervals of $10°$. In total there are 9 CLEO Cone variables and we define them from CC1, which is calculated for the closest sector to the thrust axis, to CC9, which is calculated for the most perpendicular sector to the thrust axis. These variables are calculated using particles from other than signal to avoid a possible bias on signal.

- Missing mass squared.
  The missing mass squared is calculated as $(2E_{beam} - \sum_i E_i)^2 - |\sum_i \vec{p}_i|^2$, where E_{beam} is the beam energy and $(\vec{p}_i, E_i)$ is the reconstructed particles in the CMS frame. Events which have more than one neutrino tends to have larger value.

- Visible energy in the CM frame.
  The visible energy is calculated from the reconstructed particle in the CM frame. Since events which have neutrino tends to have lower value, this is useful to suppress the semi-leptonic backgrounds.

- Distance between two leptons along the beam axis.
  The distance between two leptons along the beam axis is calculated from each distance from the beam spot. In the signal events, two leptons are originating from same vertex of B. On the other hand, the double-semi-leptonic events (e.g. $B \to D\ell\nu- > K\ell\nu\ell\nu$) have a difference on the production points of leptons.

- Cosine of θ of B at the centre-of-mass frame.
  The cosine of the angle between the B and the beam axis, $\cos\theta_B^*$. The spin-1 $\Upsilon(4S)$ particle decaying into two spin-0 B meson results in $(1 - \cos\theta_B^*)$ distribution, while the continuum events have random distribution.

- `chiProb`$_B$, The χ^2 probability of B vertex.
  As discussed in previous section 6.2.1, the χ^2 probability of the vertex fit of B meson is used as an input variable of FastBDT. The decay products of signal B meson are produced at the same decay vertex of the B meson, while the semi-leptonic backgrounds have a secondary vertex. Thus, the probability tends to lower in background events compared with the signal events.

- Likelihood of ΔE.
  The likelihood of ΔE is used instead of the original ΔE distributions. The shape of ΔE distributions depends on the existence of π^0. To take into account this dependency, the PDF of ΔE is prepared from MC samples for 2 cases, namely modes without π^0 and modes with π^0. These distributions are fitted with two Crystal Ball functions. Crystal Ball function is defined as,

$$f(x) = \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \frac{(x-\mu)}{\sigma} > -\alpha \\ A\left(B - \frac{x-\mu}{\sigma}\right)^{-n} & \frac{(x-\mu)}{\sigma} < -\alpha \end{cases} \tag{6.21}$$

with

$$A = \left(\frac{n}{|\alpha|}\right)^2 \exp(-\frac{|\alpha|^2}{2}), \tag{6.22}$$

$$B = \frac{n}{|\alpha|} - |\alpha|. \tag{6.23}$$

Figure 6.8 and 6.9 show the distributions with fitting function for electron modes and for muon modes, respectively. Because two different PDFs are used to obtain the likelihood, the distributions of the likelihood of ΔE are discontinuous.

(a) without π^0 (b) with π^0
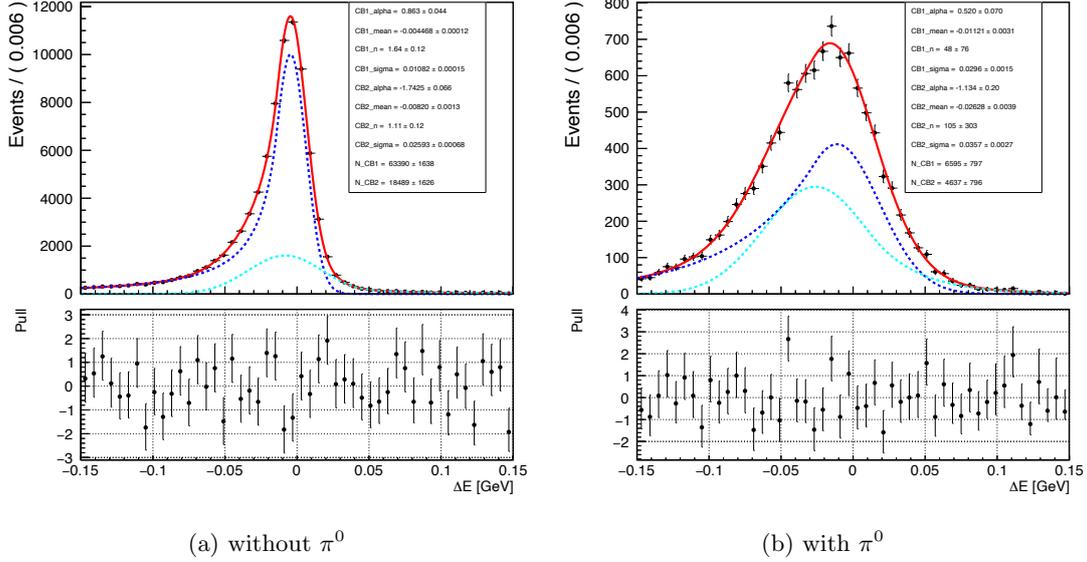
FIG 6.8: Top : ΔE distributions of $B \to X_s e^+ e^-$ with fitting functions. Black makers with error bar show the distribution of MC samples. Red line shows the total likelihood distribution and blue and cyan lines show each component of Crystal Ball function. Bottom : Pull (=[data - fit]/[statistical uncertainty]) distribution of ΔE.

TABLE 6.2 summarizes the description of the input variables to the FastBDT. Figure 6.10 - 6.11 show the distributions of the input variables for $B \to X_s e^+ e^-$ and Figure 6.12 - 6.13 are for $B \to X_s \mu^+ \mu^-$ in the MC samples.

The 8 classifiers are trained by using corresponding MC samples. The output is denoted as F. for example, $F_{e^+ e^-, low}^{B\bar{B}}$ is the FastBDT output of $B\bar{B}$ classifier in $[e^+ e^-, M_{X_s} < 1.1 \, \text{GeV}/c^2]$ category events. Figure 6.14 shows the output distributions of the MC samples.

The selection criteria on $(F^{B\bar{B}}, F^{q\bar{q}})$ are optimized with the Figure of Merit (FOM) given by $S/\sqrt{S + B}$ where S is the number of signals and B is the number of backgrounds in a signal box. The signal box is defined as $5.27 < M_{bc} < 5.29 [\, \text{GeV}/c^2]$. Since each candidate has two outputs $F^{B\bar{B}}$ and $F^{q\bar{q}}$, the optimization is performed from two-dimensional distribution of $(F^{B\bar{B}}, F^{q\bar{q}})$. The classification of samples and the criteria on the output value is summarized in Table 6.3

6.2.5   Best candidate selection

If there are more than one candidate in an event, the best candidate is selected on the basis of the FastBDT output for $B\bar{B}$, $F^{B\bar{B}}$. Average number of candidates per event after the background suppression with FastBDT is 1.13 and 1.09 for $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$, respectively.

6.2.6   Summary

The cut flow tables of the background suppression is shown in TABLE 6.4 and 6.5. The M_{bc} distributions after the background suppression are shown in FIG 6.15. The numbers and histograms are scaled to the

(a) without π^0

(b) with π^0
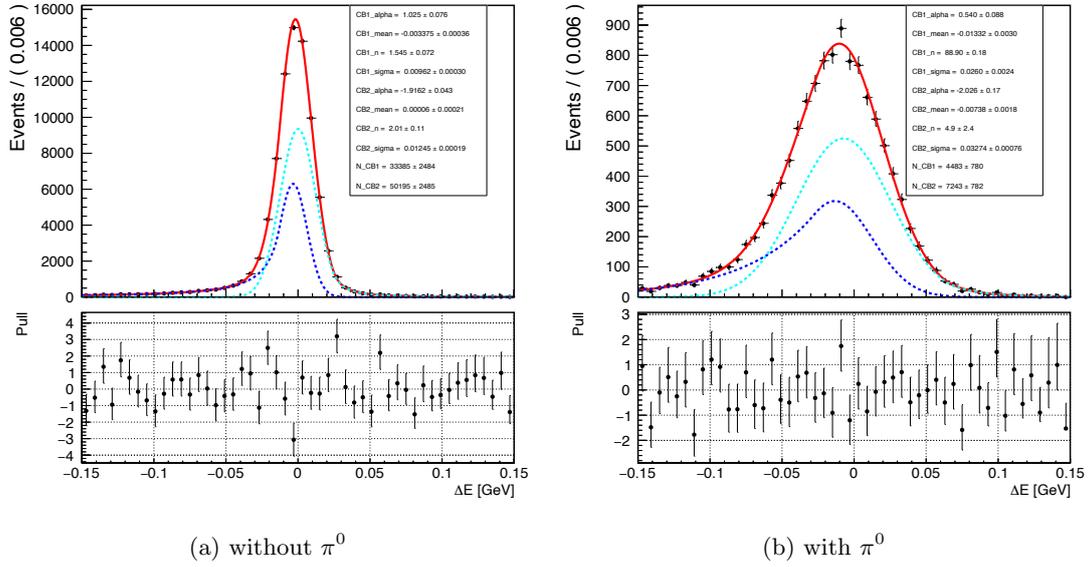
FIG 6.9: Top : ΔE distributions of $B \to X_s \mu^+ \mu^-$ with fitting functions. Black makers with error bar show the distribution of MC samples. Red line shows the total likelihood distribution and blue and cyan lines show each component of Crystal Ball function. Bottom : Pull (=[data - fit]/[statistical uncertainty]) distribution of ΔE.

TABLE 6.2: The list of input variables of FastBDT

Variables	Explanation
$h_{c0}^{so}, h_{c1}^{so}, h_{c2}^{so}, h_{c3}^{so}, h_{c4}^{so},$ $h_{n0}^{so}, h_{n2}^{so}, h_{n4}^{so},$ $h_{m0}^{so}, h_{m2}^{so}, h_{m4}^{so},$ $h_0^{oo}, h_1^{oo}, h_2^{oo}, h_3^{oo}, h_4^{oo},$ P_t, mm^2	18 KSFW variables.
Thrust$_{\text{ROE}}$	Thrust magnitude of ROE.
$\cos\theta_{\text{TBTO}}$	Cosine of angle between the thrust axis of B and that of ROE.
$\cos\theta_{\text{TBz}}$	Cosine of angle between the thrust axis of B and the beam axis.
CC1, CC2, CC3, CC4, CC5, CC6, CC7, CC8, CC9	9 CLEO Cone variables calculated from ROE.
M_{missing}^2	Missing mass squared.
E_{visible}^*	Visible energy in the centre-of-mass frame.
$\Delta z_{\ell^+\ell^-}$	Distance between two leptons along the beam axis.
$\cos\theta_B^*$	Cosine of θ of B at the centre-of-mass frame.
chiProb$_B$	The χ^2 probability of B vertex calculated by the vertex fit using all the charged daughters, excluding the K_s^0 daughters.
Likelihood of ΔE	Likelihood of ΔE.
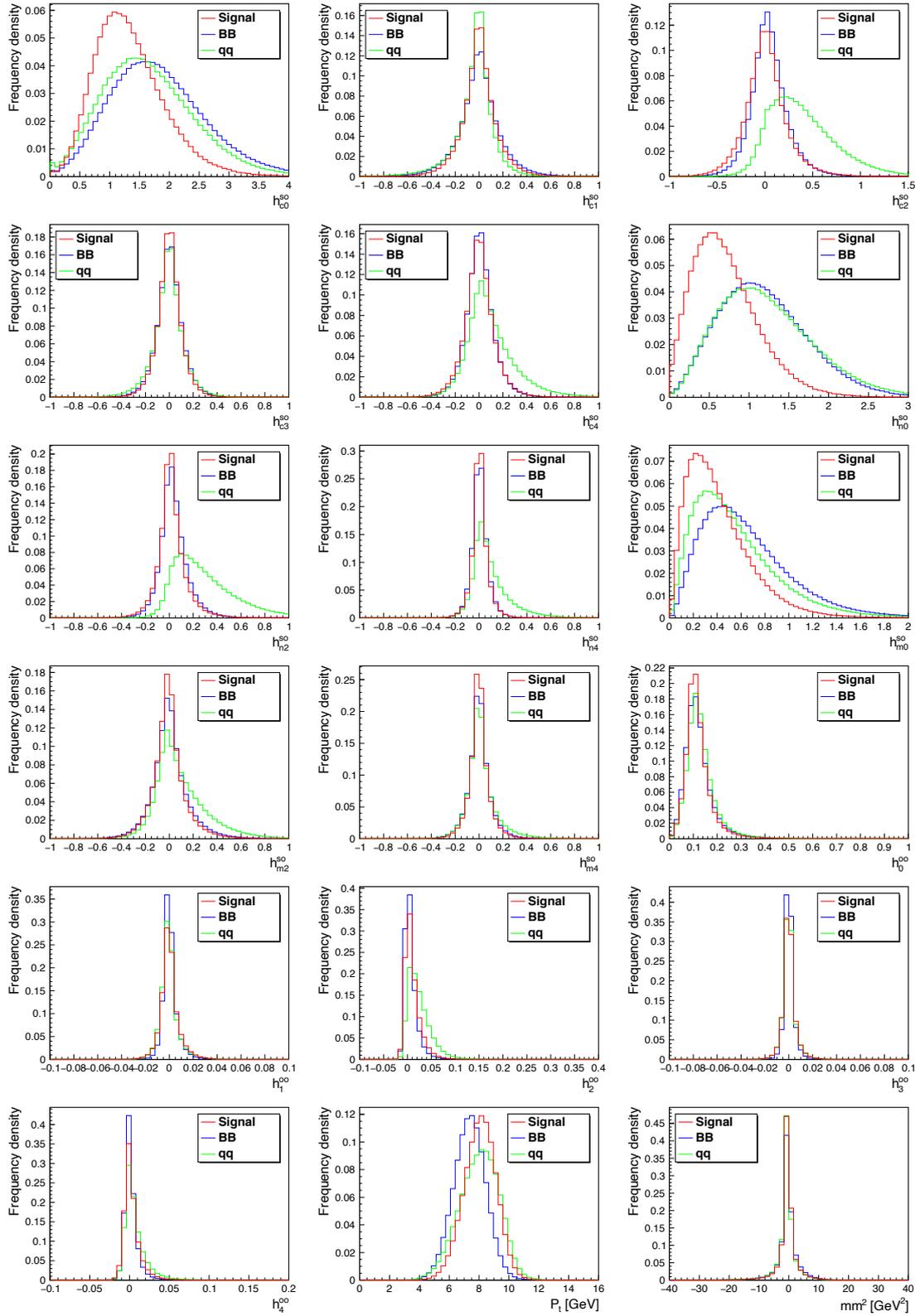
FIG 6.10: FastBDT input variables of $B \to X_s e^+ e^-$ (1) KSFW variables. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events.
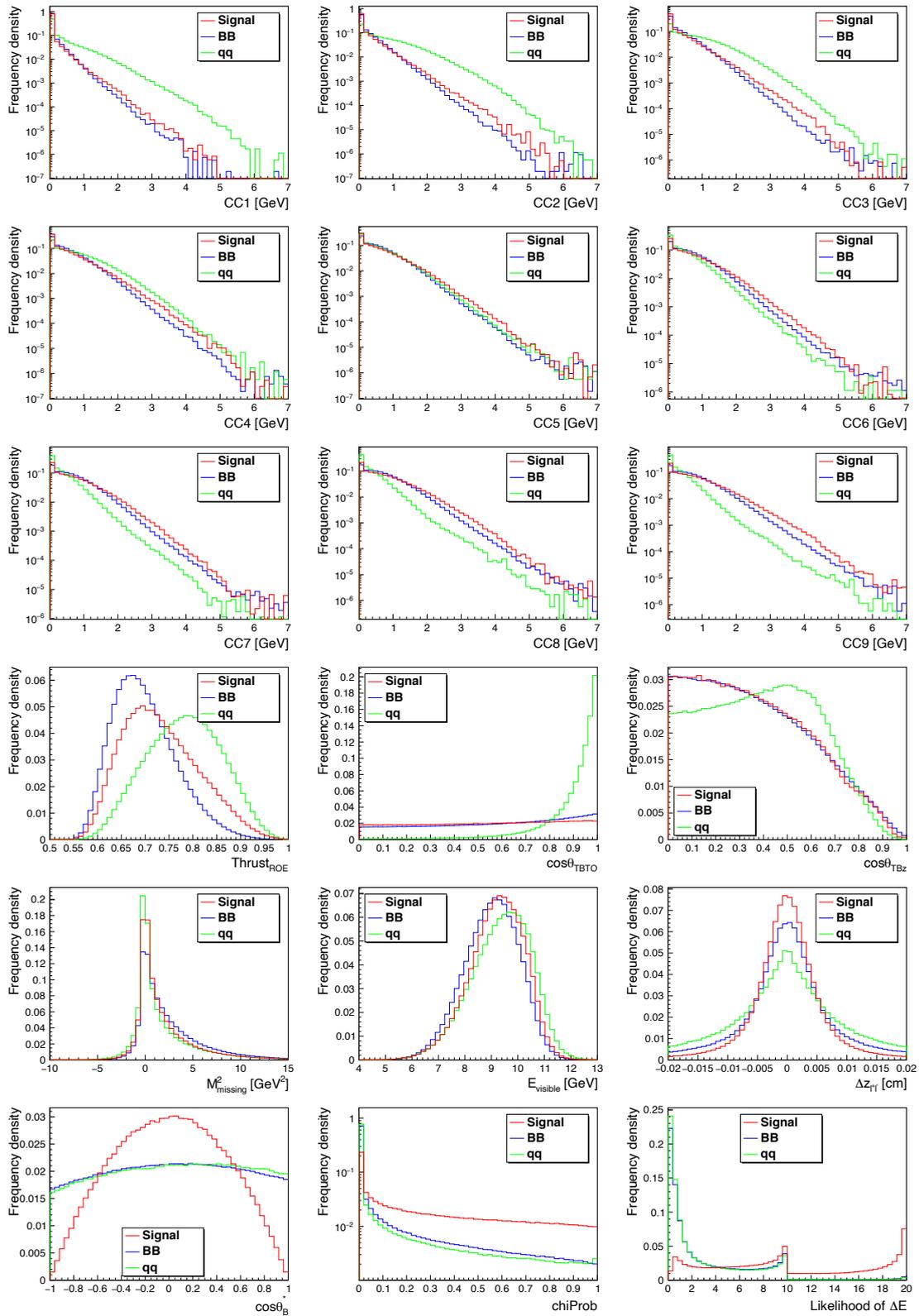
FIG 6.11: FastBDT input variables of $B \to X_s e^+ e^-$ (2) CLEO Cone variables and other variables. Red line shows the signal events, blue line shows background events from $B\overline{B}$ events, and green line shows background events from $q\overline{q}$ continuum events.
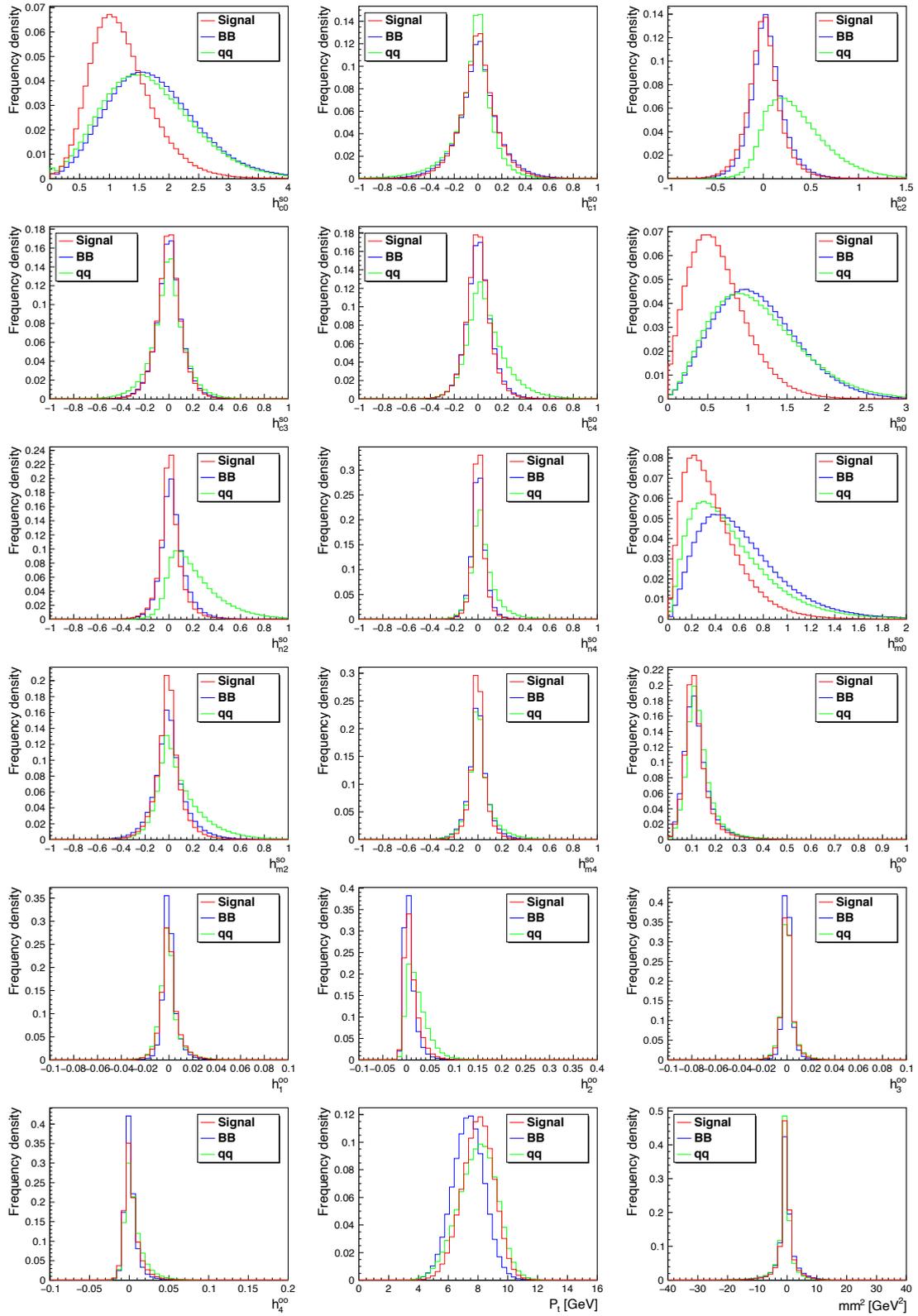
FIG 6.12: FastBDT input variables of $B \to X_s \mu^+ \mu^-$ (1) KSFW variables. Red line shows the signal events, blue line shows background events from $B\overline{B}$ events, and green line shows background events from $q\overline{q}$ continuum events.
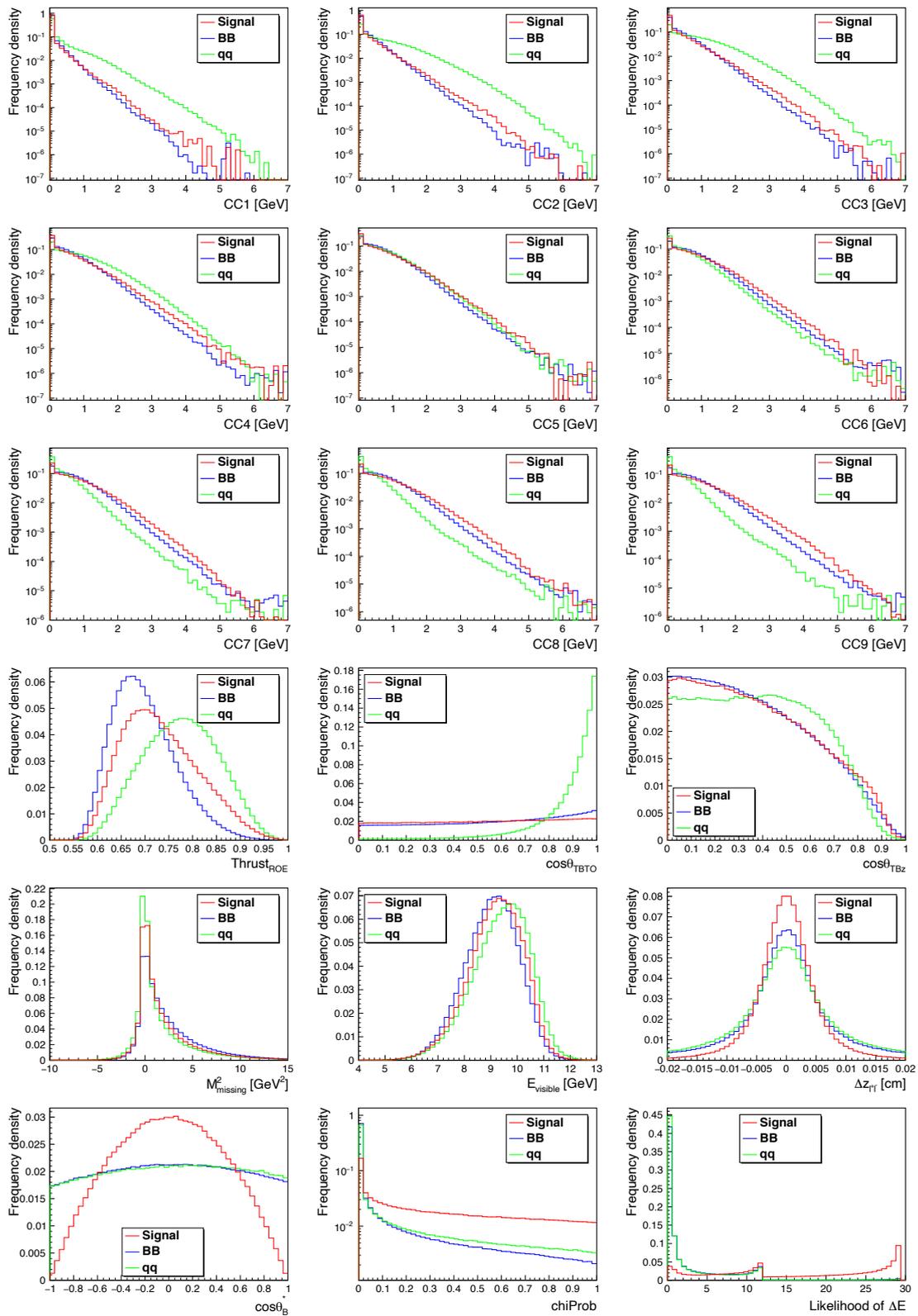
FIG 6.13: FastBDT input variables of $B \to X_s \mu^+ \mu^-$ (2) CLEO Cone variables and other variables. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events.

(a) $F^{B\bar{B}}_{e^+e^-,low}$

(b) $F^{q\bar{q}}_{e^+e^-,low}$

(c) $F^{B\bar{B}}_{e^+e^-,high}$

(d) $F^{q\bar{q}}_{e^+e^-,high}$

(e) $F^{B\bar{B}}_{\mu^+\mu^-,low}$

(f) $F^{q\bar{q}}_{\mu^+\mu^-,low}$

(g) $F^{B\bar{B}}_{\mu^+\mu^-,high}$

(h) $F^{q\bar{q}}_{\mu^+\mu^-,high}$
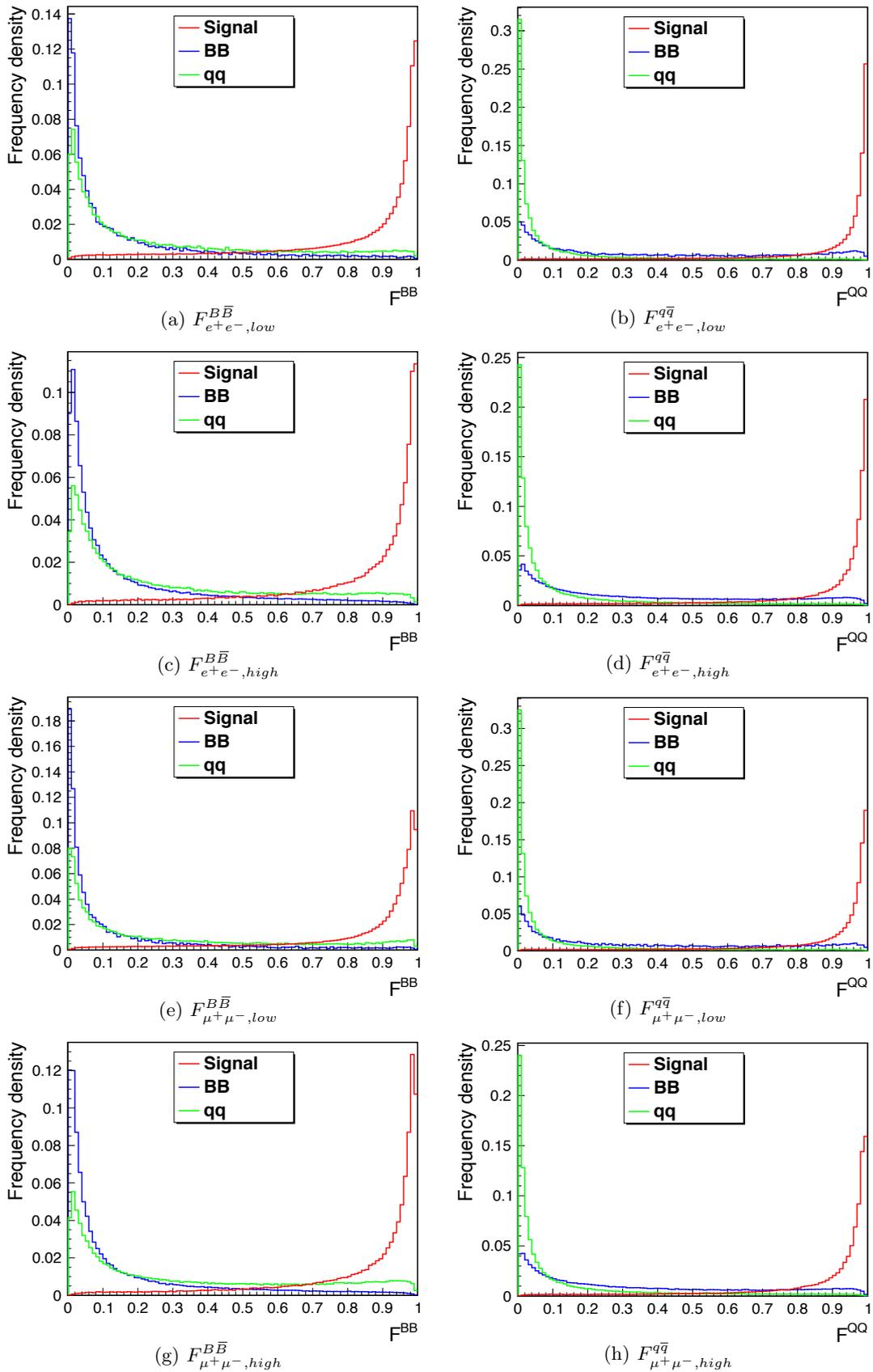
FIG 6.14: FastBDT output distributions. Red line shows the signal events, blue line shows background events from $B\bar{B}$ events, and green line shows background events from $q\bar{q}$ continuum events.

TABLE 6.3: Classification of samples for FastBDT training and criteria on the output value.

FastBDT output	Signal sample	Background sample	M_{X_s} range	Criteria
$F^{B\bar{B}}_{e^+e^-,low}$	$X_s e^+e^-$	$B\bar{B}$	$[0, 1.1]$	> 0.68
$F^{q\bar{q}}_{e^+e^-,low}$	$X_s e^+e^-$	$q\bar{q}$	$[0, 1.1]$	> 0.94
$F^{B\bar{B}}_{e^+e^-,high}$	$X_s e^+e^-$	$B\bar{B}$	$[1.1, 2.0]$	> 0.96
$F^{q\bar{q}}_{e^+e^-,high}$	$X_s e^+e^-$	$q\bar{q}$	$[1.1, 2.0]$	> 0.96
$F^{B\bar{B}}_{\mu^+\mu^-,low}$	$X_s \mu^+\mu^-$	$B\bar{B}$	$[0, 1.1]$	> 0.74
$F^{q\bar{q}}_{\mu^+\mu^-,low}$	$X_s \mu^+\mu^-$	$q\bar{q}$	$[0, 1.1]$	> 0.90
$F^{B\bar{B}}_{\mu^+\mu^-,high}$	$X_s \mu^+\mu^-$	$B\bar{B}$	$[1.1, 2.0]$	> 0.94
$F^{q\bar{q}}_{\mu^+\mu^-,high}$	$X_s \mu^+\mu^-$	$q\bar{q}$	$[1.1, 2.0]$	> 0.94

$34.6\ \mathrm{fb}^{-1}$.

TABLE 6.4: Cut flow table of $X_s e^+e^-$. Number of events satisfying each selection, signal efficiency and FOM are shown. Note that the M_{bc} cut, $5.27 < M_{\mathrm{bc}} < 5.29\,\mathrm{GeV}/c^2$ is not imposed. The numbers are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

Selection	Signal	$B\bar{B}$	$q\bar{q}$	Signal Efficiency	$S/\sqrt{S+B}$
Generated	318.7	3.63×10^7	1.29×10^8	100%	0.025
Reconstructed	40.00	9.89×10^4	3.84×10^4	12.55%	0.108
Vertex fit of B meson	38.78	7.91×10^4	2.72×10^4	12.17%	0.119
$M_{X_s} < 2.0$ GeV	37.95	5.76×10^4	2.12×10^4	11.91%	0.135
ΔE cuts	35.72	3.17×10^4	1.09×10^4	11.21%	0.173
$M_{e^+e^-} > 0.2$ GeV	29.22	3.16×10^4	7.68×10^3	9.167%	0.147
Charmonium veto	19.54	8.67×10^3	5.46×10^3	6.131%	0.164
D veto	18.16	6.45×10^3	4.58×10^3	5.697%	0.173
FastBDT	7.836	23.5	11.1	2.458%	1.20
Best candidate selection	7.705	21.5	10.0	2.418%	1.23

6.3   Peaking Backgrounds

To obtain the signal yields, the peaking backgrounds, which make a peak at $M_{\mathrm{bc}} = 5.28\,\mathrm{GeV}/c^2$, should be estimated correctly. Three sources of peaking backgrounds are considered in this study; namely, (i) Double mis-ID background, (ii) Swapped mis-ID background and (iii) Charmonium background.

6.3.1   Double mis-ID background

Hadronic B decay events, such as $B \to K\pi\pi$, can pass the background suppression if two charged hadrons are mis-identified as two leptons. These backgrounds are denoted as double mis-ID backgrounds. The double mis-ID background is estimated from $X_s h^+h^-$ events of data by applying a weight of the mis-ID rate from hadron to lepton, where h denotes a hadron. The $X_s h^+h^-$ events are reconstructed with the method except for the PID selection on the two leptons. Events passing the background suppression criteria are weighted with the mis-ID rate from hadron to lepton, f_ℓ, for each lepton. The weight factor w is

$$w = f_{\ell^+} \cdot f_{\ell^-}. \tag{6.24}$$

where f_ℓ is the mis-ID rate from hadron to lepton. The mis-ID rate is evaluated in the Belle II performance study as described in Section 3.4.

TABLE 6.5: Cut flow table of $X_s\mu^+\mu^-$. Number of events satisfying each selection, signal efficiency and FOM are shown. Note that the M_{bc} cut, $5.27 < M_{bc} < 5.29\,\mathrm{GeV}/c^2$ is not imposed. The numbers are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

Selection	Signal	$B\bar{B}$	$q\bar{q}$	Signal Efficiency	$S/\sqrt{S+B}$
Generated	316.1	3.63×10^7	1.29×10^8	100%	0.025
Reconstructed	32.31	1.13×10^5	4.68×10^4	10.22%	0.081
Vertex fit of B meson	31.61	9.01×10^4	3.53×10^4	10.00%	0.089
$M_{X_s} < 2.0$ GeV	31.11	6.69×10^4	2.74×10^4	9.842%	0.101
ΔE cuts	28.41	2.45×10^4	9.14×10^3	8.987%	0.155
Charmonium veto	21.19	8.02×10^3	7.39×10^3	6.703%	0.171
D veto	19.75	5.94×10^3	6.17×10^3	6.249%	0.179
FastBDT	9.821	38.1	37.4	3.107%	1.06
Best candidate selection	9.731	36.0	35.6	3.078%	1.08

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$

(c) $X_s \ell^+ \ell^-$
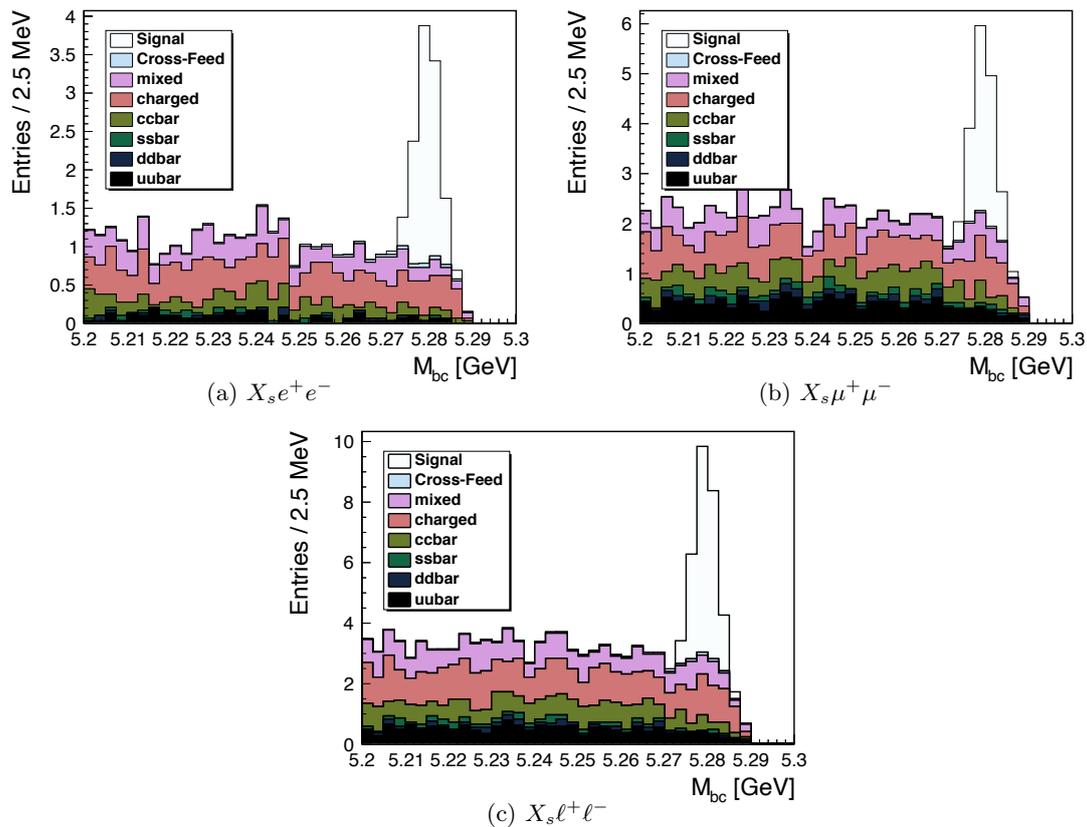
FIG 6.15: M_{bc} distributions after the background suppression. The colors indicate different MC samples. White histogram shows signal events. The histograms are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

The numbers of estimated events are 0.11 ± 0.08 and 1.3 ± 1.1 for $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$, respectively, while the numbers of events expected from the simulation are 0.10 ± 0.06 and 1.70 ± 0.24. The estimated and expected numbers of events are consistent within uncertainties. Figure 6.16 shows the M_{bc} distributions of the double mis-ID backgrounds of the MC samples.
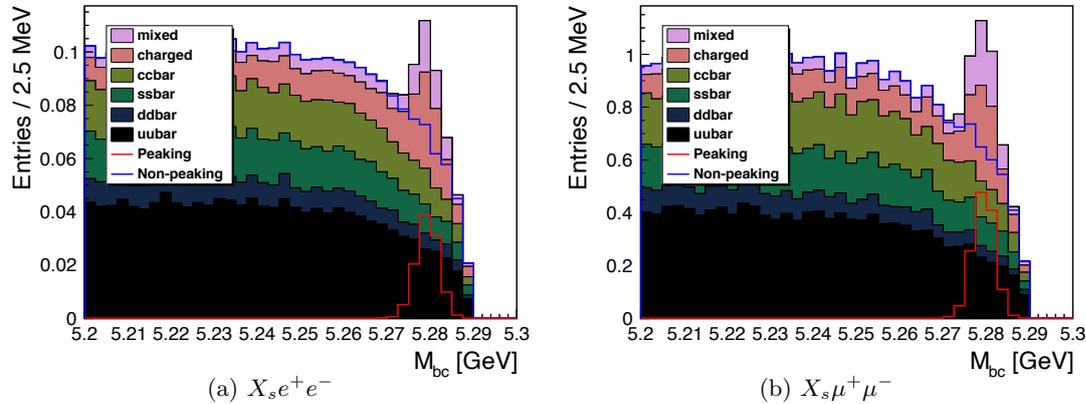
FIG 6.16: Estimated M_{bc} distributions of the double mis-ID backgrounds. MC samples are identified by color code. Red line shows total peaking component and blue shows non-peaking component. The histograms are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

6.3.2   Swapped mis-ID background

The $X_s J/\psi (\to \ell^+ \ell^-)$ events can pass the event selection if a lepton is mis-identified as a hadron and a hadron which is daughter of X_s is mis-identified as a lepton. These backgrounds are denoted as swapped mis-ID backgrounds. The swapped mis-ID background is estimated from $X_s J/\psi$, $X_s \psi(2S)$ events of data. Firstly, $X_s J/\psi$, $X_s \psi(2S)$ events are reconstructed and selected by the Charmonium veto (Section 6.2.2). Then kinematics of a candidate are recalculated swapping a lepton and a hadron. The background suppression is applied on the re-calculated kinematics and events passing the selection are weighted by the factor as defined in the following equation.

$$ w = \frac{f_\ell}{\epsilon_h} \cdot \frac{f_h}{\epsilon_\ell}. \tag{6.25} $$

where ϵ_ℓ is the lepton-ID efficiency, f_h is the mis-ID rate from lepton to hadron, and ϵ_h is the hadron-ID efficiency. The mis-ID rate and identification efficiency are evaluated in the Belle II performance studies as described in Section 3.4.

The estimation of the swapped mis-ID backgrounds is validated using MC samples. The numbers of estimated events are 0.006 ± 0.004 and 0.56 ± 0.50 for $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$, respectively, while the numbers of events expected from the simulation are 0 and 0.41 ± 0.12. The estimated and expected numbers of events are consistent within uncertainties. Figure 6.17 shows the M_{bc} distributions of the swapped mis-ID backgrounds estimated with the MC samples.

6.3.3   Charmonium background

Although the most of $X_s J/\psi$, $X_s \psi(2S)$ events are rejected by the Charmonium veto in Section 6.2.2, contamination from these events is unavoidable. Since it is difficult to estimate the background from data, another set of the generic MC samples are used. Events are reconstructed and selected with the usual method and the peaking backgrounds are chosen with the MC-truth information. If two lepton's mother is J/ψ and $\psi(2S)$, the events are recognized as the Charmonium background. The numbers of estimated events in whole M_{bc} range are 3.49 ± 0.35 and 2.01 ± 0.26 for $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$, respectively.

Figure 6.18 shows the M_{bc} distributions of the Charmonium background estimated from MC samples.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
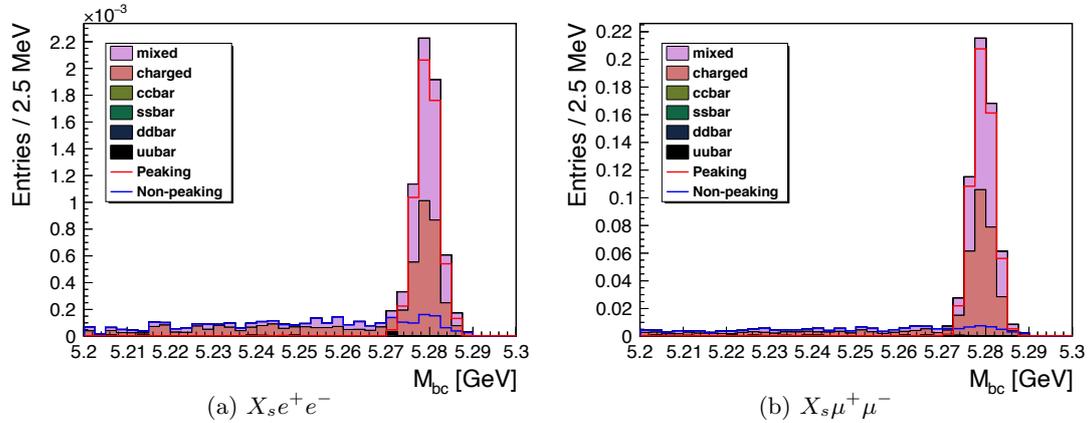
FIG 6.17: Estimated M_{bc} distributions of the swapped mis-ID backgrounds. MC samples are identified by color code. Red line shows total peaking component and blue shows non-peaking component. The histograms are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
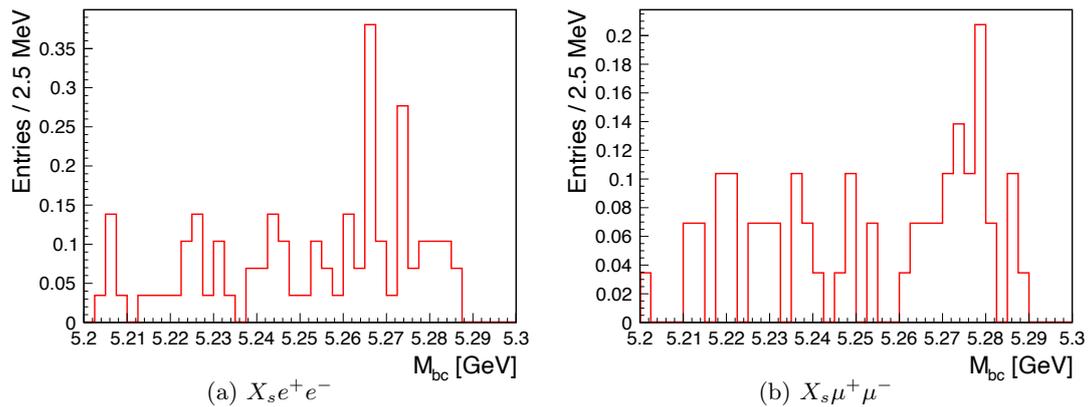
FIG 6.18: Estimated M_{bc} distributions of the Charmonium backgrounds. The histograms are scaled for an integrated luminosity of $34.6\,\mathrm{fb}^{-1}$.

Chapter 7

Extraction of the branching fraction

The signal yield N_{signal} is extracted from the M_{bc} distributions using the extended maximum likelihood fit. Then, the branching fraction is calculated with the following function.

$$\mathcal{B}(B \to X_s \ell^+ \ell^-) = \frac{N_{\mathrm{signal}}}{2 N_{B\bar{B}} \times \epsilon} \tag{7.1}$$

where $N_{B\bar{B}}$ is the number of B meson pairs and ϵ is the reconstruction efficiency of $B \to X_s \ell^+ \ell^-$. Actually, instead of the signal yields, the branching fraction is used as the floating parameter in the fitting. The branching fraction of $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$ is measured separately. Moreover, the branching fraction of $B \to X_s \ell^+ \ell^-$ is calculated from the simultaneous fitting of the M_{bc} distributions of $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$ assuming the lepton flavor universality.

7.1 Probability Density Function (PDF)

The likelihood function $\mathcal{L}$ for the extended maximum likelihood fit it expressed as the following function.

$$\mathcal{L} = \frac{\exp\left(-\sum_j N_j\right)}{N!} \prod_i^N \left(\sum_j N_j P_j\right) \tag{7.2}$$

where i runs over all events, j runs over the categories of events, N_j is the yield of j-th category, N is the total number of events, and P_j is the j-th probability density function.

For the analysis, six categories of the probability density function are considered, (i) Signal, (ii) Self cross-feed, (iii) Non-peaking backgrounds, and (iv - vi) Three peaking backgrounds. The functions and parameters are summarized in TABLE 7.1. Each component is described in the following sections.

TABLE 7.1: Summary of the probability functions and parameters.

Component (notation)	Function	Parameters (fix or float)
Signal (sig)	Gaussian	Branching fraction (Yield) : float
		Shape parameters : fix
Self cross-feed (scf)	histogram PDF	N_{scf}/N_{sig} : fix
Non-peaking background (bkg)	ARGUS function	Yield : float
		Shape : float
		End point : fix
Peaking background (pkg)	histogram PDF	Yield : fix

7.1.1 Signal

The signal PDF is modeled by a Gaussian. Mean (μ) and width (σ) are defined by fitting the M_{bc} distributions of $X_s J/\psi$ control samples in data as a Gaussian and an ARGUS function. The control

samples are reconstructed in the same way of $X_s\ell^+\ell^-$ except for the Charmonium veto. The veto condition for the J/ψ is flipped to select the events. Figure 7.1 shows the M_{bc} distributions of the control samples with the fitting function. In the fitting, the end point parameter of the ARGUS function is fixed at the value defined in the Section 7.1.3. TABLE 7.2 shows the shape parameters obtained by the fitting.

(a) $X_s e^+ e^-$
(b) $X_s \mu^+ \mu^-$
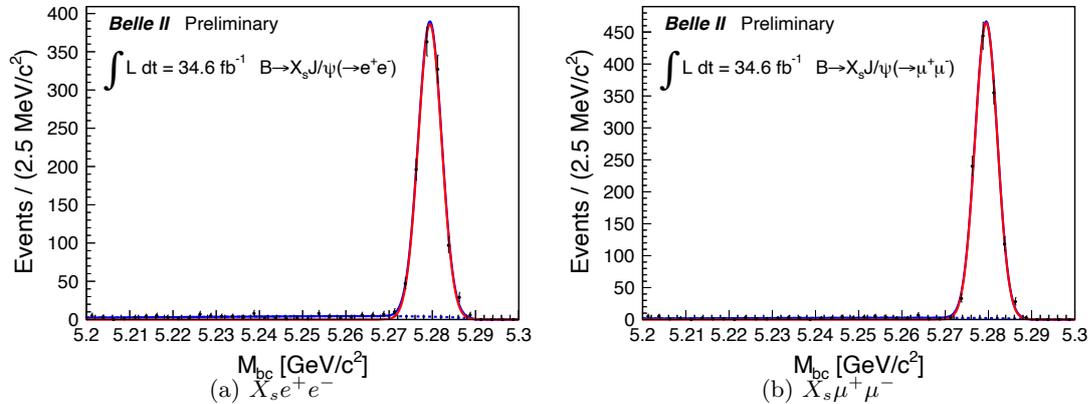
FIG 7.1: M_{bc} distributions of the $B \to X_s J/\psi$ control samples in data. The fit contained the following components: a Gaussian for the signal (red line) and an ARGUS function to model background from the continuum and combinatorial B decays (dashed blue line).

TABLE 7.2: Shape parameters of the signal PDF.

Mode	Parameter	Value
$B \to X_s e^+ e^-$	Mean (μ)	5.279424 ± 0.000089 GeV
	Width (σ)	2.679 ± 0.076 MeV
$B \to X_s \mu^+ \mu^-$	Mean (μ)	5.279531 ± 0.000080 GeV
	Width (σ)	2.602 ± 0.062 MeV

7.1.2   Self cross-feed

The events originating from $B \to X_s\ell^+\ell^-$ which are wrongly reconstructed, for example mis-identification of $K^+\pi^-$ as π^+K^-, are denoted as the self cross-feed. The function of the self cross-feed is constructed from the signal MC samples. Yield of the self cross-feed should be proportional to the signal. The ratio of the self cross-feed to the signal is fixed to the value which are estimated by the simulation.

Figure 7.2 shows the histogram PDF of the self cross-feed. TABLE 7.3 shows the ratio of self cross-feed to the signal.

TABLE 7.3: The ratio of the self cross-feed to the signal estimated by the simulation.

Mode	Parameter	Value
$B \to X_s e^+ e^-$	Ratio (N_{scf}/N_{sig})	0.1211
$B \to X_s \mu^+ \mu^-$	Ratio (N_{scf}/N_{sig})	0.0786

7.1.3   Non-peaking background

The continuum backgrounds and some $B\overline{B}$ backgrounds have not a peak on the M_{bc} distribution. These backgrounds are denoted as the non-peaking background. PDF of the non-peaking backgrounds is mod-

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$
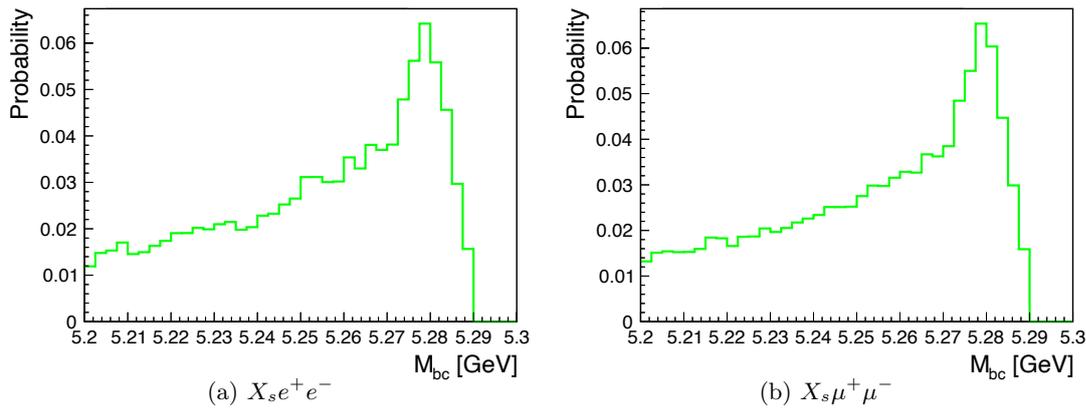
FIG 7.2: M_{bc} PDF of the self cross-feed.

eled by an ARGUS function:

$$f_{\mathrm{ARGUS}}(M_{\mathrm{bc}}) \propto M_{\mathrm{bc}} \cdot \sqrt{t} \cdot \exp(-at), \tag{7.3}$$

$$t \equiv 1 - \left(\frac{M_{\mathrm{bc}}}{E_{beam}}\right)^2. \tag{7.4}$$

where E_{beam} is an endpoint parameter and a is a shape parameter. The endpoint of the ARGUS function is obtained by fitting the control samples of $B^0 \to D^- \pi^+, D^- \to K^+ \pi^- \pi^-$ and $B^+ \to \overline{D}^0 \pi^+, \overline{D}^0 \to K^+ \pi^-$. The particle selection of $K^\pm$ and $\pi^\pm$ is same for the signal events. The D meson is required to have nominal mass with the following criteria; $1.85 < M < 1.89$ GeV. The B meson is selected by the following criteria; $-0.15 < \Delta E < 0.15$ GeV and $5.2 < M_{\mathrm{bc}} < 5.3$ GeV. TABLE 7.4 summarizes the selection criteria of the control samples.

TABLE 7.4: Summary of the selection criteria for the $B \to D\pi$ control samples.

Particle	Selection Criteria
$K^\pm$	$dr < 0.5\,\mathrm{cm}, |dz| < 2.0\,\mathrm{cm}, PID_K > 0.6$, nCDCHits$> 20$
$\pi^\pm$	$dr < 0.5\,\mathrm{cm}, |dz| < 2.0\,\mathrm{cm}, PID_\pi > 0.6$, nCDCHits$> 20$
D	$1.85 < M < 1.89$ GeV/c^2
B	$-0.15 < \Delta E < 0.15\,\mathrm{GeV}, 5.2 < M_{\mathrm{bc}} < 5.3\,\mathrm{GeV}/c^2$

Figure 7.3 shows the M_{bc} distributions of the control samples with the fitting function. TABLE 7.5 shows the end point parameter obtained by the fitting.

TABLE 7.5: Shape parameters of the background PDF.

Parameter	Value
End point	5.28973 ± 0.00015 GeV

7.1.4 Peaking background

Three sources of the peaking backgrounds are considered as explained in Section 6.3: (i) Double mis-ID background, (ii) Swapped mis-ID background and (iii) Charmonium background.

Figure 7.4 - 7.6 show the histogram PDF. The shape and yield of (i) and (ii) are estimated from data and these of (iii) are estimated from the MC samples. The yields of these peaking backgrounds are summarized in TABLE 7.6.
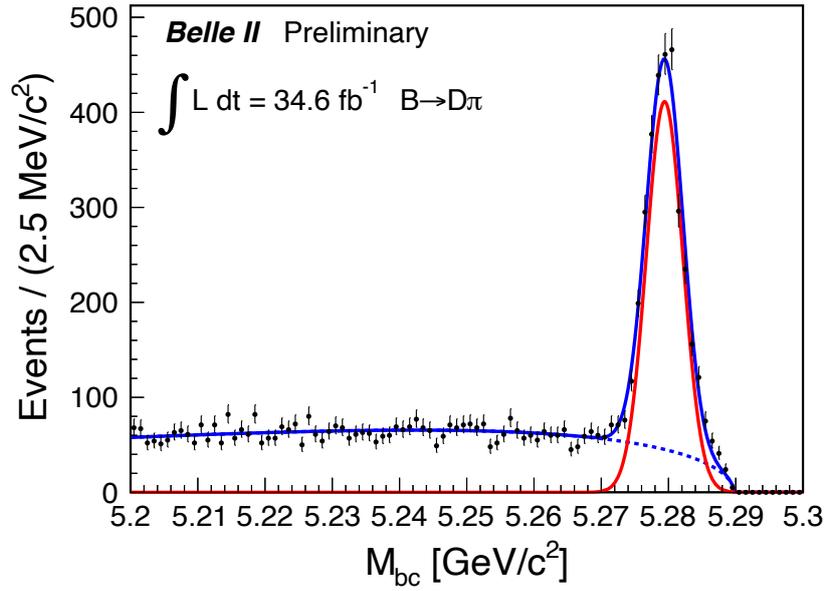
FIG 7.3: M_{bc} distributions of the $B \to D\pi$ control samples in data. The fit contained the following components: a Gaussian for the signal (red line) and an ARGUS function to model background from the continuum and combinatorial B decays (dashed blue line).

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$
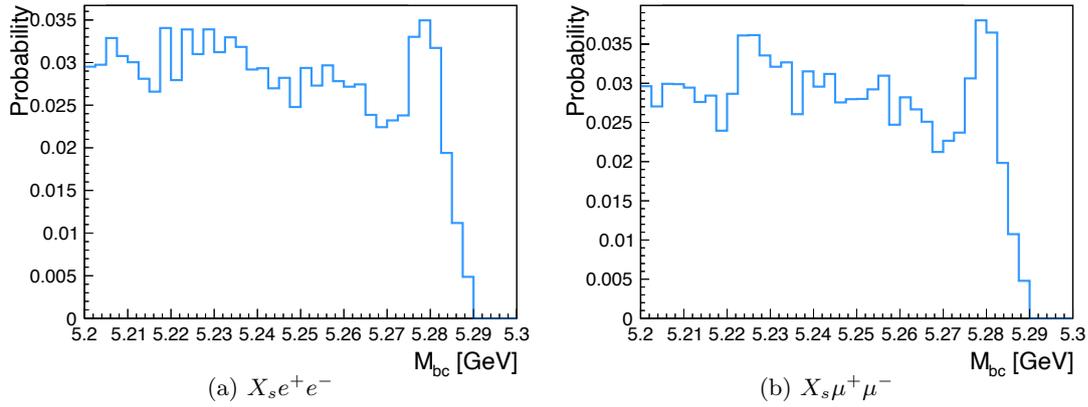
FIG 7.4: M_{bc} PDF of the double mis-ID background.

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$
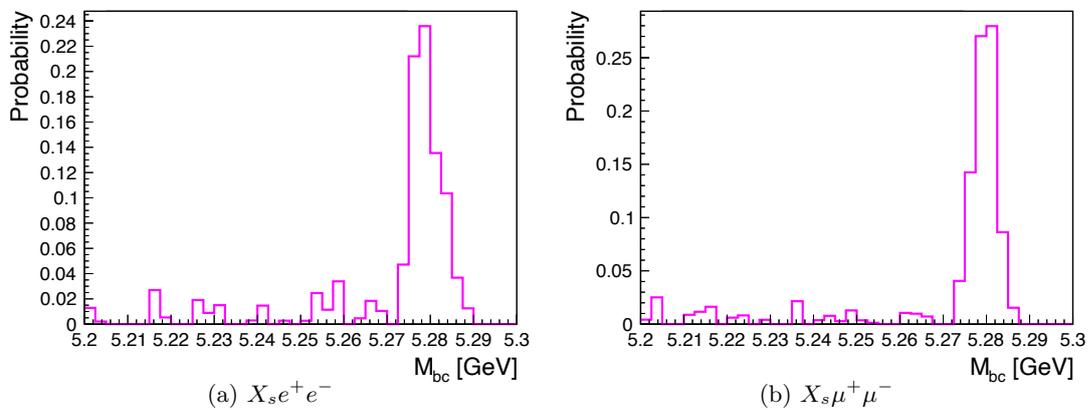
FIG 7.5: M_{bc} PDF of the swapped mis-ID background.

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$
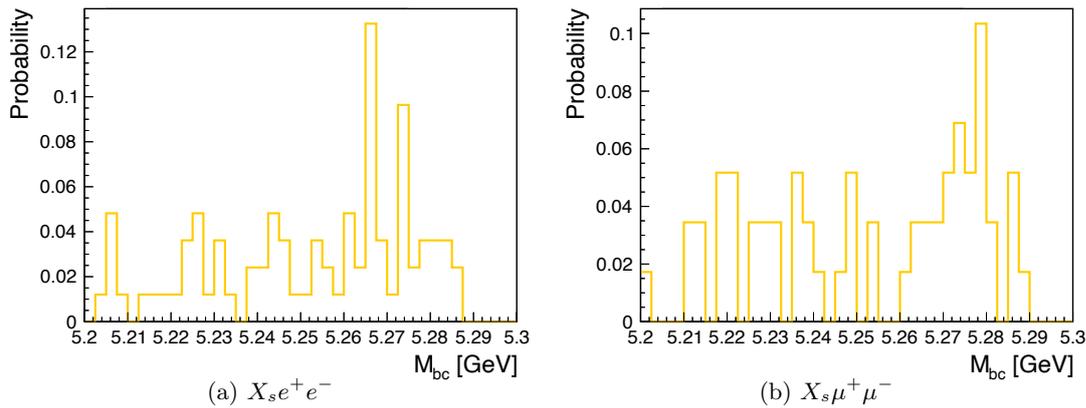
FIG 7.6: $M_{\rm bc}$ PDF of the Charmonium background.

TABLE 7.6: The yields of the peaking backgrounds.

Mode	Parameter	Value
$B \to X_s e^+ e^-$	Yield of the double mis-ID	11.7 ± 3.1
	Yield of the swapped mis-ID	0.015 ± 0.006
	Yield of the Charmonium	2.87 ± 0.32
$B \to X_s \mu^+ \mu^-$	Yield of the double mis-ID	22.0 ± 4.0
	Yield of the swapped mis-ID	1.16 ± 0.19
	Yield of the Charmonium	2.01 ± 0.26

7.2 Fitter Check

The validation of the fitter is performed with a toy MC test. A pull distribution for a toy MC test is useful sign to check relevance of the fitting method. A pull value of the signal yield is defined as following calculation;

$$pull = \frac{N_{input} - N_{observed}}{\delta N}. \tag{7.5}$$

where N_{input} is a number of generated events, $N_{observed}$ is an extracted number of events from fitting, and δN is an error of fitted parameter $N_{observed}$. When a pull distribution is fitted by Gaussian, a result with mean equal to 0 and width equal to 1 shows a relevance of fitting. The test samples are generated from the PDFs by fluctuating the number of each component with a Poisson distribution around the expected number of events. 1000 MC samples are produce to make the pull distribution. Figure 7.7 shows pull distributions of signal yields. The fit results of mean μ_{pull} and width σ_{pull} are as followings,

$$\mu_{pull, X_s e^+ e^- - mode} = -0.1985 \pm 0.035, \tag{7.6}$$

$$\sigma_{pull, X_s e^+ e^- - mode} = 1.096 \pm 0.025, \tag{7.7}$$

$$\mu_{pull, X_s \mu^+ \mu^- - mode} = -0.1575 \pm 0.033, \tag{7.8}$$

$$\sigma_{pull, X_s \mu^+ \mu^- - mode} = 1.047 \pm 0.023, \tag{7.9}$$

$$\tag{7.10}$$

The obtained mean is significantly lower than 0 and the obtained width is also larger than unity. This indicates that the fitter might induce a bias on the number of signal events. These effects are included in the systematic uncertainty that is discussed in the following section (8.3).

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
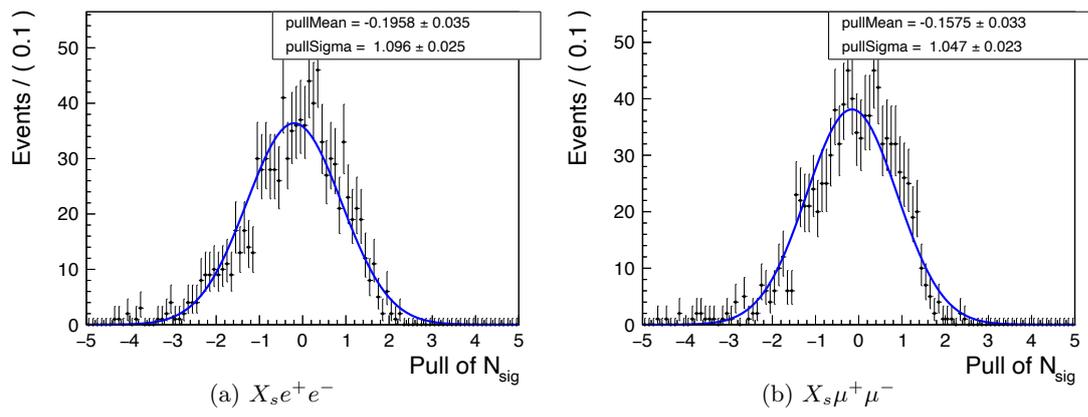
FIG 7.7: Pull distributions of signal yields.

Chapter 8

Systematic uncertainty

8.1 Number of B meson pairs

The number of B meson pairs are estimated from the number of hadronic events at the $\Upsilon(4S)$ resonance operation subtracting the number at the out of $\Upsilon(4S)$ resonance operation. The hadronic events, such as $B\overline{B}$ events and continuum events, can be distinguished from QED events, such as Bhabha and two-photon events, by requiring the number of tracks and clusters. The continuum events have similar cross-section in both on the $\Upsilon(4S)$ resonance and out of the resonance, while the cross-section of $B\overline{B}$ events drastically decrease out of the $\Upsilon(4S)$ resonance. The number of $B\overline{B}$ events $N_{B\overline{B}}$ are estimated using following equation,

$$N_{B\overline{B}} = \frac{N_{\text{had}}^{\text{on-res}} - R_{\text{lumi}} \times N_{\text{had}}^{\text{off-res}} \times k}{\epsilon_{B\overline{B}}}. \tag{8.1}$$

where $N_{\text{had}}^{\text{on-res}}$ is the number of hadronic events in on-resonance of $\Upsilon(4S)$, $N_{\text{had}}^{\text{off-res}}$ is the number of hadronic events in off-resonance $\Upsilon(4S)$ ($\sqrt{s} = 10.519\,\text{GeV}$), R_{lumi} is the luminosity ratio between the on-resonance data and off-resonance data, k is the correction factor of non-$B\overline{B}$ event cross-section for different collision energy, and $\epsilon_{B\overline{B}}$ is the selection efficiency of hadronic events. The number of B meson pairs in the data set has been determined to be $N_{B\overline{B}} = (37.7 \pm 0.6) \times 10^6$ [70].

8.2 Efficiency correction

The reconstruction efficiency estimated with the MC simulation is 2.418% for $X_s e^+ e^-$ and 3.078% for $X_s \mu^+ \mu^-$. Since there are discrepancies between data and MC on the selection efficiency for each particle, the reconstruction efficiency should be corrected. The correction factors for the particle selections are estimated with data-driven analyses. The uncertainties on the correction factors are propagated to the systematic uncertainty.

8.2.1 Charged track reconstruction efficiency

The track reconstruction efficiency in high momentum region ($p_T > 200\,\text{MeV}/c$) is evaluated using $e^- e^- \to \tau^+ \tau^-$ events. The τ-pair production has large cross section at the $\Upsilon(4S)$ resonance energy and provides good opportunity to investigate the tracking performance at Belle II. The target process is $e^- e^- \to \tau^+ \tau^-$ where one tau lepton decays leptonically ($\tau \to \ell \nu_\ell \nu_\tau$) while the other decays hadronically into three charged pions ($\tau \to 3\pi^{\pm}\nu_\tau + n\pi^0$). The τ-pair events are tagged from three good quality tracks. Then the existence of an additional track is inferred. The tracking efficiency is calculated from the fraction of the number of 4-tracks events over 3-tracks + 4-tracks events. The tracking efficiency evaluated with data is consistent with that of MC within the uncertainty of 0.80%. The systematic uncertainty of 0.80% is assigned on each track [71].

The tracking efficiency of low momentum track ($p_T < 200\,\text{MeV}/c$) is investigated using slow-pion decayed from D^*; the efficiency is estimated with $B \to D^*\pi$ and $B \to D^*\rho$. The slow tracking efficiency on data is consistent with that of MC within the uncertainty of 9.87%. For each slow track, the uncertainty of 9.87% is assigned.

The total uncertainties on the $X_s\ell^+\ell^-$ reconstruction efficiency due to the tracking efficiency is 3.8% for $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$.

8.2.2   Lepton identification efficiency

The lepton identification efficiency and its ratio between data and MC are evaluated using calibration samples as functions of momentum and polar angle of track. The detail is described in Section 3.4. The lepton identification efficiency correction factor on the $X_s\ell^+\ell^-$ reconstruction efficiency is $(96.3^{+2.7}_{-2.1})\%$ for $X_s e^+ e^-$ and $(85.7^{+4.7}_{-2.5})\%$ for $X_s \mu^+ \mu^-$.

8.2.3   Hadron $(K^\pm, \pi^\pm)$ identification efficiency

The hadron identification performance is studied as described in Section 3.4. The correction factors between data and MC are evaluated as functions of momentum and polar angle [15]. The reconstruction efficiency of $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$ is corrected by factor of $98.3 \pm 1.2\%$ and $98.5 \pm 1.3\%$ due to kaon and $98.0 \pm 0.8\%$ and $98.1 \pm 0.8\%$ due to pion, respectively.

8.2.4 K_S^0 reconstruction efficiency

The K_S^0 reconstruction efficiency is evaluated as function of distance between the interaction point and the vertex position of K_S^0. There is no strong deviation from unity on the efficiency ratio between data and MC. The systematic uncertainty is assigned on each K_S^0 candidate depending on the vertex distance.. The total uncertainty of 1.1% and 1.0% is assigned on the $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$ reconstruction efficiency, respectively.

8.2.5 π^0 reconstruction efficiency

The correction factor of π^0 reconstruction efficiency is estimated by using $\eta \to \gamma\gamma$ and $\eta \to 3\pi^0$. By assuming the data-MC efficiency ratio of $\pi^0 \to \gamma\gamma$ and that of $\eta \to \gamma\gamma$ is same, the π^0 reconstruction efficiency is extracted as follows;

$$
\begin{aligned}
\frac{\epsilon_{\text{data}}(2\pi^0)}{\epsilon_{\text{MC}}(2\pi^0)} &= \frac{N_{\text{data}}(\eta \to 3\pi^0)/N_{\text{MC}}(\eta \to 3\pi^0)}{N_{\text{data}}(\eta \to \gamma\gamma)/N_{\text{MC}}(\eta \to \gamma\gamma)}, \\
\frac{\epsilon_{\text{data}}(\pi^0)}{\epsilon_{\text{MC}}(\pi^0)} &= \sqrt{\frac{\epsilon_{\text{data}}(2\pi^0)}{\epsilon_{\text{MC}}(2\pi^0)}}.
\end{aligned}
\tag{8.2}
$$

For each π^0 candidate, the correction factor of $93.2 \pm 3.4\%$ is assigned. In total, the $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$ efficiency is corrected by factor of $99.5 \pm 0.2\%$ and $99.8 \pm 0.1\%$, respectively.

8.2.6   FastBDT selection efficiency

The FastBDT is trained to suppress large backgrounds by using MC samples. Even though there are no large difference on input variables between data and MC, the efficiency correction should be evaluated. The FastBDT efficiency is evaluated from $B \to X_s J/\psi(\to \ell^+\ell^-)$ samples. By fitting the M_{bc} distribution with Gaussian and Argus function before and after the FastBDT selection, the efficiency is calculated. The efficiency correction factor is $107.2 \pm 4.4\%$ for $X_s e^+ e^-$ and $103.7 \pm 4.0\%$ for $X_s \mu^+ \mu^-$.

8.2.7   Summary of the efficiency correction

The efficiency correction factors from each particle selection is summarized in TABLE 8.1.

8.3   Fitter bias

The systematic uncertainty due to the fitter bias on the signal yields is estimated from the pull distribution and the linearity check in Section 7.2. The shift of the mean of the pull distribution would indicate the bias on the signal yield and the large width of the pull would indicate that the statistical uncertainty is

TABLE 8.1: The efficiency correction factors on the $B \to X_s \ell^+ \ell^-$ reconstruction.

Source	$B \to X_s e^+ e^-$ [%]	$B \to X_s \mu^+ \mu^-$ [%]
Tracking efficiency	100 ± 3.8	100 ± 3.8
Electron ID	$96.3^{+2.7}_{-2.1}$	-
Muon ID	-	$85.7^{+4.7}_{-2.5}$
Kaon ID	98.3 ± 1.2	98.5 ± 1.3
Pion ID	98.0 ± 0.8	98.1 ± 0.8
K_s^0 reconstruction	100 ± 1.1	100 ± 1.0
π^0 reconstruction	99.5 ± 0.2	99.8 ± 0.1
FastBDT efficiency	107.2 ± 4.4	103.7 ± 4.0

under estimated. The systematic uncertainty due to the pull mean $\sigma_{\text{syst mean}}$, and that due to the pull width, $\sigma_{\text{syst width}}$ are evaluated by the following equation.

$$\sigma_{\text{syst mean}} = \sigma_{\text{stat}} \cdot \mu_{pull}, \tag{8.3}$$

$$\sigma_{\text{syst width}} = \sigma_{\text{stat}} \cdot (\sigma_{pull} - 1.). \tag{8.4}$$

where μ_{pull} and σ_{pull} are mean and width of the pull distribution and σ_{stat} is the statistical uncertainty.

8.4 PDF uncertainty

The systematic uncertainties from the fixed parameters in the PDF are estimated by varying the parameters by these uncertainties.

8.4.1 Uncertainty of signal shape

The signal PDF is modeled by a Gaussian and the shape parameters are fixed. The systematic uncertainty due to fixed shape parameters are estimated by varying them by $\pm 1\sigma$.

8.4.2 Uncertainty of self cross-feed ratio

The self cross-feed PDF is modeled by a histogram PDF estimated with the MC samples. The systematic uncertainty due to the self cross-feed is estimated by varying the ratio of the self cross-feed and the signal yield by 100%.

8.4.3 Uncertainty of peaking background yields

The double mis-ID and swapped mis-ID backgrounds are estimated using the mis-ID probabilities and PID efficiencies. To take into account the uncertainties of the mis-ID probabilities and PID efficiencies, the uncertainties of the yields are calculated. Then the systematic uncertainties are estimated by varying the yields by calculated uncertainties.

The Charmonium background is estimated using the MC samples. The systematic uncertainty due to the Charmonium backgrounds is estimated by varying the yield by 100%, conservatively.

8.5 Signal modeling of non-resonant X_s

The $B \to X_s \ell^+ \ell^-$ samples are generated with EVTGEN [54], PYTHIA [55] [56], and PHOTOS [57]. The non-resonant $X_s \ell^+ \ell^-$ samples rely on the EvtBtoXsll decay model which is based on the following papers [60] [61] [35] [62] [63]. The decay model has several parameters, K^*-X_s transition point, b-quark mass, and the Fermi motion momentum. These parameters might change the di-lepton mass distribution and X_s mass distribution and thus the reconstruction efficiency. The systematic uncertainty is estimated by varying these parameters within proper range. The hadronization of X_s relies on the PYTHIA. The fragmentation of X_s modes are corrected according to the Belle result.

8.5.1 K^*-X_s transition point

To estimate the reconstruction efficiency, the M_{X_s} is required to be greater than 1.1 GeV/c^2 for the non-resonant $X_s\ell^+\ell^-$ samples. The value of the point is same with the previous study [2]. The transition position is varied by ±0.1 GeV/c^2 and then the reconstruction efficiency is recalculated. The systematic uncertainty to the branching fraction is provided from the range of reconstruction efficiency.

8.5.2 b-quark mass

The b-quark mass, m_b, is assumed to be 4.8 GeV/c^2 to generate the $X_s\ell^+\ell^-$ samples. The b-quark mass is conservatively varied in the range of ±0.15 GeV/c^2. Figure 8.1 shows M_{X_s} and Figure 8.2 shows $M_{\ell^+\ell^-}$ distributions when the b-quark mass is varied. The effects on the M_{X_s} and $M_{\ell^+\ell^-}$ due to the b-quark mass variation is very small as shown in FIG 8.1 and FIG 8.2.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
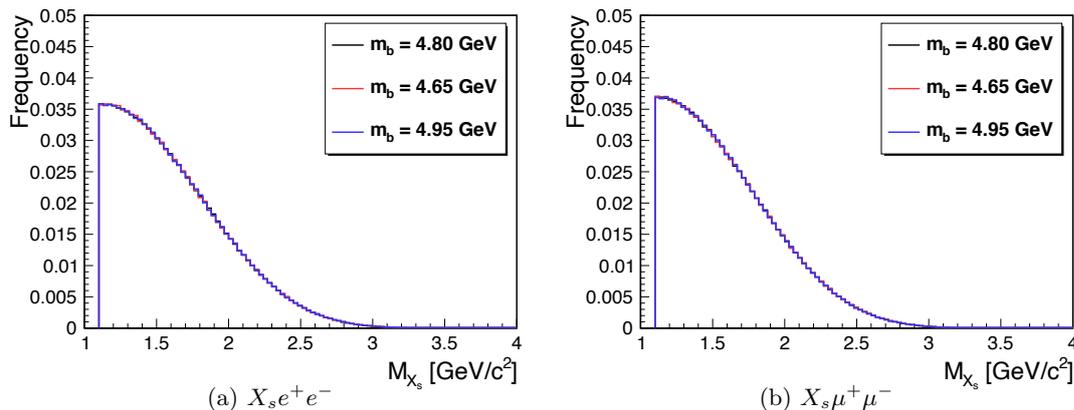
FIG 8.1: M_{X_s} distributions for various b-quark mass. Black line shows the distribution of $m_b = 4.80$ GeV/c^2, red line shows that of $m_b = 4.65$ GeV/c^2, and blue line shows that of $m_b = 4.95$ GeV/c^2.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
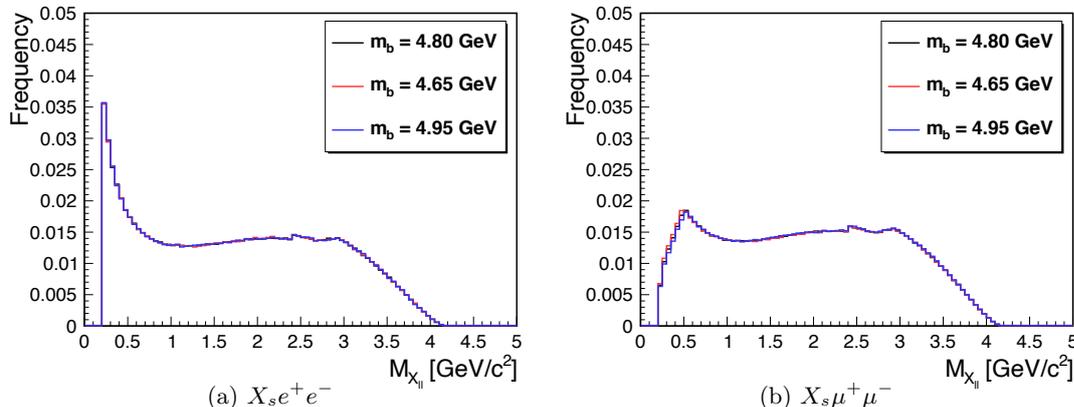
FIG 8.2: $M_{\ell^+\ell^-}$ distributions for various b-quark mass. Black line shows the distribution of $m_b = 4.80$ GeV/c^2, red line shows that of $m_b = 4.65$ GeV/c^2, and blue line shows that of $m_b = 4.95$ GeV/c^2.

8.5.3 Fermi motion momentum

The Fermi motion momentum p_F of b-quark in the B meson is evaluated with hadronic moments measurements in semi-lepton B decay and the photon spectrum in $B \to X_s\gamma$ decay. Recent result on p_F is $pF = 0.461^{+0.036}_{-0.039}$ GeV [1]. To calculate the reconstruction efficiency, the parameter is assumed to be $p_F = 0.410$ GeV. The efficiency correction is calculated using MC samples in which p_F is set at 0.461 GeV

and a systematic uncertainty is estimated by varying in the range of $0.422 < p_F < 0.498\,[\text{GeV}]$. Figure 8.3 shows M_{X_s} and Figure 8.4 shows $M_{\ell^+\ell^-}$ distributions when p_F is varied.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
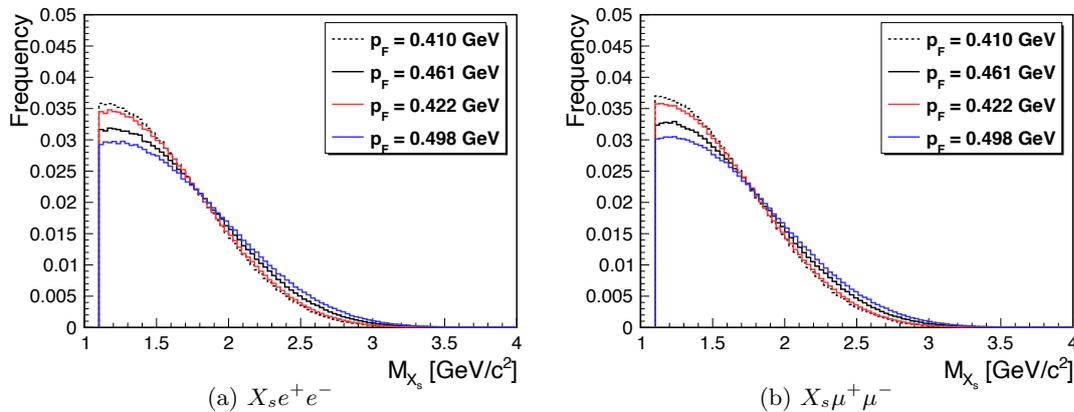
FIG 8.3: M_{X_s} distributions for various Fermi motion momentum. Black dashed line shows the distribution of $p_F = 0.410\,\text{GeV}$, black line shows the distribution of $p_F = 0.461\,\text{GeV}$, red line shows the distribution of $p_F = 0.422\,\text{GeV}$, and blue line shows the distribution of $p_F = 0.498\,\text{GeV}$.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
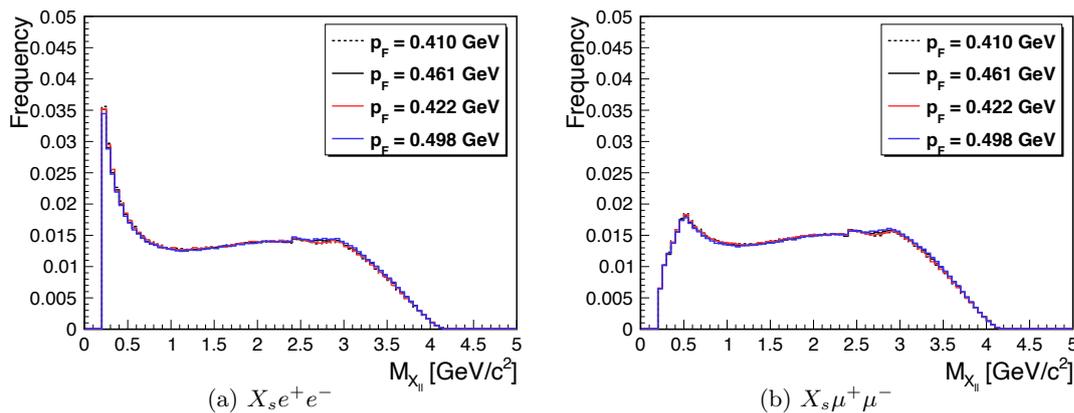
FIG 8.4: $M_{\ell^+\ell^-}$ distributions for various Fermi motion momentum. Black dashed line shows the distribution of $p_F = 0.410\,\text{GeV}$, black line shows the distribution of $p_F = 0.461\,\text{GeV}$, red line shows the distribution of $p_F = 0.422\,\text{GeV}$, and blue line shows the distribution of $p_F = 0.498\,\text{GeV}$.

8.5.4 Fragmentation and missing modes of X_s

The reconstruction efficiency depends on the X_s decay modes. The prediction of the fragmentation of the X_s modes in the MC samples should be corrected by using real data. To keep the fraction of $K^{(*)}\ell^+\ell^-$ and non-resonant $X_s\ell^+\ell^-$, an overall normalization factor of 0.986 is applied in on non-resonant $X_s\ell^+\ell^-$ modes. The uncertainty due to the fraction of $K^{(*)}\ell^+\ell^-$ and non-resonant $X_s\ell^+\ell^-$ is described in Section 8.5.5. Since no yields are observed in the $K4\pi$ and $3K$ modes, only the normalization factor is applied. The correction factor is estimated with $B \to X_s J/\psi$ at the Belle experiment which is shown in TABLE 8.2. Since there might be some resonant contributions in the $K\pi$ modes, the correction factor on these mode is larger compared with other modes. By applying the correction and keeping the fraction of $K^{(*)}\ell^+\ell^-$ and non-resonant $X_s\ell^+\ell^-$, the fraction of missing modes which includes more than 4 pions, or hadrons other than kaon and pion is also modified to follow the data. The fragmentation of X_s modes will be measured with $B \to X_s\ell^+\ell^-$ when the Belle II collects enough data to measure the yield of each mode.

TABLE 8.2: The fragmentation correction factors.

Mode	Factor
$K^{\pm}\ \pi^{\mp}$	0.528 ± 0.016
$K^0_S\ \pi^{\mp}$	0.615 ± 0.034
$K^{\pm}\ \pi^0$	0.718 ± 0.055
$K^0_S\ \pi^0$	0.539 ± 0.090
$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}$	1.145 ± 0.031
$K^0_S\ \pi^{\pm}\ \pi^{\mp}$	0.835 ± 0.049
$K^{\pm}\ \pi^{\mp}\ \pi^0$	0.941 ± 0.055
$K^0_S\ \pi^{\pm}\ \pi^0$	0.888 ± 0.106
$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}$	1.583 ± 0.136
$K^0_S\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}$	1.557 ± 0.293
$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^0$	1.325 ± 0.191
$K^0_S\ \pi^{\pm}\ \pi^{\mp}\ \pi^0$	1.067 ± 0.292
$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}$	0.986
$K^0_S\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}$	0.986
$K^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^{\mp}\ \pi^0$	0.986
$K^0_S\ \pi^{\pm}\ \pi^{\mp}\ \pi^{\pm}\ \pi^0$	0.986
$K^{\pm}\ K^{\mp}\ K^{\pm}$	0.986
$K^{\pm}\ K^{\mp}\ K^0_S$	0.986

8.5.5   Fraction of $B \to K\ell^+\ell^-$, $B \to K^*\ell^+\ell^-$, and non-resonant $B \to X_s\ell^+\ell^-$

Fraction of three kinds of signal MC sample, $B \to K\ell^+\ell^-$, $B \to K^*\ell^+\ell^-$ and non-resonant $B \to X_s\ell^+\ell^-$, is based on the SM prediction on the branching fraction in this paper as discussed in Section 4.3.1. Since the reconstruction efficiency depends on the X_s decay modes, the uncertainty of the fraction of X_s modes is an important source of systematic uncertainty on the branching fraction. The systematic uncertainty is estimated by varying the fraction according to the prediction uncertainty on $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$.

Chapter 9

Validation with control modes

In this chapter, the measurements of the branching fraction on the $B \to X_s J/\psi$ are discussed. Since the measurement method of the control modes is almost same with that of $B \to X_s \ell^+ \ell^-$, the measurement results can validate the method on the $B \to X_s \ell^+ \ell^-$, such as the reconstruction efficiency estimation and the signal extraction. Four exclusive $B \to X_s J/\psi$ modes are used for the validation, $B^+ \to K^+ J/\psi$, $B^0 \to K^0 J/\psi$, $B^0 \to K^{*0} J/\psi$, and $B^+ \to K^+ \pi^- \pi^+ J/\psi$, in which the branching fractions have been measured in previous studies.

9.1 Reconstruction of $B \to X_s J/\psi \, (\to \ell^+ \ell^-)$

The selection criteria on $e^\pm, \mu^\pm, K^\pm, \pi^\pm$ and K^0_S are exactly same as those discussed in Section 5.1.

A K^{*0} meson is reconstructed from $K^+ \pi^-$ for the $B^0 \to K^{*0} J/\psi$ mode. The K^{*0} candidates are required to satisfy the mass criterion, $|M_{K^+\pi^-} - M_{K^{*0}}| < 75 \, \mathrm{MeV}/c^2$ where $M_{K^{*0}}$ is the nominal K^{*0} meson mass. For the $B \to K^+\pi^-\pi^+ J/\psi$ mode, candidates are required to satisfy $M_{K^+\pi^-\pi^+} < 2.0 \, \mathrm{GeV}/c^2$ to suppress combinatorial backgrounds.

A J/ψ meson is reconstructed from $e^+ e^-$ or $\mu^+ \mu^-$. The J/ψ candidates are required to satisfy the mass criterion which is used to veto J/ψ for the $B \to X_s \ell^+ \ell^-$ analysis discussed in Section 6.2.2.

$$-0.40 \, \mathrm{GeV}/c^2 < M_{e^+(\gamma)e^-(\gamma)} - M_{J/\psi} < 0.15 \, \mathrm{GeV}/c^2 \tag{9.1}$$

$$-0.25 \, \mathrm{GeV}/c^2 < M_{\mu^+\mu^-} - M_{J/\psi} < 0.10 \, \mathrm{GeV}/c^2 \tag{9.2}$$

Then B meson is reconstructed by combining X_s final states and J/ψ with following criteria.

$$5.2 \, \mathrm{GeV}/c^2 < M_{\mathrm{bc}} < 5.3 \, \mathrm{GeV}/c^2, \tag{9.3}$$

$$-0.15 \, \mathrm{GeV} < \Delta E < 0.15 \, \mathrm{GeV}. \tag{9.4}$$

Compared with $B \to X_s \ell^+ \ell^-$, the $B \to X_s J/\psi \, (\to \ell^+ \ell^-)$ samples are very clean thanks to the large cross-section. To keep most of signal events, only the pre-selection and the D veto (Section 6.2) are applied.

The reconstruction efficiency is estimated with MC samples for each modes. The estimated efficiency is summarized in TABLE 9.1.

9.1.1 Signal MC for $B^+ \to K^+\pi^-\pi^+ J/\psi$

The MC samples for all modes except for $B^+ \to K^+\pi^-\pi^+ J/\psi$ are extracted from the generic MC samples which is described in Section 4.3. Since $B^+ \to K^+\pi^-\pi^+ J/\psi$ has several resonances and non-resonant components, the signal MC samples are produced according to the study of $K^+\pi^-\pi^+$ final states at the Belle experiment [72].

Figure 9.1 shows the $M_{K^+\pi^-\pi^+}$ distribution at the MC truth level used to estimated the efficiency. Following five resonances and non-resonant components are considered, $K_1(1270), K_1(1400), K^*(1410), K^*_2(1430),$ and $K^*(1680)$.

TABLE 9.1: Reconstruction efficiency on the control modes $B \to X_s J/\psi$ estimated in MC samples.

Modes	Efficiency
$B^+ \to K^+ J/\psi (\to e^+ e^-)$	2.36%
$B^0 \to K^0 J/\psi (\to e^+ e^-)$	0.633%
$B^0 \to K^{*0} J/\psi (\to e^+ e^-)$	0.640%
$B^+ \to K^+ \pi^- \pi^+ J/\psi (\to e^+ e^-)$	0.553%
$B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$	2.59%
$B^0 \to K^0 J/\psi (\to \mu^+ \mu^-)$	0.668%
$B^0 \to K^{*0} J/\psi (\to \mu^+ \mu^-)$	0.702%
$B^+ \to K^+ \pi^- \pi^+ J/\psi (\to \mu^+ \mu^-)$	0.627%
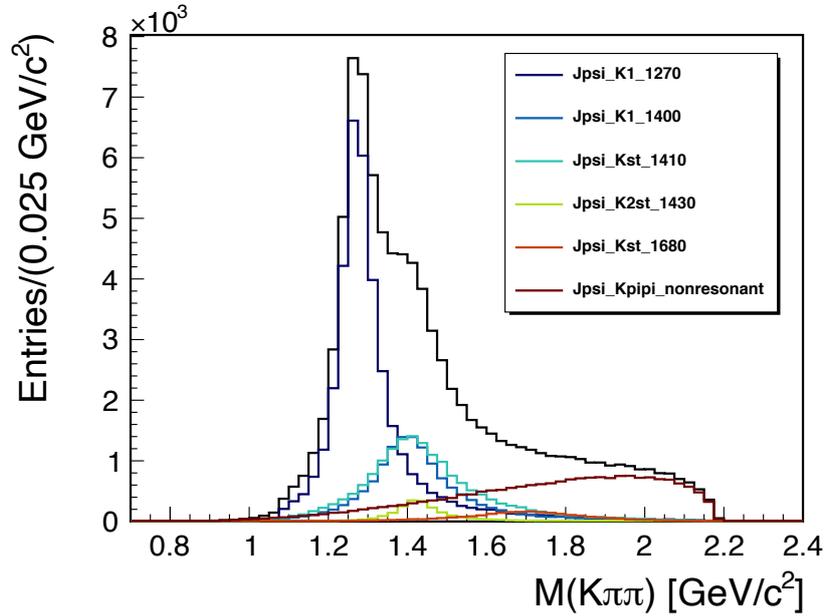
FIG 9.1: $M_{K^+\pi^-\pi^+}$ distribution at the MC truth level which is used to estimated the reconstruction efficiency. Black line shows the total distribution. Contributions from each process is identified by color code.

9.2 Branching fraction extraction for $B^+ \to X_s J/\psi$

The branching fraction is obtained by fitting the $M_{\rm bc}$ distributions. The branching fraction for B^+ modes and B^0 modes can be calculated from the signal yields with the following function.

$$\mathcal{B}(B^+ \to X^+) = \frac{N_{\rm signal}}{2N_{B\bar{B}} \times f_{+-} \times \epsilon} \tag{9.5}$$

$$\mathcal{B}(B^0 \to X^0) = \frac{N_{\rm signal}}{2N_{B\bar{B}} \times f_{00} \times \epsilon} \tag{9.6}$$

Here f_{+-} and f_{00} are the branching fraction of $\Upsilon(4S) \to B^+B^-$ and $\Upsilon(4S) \to B^0\bar{B}^0$, respectively.

The probability density function is constructed from two components, a Gaussian for signal and an ARGUS function for non-peaking backgrounds. Since the control modes are very clean and signal peak is clear, peaking background components are omitted and all parameters of the signal Gaussian are floated. The endpoint of the ARGUS fiction is fixed at the value estimated with $B \to D\pi$ samples shown in TABLE 7.5.

9.3 Systematic uncertainty for $B^+ \to X_s J/\psi$

Relevant sources of systematic uncertainty which are described in Chapter 8 are considered. Additional systematic uncertainty is due to f_{+-} and f_{00}. f_{+-} and f_{00} are measured by BaBar, Belle, and CLEO experiments, $f_{+-} = 0.514 \pm 0.006$ and $f_{00} = 0.486 \pm 0.006$. The uncertainty of f_{+-} or f_{00} is included as the systematic uncertainty.

Efficiency correction

The reconstruction efficiency estimated with MC samples is corrected with correction factors which are discussed in Section 8.2. All correction factors except for the FastBDT efficiency are considered, since the FastBDT is not applied. The efficiency correction on the $B \to X_s J/\psi$ is summarized in TABLE 9.2.

TABLE 9.2: The efficiency correction factors on the $B \to X_s J/\psi$.

Source	$K^+ J/\psi\,(e^+e^-)$ [%]	$K^0 J/\psi\,(e^+e^-)$ [%]	$K^{*0} J/\psi\,(e^+e^-)$ [%]	$K^+\pi^-\pi^+ J/\psi\,(e^+e^-)$ [%]
Tracking efficiency	100 ± 2.4	100 ± 3.9	100 ± 4.2	100 ± 2.0
Electron ID	96.3 ± 3.1	96.3 ± 3.2	96.1 ± 3.1	96.4 ± 3.1
Kaon ID	95.6 ± 1.4	-	99.5 ± 1.5	100.5 ± 1.6
Pion ID	-	-	97.1 ± 1.4	94.4 ± 2.9
K_S^0 reconstruction	-	100 ± 7.8	-	-
Source	$K^+ J/\psi\,(\mu^+\mu^-)$ [%]	$K^0 J/\psi\,(\mu^+\mu^-)$ [%]	$K^{*0} J/\psi\,(\mu^+\mu^-)$ [%]	$K^+\pi^-\pi^+ J/\psi\,(\mu^+\mu^-)$ [%]
Tracking efficiency	100 ± 2.4	100 ± 3.9	100 ± 4.2	100 ± 2.0
Muon ID	86.4 ± 4.0	86.4 ± 4.0	86.1 ± 4.5	86.1 ± 4.1
Kaon ID	95.8 ± 1.4	-	99.5 ± 1.5	100.3 ± 1.5
Pion ID	-	-	97.1 ± 1.4	94.3 ± 2.9
K_S^0 reconstruction	-	100 ± 8.0	-	-

Fitter bias

The systematic uncertainty due to the fitter bias is estimated from the pull distribution. Test samples are generated from the PDF by fluctuating the number of events with a Poisson distribution and 1000 test samples are produced. Figure 9.2 and 9.3 show the pull distribution which are obtained by toy MC tests. Obtained results of mean μ_{pull} and width σ_{pull} are summarized in TABLE 9.3. The obtained width is consistent with unity within the uncertainty in all decay modes and most of mean is also consistent with 0. The systematic uncertainty due to the fitter bias is only assigned to the modes in which the discrepancy of mean from 0 is at more than 2σ level.

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

(b) $B^0 \to K^0 J/\psi \, (\to e^+ e^-)$

(c) $B^+ \to K^{*0} J/\psi \, (\to e^+ e^-)$
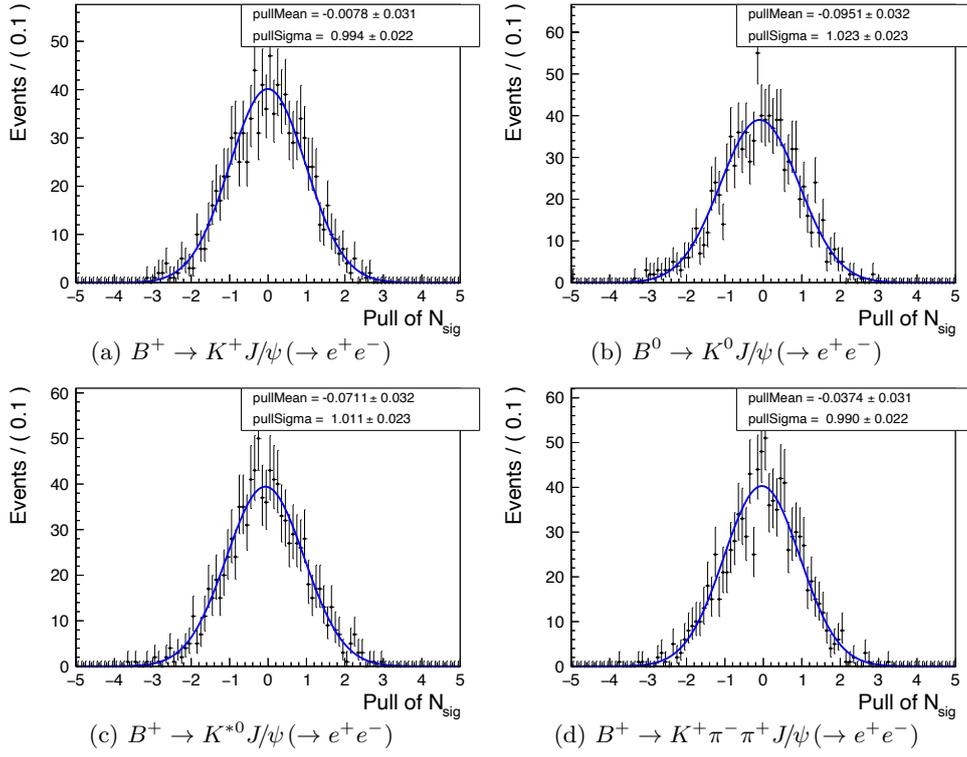
(d) $B^+ \to K^+ \pi^- \pi^+ J/\psi \, (\to e^+ e^-)$

FIG 9.2: Pull of number of signal distributions of the $B \to X_s J/\psi \, (\to e^+ e^-)$.

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

(b) $B^0 \to K^0 J/\psi \, (\to \mu^+ \mu^-)$

(c) $B^0 \to K^{*0} J/\psi \, (\to \mu^+ \mu^-)$

(d) $B^+ \to K^+ \pi^- \pi^+ J/\psi \, (\to \mu^+ \mu^-)$
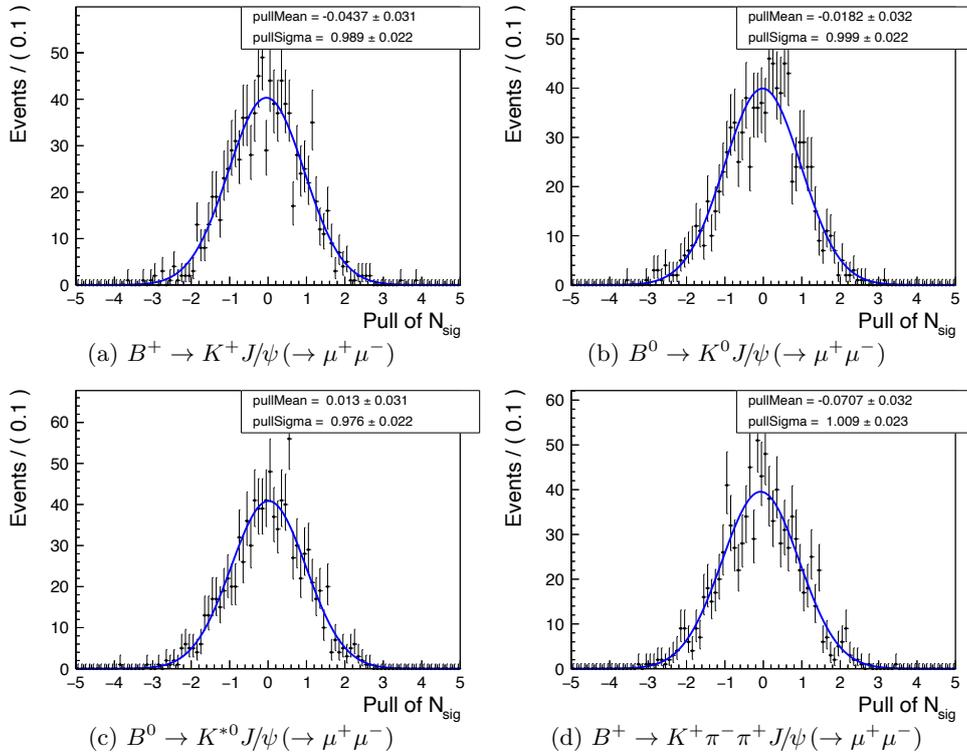
FIG 9.3: Pull of number of signal distributions of the $B \to X_s J/\psi \, (\to \mu^+ \mu^-)$.

TABLE 9.3: Obtained mean and width of the pull distributions on $B \to X_s J/\psi$ control samples.

Modes	μ_{pull}	σ_{pull}
$B^+ \to K^+ J/\psi (\to e^+ e^-)$	-0.008 ± 0.031	0.994 ± 0.022
$B^0 \to K^0 J/\psi (\to e^+ e^-)$	-0.095 ± 0.032	1.023 ± 0.023
$B^0 \to K^{*0} J/\psi (\to e^+ e^-)$	-0.071 ± 0.032	1.011 ± 0.023
$B^+ \to K^+ \pi^- \pi^+ J/\psi (\to e^+ e^-)$	-0.037 ± 0.031	0.990 ± 0.022
$B^+ \to K^+ J/\psi (\to \mu^+ \mu^-)$	-0.044 ± 0.031	0.989 ± 0.022
$B^0 \to K^0 J/\psi (\to \mu^+ \mu^-)$	-0.018 ± 0.032	0.999 ± 0.022
$B^0 \to K^{*0} J/\psi (\to \mu^+ \mu^-)$	-0.013 ± 0.031	0.976 ± 0.022
$B^+ \to K^+ \pi^- \pi^+ J/\psi (\to \mu^+ \mu^-)$	-0.071 ± 0.032	1.009 ± 0.023

9.4 Results on the branching fraction of $B^+ \to X_s J/\psi$

Figure 9.4 and Figure 9.5 shows the M_{bc} distributions with fitting functions for $B \to X_s J/\psi (\to e^+ e^-)$ and $B \to X_s J/\psi (\to \mu^+ \mu^-)$, respectively. The results of the branching fractions are summarized in TABLE 9.4. The systematic uncertainty is summarized in TABLE 9.5.
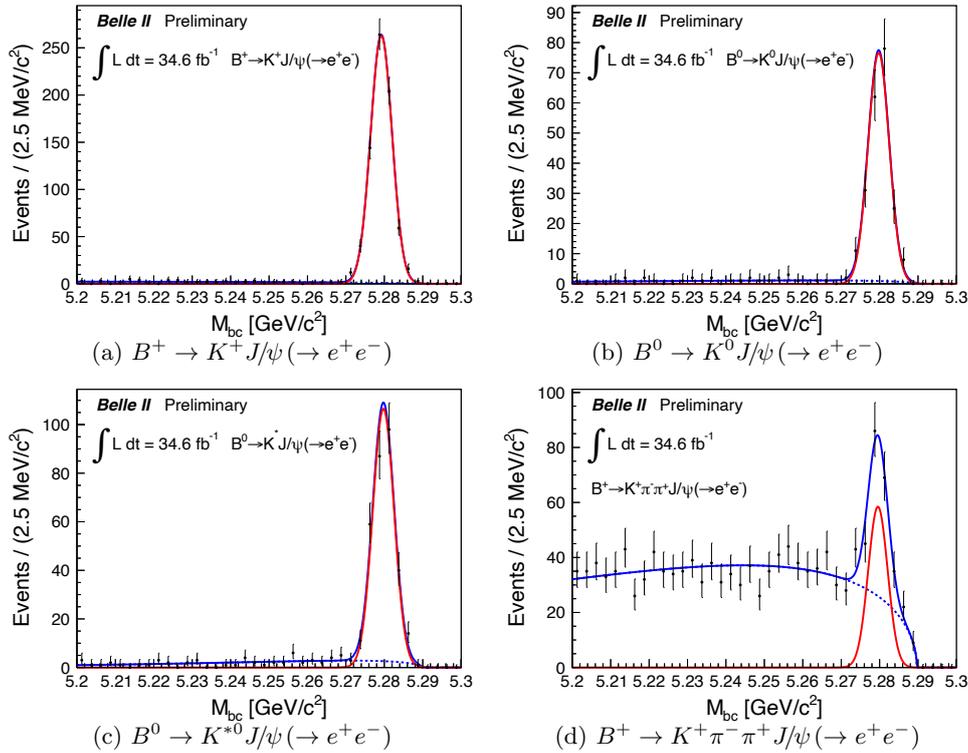
FIG 9.4: M_{bc} distributions of the $B \to X_s J/\psi (\to e^+ e^-)$ with fitting function. The fit contained the following components: a Gaussian for the signal (red line) and an ARGUS function to model background from the continuum and combinatorial B decays (dashed blue line). The total fit is the solid blue line and the data are overlaid as black makers.

These results are compared with measurements by other experiments. Figure 9.6 shows comparison on the branching fraction between the world average and our results. Most of our measurements are consistent with the world average within 1σ. Result on $B^+ \to K^+ J/\psi$ with electron modes differs from with the world average, $[9.97 \pm 0.30] \times 10^{-4}$, by 2.1σ which is the largest value among all results. We have confirmed that the analysis procedure of the branching fraction measurement is validated from these results.

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

(b) $B^0 \to K^0 J/\psi \, (\to \mu^+ \mu^-)$

(c) $B^0 \to K^{*0} J/\psi \, (\to \mu^+ \mu^-)$

(d) $B^+ \to K^+ \pi^- \pi^+ J/\psi \, (\to \mu^+ \mu^-)$
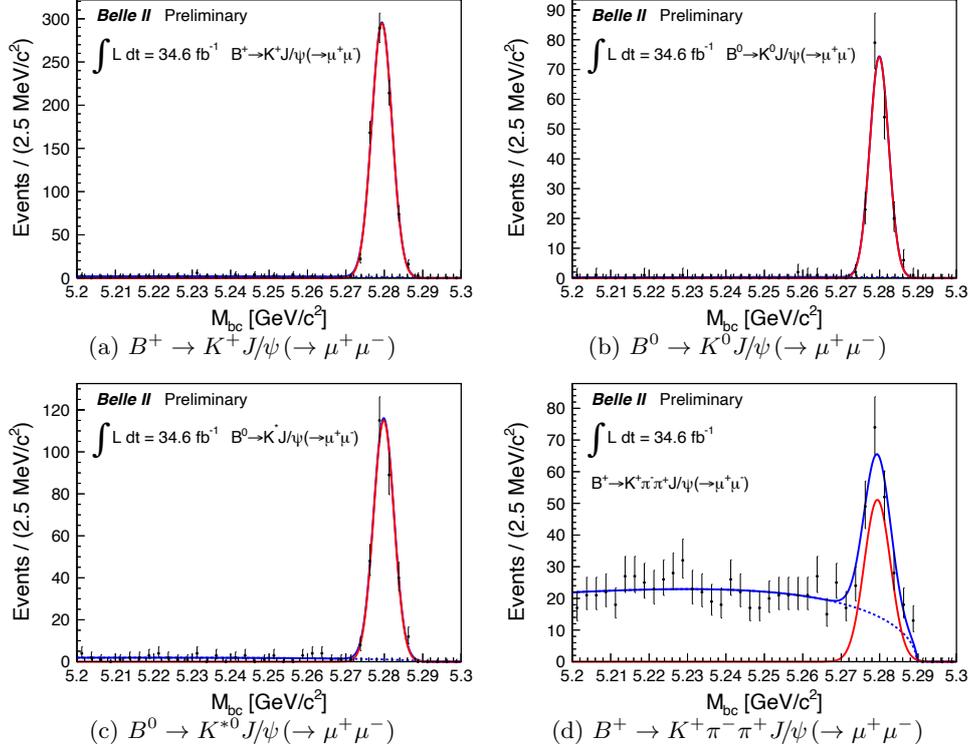
FIG 9.5: $M_{\rm bc}$ distributions of the $B \to X_s J/\psi \, (\to \mu^+ \mu^-)$ with fitting function. The fit contained the following components: a Gaussian for the signal (red line) and an ARGUS function to model background from the continuum and combinatorial B decays (dashed blue line). The total fit is the solid blue line and the data are overlaid as black makers.

TABLE 9.4: The branching fractions and corresponding signal yields on $B \to X_s J/\psi$ control samples.

Modes	Branching fraction $[10^{-4}]$	Signal yields
$B^+ \to K^+ J/\psi \, (\to e^+ e^-)$	$8.69 \pm 0.33(\text{stat}) \pm 0.41(\text{syst})$	734 ± 27
$B^0 \to K^0 J/\psi \, (\to e^+ e^-)$	$9.40 \pm 0.68(\text{stat}) \pm 0.90(\text{syst})$	211 ± 15
$B^0 \to K^{*0} J/\psi \, (\to e^+ e^-)$	$13.56 \pm 0.85(\text{stat}) \pm 0.81(\text{syst})$	268 ± 18
$B^+ \to K^+ \pi^- \pi^+ J/\psi \, (\to e^+ e^-)$	$7.54 \pm 1.11(\text{stat}) \pm 1.60(\text{syst})$	148 ± 21
$B^+ \to K^+ J/\psi \, (\to \mu^+ \mu^-)$	$9.35 \pm 0.34(\text{stat}) \pm 0.54(\text{syst})$	780 ± 28
$B^0 \to K^0 J/\psi \, (\to \mu^+ \mu^-)$	$8.64 \pm 0.66(\text{stat}) \pm 0.88(\text{syst})$	183 ± 14
$B^0 \to K^{*0} J/\psi \, (\to \mu^+ \mu^-)$	$14.28 \pm 0.84(\text{stat}) \pm 1.04(\text{syst})$	306 ± 18
$B^+ \to K^+ \pi^- \pi^+ J/\psi \, (\to \mu^+ \mu^-)$	$8.87 \pm 1.04(\text{stat}) \pm 1.83(\text{syst})$	176 ± 19
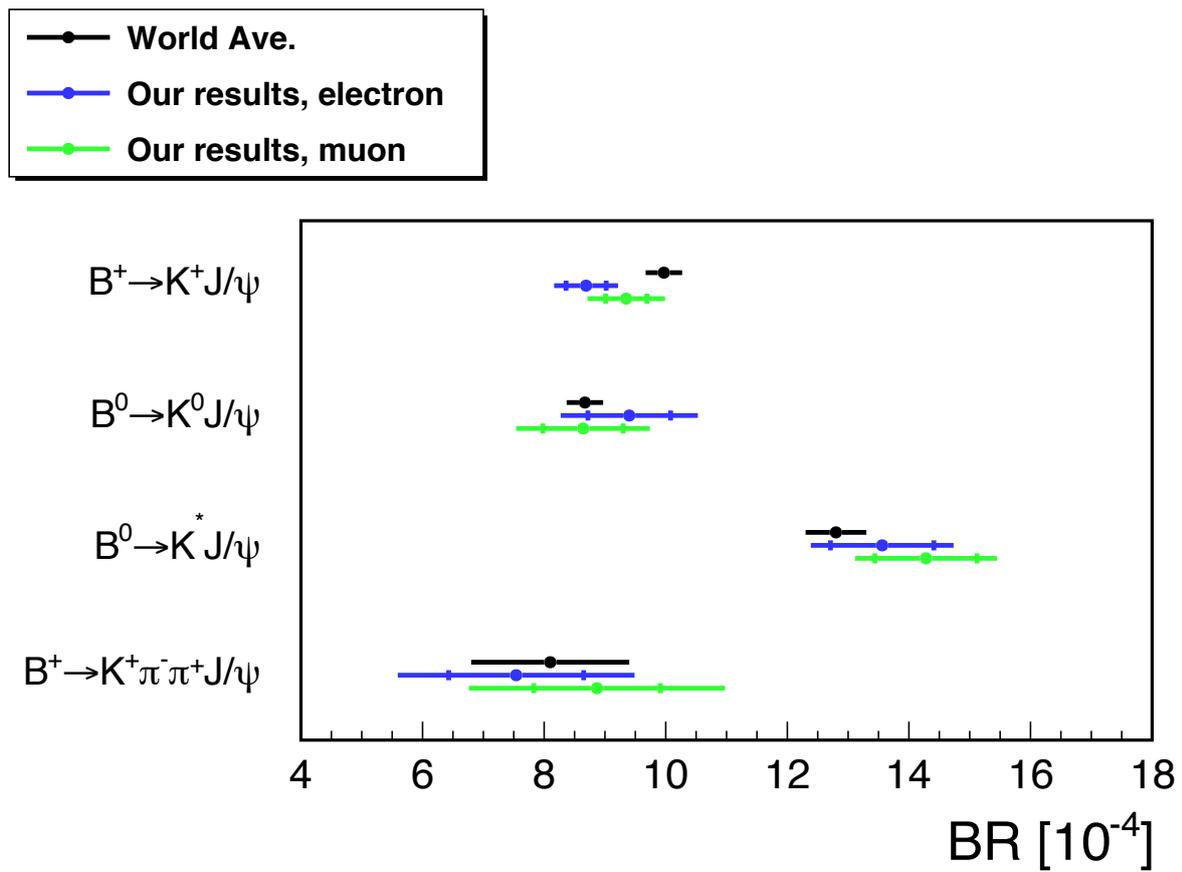
FIG 9.6: Measurements results on the branching fractions of $B \to X_s J/\psi$ modes. Black dots show the world average, blue dots show the our results with electron modes, and green dots show the our results with muon modes. The internal error bars show only the statistical uncertainty and outer ones show the uncertainty combined statistical and systematic one.

TABLE 9.5: Systematic uncertainties on the branching fraction in unit of 10^{-4}. The total is obtained from the sum in quadrature of all contributions.

Source	$K^+ J/\psi \, (e^+ e^-)$	$K^0 J/\psi \, (e^+ e^-)$	$K^{*0} J/\psi \, (e^+ e^-)$	$K^+ \pi^- \pi^+ J/\psi \, (e^+ e^-)$
Number of B meson pairs	± 0.14	± 0.15	± 0.22	± 0.12
f_{+-} or f_{00}	± 0.11	± 0.11	± 0.16	± 0.09
Tracking efficiency	± 0.21	± 0.37	± 0.57	± 1.56
Electron ID	± 0.28	± 0.31	± 0.44	± 0.25
Kaon ID	± 0.13	-	± 0.20	± 0.12
Pion ID	-	-	± 0.19	± 0.23
K_S^0 reconstruction	-	± 0.74	-	-
Fitter bias	-	± 0.06	± 0.06	-
Source	$K^+ J/\psi \, (\mu^+ \mu^-)$	$K^0 J/\psi \, (\mu^+ \mu^-)$	$K^{*0} J/\psi \, (\mu^+ \mu^-)$	$K^+ \pi^- \pi^+ J/\psi \, (\mu^+ \mu^-)$
Number of B meson pairs	± 0.15	± 0.14	± 0.23	± 0.14
f_{+-} or f_{00}	± 0.12	± 0.10	± 0.17	± 0.10
Tracking efficiency	± 0.22	± 0.33	± 0.60	± 1.75
Muon ID	± 0.43	± 0.40	± 0.74	± 0.43
Kaon ID	± 0.14	-	± 0.21	± 0.14
Pion ID	-	-	± 0.21	± 0.28
K_S^0 reconstruction	-	± 0.69	-	-
Fitter bias	-	-	-	± 0.07

Chapter 10

Results and Discussion

10.1   Results on $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$

Figure 10.1 shows M_{bc} distributions with the fitting function. The central values of the branching fractions and corresponding signal yields are summarized in TABLE 10.1.

(a) $X_s e^+ e^-$ (b) $X_s \mu^+ \mu^-$
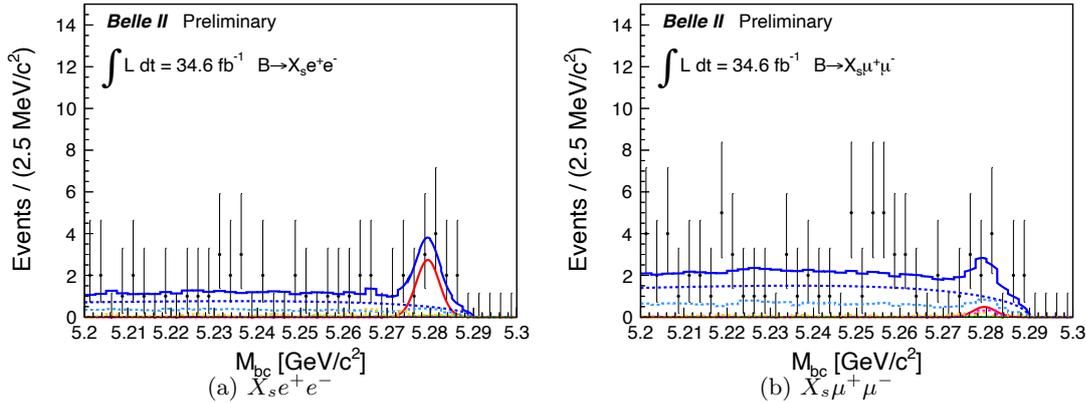
FIG 10.1: M_{bc} distributions of the $B \to X_s \ell^+ \ell^-$ with the fitting function. The fit contained the following components: a Gaussian for the signal (red line), an ARGUS function to model background from the continuum and combinatorial B decays (dashed blue line), a histogram PDF to describe self cross-feed (dashed green line), and three histogram PDFs to describe peaking backgrounds, (i) double mis-ID background (dashed light-blue line), (ii) swapped mis-ID background (dashed magenta line), (iii) Charmonium background (dashed orange line). The total fit result is the solid blue line and the data are overlaid as black makers.

TABLE 10.1: Central values of the branching fractions and corresponding signal yileds.

Modes	Branching fraction [10^{-6}]	Signal yields
$B \to X_s e^+ e^-$	$4.86^{+2.75}_{-2.42}(\mathrm{stat})^{+1.02}_{-0.92}(\mathrm{syst})$	$8.25^{+4.69}_{-4.09}(\mathrm{stat})^{+1.33}_{-1.25}(\mathrm{syst})$
$B \to X_s \mu^+ \mu^-$	$0.78^{+2.21}_{-1.85}(\mathrm{stat})^{+0.43}_{-0.38}(\mathrm{syst})$	$1.37^{+3.87}_{-3.24}(\mathrm{stat})^{+0.74}_{-0.65}(\mathrm{syst})$

The systematic uncertainties which are discussed in Chapter 8 are summarized in TABLE 10.2. The uncertainties are categorized as additive uncertainty which might induce a bias on signal yields and multiplicative uncertainty which affect the reconstruction efficiency or the number of $B\overline{B}$ pairs. Total uncertainty is obtained from the sum in quadrature of contributions.

TABLE 10.2: Systematic uncertainties on the branching fraction in unit of 10^{-6}. The uncertainties are categorized as additive (A) or multiplicative (M).

Source	$B \to X_s e^+ e^-$	$B \to X_s \mu^+ \mu^-$
Number of B meson pairs (M)	± 0.08	± 0.01
Tracking efficiency (M)	$+0.19 \atop -0.18$	± 0.02
Electron ID (M)	$+0.13 \atop -0.10$	-
Muon ID (M)	-	$+0.02 \atop -0.03$
Kaon ID (M)	± 0.06	± 0.007
Pion ID (M)	$+0.02 \atop -0.04$	± 0.004
K_S^0 reconstruction (M)	$+0.03 \atop -0.04$	$+0.005 \atop -0.006$
π^0 reconstruction (M)	$+0.007 \atop -0.010$	± 0.001
FastBDT efficiency (M)	$+0.21 \atop -0.19$	± 0.02
Efficiency correction	$+0.32 \atop -0.29$	$+0.036 \atop -0.043$
Pull bias (A)	$+0.54 \atop -0.48$	$+0.35 \atop -0.29$
Pull width (A)	$+0.26 \atop -0.23$	$+0.10 \atop -0.09$
Fitter bias	$+0.60 \atop -0.53$	$+0.36 \atop -0.30$
Signal mean (A)	± 0.08	$+0.02 \atop -0.03$
Signal width (A)	± 0.16	$+0.008 \atop -0.004$
Self cross-feed ratio (A)	$+0.46 \atop -0.47$	± 0.05
Double mis-ID yields (A)	± 0.10	$+0.16 \atop -0.15$
Swapped mis-ID yields (A)	± 0.002	± 0.09
Charmonium yields (A)	± 0.04	± 0.10
Fixed parameters	± 0.51	$+0.22 \atop -0.21$
K^*-X_s transition (M)	$+0.02 \atop -0.13$	$+0.01 \atop -0.02$
b-quark mass (M)	± 0.03	± 0.003
Fermi motion momentum (M)	$+0.07 \atop -0.06$	$+0.005 \atop -0.011$
Fragmentation and missing modes of X_s (M)	$+0.15 \atop -0.14$	± 0.02
Fraction of $B \to K\ell^+\ell^-$ (M)	$+0.53 \atop -0.41$	$+0.09 \atop -0.07$
Fraction of $B \to K^*\ell^+\ell^-$ (M)	$+0.12 \atop -0.10$	$+0.010 \atop -0.004$
Signal modeling	$+0.57 \atop -0.47$	$+0.094 \atop -0.077$
Total additive	$+0.78 \atop -0.74$	$+0.42 \atop -0.37$
Total multiplicative	$+0.65 \atop -0.55$	$+0.100 \atop -0.087$
Total	$+1.02 \atop -0.92$	$+0.43 \atop -0.38$

10.2 Results on $B \to X_s \ell^+ \ell^-$

The branching fraction of $B \to X_s \ell^+ \ell^-$ is obtained by simultaneous fitting of M_{bc} distributions of $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$ assuming the lepton flavor universality. M_{bc} distributions and the fitting function are shown in FIG 10.2. The central value of the branching fraction is shown in TABLE 10.3.
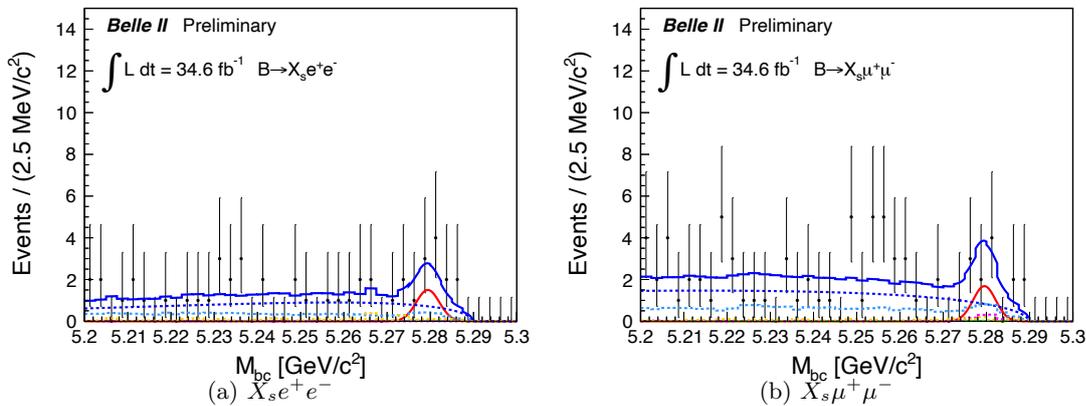

(a) $X_s e^+ e^-$

(b) $X_s \mu^+ \mu^-$

FIG 10.2: M_{bc} distributions of the $B \to X_s \ell^+ \ell^-$ with the fitting function when the simultaneous fitting is applied. The fit contained the following components: a Gaussian for the signal (red line), an ARGUS function to model background from the continuum and combinatorial B decays (dashed blue line), a histogram PDF to describe self cross-feed (dashed green line), and three histogram PDFs to describe peaking backgrounds, (i) double mis-ID background (dashed light-blue line), (ii) swapped mis-ID background (dashed magenta line), (iii) Charmonium background (dashed orange line). The total fit is the solid blue line and the data are overlaid as black makers.

TABLE 10.3: Central value of the branching fractions.

Modes	Branching fraction [10^{-6}]
$B \to X_s \ell^+ \ell^-$	$2.78^{+1.82}_{-1.65}(\mathrm{stat})^{+0.66}_{-0.61}(\mathrm{syst})$

The systematic uncertainty for the simultaneous fitting is summarized in TABLE 10.4. Systematic uncertainties due to the number of $B\overline{B}$ pairs, efficiency corrections, and signal modeling are assumed to be correlated between electron modes and muon modes.

10.3 Discussion and prospect

The statistical significance of the observation of the signal events is computed from the $\sqrt{-2\log(\mathcal{L}_0/\mathcal{L}_{max})}$, where $\mathcal{L}_{max}$ and $\mathcal{L}_0$ is the likelihood from the nominal fit and from the fit omitting the signal component, respectively. The obtained statistical significances are 2.1σ, 0.40σ, and 1.7σ for the $B \to X_s e^+ e^-$, $B \to X_s \mu^+ \mu^-$, and $B \to X_s \ell^+ \ell^-$, respectively. The negative log likelihood profiles are shown in FIG 10.3. The statistical likelihood function is, then, convolved with a Gaussian function of the width given by the additive systematic uncertainty which might induce a bias on signal yields. The total significances are 2.0σ, 0.40σ, and 0.17σ for the $B \to X_s e^+ e^-$, $B \to X_s \mu^+ \mu^-$, and $B \to X_s \ell^+ \ell^-$, respectively.

Since the significance of the signals are below 2σ for $B \to X_s \mu^+ \mu^-$ and $B \to X_s \ell^+ \ell^-$, upper limits at 90% and 95% confidence levels (CL) are determined for the branching fraction. The limit of the branching fraction at 90% (95%) CL, $\mathcal{B}_{90(95)}$, is obtained from the following equation.

$$0.90(0.95) = \frac{\int_0^{\mathcal{B}_{90(95)}} L(\mathcal{B})d\mathcal{B}}{\int_0^{\infty} L(\mathcal{B})d\mathcal{B}} \tag{10.1}$$

TABLE 10.4: Systematic uncertainties on the branching fraction in unit of 10^{-6}. The uncertainties are categorized as additive (A) or multiplicative (M).

| Source	$B \to X_s \ell^+ \ell^-$
Number of B meson pairs (M)	± 0.04
Tracking efficiency (M)	$+0.11$ -0.10
Electron ID (M)	$+0.006$ -0.012
Muon ID (M)	$+0.07$ -0.13
Kaon ID (M)	± 0.03
Pion ID (M)	$+0.01$ -0.02
K_S^0 reconstruction (M)	± 0.02
π^0 reconstruction (M)	$+0.001$ -0.005
FastBDT efficiency (M)	$+0.10$ -0.09
Efficiency correction	$+0.17$ -0.19
Pull bias (A)	$+0.46$ -0.41
Pull width (A)	$+0.19$ -0.18
Fitter bias	$+0.50$ -0.45
Signal mean (A)	± 0.05
Signal width (A)	± 0.06
Self cross-feed ratio (A)	$+0.15$ -0.17
Double mis-ID yields (A)	± 0.12
Swapped mis-ID yields (A)	$+0.03$ -0.04
Charmonium yields (A)	$+0.07$ -0.09
Fixed parameters	$+0.22$ -0.24
K^*-X_s transition (M)	$+0.04$ -0.08
b-quark mass (M)	± 0.01
Fermi motion momentum (M)	± 0.04
Fragmentation and missing modes of X_s (M)	$+0.09$ -0.08
Fraction of $B \to K\ell^+\ell^-$ (M)	$+0.32$ -0.25
Fraction of $B \to K^*\ell^+\ell^-$ (M)	$+0.04$ -0.03
Signal modeling	$+0.34$ -0.28
Total additive	$+0.54$ -0.51
Total multiplicative	$+0.38$ -0.34
Total	$+0.66$ -0.61
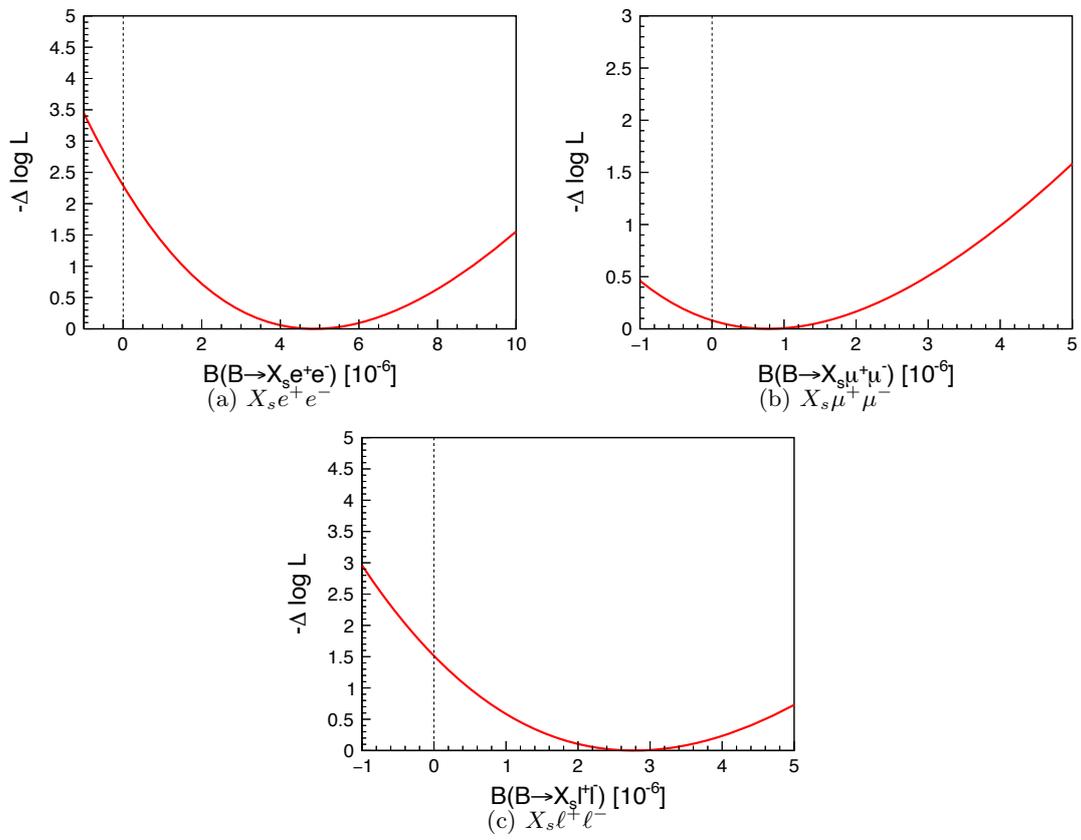
FIG 10.3: Negative log-likelihood statistical-only profiles for the branching fraction of $B \to X_s\ell^+\ell^-$.

where L is the likelihood function convolved with a Gaussian function of the width given by the total systematic uncertainty. TABLE 10.5 shows the results of the branching fractions.

TABLE 10.5: Results on the branching fractions.

Modes	Branching fraction [10^{-6}]
$B \to X_s e^+ e^-$	$4.86^{+2.75}_{-2.42}(\text{stat})^{+1.02}_{-0.92}(\text{syst})$
$B \to X_s \mu^+ \mu^-$	< 4.67 (5.61) at 90% (95%) CL
$B \to X_s \ell^+ \ell^-$	< 5.54 (6.30) at 90% (95%) CL

The results on the branching fractions are compared with other measurements in FIG 10.4. For $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$, the results are consistent with the world average, $(6.67 \pm 0.82) \times 10^{-6}$ and $(4.27^{+0.98}_{-0.91}) \times 10^{-6}$, within 1σ of the uncertainty. Moreover, our results are consistent with other measurements and the SM predictions.

The result of $B \to X_s \ell^+ \ell^-$ is also consistent with the world average, Belle measurement and the SM prediction. Compared with the BaBar measurement, $(6.73^{+0.92}_{-0.87}) \times 10^{-6}$, our results are smaller and the difference is at level 1.4 standard deviation of BaBar measurement.
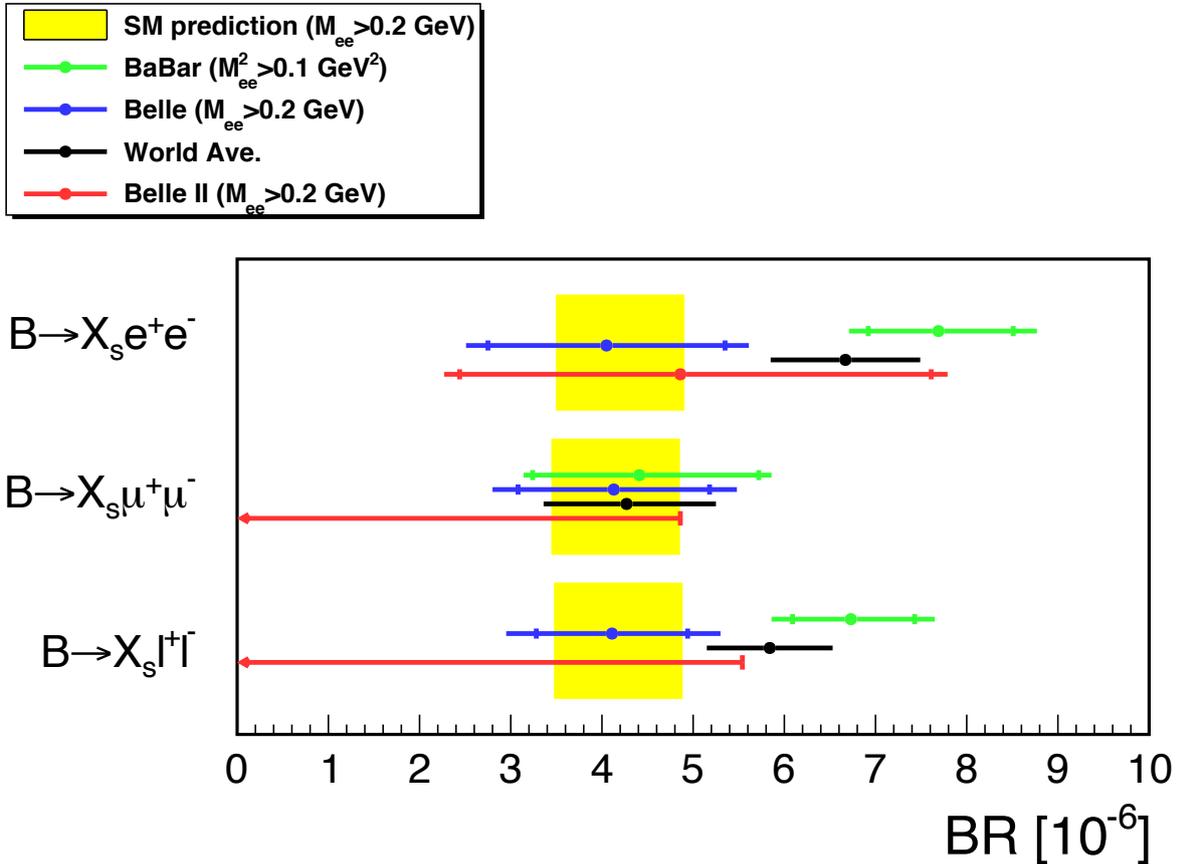
FIG 10.4: Measurement results of the branching fractions of $B \to X_s \ell^+ \ell^-$. Red lines show results of this analysis, green lines show BaBar measurements [16], blue lines show Belle measurement [17], and black lines show the world average [1]. The internal error bars show only the statistical uncertainty and outer ones show the uncertainty combined statistical and systematic one. Yellow band show the SM prediction 2.2.

Since the signal yields are less significance and thus the statistical uncertainty is large, the measurement of R_{X_s} is not performed in this thesis. The R_{X_s} will be determined with very similar way which is shown in this thesis in near future.

Systematic uncertainty

One of the dominant sources of the systematic uncertainties is due to the fitter bias discussed in Section 7.2, which is proportional to the statistical uncertainty. Thus, this uncertainty decreases as the integrated luminosity increases. Another dominant source is the signal modeling of non-resonant X_s, especially the fraction of $B \to K\ell^+\ell^-$ (Section 8.5.5). The uncertainty is caused by the large uncertainty of the SM prediction on the $B \to K\ell^+\ell^-$ branching fraction. The fraction will be directly determined by the measurement of M_{X_s} distribution, $d\Gamma/dM_{X_s}$, in $B \to X_s\ell^+\ell^-$ decays using large statistics. Belle II will also perform the precise measurements of branching fraction on $B \to K^{(*)}\ell^+\ell^-$ which contribute to reduce this uncertainty. Moreover, the fragmentation of X_s can be measured in Belle II and the uncertainty due to that will decrease. Calibration of the Belle II detector is still at early stage. The difference between data and MC on the particle ID and tracking efficiency will be understood and thus the systematic uncertainty due to the efficiency correction and peaking backgrounds will be under better control. Systematic uncertainty due to the FastBDT which a dominant source is limited by the statistics of $B \to X_s J/\psi$ control samples. Thus, this component will decrease when the statistics increase.

Prospect

This thesis is based on the integrated luminosity of $34.6\,\text{fb}^{-1}$ including 37.7×10^6 $B\overline{B}$ pairs. Belle II continues to accumulate data and an integrated luminosity of $600\,\text{fb}^{-1}$ and that of $1.1\,\text{ab}^{-1}$ are expected by 2021 summer and 2022 summer, respectively. The statistical uncertainty on the branching fraction is expected to be reduced by a factor of 4.2 and 5.6 for each stage. Some sources of the systematic uncertainty will also decrease thanks to the large statistics and further calibration as discussed above. Compared with BaBar measurements based on an integrated luminosity of $424\,\text{fb}^{-1}$ [16] and Belle measurements based on partial date corresponding to an integrated luminosity of $140\,\text{fb}^{-1}$ [17], the most precise measurements of the branching fraction will be performed at early stage of Belle II.

In addition, the first measurement of R_{X_s} will be performed for the further test of LFU. The BaBar measurements of the branching fraction correspond to $R_{X_s} = 0.57^{+0.18}_{-0.17}$ assuming the systematic uncertainties are negligible in the ratio. Using the data set corresponding to an integrated luminosity of $1.1\,\text{ab}^{-1}$, the statistical uncertainty is expected to be reduced by a factor of 1.6 compared with the BaBar measurements.If the true values of the branching fraction are at the BaBar measurements, the Belle II is expected to observe a discrepancy on the R_{X_s} from the unity at 3.8σ level. Furthermore, Belle II will continue data taking aiming $50\,\text{ab}^{-1}$ which is one hundred over larger than BaBar statistics. The R_{X_s} measurements using upcoming Belle II full data set will provide decisive information to the lepton flavor universality test.

Chapter 11

Conclusion

The inclusive $B \to X_s \ell^+ \ell^-$ decays are a great probe for physics beyond the standard model of particle physics. Moreover, $B \to X_s \ell^+ \ell^-$ decays provide complementary information with better theoretical uncertainty to the exclusive $B \to K^{(*)} \ell^+ \ell^-$ decays in which tensions from the SM prediction have been observed. Belle II is a unique experiment to explore the process with large statistics to shed light on the anomalies.

We performed the measurements of the branching fractions of $B \to X_s \ell^+ \ell^-$ decay using the data set accumulated by Belle II experiment which corresponds to 37.7 million $B\bar{B}$ pairs. The obtained results are

$$\mathcal{B}(B \to X_s e^+ e^-) = [4.86^{+2.75}_{-2.42}(\text{stat})^{+1.02}_{-0.92}(\text{syst})] \times 10^{-6} \tag{11.1}$$

$$\mathcal{B}(B \to X_s \mu^+ \mu^-) < 4.67(5.61) \times 10^{-6} at 90\%(95\%)CL \tag{11.2}$$

$$\mathcal{B}(B \to X_s \ell^+ \ell^-) < 5.54(6.30) \times 10^{-6} at 90\%(95\%)CL \tag{11.3}$$

Because the statistical significance on $B \to X_s \mu^+ \mu^-$ and $B \to X_s \ell^+ \ell^-$ is less than 2σ, the upper limit on the branching fraction is set for these modes. The branching fraction of $B \to X_s e^+ e^-$ and $B \to X_s \mu^+ \mu^-$ is consistent with previous measurements and the SM prediction. Result of $B \to X_s \ell^+ \ell^-$ is consistent with the world average, Belle measurement and the SM prediction, while the difference from BaBar is at 1.4σ level.

The analysis procedure of $B \to X_s \ell^+ \ell^-$ decays at Belle II experiment well established and we have got ready to lead decisive conclusions regarding the anomalies which are observed in the exclusive $B \to K^{(*)} \ell^+ \ell^-$ decays with upcoming Belle II data.

Appendix A

MC Calibration

In the comparison of the FastBDT input variables between data and MC, a discrepancy is observed in the ΔE distribution. The MC samples were tuned to adequately reproduce the observed data distribution.

The ΔE distributions of the control samples of $B^+ \to K^+\ell^+\ell^-$ and $B^+ \to K^+\pi^0\ell^+\ell^-$ are fitted with Crystal Ball function and the first order polynomial function. Figure A.1 and A.2 show the distributions with the fitting functions. The mean and width obtained by the fitting is summarized in the FIG A.3 and A.4.

To calibrate the MC samples, the ΔE of MC samples are shifted by the mean differences depending the existence of π^0. Moreover, the ΔE distribution of MC samples are smeared with a Gaussian of width given by $\sqrt{\sigma_{\mathrm{data}}^2 - \sigma_{\mathrm{MC}}^2}$.

(a) $K^+e^+e^-$ in MC samples

(b) $K^+\mu^+\mu^-$ in MC samples

(c) $K^+e^+e^-$ in data

(d) $K^+\mu^+\mu^-$ in data

FIG A.1: ΔE distributions of $B \to K^+\ell^+\ell^-$ with the fitting function

(a) $K^+\pi^0 e^+ e^-$ in MC samples

(b) $K^+\pi^0 \mu^+ \mu^-$ in MC samples

(c) $K^+\pi^0 e^+ e^-$ in data

(d) $K^+\pi^0 \mu^+ \mu^-$ in data
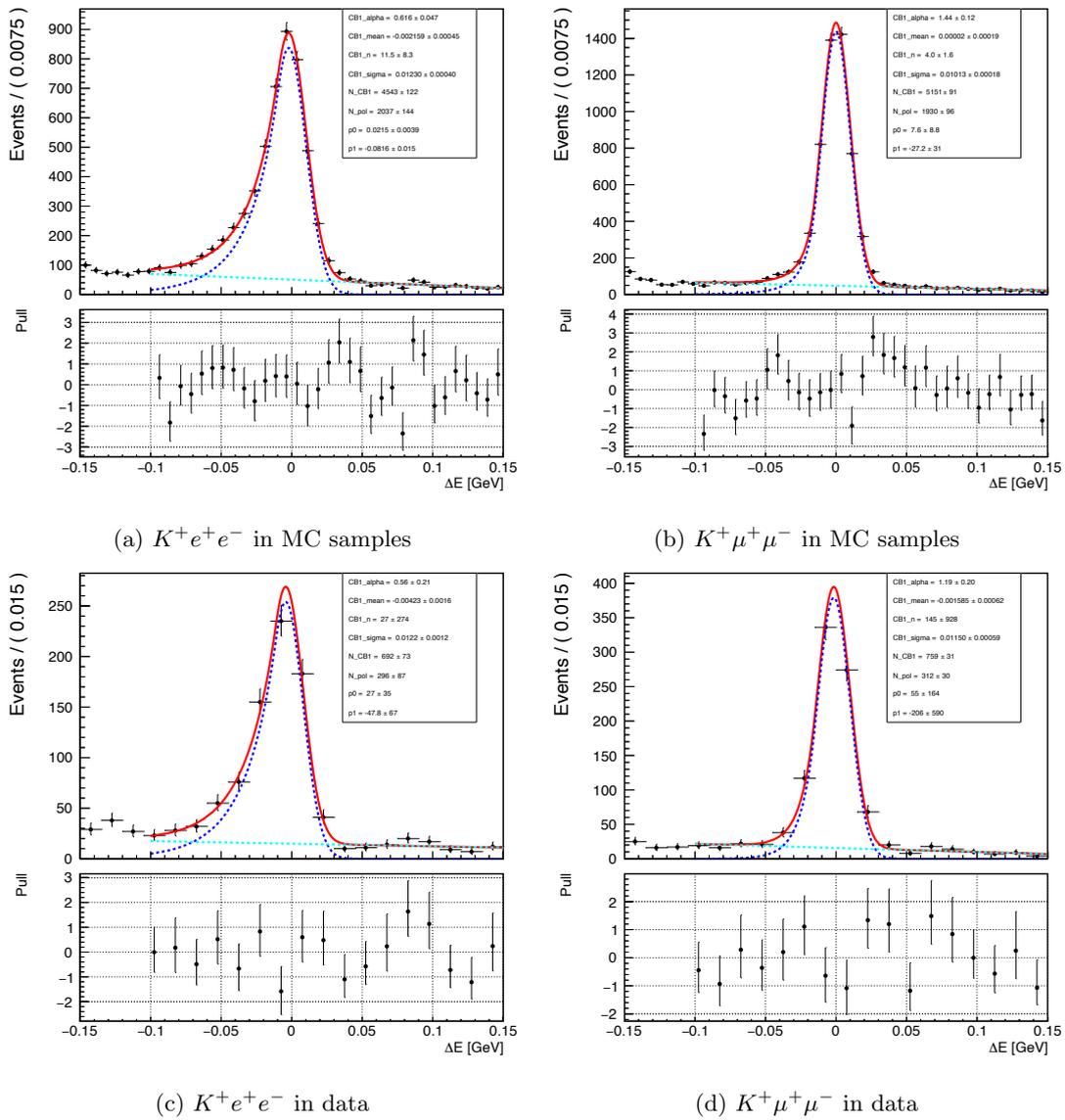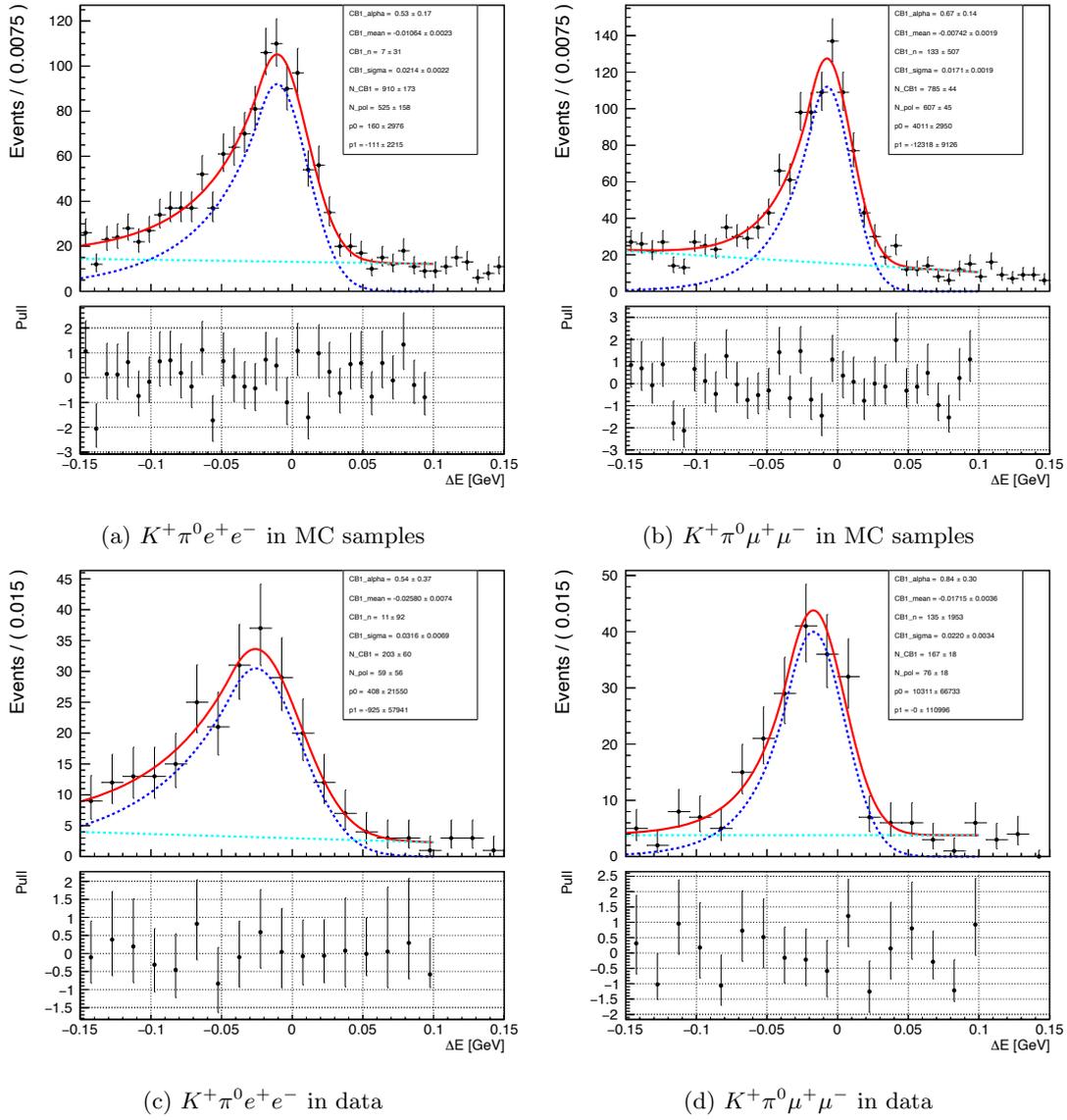
FIG A.2: ΔE distributions of $B \to K^+\pi^0\ell^+\ell^-$ with the fitting function

(a) Mean

(b) Difference of mean
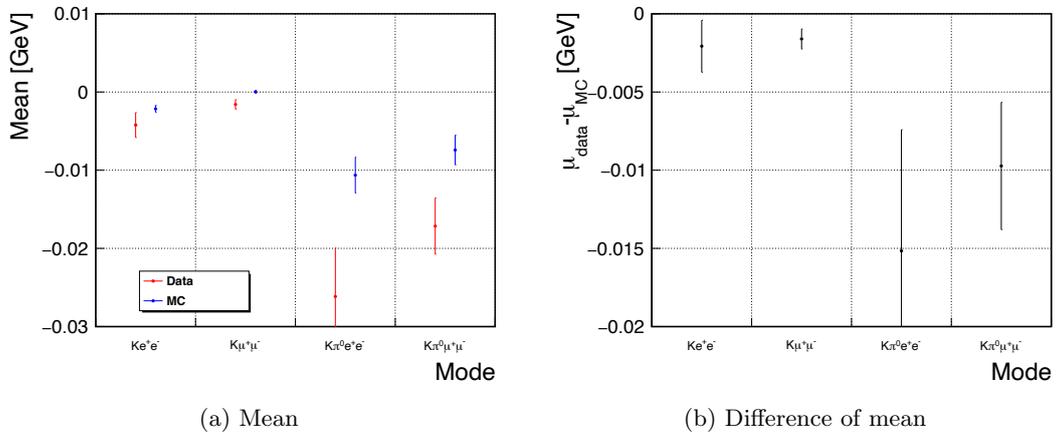
FIG A.3: Mean of the Crystal Ball function and its difference between data and MC
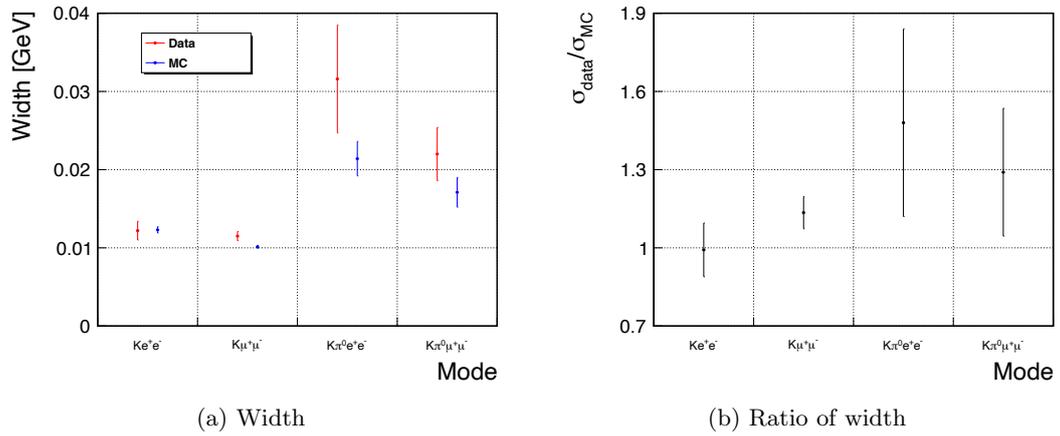
(a) Width

(b) Ratio of width

FIG A.4: Width of the Crystal Ball function and its difference between data and MC

Appendix B

Lepton ID efficiency correction and uncertainty with two-photon events

B.1 Introduction

This chapter describes the method and results of the correction and the systematic uncertainty of the lepton identification (ID) efficiency.

The reconstruction efficiency calculated in Monte-Carlo (MC) sample can be different from the data due to various reasons. It is necessary to know the difference of the efficiency for a better estimation of the branching fraction. Particle ID efficiency should be known precisely for the flavor physics. Especially, the lepton ID efficiency is very important for the lepton-flavor-universality test with $B \to X_s \ell^+ \ell^-$.

The two-photon process $e^+ e^- \to e^+ e^- \ell^+ \ell^-$ is used as signal channel to study the lepton ID efficiency. This process has only two leptons ($\ell^+ \ell^-$) in the detector region and electron-positron ($e^+ e^-$) escapes into the beam pipe. One can use a lepton to probe the lepton ID efficiency by tagging the other lepton, what we call the tag-and-probe method.

B.2 Data set

B.3 Data sample

Data accumulated on the $\Upsilon(4S)$ resonance in the Phase3 of Belle II are used for this study. The data sample are equivalent with that of $B \to X_s \ell^+ \ell^-$ analysis. The amount of data corresponds to an integrated luminosity of $34.6 \, \mathrm{fb}^{-1}$ which.

B.3.1 Monte-Carlo simulation sample

For the simulation study, we have produced MC samples listed in TABLE B.1. TABLE B.1 shows the number of events and the equivalent luminosity of each sample. Phase3 Belle II detector geometry has been used for the detector simulation.

B.4 Selection

B.4.1 Event selection

A signal candidate is reconstructed from two tracks satisfying the following requirements, $p > 0.3 \, \mathrm{GeV}/c, dr < 2.0 \, \mathrm{cm}, |dz| < 5.0 \, \mathrm{cm}$. The mass of electron and muon is assigned to the track in the electron ID study and the muon ID study, respectively. To suppress hadronic events and $\tau^+ \tau^-$ events, the following selection is then applied:

- $p > 0.4$ GeV/c

TABLE B.1: Summary of the MC samples used in this analysis.

Sample	Number of events	Equivalent luminosity $[\mathrm{fb}^{-1}]$
$e^+e^-e^+e^-$	39.74×10^8	100
$e^+e^-\mu^+\mu^-$	18.87×10^8	100
$e^+e^-\pi^+\pi^-$	18.95×10^6	10
$e^+e^-K^+K^-$	0.798×10^6	100
$e^+e^-p\bar{p}$	0.117×10^6	100
$\tau^+\tau^-$	91.9×10^6	100
$e^+e^-(\gamma)$	29.4×10^8	10
$\mu^+\mu^-(\gamma)$	0.574×10^8	50

- $\cos\theta_{\mathrm{open}} > -0.997$, θ_{open} is an opening angle of two tracks in the lab frame. This selection helps to veto cosmic events.

- $|\vec{p}^{\,*}_{z+} + \vec{p}^{\,*}_{z-}| < 2.5$ GeV.

- $E^*_{\mathrm{vis}} < 6.0$ GeV, E_{vis} is the sum of the cluster energies in the ECL

- $|\vec{p}^{\,*}_{t+} + \vec{p}_{t-^*}| < 0.15$ GeV

- $M_{\ell^+\ell^-} < 3.0\,\mathrm{GeV}$

The convention X^* means that the X variable is expressed in the center-of-mass system, $X_{+/-}$ means that the X of positive/negative charge track. Moreover, we required there are only one candidate in a event. The selection $M_{\ell^+\ell^-} < 3.0\,\mathrm{GeV}$ is introduced to suppress the Bhabha events which are not well simulated in MC due to the trigger condition. The detail will be described in the following sections.

B.4.2 Trigger selection

Because the two-photon process has only two tracks in the detector region, the L1 Trigger has a large effect on the reconstruction efficiency. Moreover, the L1 trigger might induce a bias on the lepton ID efficiency, especially the electron ID, if the trigger requires a certain energy deposit in the ECL. Thus, the CDC trigger is preferred for this study. We select candidates which are triggered by CDC-based two-track trigger, which require two full tracks penetrating CDC.

B.4.3 Tag selection

In this method, a lepton is used to suppress backgrounds and identify the lepton flavor and the other lepton is used to probe the efficiency. We applied the following tag selection on a track to select $e^+e^-e^+e^-$ or $e^+e^-\mu^+\mu^-$.

- electron : $PID_e > 0.95$

- muon : $PID_\mu > 0.95$ and $p > 0.7$ GeV/c

The other lepton track is used as a probe particle.

B.5 Lepton ID efficiency calculation

The lepton ID efficiency is calculated using the probe particle which is obtained by the tag selection. The number of events selected with the tag selection is denoted as N_{tag}. Then, we applied the following probe selections on the probe track.

- electron ID efficiency : $PID_e > 0.9$

- muon ID efficiency: $PID_\mu > 0.9$

The number of event satisfying the probe selection is denoted as N_{probe}.

In the MC study, the LID efficiency is calculated from just a fraction of N_{tag} and N_{probe} using only signal samples.

$$\epsilon^{\text{MC}} = \frac{N_{\text{probe}}}{N_{\text{tag}}}. \tag{B.1}$$

On the other hand, N_{tag} and N_{probe} must contain some background events in the data. The contribution of the backgrounds should be subtracted to obtain the correct LID efficiency. We define the LID efficiency in the data as the following.

$$\epsilon^{\text{Data}} = \frac{N_{\text{probe}} - f \cdot \sum_T \sum_P n_{\text{probe}}^{T,P} \cdot r_T \cdot r_P}{N_{\text{tag}} - f \cdot \sum_T \sum_P n_{\text{tag}}^{T,P} \cdot r_T} \tag{B.2}$$

where f is the fraction of number of events between data and MC before tag and probe selections. $n_{\text{tag/probe}}^{T,P}$ is the number of background events estimated in MC and $r_{T(P)}$ is a correction of mis-ID probability or efficiency between data and MC. T and P are the type of particle on the tag and probe side track, respectively ($T, P = e, \mu, \pi, K, p$). Note that, we eliminate $T = P = e$ and $T = P = \mu$ cases for the electron, muon efficiency, respectively.

Momentum of the probe particle and $M_{\ell^+\ell^-}$ distributions of the data samples and background events estimated with MC sample are shown in FIG B.1 and FIG B.2 for electron ID efficiency study and muon ID efficiency study, respectively. Dominant source of the backgrounds is other two photon events, mainly $e^+e^-\pi^+\pi^-$ events.
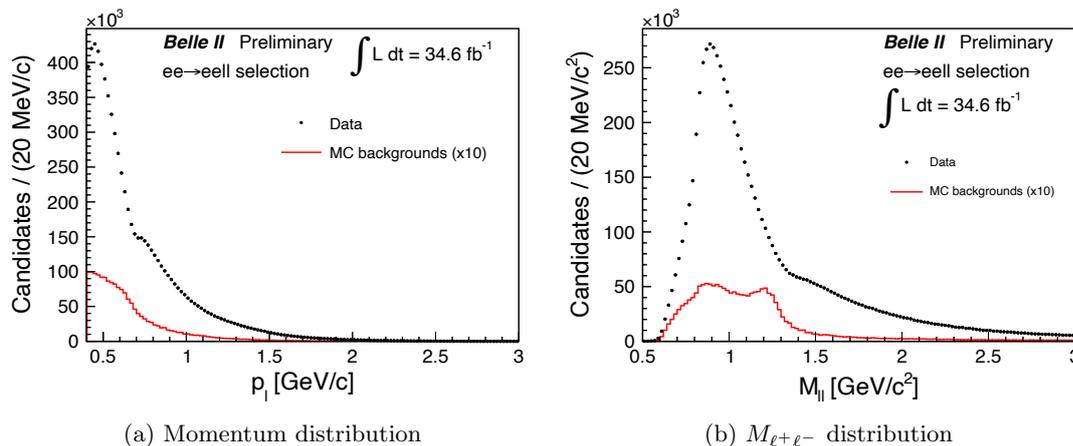

(a) Momentum distribution

(b) $M_{\ell^+\ell^-}$ distribution

FIG B.1: Momentum of the probe particle and $M_{\ell^+\ell^-}$ distributions after the electron ID probe selection. Black line shows the data sample and red line shows background events estimated with MC. MC distributions are multiplied by 10 for visualization purpose.

B.6 Systematic uncertainty

B.6.1 Correction between data and MC

The backgrounds estimated the MC samples are corrected using correction factors with hadron mis-ID probability obtained with data driven analysis []. The uncertainty of the correction factors are included as systematic uncertainty.

B.6.2 Generator effects

The generation of MC samples of hadronic two photon events, such as $e^+e^-\pi^+\pi^-$, has irreducible uncertainties due to the less knowledge of intermediate states in the process. The uncertainty is typically up to 10%. Systematic uncertainty because of the generator uncertainty is calculated by varying $n^{T,P}(T, P = \pi, K, p)$ from $0.9 \cdot n^{T,P}$ to $1.1 \cdot n^{T,P}$. The interval of ϵ^{Data} is taken into account the systematic uncertainty.

(a) Momentum distribution
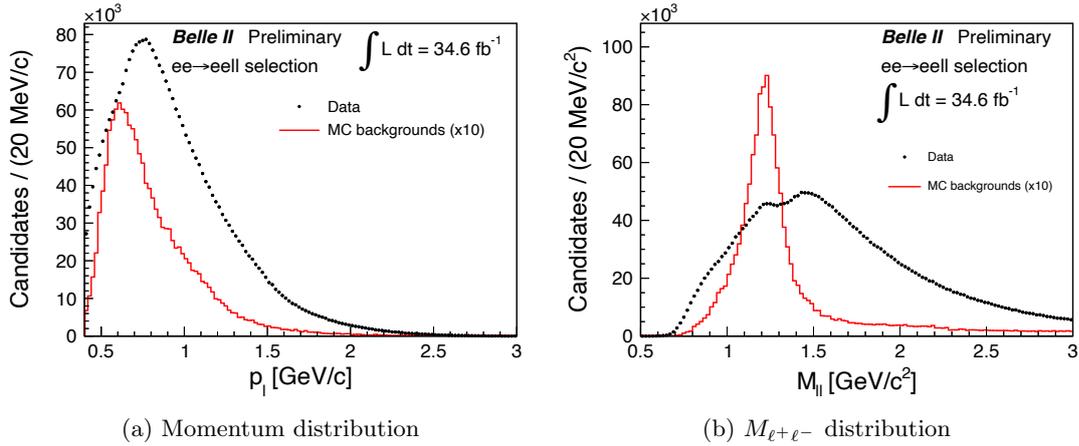
(b) $M_{\ell^+\ell^-}$ distribution

FIG B.2: Momentum of the probe particle and $M_{\ell^+\ell^-}$ distributions after the muon ID probe selection. Black line shows the data sample and red line shows background events estimated with MC. MC distributions are multiplied by 10 for visualization purpose.

B.7 Binning definition

The lepton ID efficiency is calculated as function of momentum and polar angle of lepton. TABLE B.2 and TABLE B.3 show the definition binning of electronID, muonID efficiency. The bins are chosen according to the ECL and KLM geometry.

The lepton efficiency is calculated for bins in which there are at least 50 events before the tag selection.

TABLE B.2: Definition of binning for electronID study.

p_{LAB} bins [GeV]	θ bins [rad.]	
$0.4 - 0.5$	$0.22 - 0.56$	(FWD end-cap)
$0.5 - 1.0$	$0.56 - 1.13$	(FWD barrel)
$1.0 - 1.5$	$1.13 - 1.57$	(Central FWD barrel)
$1.5 - 2.0$	$1.57 - 1.88$	(Central BWD barrel)
$2.0 - 2.5$	$1.88 - 2.23$	(BWD barrel)
$2.5 - 3.0$	$2.23 - 2.71$	(BWD end-cap)

B.8 Results

B.8.1 electron ID efficiency

Figure B.3 and B.4 show the electron ID efficiency and the correction between data and MC for $PID_e > 0.9$.

The efficiency at $> 1\,\mathrm{GeV}$ regions is around 98%. In the region of $0.5 < p < 1.0\,\mathrm{GeV}/c$, the CDC dE/dx band of electron and that of kaon are crossing and thus the efficiency is less than other region.

The efficiency ratio in central barrel region is stable at about 0.98. Since tracks pass through diagonally to the CDC wires in the forward and backward region, the calibration of dE/dx is challenging. Fluctuation of the ratio in the region will be decreased by dedicated calibration study.

muon ID efficiency

Figure B.5 and B.6 show results of the muon ID efficiency for $PID_\mu > 0.9$.

TABLE B.3: Definition of binning for muonID study.

p_{LAB} bins [GeV]	θ bins [rad.]	
$0.4 - 0.5$	$0.40 - 0.64$	(FWD EKLM)
$0.5 - 0.7$	$0.64 - 0.82$	(FWD EKLM-BKLM separation)
$0.7 - 1.0$	$0.82 - 1.16$	(FWD BKLM)
$1.0 - 1.5$	$1.16 - 1.46$	(Central BKLM)
$1.5 - 2.0$	$1.46 - 1.78$	(BWD BKLM)
$2.0 - 2.5$	$1.78 - 2.13$	(BKLM-solenoid-chimney)
$2.5 - 3.0$	$2.13 - 2.22$	(BWD EKLM-BKLM separation)
	$2.22 - 2.60$	(BWD EKLM)
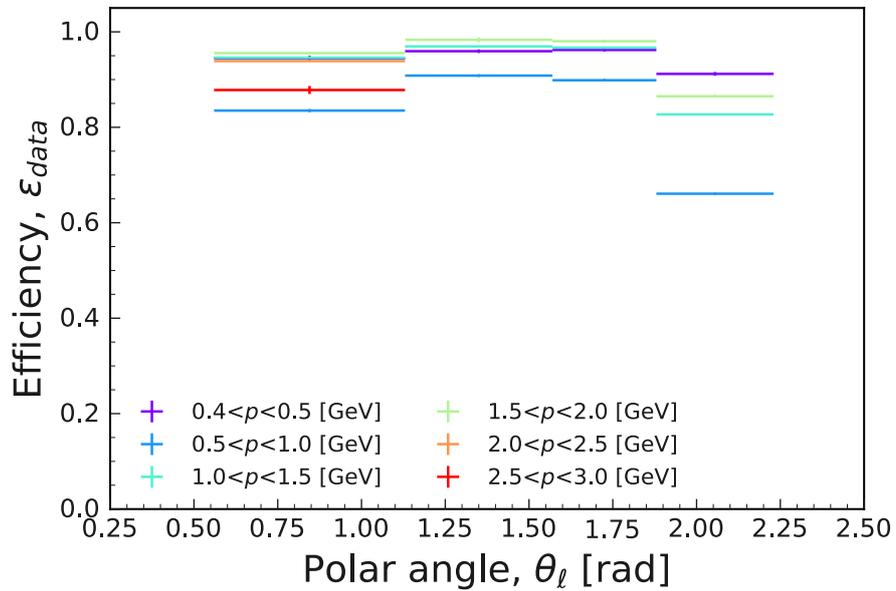
FIG B.3: Electron ID efficiency of data for $PID_e > 0.9$.
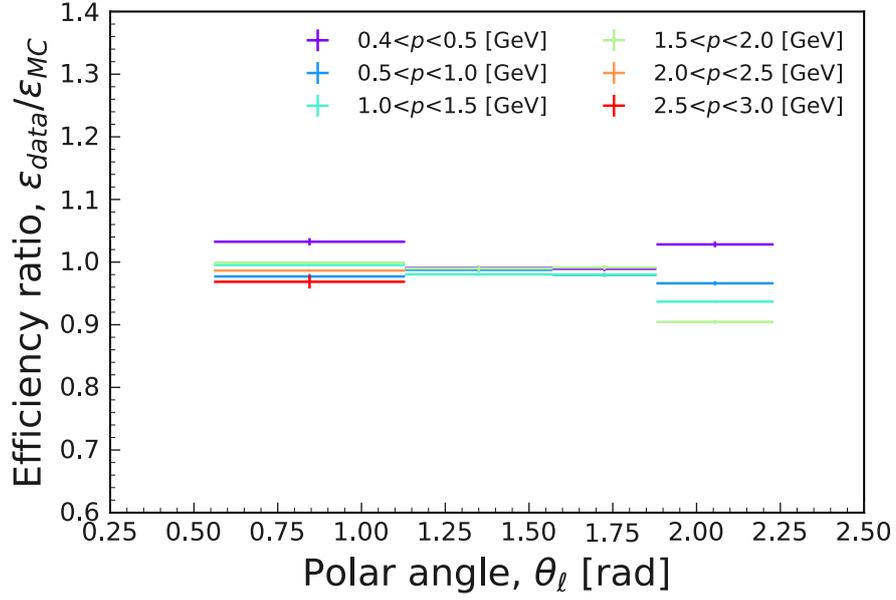
FIG B.4: Ratio of electron ID efficiency between data and MC for $PID_e > 0.9$.

Since low momentum tracks cannot reach the KLM, the efficiency in the region of $p < 0.7\,\mathrm{GeV}/c$ is lower than the other regions. The efficiency is increasing as a function of p and that at $p > 1.5$ GeV is around 90%.

The efficiency ratio in the low momentum regions is smaller than unity. In the high momentum region, the ratio is around 0.95.
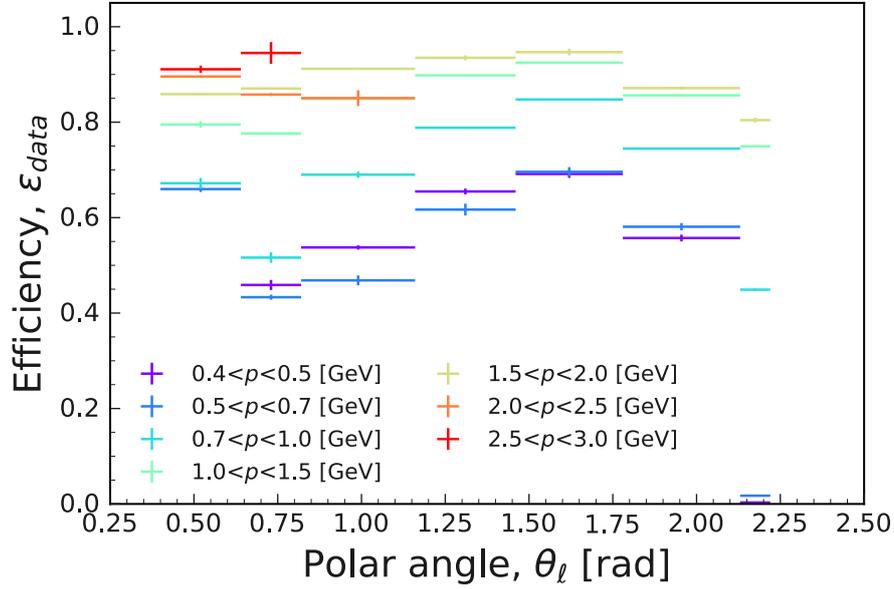
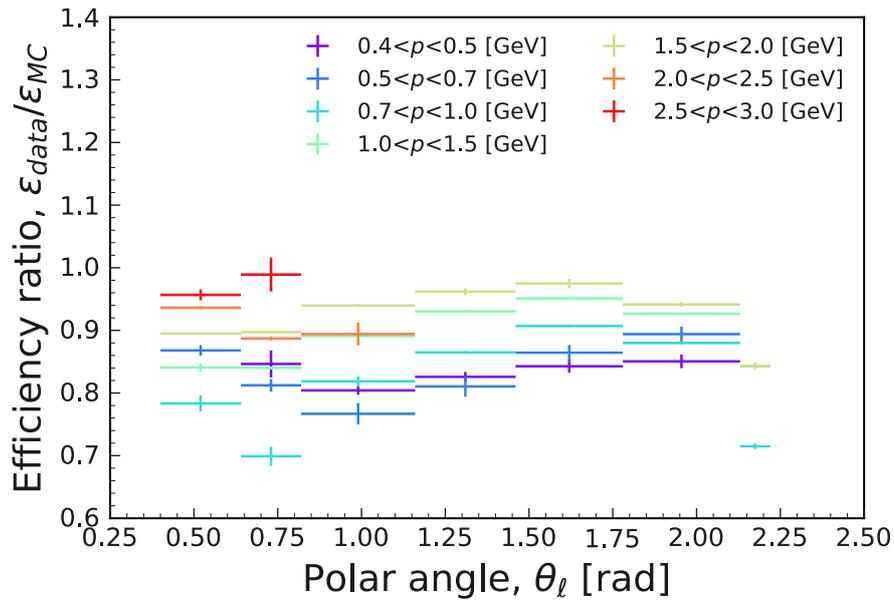FIG B.5: Electron ID efficiency of data for $PID_\mu > 0.9$.

FIG B.6: Ratio of electron ID efficiency between data and MC for $PID_\mu > 0.9$.

Bibliography

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

[2] Y. Sato et al., Belle, *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) no. 3, 032008, `arXiv:1402.7134 [hep-ex]`. [Addendum: Phys.Rev.D 93, 059901 (2016)].

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

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

[5] B. Capdevila, S. Descotes-Genon, J. Matias, and J. Virto, *Assessing lepton-flavour non-universality from $B \to K^*\ell\ell$ angular analyses*, JHEP **10** (2016) 075, `arXiv:1605.03156 [hep-ph]`.

[6] M. Algueró, B. Capdevila, A. Crivellin, S. Descotes-Genon, P. Masjuan, J. Matias, M. Novoa Brunet, and J. Virto, *Emerging patterns of New Physics with and without Lepton Flavour Universal contributions*, Eur. Phys. J. C **79** (2019) no. 8, 714, `arXiv:1903.09578 [hep-ph]`. [Addendum: Eur.Phys.J.C 80, 511 (2020)].

[7] W. Altmannshofer et al., Belle-II, *The Belle II Physics Book*, PTEP **2019** (2019) no. 12, 123C01, `arXiv:1808.10567 [hep-ex]`. [Erratum: PTEP 2020, 029201 (2020)].

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

[9] T. Abe, Belle II Collaboration, *Belle II Technical Design Report*, `arXiv:1011.0352 [physics.ins-det]`.

[10] Y. Maeda, Belle-II TOP Group, *Status of installation and commissioning for the Belle II time-of-propagation counter*, JINST **12** (2017) no. 08, C08005, `arXiv:1705.09895 [physics.ins-det]`.

[11] M. Bessner, Belle-IIiTOP Group, *Performance of the Belle II imaging Time-Of-Propagation (iTOP) detector in first collisions*, Nucl. Instrum. Meth. A **958** (2020) 162318.

[12] Y. Masanobu, *Particle Identification using the Aerogel RICH Counter at the Belle II Experiment*, Doctoral Thesis, Tokyo Metropolitan University (2020) .

[13] Y. Lai et al., *Performance of the Belle II aerogel-based ring-imaging Cherenkov counter system in SuperKEKB 2019 phase 3 operation*, JINST **15** (2020) no. 07, C07039.

[14] The Belle II collaboration, *Muon and electron identification efficiencies and hadron-lepton mis-identification probabilities*, .

[15] The Belle II Collaboration, *Kaon and Pion Identification Performances in Phase III data*, .

[16] J. Lees et al., BaBar, *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 [hep-ex]`.

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

[18] G. Aad et al., ATLAS, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1–29, `arXiv:1207.7214 [hep-ex]`.

[19] S. Chatrchyan et al., CMS, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30–61, `arXiv:1207.7235 [hep-ex]`.

[20] Y. Golfand and E. Likhtman, *Extension of the Algebra of Poincare Group Generators and Violation of p Invariance*, JETP Lett. **13** (1971) 323–326.

[21] D. Volkov and V. Akulov, *Is the Neutrino a Goldstone Particle?*, Phys. Lett. B **46** (1973) 109–110.

[22] J. Wess and B. Zumino, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. B **70** (1974) 39–50.

[23] J. Wess and B. Zumino, *Supergauge Invariant Extension of Quantum Electrodynamics*, Nucl. Phys. B **78** (1974) 1.

[24] S. Ferrara and B. Zumino, *Supergauge Invariant Yang-Mills Theories*, Nucl. Phys. B **79** (1974) 413.

[25] A. Salam and J. Strathdee, *Supersymmetry and Nonabelian Gauges*, Phys. Lett. B **51** (1974) 353–355.

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

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

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

[29] K. Abe et al., Belle, *Observation of the decay $B \to K \ell^+ \ell^-$*, Phys. Rev. Lett. **88** (2002) 021801, `arXiv:hep-ex/0109026`.

[30] G. Altarelli and L. Maiani, *Octet Enhancement of Nonleptonic Weak Interactions in Asymptotically Free Gauge Theories*, Phys. Lett. B **52** (1974) 351–354.

[31] M. K. Gaillard and B. W. Lee, *$\Delta I = 1/2$ Rule for Nonleptonic Decays in Asymptotically Free Field Theories*, Phys. Rev. Lett. **33** (1974) 108.

[32] E. Witten, *Short Distance Analysis of Weak Interactions*, Nucl. Phys. B **122** (1977) 109–143.

[33] A. J. Bevan et al., BaBar, Belle, *The Physics of the B Factories*, Eur. Phys. J. C **74** (2014) 3026, `arXiv:1406.6311 [hep-ex]`.

[34] H. H. Asatrian, H. M. Asatrian, C. Greub, and M. Walker, *Two loop virtual corrections to $B \to X_s l^+ l^-$ in the standard model*, Phys. Lett. B **507** (2001) 162–172, `arXiv:hep-ph/0103087`.

[35] 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`.

[36] B. Grinstein, M. J. Savage, and M. B. Wise, *$B \longrightarrow X(s) e+ e-$ in the Six Quark Model*, Nucl. Phys. B **319** (1989) 271–290.

[37] A. Ali, T. Mannel, and T. Morozumi, *Forward backward asymmetry of dilepton angular distribution in the decay $b \longrightarrow s l+ l-$*, Phys. Lett. B **273** (1991) 505–512.

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

[39] G. Hiller and M. Schmaltz, *Diagnosing lepton-nonuniversality in $b \to s\ell\ell$*, JHEP **02** (2015) 055, `arXiv:1411.4773 [hep-ph]`.

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

[41] B. Aubert et al., BaBar, *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`.

[42] 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`.

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

[44] B. Aubert et al., BaBar, *Direct CP, Lepton Flavor and Isospin Asymmetries in the Decays $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. Lett. **102** (2009) 091803, `arXiv:0807.4119 [hep-ex]`.

[45] J. Lees et al., BaBar, *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 [hep-ex]`.

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

[47] T. Aaltonen et al., CDF, *Observation of the Baryonic Flavor-Changing Neutral Current Decay $\Lambda_b \to \Lambda\mu^+\mu^-$*, Phys. Rev. Lett. **107** (2011) 201802, `arXiv:1107.3753 [hep-ex]`.

[48] V. Khachatryan et al., CMS, LHCb, *Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68–72, `arXiv:1411.4413 [hep-ex]`.

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

[50] M. Aaboud et al., ATLAS, *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 [hep-ex]`.

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

[52] J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl, and D. M. Straub, *B-decay discrepancies after Moriond 2019*, Eur. Phys. J. C **80** (2020) no. 3, 252, `arXiv:1903.10434 [hep-ph]`.

[53] E. Mulyani and J. Flanagan, *Calibration of X-ray Monitor during the Phase I of SuperKEKB commissioning*, in *5th International Beam Instrumentation Conference*, p. TUPG72. 2017.

[54] D. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152–155.

[55] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177, `arXiv:1410.3012 [hep-ph]`.

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

[57] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97–107, `arXiv:hep-ph/0506026`.

[58] S. Agostinelli et al., GEANT4, *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth. **A506** (2003) 250–303.

[59] A. Ali, P. Ball, L. 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`.

[60] F. Kruger and L. Sehgal, *Lepton polarization in the decays $b \longrightarrow X(s)$ mu+ mu- and $B \longrightarrow X(s)$ tau+ tau-*, Phys. Lett. B **380** (1996) 199–204, `arXiv:hep-ph/9603237`.

[61] A. Ali, G. Hiller, L. Handoko, and T. Morozumi, *Power corrections in the decay rate and distributions in $B \to X_s l^+ l^-$ in the Standard Model*, Phys. Rev. D **55** (1997) 4105–4128, `arXiv:hep-ph/9609449`.

[62] H. Asatryan, H. Asatrian, C. Greub, and M. Walker, *Calculation of two loop virtual corrections to $b \to s l^+ l^-$ in the standard model*, Phys. Rev. D **65** (2002) 074004, `arXiv:hep-ph/0109140`.

[63] C. Bobeth, M. Misiak, and J. Urban, *Photonic penguins at two loops and m_t dependence of $BR[B \to X_s l^+ l^-]$*, Nucl. Phys. B **574** (2000) 291–330, `arXiv:hep-ph/9910220`.

[64] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, Comput. Softw. Big Sci. **1** (2017) no. 1, 2.

[65] J. H. Friedman, *Stochastic gradient boosting*, Comput. Stat. Data Anal. **38** (2002) 367–378.

[66] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, `arXiv:physics/0703039`.

[67] S. Lee et al., Belle, *Evidence for $B^0 \to \pi^0 \pi^0$*, Phys. Rev. Lett. **91** (2003) 261801, `arXiv:hep-ex/0308040`.

[68] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes*, Phys. Rev. Lett. **41** (1978) 1581.

[69] D. Asner et al., CLEO, *Search for exclusive charmless hadronic B decays*, Phys. Rev. D **53** (1996) 1039–1050, `arXiv:hep-ex/9508004`.

[70] The Belle II Collaboration, *Approved B counting analysis results with ICHEP2020 dataset*, .

[71] The Belle II Collaboration, *Measurement of the tracking efficiency and fake rate with e+e- -> tau+tau- events*, .

[72] H. Guler et al., Belle, *Study of the $K^+ \pi^+ \pi^-$ Final State in $B^+ \to J/\psi K^+ \pi^+ \pi^-$ and $B^+ \to \psi' K^+ \pi^+ \pi^-$*, Phys. Rev. D **83** (2011) 032005, `arXiv:1009.5256 [hep-ex]`.