Ph.D. Thesis

Measurement of Forward-Backward Asymmetry in the $B \to X_s \ell^+ \ell^-$ Decay with Semi-inclusive Method

(準包括的再構成法を用いた$B \to X_s \ell^+ \ell^-$ 崩壊における レプトン前後非対称度の測定)

Department of Physics, Tohoku University

Yutaro Sato

February 2013

Abstract

A $b \to s\ell^+\ell^-$ decay is a Flavor Changing Neutral Current (FCNC) process, which proceeds via a loop diagram in the Standard Model (SM). Since new particle may appear in the loop, the $b \to s\ell^+\ell^-$ decay is a good probe for searching new physics particles beyond the Standard model.

We report measurement of the forward-backward asymmetry (A_{FB}) in $B \to X_s\ell^+\ell^-$ decays with semi-inclusive reconstruction method for the first time, where ℓ is an electron or a muon and X_s is a hadronic system containing an s-quark. We measure the A_{FB} as

$$\begin{array}{rcl}
A_{FB}(\qquad q^2 < 4.3\ \mathrm{GeV}^2/c^4 \qquad) & = & 0.34 \pm 0.24(\mathrm{stat}) \pm 0.02(\mathrm{syst}) \\
A_{FB}(\ 4.3\ \mathrm{GeV}^2/c^4 < q^2 < J/\psi\ \text{veto region}\ ) & = & 0.04 \pm 0.31(\mathrm{stat}) \pm 0.05(\mathrm{syst}) \\
A_{FB}(q^2\ \text{between}\ J/\psi \text{and}\ \psi(2S) \text{veto regions}) & = & 0.28 \pm 0.21(\mathrm{stat}) \pm 0.01(\mathrm{syst}) \\
A_{FB}(\qquad q^2\ \text{above}\ \psi(2S)\ \text{veto region} \qquad) & = & 0.28 \pm 0.15(\mathrm{stat}) \pm 0.01(\mathrm{syst}),
\end{array}$$

where $q^2 = (p_{\ell^+} + p_{\ell^-})^2$ is the dilepton mass-square. Our result in high q^2 bin is consistent with the theoretical SM prediction and favors $C_{10}{\cdot}\mathrm{Re}(C_9)$ term to be negative with 2.3σ (97.9% C.L.). Our result in the lowest q^2 bin is 1.8σ (6.6% C.L.) away from the theoretical SM prediction. Our result is used to constrain the new physics beyond the Standard Model. The results are based on a data sample of 711 fb^{-1} recorded with the Belle detector at the KEKB e^+e^- storage ring, corresponding to 772×10^6 $B\bar{B}$ pairs at the $\Upsilon(4S)$ resonance.

Acknowledgement

First, I would like to thank my advisor, Prof. Hitoshi Yamamoto, for giving me a chance on a research in high energy physics in such a great place.

I greatly appreciate Dr. Akimasa Ishikawa for his helpful advice and bright ideas for the analyses. I could have never finished my thesis without his support.

My thanks go to all the members of the Belle and KEKB. I am especially grateful to Prof. Y. Sakai, Dr. M. Nakao, Y. Kwon, Dr. T. Karim Dr. S. Nishida, A. Rostomyan, S. Vahsen, Dr. T. Kawasaki, T. Kuhr, J. Wicht, K. Kinoshita, Prof. S. Uno, and Dr. N. Taniguchi for their help and valuable suggestions.

I also owe a lot to the members of Tohoku University. Dr. T. Sanuki, Dr. T. Nagamine, Dr. Y. Onuki, Dr. Y. Takubo, and Dr. Y. Hori gave me useful comments and warm encouragement. I owe a great deal to my colleagues, with whom I shared a life in the laboratory. I would like to thank secretary Ms. K. Kobayashi for her help.

Last but not the least, I would like to express my gratitude to my parents. The project would not have been completed without their moral support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The goal of the elementary particle physics is to understand the fundamental elements and their interactions. Many theoretical and experimental physicists have made great efforts to give an answer to this question. These efforts lead to a construction of the standard model (SM). In the SM, the matter consists of quarks and leptons which are spin-1/2 fermions. The electromagnetic, weak and strong interactions, which are mediated by spin-1 bosons, are described by a Gauge field theory. The SM explains almost all the phenomena of the elemental particles measured so far. A boson consistent with the Higgs boson, which was the last unobserved particle of the SM, has been observed by the ATLAS [1] and CMS [2] experiments at the Large Hadron Collider (LHC) at CERN in 2012. The SM has been very successful, although there are many unresolved problems that cannot be answered by the SM. For example, the SM does not have a candidate for dark matter, not incorporate the full theory of gravitation, and so on. It is generally believed that there still exists some hidden new physics beyond the SM.

A search for the new physics is within the main focus of particle physics. In principle, there are two ways to search for the new physics. One is direct approach at the energy frontier, such as ATLAS and CMS experiments, in which the new particles could be produced and observed directly as real particles with specific signatures. The other is indirect search by the intensity frontier, such as Belle, BABAR and LHCb, in which the new particles appear as virtual particles in loop processes, leading to measure deviations from the pure SM expectations. The direct and indirect approaches are complementary. The LHC experiment accumulates data well, however, evidence of the new physics has not been found so far. Figure 1.1 shows an example of the new physics search result. The new physics at TeV scale needs to have a flavor structure to provide the suppression mechanism for already observed flavor changing neutral current (FCNC) processes. Once the new physics is discovered, it is important to measure this structure by the indirect search. If the evidence of new physics is not observed by the direct search from now on, the indirect search is more important. The sensitivity of the indirect search to the new physics depends on the strength of the flavor violating coupling of the new physics. The mass reach for new particle can be as high as $\mathcal{O}$ (100 TeV/c^2), if the couplings are enhanced compared to the SM.

One of most effective modes in indirect search is rare decays that are forbidden or suppressed in the SM but may be enhanced in the non-SM physics. In the SM, the FCNC, such as $b \to s$ (or d) transition, are forbidden at a tree level. As the lowest order of diagram, these transitions are possible proceeding with a loop. Instead of a virtual

W, a charged Higgs or non-SM particles can mediate the loop. Therefore, it is excellent probe to new physics.

In this thesis, we report a measurement of the forward-backward asymmetry in inclusive $B \to X_s \ell^+\ell^-$ decay, which is one of the FCNC processes, where X_s is a hadronic system including an s quark. We use a data sample containing 772×10^6 $B\bar{B}$ pairs, which is recorded at the $\Upsilon(4S)$ resonance with Belle detector at the KEKB e^+e^- collider.

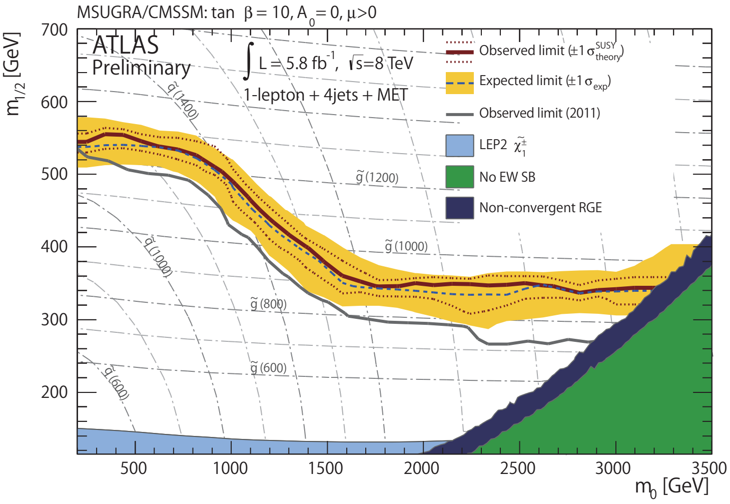


Figure 1.1: 95% CL exclusion limit presented in the m_0 - $m_{1/2}$ MSUGRA/CMSSM plane with tan$\beta = 10$, $A_0 = 0$ and $\mu > 0$ [3]. The blue dashed line show the expected limits at 95% CL, with the light (yellow) bands indicating the $\pm 1\sigma$ variation on the median expected limit due to the experimental uncertainties. $\pm 1\sigma$ variation on the median expected limit due to the experimental uncertainties. The observed nominal limit is indicated by a solid dark red line with the dark red dotted lines being obtained by varying the signal cross section by the scale and PDF uncertainties. The observed limit set by the previous ATLAS analysis [4] using 7 TeV data is shown as a solid grey line. Some regions are theoretically excluded because there is no electroweak symmetry breaking (green area) or because the numerical algorithms solving the renormalization group equations do not converge (dark blue area).

Chapter 2

Theoretical Framework and Physics Motivation

2.1 Flavor Changing Neutral Current (FCNC)

Most of the B decays are $b \to c$ transitions at the quark level. Compared to the $b \to c$ transitions, $b \to u$ transitions are suppressed by the CKM factor ($|V_{ub}/V_{cb}|^2 \sim 0.01$). In the SM, $b \to c$ or u transitions are allowed via tree diagram as shown in Fig. 2.1. These decays are called flavor changing charged currents. On the other hand, $b \to s$ or d transitions are forbidden at tree level, and can only be generated via a loop diagram. These decays are called flavor changing neutral currents (FCNC). The rate for the $b \to s$ transitions are comparable to CKM-suppressed $b \to u$ transitions.

The FCNC process all involve the emission of a neutral boson, either a gluon, photon, or Z boson. Figure 2.2 shows examples of the FCNC processes. Figure 2.2(a) shows a radiative penguin [1] process in which an external real photon is emitted. Figure 2.2(b) shows electroweak penguin process in which an emitted virtual photon or Z boson produce lepton pairs. Figure 2.2(c) and 2.2(d) show the annihilation penguin processes in which the loop connects the two quarks in the B meson. Figure 2.2(e) shows the gluonic process in which a gluon is emitted from the loop. The diagrams as shown here are just one example. Box diagram as shown in Fig. 2.2(f) also contributes to the radiative process with lepton pairs and the annihilation penguin process.

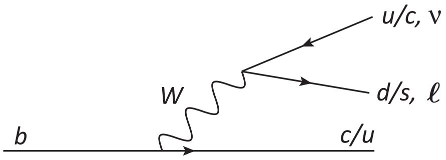
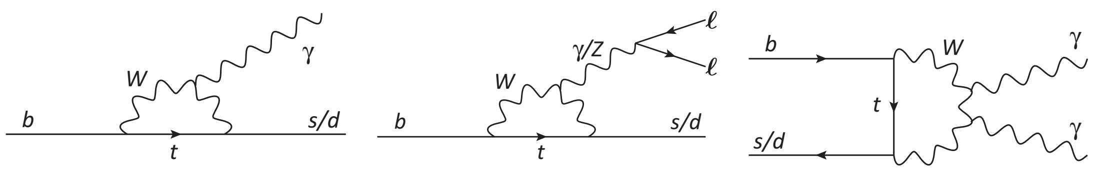


Figure 2.1: $b \to c$ or u transition.

If non-SM particles exist, they can appear in the loop. For examples in the supersymmetry scenarios, (i) the up-type quarks and the charged Higgs, (ii) the up-type squarks and charginos (iii) the down-type squarks and neutralinos, and (iv) the down-type squarks

---

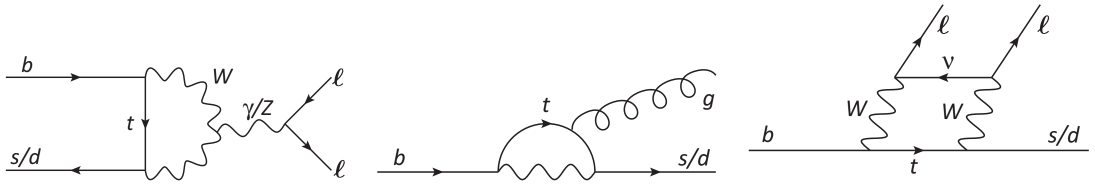
[1] The name penguin decays was first introduced in Ref. [6] as the results of a bet. A more detailed account of the name can be found in Ref. [7].



(a) $b \to s\gamma, d\gamma$.

(b) $b \to s\ell^+\ell^-, d\ell^+\ell^-$.

(c) $B \to \gamma\gamma$.



(d) $B \to \ell^+\ell^-$.

(e) $b \to sg, dg$.

(f) Box diagram for $b \to s\ell^+\ell^-, d\ell^+\ell^-$ or $B \to \ell^+\ell^-$.

Figure 2.2: Examples of FCNC processes.

and gluinos, can appear in the loop [5]. We can obtain information on the masses and couplings of the virtual particles running in the loops by measuring the FCNC processes precisely.

The $b \to d\gamma$ and $b \to s\ell^+\ell^-$ are suppressed with respect to $b \to s\gamma$ in the SM by two orders of magnitude mainly due to additional $|V_{td}/V_{ts}|^2$ and α_{em} factors, respectively. The $b \to d\ell^+\ell^-$ is suppressed by four orders of magnitude due to both factors. The annihilation penguin processes is highly suppressed compared to the radiative and electroweak penguin processes, due to the extra α_{QED} suppression for $B \to \gamma\gamma$ and helicity-suppression for $B \to \ell^+\ell^-$. In addition, since the two quarks must find each other, the $\mathcal{O}(f_B^2/m_B^2)$ suppression factor is added.

2.2 Effective Hamiltonian Approach

The low-energy effective Hamiltonian constructed using the operator product expansion (OPE) plays a crucial role to analyze B decays. The relevant length scale of B meson decay is of the order of $\mathcal{O}(1/m_b)$ while the relevant length scale of loops leading to the FCNC processes is of the order of $\mathcal{O}(1/m_W)$, where m_b and m_W are the mass of the b quark and W boson, respectively. The OPE describes effective Hamiltonian, separating two different scales. The "long-distance" contributions contained in the operator matrix elements are described by $\mathcal{O}_i(\mu)$ and the "short distance" physics is described by the Wilson coefficients $C_i(\mu)$. The $\mathcal{O}_i$ and C_i, intuitively speaking, are considered as effective vertices and the corresponding coupling constants.

The effective Hamiltonian in the SM inducing the $b \to s\ell^+\ell^-$ and $b \to s\gamma$ transitions can be expressed as follows [8]:

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \left(\sum_{i=1}^{10} \left(C_i(\mu)\mathcal{O}_i(\mu) + C_i'(\mu)\mathcal{O}_i'(\mu) \right) \right), \tag{2.1}$$

where $\mathcal{O}_i(\mu)$ are dimension-six operators at the scale μ, $C_i(\mu)$ are the corresponding Wilson coefficients, G_F is the Fermi coupling constant, and the CKM dependence has been made explicit. The primed operators are chirality flipped compared to the non-primed operators, and they are highly suppressed in the SM.

$\mathcal{O}_1$ and $\mathcal{O}_2$ are current-current operator, $\mathcal{O}_{3-6}$ are QCD penguin operator, $\mathcal{O}_7$ and $\mathcal{O}_8$ are the electromagnetic and chromomagnetic operator, and $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the vector component and axial component of electroweak penguin operator, respectively. The operators are described as:

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

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

$$\mathcal{O}_3 = (\bar{s}_L\gamma_\mu b_L)\sum_q(\bar{q}\gamma^\mu q)\,, \tag{2.4}$$

$$\mathcal{O}_4 = (\bar{s}_L\gamma_\mu T^a b_L)\sum_q(\bar{q}\gamma^\mu T^a q)\,, \tag{2.5}$$

$$\mathcal{O}_5 = (\bar{s}_L\gamma_{\mu_1}\gamma_{\mu_2}\gamma_{\mu_3} b_L)\sum_q(\bar{q}\gamma^{\mu_1}\gamma^{\mu_2}\gamma^{\mu_3} q)\,, \tag{2.6}$$

$$\mathcal{O}_6 = (\bar{s}_L\gamma_{\mu_1}\gamma_{\mu_2}\gamma_{\mu_3} T^a b_L)\sum_q(\bar{q}\gamma^{\mu_1}\gamma^{\mu_2}\gamma^{\mu_3} T^a q)\,, \tag{2.7}$$

$$\mathcal{O}_7 = \frac{e}{g_s^2} m_b\,(\bar{s}_L\sigma^{\mu\nu} b_R)\,F_{\mu\nu}, \tag{2.8}$$

$$\mathcal{O}_8 = \frac{1}{g_s} m_b\,(\bar{s}_L\sigma^{\mu\nu} T^a b_R)\,G^a_{\mu\nu}, \tag{2.9}$$

$$\mathcal{O}_9 = \frac{e^2}{g_s^2}(\bar{s}_L\gamma_\mu b_L)\sum_\ell\left(\bar{\ell}\gamma^\mu\ell\right)\,, \tag{2.10}$$

$$\mathcal{O}_{10} = \frac{e^2}{g_s^2}(\bar{s}_L\gamma_\mu b_L)\sum_\ell\left(\bar{\ell}\gamma^\mu\gamma_5\ell\right)\,, \tag{2.11}$$

where the subscripts L and R refer to left- and right- handed components of the fermion fields, e and g_s represent the electromagnetic and strong coupling constants, respectively, and $F_{\mu\nu}$ and $G^{\mu\nu}$ denote QED and QCD field strength tensors, respectively.

The $C_i(\mu)$ are evaluated perturvatively at the high energy scale $\mathcal{O}(m_W)$. To obtain the low energy theory, the $C_i(\mu)$ are evolved down to the scale $\mathcal{O}(m_b)$ using renormalization group equation (RGE). The RGE guarantees that the μ dependence of $C_i(\mu)$ is canceled by the μ dependence of $\mathcal{O}(\mu)$, thus observables should not depend on the renormalization scale μ. Using this method, model-independent calculations are possible. If non-SM physics contributes to the additional diagrams, they modify the relevant Wilson coefficients or add new operators. In other words, we can perform searches for new physics by measuring the Wilson coefficients in experiments and comparing them with the prediction of the SM.

2.3 Analysis Methods

For the radiative and electroweak penguin processes, two analyses method have been performed : exclusive reconstruction method and inclusive measurement method.

2.3.1 Exclusive Reconstruction Method

A B meson decaying into an exclusive final state is reconstructed by measuring all long-lived decay products ($\pi^{\pm}$, $K^{\pm}$, $e^{\pm}$, $\mu^{\pm}$ and γ), selecting intermediate states of certain invariant mass. The exclusive reconstruction method has an advantage of having strong kinematic discrimination against background. In the B-factories, the two independent kinematic variables are used : the beam constrained mass ($M_{\rm bc}$) and the energy difference (ΔE), which will be defined in Eq. 4.6 and 4.7. Although the exclusive reconstruction method has an advantage of having strong kinematic discrimination against backgrounds by using these variables, the exclusive B reconstruction method has a disadvantage of having sizable theoretical uncertainties associated with the form-factor of the hadronic B to K meson transition process, in general.

2.3.2 Fully- and Semi-Inclusive Measurement Method

In fully inclusive measurement method for the $b \to s\ell^+\ell^-$ ($b \to s\gamma$), the system recoiling against the emitted dilepton (photon) is not reconstructed. The inclusive measurement method has an advantage of less theoretical uncertainties than the exclusive reconstruction method. However, the background rejection is challenging. So far, the full-inclusive measurement method of $b \to s\gamma$ has been performed. However, the full-inclusive measurement methods of $b \to s\ell^+\ell^-$ has not been performed due to huge backgrounds from the random combination of semi-leptonic B decays and its small branching fraction.

An alternative method is to measure as many exclusive modes as possible and then sum them up, which called *semi-inclusive reconstruction method*. The semi-inclusive reconstruction methods can be performed with lower background than inclusive measurement method, and can be performed with less theoretical uncertainty than exclusive reconstruction method.

2.4 Measurement

Several scenarios have been proposed for the new physics. The FCNC processes provide strong constraints on the new physics models. The model independent study using FCNC processes has also been performed. In this section, we review the $b \to s\gamma$, $b \to s\ell^+\ell^-$, and $B_s \to \mu\mu$.

2.4.1 $b \to s\gamma$

The $b \to s\gamma$ is mainly generated through the electromagnetic penguin diagram. Therefore this process has sensitivity to the C_7. The exclusive $b \to s\gamma$ transition, $B \to K - *\gamma$ was first observed by the CLEO experiment [9]. The matrix element and decay width of $b \to s\gamma$ decay is as follows [10]:

$$\mathcal{M}(b \to s\gamma) = \frac{4G_F}{\sqrt{2}}\frac{e}{16\pi^2}V_{ts}^*V_{tb}m_bC_7^{eff}(\bar{s}\sigma_{\mu\nu}b_R)F^{\mu\nu} \qquad (2.12)$$

$$\Gamma(b \to s\gamma) = \frac{G_F^2\alpha_{\rm em}m_b^5}{32\pi^4}\left|C_7\right|^2\left|V_{tb}V_{ts}^*\right|^2 . \qquad (2.13)$$

The measurement and theoretical calculations for $B \to X_s\gamma$ branching fraction are shown in Fig 2.3. The results are in good agreement with the Standard Model prediction. These results implies very stringent constraints on the new physics models. For example, the charged Higgs mass in the two-Higgs doublet model is constrained to be above 295 GeV at 95% C.L. [11] and the inverse compactification radius ($1/R$) of the minimal universal extra dimension model is constrained to be above 600 GeV at 95% C.L. [12].

In the SM, the emitted photon in $b \to s\gamma$ decays are predominantly left-handed, namely $C_7' = \frac{m_s}{m_b}C_7$. Large contributions from right-handed photon indicate new physics. However, the branching fraction is sensitive to only the $|C_7|^2 + |C_7'|^2$. To extract the ratio C_7'/C_7, the time-dependent CP asymmetry for $b \to s\gamma$ has been measured. The time-dependent CP asymmetry is suppressed by the quark mass ratio ($2m_s/m_b$). The expected mixing-induced CP asymmetry parameter (S_{CP}) is $\mathcal{O}(3\%)$, and the direct CP asymmetry parameter (A_{CP}) is $\mathcal{O}(0.6\%)$ [13]. The measurements of the time-dependent CP asymmetry of $b \to s\gamma$ are summarized in Fig. 2.4. All results are compatible with null asymmetry with errors that are still not small enough to provide nontrivial constraints on right-handed currents, but this observable will be one of the best ways to search for the new physics in future experiments.

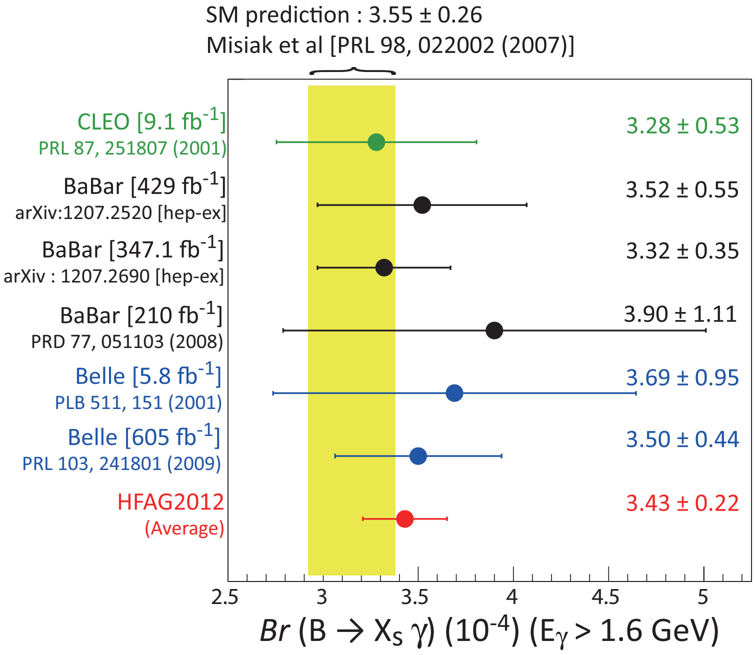
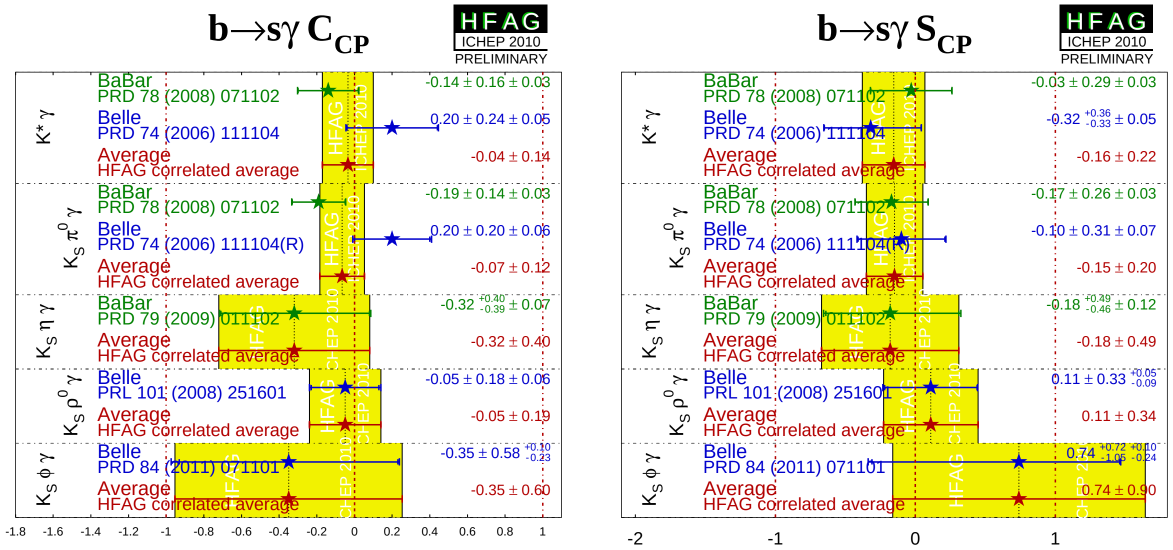


Figure 2.3: The measurements and theoretical calculations for $B \to X_s\gamma$ branching fraction.



(a) Direct CP asymmetry C_{CP}.

(b) Mixing-induced CP asymmetry S_{CP}.

Figure 2.4: CP asymmetry terms measured with $b \to s\gamma$ decays.

2.4.2 $b \to s\ell^+\ell^-$

The exclusive $b \to s\ell^+\ell^-$ transition, $B \to K\ell^+\ell^-$ was first observed by the Belle experiment [14]. The matrix element of $b \to s\ell^+\ell^-$ decay is written as follows [15]

$$\mathcal{M}(b \to s\ell^+\ell^-) = \frac{G_F\alpha_{\text{EM}}}{\sqrt{2}\pi}V_{ts}^*V_{tb}\left[(C_9^{eff} - C_{10})\bar{s}_L\gamma_\mu b_L\bar{\ell}_L\gamma^\mu\ell_L\right. \tag{2.14}$$

$$+(C_9^{eff} + C_{10})\bar{s}_L\gamma_\mu b_L\bar{\ell}_R\gamma^\mu\ell_R \tag{2.15}$$

$$\left.-2C_7^{eff}\bar{s}i\sigma_{\mu\nu}\frac{q^\nu}{q^2}(m_sL + m_bR)b\bar{\ell}\gamma^\mu\ell\right]. \tag{2.16}$$

The $b \to s\ell^+\ell^-$ decay has dependence on the C_9, and C_{10} besides C_7. Thus, this decay presents a complementary and more complex test of the Standard Model, compared with the $b \to s\gamma$ decay, which has sensitive to only C_7. The $b \to s\ell^+\ell^-$ has a rich observables : branching fraction, differential branching fraction, ratios of the branching fractions between lepton flavor, CP asymmetries, forward-backward asymmetries, isospin asymmetries, and angular distributions. For exclusive $b \to s\ell^+\ell^-$ decay, many analyses has been performed. However for the inclusive $b \to s\ell^+\ell^-$ decay, only the branching fraction and differential branching fraction has been measured using semi-inclusive method.

The invariant dilepton mass distribution for the inclusive decay $B \to X_s\ell^+\ell^-$ at lowest order can be written as follows [16]

$$\frac{d\Gamma(b \to s\ell^+\ell^-)}{dq^2} = \Gamma_0 m_b^3(1-s)^2\left[\left(|C_9|^2 + C_{10}^2\right)(1+2s)\right. \tag{2.17}$$

$$\left.+\frac{4}{s}|C_7|^2(2+s) + 12\text{Re}(C_7^*C_9)\right], \tag{2.18}$$

where $q^2 = (p_{\ell^+} + p_{\ell^-})^2$ is the dilepton mass-square, $s = q^2/m_b^2$, and

$$\Gamma_0 = \frac{G_F^2}{48\pi^3}\frac{\alpha_{\rm em}^2}{16\pi^2}\left|V_{tb}V_{ts}^2\right|^2 . \tag{2.19}$$

Table 2.1 shows the theoretical calculations of the branching fraction for the exclusive decay process of $B \to K\ell^+\ell^-$ and $B \to K^*\ell^+\ell^-$ ($\ell = e$, μ) and inclusive $B \to X_s\ell^+\ell^-$ process. As for the experimental results, both the Belle and BABAR collaborations have observed exclusive $B \to K^{(*)}\ell^+\ell^-$ decays [17] [18] and have measured inclusive $B \to X_s\ell^+\ell^-$ decay [19] [20]. The HFAG world averages are shown in Table 2.2. The results are consistent with Standard Model prediction. The exclusive decay has less prediction power than the inclusive decay due to the irreducible form factor uncertainty in the theoretical calculation. The results of branching fraction of the inclusive $B \to X_s\ell^+\ell^-$ decay is strongly disfavor the case with the flipped sign of C_7 [21].

Table 2.1: Recent theoretical branching fraction calculations for exclusive $B \to K^{(*)}\ell^+\ell^-$ and inclusive $B \to X_s\ell^+\ell^-$ decays [8] [22].

Mode	e^+e^- mode [$\times\ 10^{-6}$]	$\mu^+\mu^-$ mode [$\times\ 10^{-6}$]
$B \to K\ell^+\ell^-$	0.35 ± 0.12	0.35 ± 0.12
$B \to K^*\ell^+\ell^-$	1.58 ± 0.49	1.19 ± 0.39
$B \to X_s\ell^+\ell^-$	6.89 ± 1.01 4.2 ± 0.7 for $M_{e^+e^-} > 0.2$ GeV/c^2	4.15 ± 0.70

Table 2.2: Experimental measured branching fraction for exclusive $B \to K^{(*)}\ell^+\ell^-$ and (semi-)inclusive $B \to X_s\ell^+\ell^-$ decays (HFAG) [23]. Unit is 10^{-6}. † Dilepton invariant mass is required to be more than 0.2 GeV/c^2.

Mode	$\ell^+\ell^- = e^+e^-$	$\ell^+\ell^- = \mu^+\mu^-$	$\ell^+\ell^- = e^+e^-/\mu^+\mu^-$(ave.)
$B \to K\ell^+\ell^-$	0.44 ± 0.06	0.48 ± 0.06	0.45 ± 0.04
$B \to K^*\ell^+\ell^-$	$1.19^{+0.17}_{-0.16}$	$1.15^{+0.16}_{-0.15}$	$1.08^{+0.12}_{-0.11}$
$B \to s\ell^+\ell^-$	$4.91^{+1.04}_{-1.06}$ †	$2.23^{+0.97}_{-0.98}$	$3.66^{+0.76}_{-0.77}$ †

Forward-backward asymmetry, which is the target in this analysis and one of the most sensitive observable to the new physics, is defined as follows

$$A_{FB} \equiv \frac{N(\cos\theta_\ell > 0) - N(\cos\theta_\ell < 0)}{N(\cos\theta_\ell > 0) + N(\cos\theta_\ell < 0)}, \tag{2.20}$$

where θ_ℓ is the angle between the $\ell^+[\ell^-]$ and the B meson three-momenta in the $\ell^+\ell^-$ center-of-mass frame in $\bar{B}^0$ or B^- [B^0 or B^+] decay. The $b \to s\ell^+\ell^-$ transition is contributed from more than one diagram: electroweak penguin diagrams mediated by γ/Z (Fig. 2.2(b)) and box diagram (Fig. 2.2(f)). The forward-backward asymmetry then is caused by the interference between vector and axial vector currents. It is the analogy of the fermion forward-backward asymmetry in $e^+e^- \to f\bar{f}$ due to the interference between diagrams mediated by γ and Z boson. The forward-backward asymmetry for the inclusive decay $B \to X_s\ell^+\ell^-$ at lowest order can be written as follows [16]

$$\frac{dA_{FB}}{dq^2} = -3\Gamma_0 m_b^3(1-s)^2 s C_{10}\mathrm{Re}\left(C_9 + \frac{2}{s}C_7\right). \tag{2.21}$$

Figure 2.5 shows the shape of the A_{FB} for the SM and other cases with non-SM parameters. In the case of the SM, there is a zero crossing point of the A_{FB}. The A_{FB} is negative at low-q^2 region by a contribution from the C_7. On the other hand, the A_{FB} is posive at the high-q^2 region by contributions from the C_9 and C_{10}. In the case of the non-SM, the zero crossing point may disappear. Figure 2.6 shows current experimental results of the A_{FB} in exclusive $B \to K^*\ell^+\ell^-$.

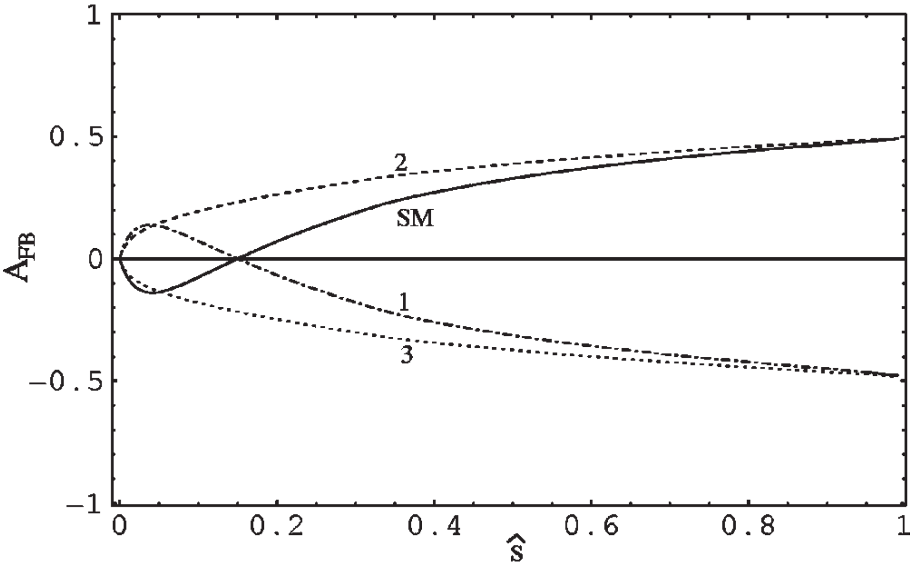


Figure 2.5: The forward-backward asymmetry, A_{FB}, for $b \to s\ell^+\ell^-$ decay theoretically calculated assuming various value of the Wilson coefficients. Line SM, 1, 2, 3 correspond to $(\Delta C_9, \Delta C_{10}) = (0,0), (1,8.5), (-9,9), (-8.5,-1)$, respectively, where ΔC_i is the deviation from the SM value.

2.4.3 $B_s \to \mu^+\mu^-$

In terms of Wilson coefficients, the branching ratio of $B \to \mu^+\mu^-$ is expressed as [25] :

$$\mathcal{B}(B_s \to \mu^+\mu^-) \propto |C_S - C_S'|^2\left(1 - \frac{4m_\mu^2}{m_{B_s}^2}\right) + \left|(C_P - C_P') + \frac{2m_\mu}{m_{B_s}}(C_{10} - C_{10}')\right|^2, \tag{2.22}$$

where C_S and C_P are the Wilson coefficients for scalar and pseudo-scalar operators arising in new physics scenarios. The $B \to \mu^+\mu^-$ decay is sensitive to contributions in the scalar/pseudo-scalar sector. In the MSSM and 2HDM models, there may be large contributions to the C_S and C_P. The contributions are proportional to the tanβ^6 and tanβ^4

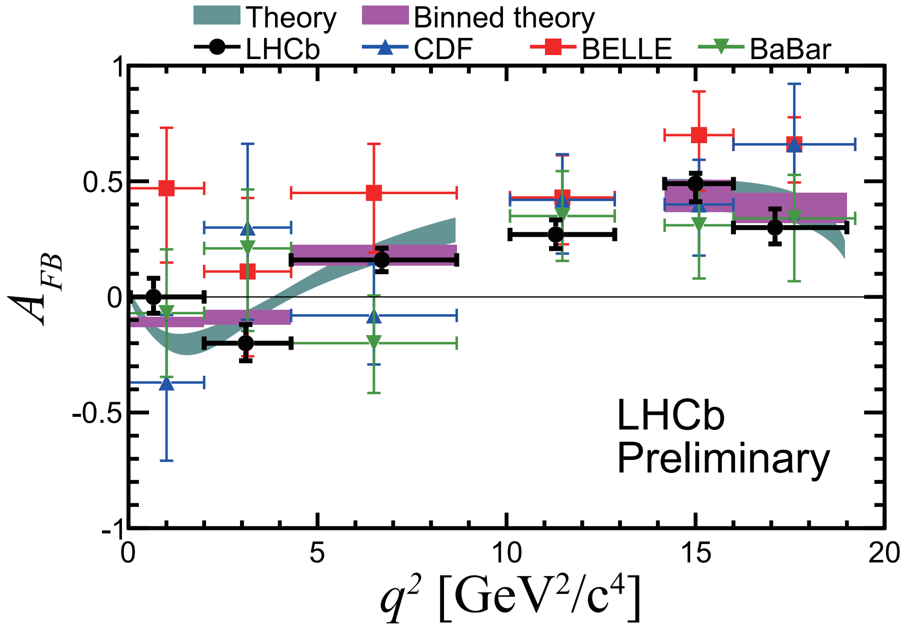


Figure 2.6: The measurement of forward-backward asymmetry, A_{FB}, as a function of q^2 in exclusive $K^*\ell^+\ell^-$. Points include both statistical and systematic uncertainties.

in the MSSM and 2HDM models, respectively. We can constrained SUSY models especially at high tanβ using the $B_s \to \mu^+\mu^-$ decay. Recent theoretical branching fraction calculations for $B_s \to \mu^+\mu^-$ is as follows [26] [27] :

$$\mathcal{B}(B_s \to \mu^+\mu^-) = (3.54 \pm 0.30) \times 10^{-9}. \tag{2.23}$$

The first evidence of the $B_s \to \mu^+\mu^-$ is found by the LHCb collaboration [28],

$$\mathcal{B}(B_s \to \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9}, \tag{2.24}$$

which is in agreement with the Standard Model prediction.

2.5 Constraints on Wilson Coefficients

The model-independent measurements of Wilson coefficients are powerful test for possible candidates of new physics sources. Figure 2.7 shows the constraints on the Wilson coefficients with the hypothesis of minimal flavor violation [29]. The constraints on the C_7 and C_8 are induced by the $b \to s\gamma$ and $b \to d\gamma$ which also has small sensitivity to the C_8 through higher order correction, in addition to the C_7. The $b \to s\ell^+\ell^-$ mainly induces the constraints on the C_7, C_9 and C_{10}. The $B_s \to \mu\mu$ gives constraints on the C_{10} and scalar contribution. If non-SM physics contribution exists, the Wilson coefficients would be deviated from values predicted in the SM.

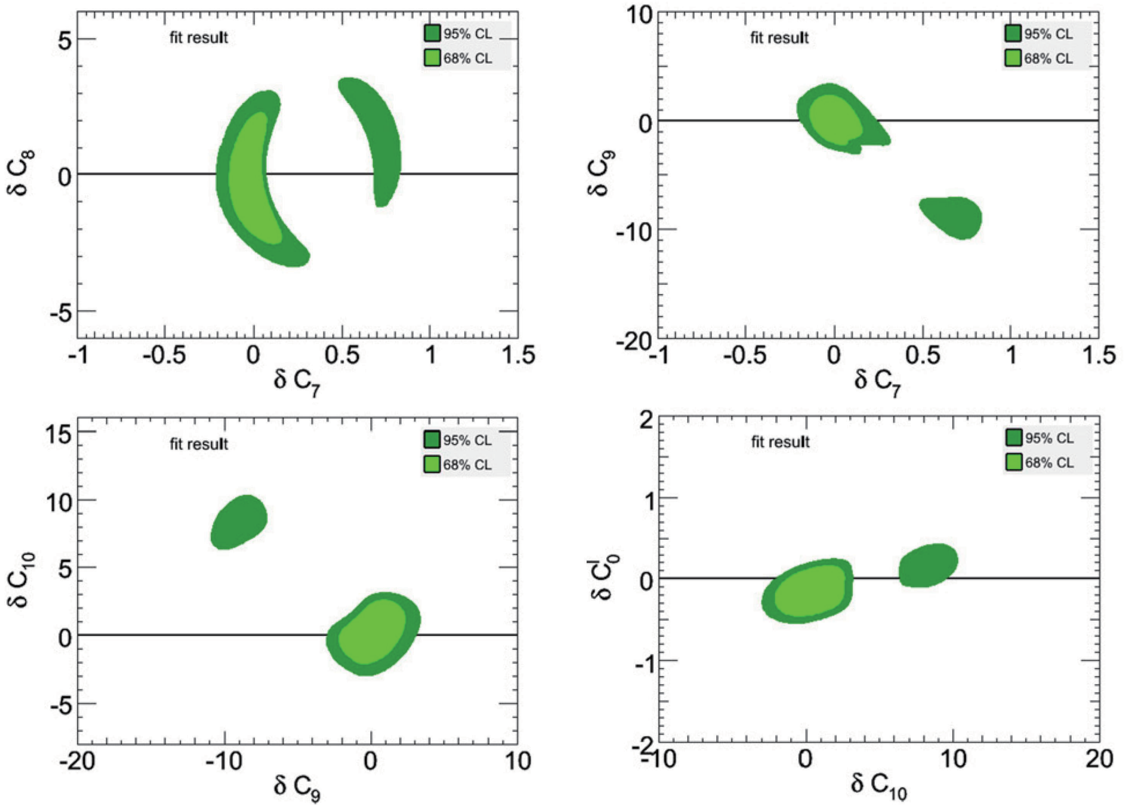


Figure 2.7: Constraints on the Wilson coefficients with the hypothesis of minimal flavor violation. δC_i is deviation from the SM prediction. δC_0^ℓ is the coefficient for scalar contribution.

Chapter 3

Experimental Apparatus and Analysis Tools

In this section, we describe the experimental apparatus of the KEK B factory which consists of the KEKB accelerator and the Belle detector. The experiment is located at High Energy Accelerator Research Organization (KEK) in Tsukuba-city, Japan.

3.1 KEKB Accelerator

KEKB [30] is a two-ring energy-asymmetric e^+e^- collider and aims to produce huge number of B and anti-B meson pairs. Figure 3.1 shows a schematic layout of KEKB accelerator. A linear accelerator (Linac) accelerates an electron and positron up to the required energy and injects them to storage rings. THE KEKB accelerator has two different storage rings: the ring for 8 GeV electrons is called the High Energy Ring (HER), and that for 3.5 GeV positrons is called the Low Energy Ring (LER). The HER and LER were constructed side by side in the tunnel used for TRISTAN experiment. The two rings cross at one point called the interaction point (IP), where electrons and positrons collide with a crossing angle of ±11 mrad. The crossing angle was one of the novel features of the KEKB design, providing effective beam separation after collision without a high detector background level.

The center-of-mass energy is designed to be

$$\sqrt{s} = 2\sqrt{E_{\mathrm{HER}} \cdot E_{\mathrm{LER}}} = 10.58\,\mathrm{GeV}, \tag{3.1}$$

which corresponds to the mass of the $\Upsilon(4S)$ resonance, just above $B\bar{B}$ production threshold. The cross-section for various processes in e^+e^- collisions at the $\Upsilon(4S)$ resonance are summarized in Table 3.1. The $b\bar{b}$ production cross-section is about 1.1 nb. The Υ resonance stand on top of large continuum background coming from light-quark pair production ($e^+e^- \to q\bar{q}$ with $q = u, d, s, c$). The e^+e^- storage rings operating at the $\Upsilon(4S)$ resonance are called B-factories. The $\Upsilon(4S)$ dominantly decay to $B^0\bar{B}^0$ and B^+B^- pairs which are created with a Lorentz boost

$$\beta\gamma = \frac{E_{\mathrm{HER}} - E_{\mathrm{LER}}}{2\sqrt{E_{\mathrm{HER}}E_{\mathrm{LER}}}} = 0.425, \tag{3.2}$$

due to the energy asymmetry. For measurement of time dependent asymmetry, the distance of the decay vertices (Δz) of the B meson pairs is measured instead of the difference

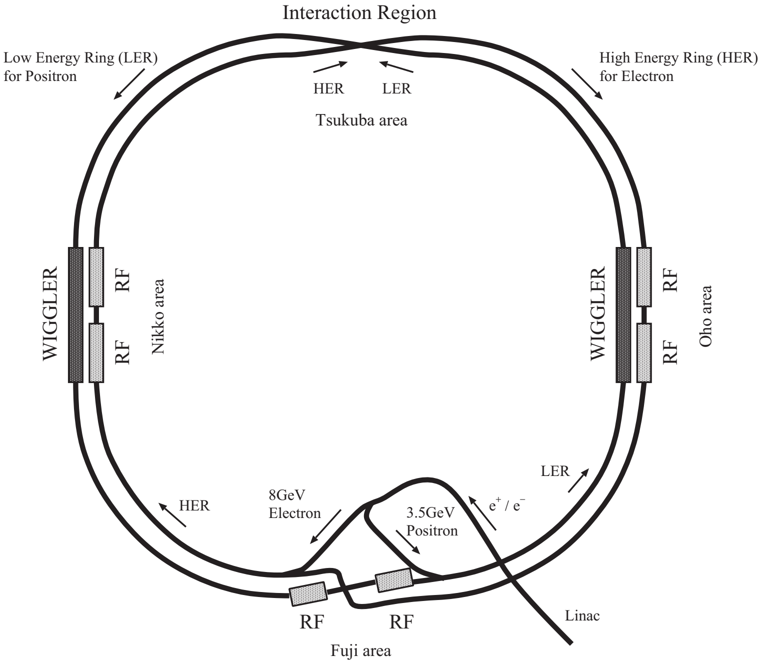


Figure 3.1: Schematic layout of KEKB accelerator

Table 3.1: Cross-section for various processes in e^+e^- collisions at $\sqrt{s} = 10.58$ GeV. QED refers to Bhabha and radiative Bhabha processes.

Process	σ [nb]
$b\bar{b}$	1.1
$c\bar{c}$	1.3
$q\bar{q}$ $(q = u, d, s)$	2.1
$\tau^+\tau^-$	0.93
QED $(25.551° < \theta < 159.94°)$	37.8
$\gamma\gamma$	11.1

of the decay time (Δt) from the relation $\Delta z \sim c\beta\gamma\Delta t$. The typical B-meson decay length is dilated from ∼20 μm to ∼200 μm by the Lorentz boost.

The design instantaneous luminosity of KEKB is $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}. However, it exceeded this goal in 2004, and the world's highest luminosity of

$$\mathcal{L} = 2.11 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1} \tag{3.3}$$

has been achieved in June 2009. Several improvements during that period increased the instantaneous luminosity. In early 2004, a new operation method called continuous injection was successfully introduced, which removes the dead time of the ordinary injection method. In early 2007, a new instrument called a crab cavity [31] was installed. In the original design of KEKB, the two beams do not collide head-on, but with a small crossing angle of ±11 mrad. The crab cavities kick the beams in the horizontal plane, and make the head-on collisions, while retaining the crossing angle of beams.

Figure 3.2 shows the history of the luminosity. The total integrated luminosity had reached 1000 fb^{-1}, which is one of the primary targets of the KEKB project, by finishing the data taking in June 2010.

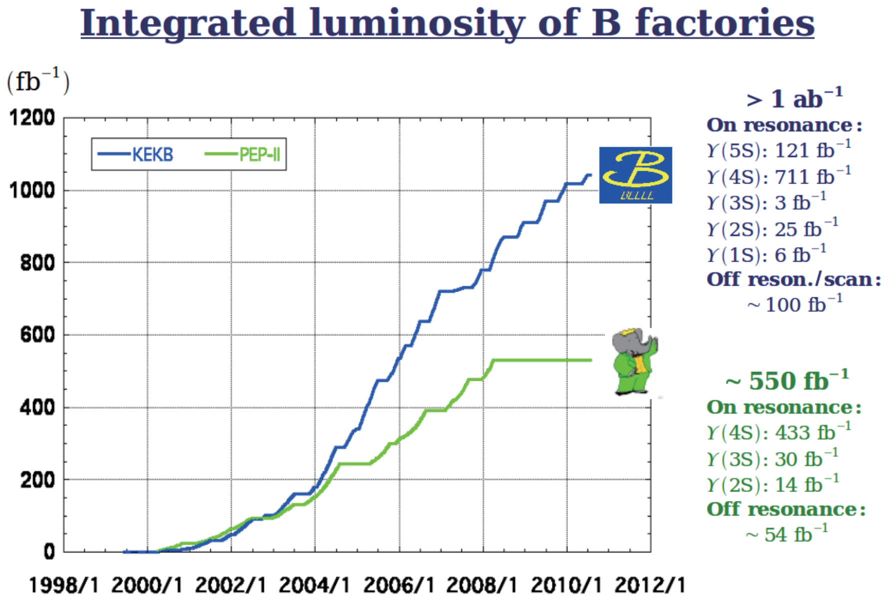


Figure 3.2: The integrated luminosity of B-factories : Belle (blue) and BABAR (green).

3.2 Belle Detector

Belle detector [32], shown in Fig. 3.3, is a general-purpose 4π detector compose of many sub-detector. The excellent performances of the particle identification and tracking system, and large angular coverage, make it very efficient to reconstruct B decays. A superconducting solenoidal magnet producing a 1.5 T field is used for the momentum measurements.

B-meson decay vertices are measured by a silicon vertex detector (SVD) situated outside of a cylindrical beryllium beam-pipe. Charged particle tracking is provided by a wire drift chamber (CDC) together with the SVD. Particle identification (PID) is provided by dE/dx measured in CDC, aerogel Cherenkov counters (ACC) and time-of-flight counters (TOF) situated radially outside of the CDC. Electromagnetic particles are detected by an array of CsI(Tl) crystals (ECL) located inside the solenoid coil. The outermost detector is the K_L and muon detector (KLM), layers of resistive plate counters instrumented in the iron flux return. A pair of BGO crystal arrays (EFC), which is placed on the surfaces of the QCS (Quadrupole Collision Superconducting magnet) cryostat, covers forward and backward regions uncovered by the other detectors.

A major detector upgrade has been performed in the summer of 2003. A 3-layer SVD with a 2 cm radius beam-pipe was used until the summer of 2003. A data sample corresponding to a integrated luminosity of 140 fb^{-1} was collected with this configuration. In the summer of 2003, a 4-layer SVD, a 1.5 cm radius beam-pipe, and a small-cell inner drift chamber were installed.

The origin of the coordinate system is defined as the position of the nominal IP. The z axis is aligned with the direction opposite to the positron beam and is parallel to the direction of the magnetic field within the solenoid. The x axis is horizontal and points towards the outside of the ring, and the y axis is vertical. The polar angle θ and azimuthal angle ϕ are measured relative to the positive z and x axes, respectively. The radial distance is defined with $r = \sqrt{x^2 + y^2}$.

The following subsections provide a more detailed description of every sub-detector.

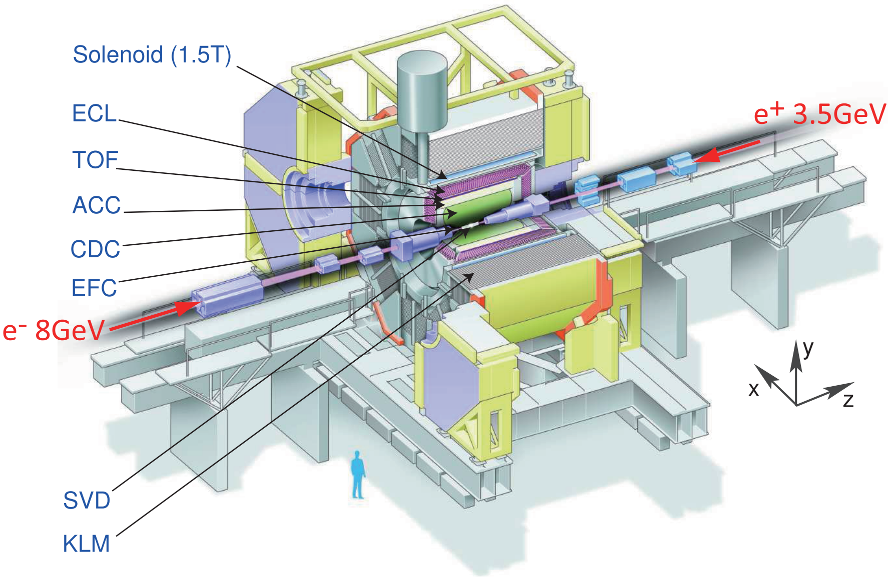


Figure 3.3: Overview of the Belle detector

3.2.1 Silicon Vertex Detector (SVD)

The Silicon Vertex Detector (SVD) provides very precise position measurements and play a crucial role in measuring time-dependent CP violation in the neutral B meson system.

Figure 3.4(a) illustrates the end and side views of SVD1, which is used to the summer of 2003. The SVD1 consists of three concentric cylindrical layers arranged to cover $23° < \theta < 139°$. Its coverage corresponds to 86% of full solid angle. The three layers at 30.0 mm, 45.5 mm and 60.5 mm radii surround the beam pipe that has a double-wall beryllium cylinder of 2.0 cm radius. There are 8/10/14 ladders along ϕ in layers 1/2/3, respectively.

In the summer of 2003, a new vertex detector, SVD2, was installed [33]. Figure 3.4(b) shows the configuration of the SVD2. The SVD2 consists of four concentric cylindrical layers and the polar angle acceptance is improved to cover $17° < \theta < 150°$, which is the same as CDC and corresponds to the 92% of the full solid angle. The four layers at 20.0 mm, 43.5 mm, 70.0 mm, and 88.0 mm radii surround the beam pipe whose radii is 1.5 cm. There are 6/12/18/18 ladders in layers 1/2/3/4, respectively.

Both SVD1 and SVD2 used a common double-sided silicon strip detectors (DSSDs) design. A DSSD is essentially a depleted *pn* junction. A charged particle passing through the junction liberates electrons from the valence band into the conduction band creating electron-hole pairs. These pairs create currents in the p^+ and n^+ strips located on the surface of the DSSD. The p^+ strips are aligned along the beam axis and therefore measure the azimuthal angle ϕ. The n^+ strips are aligned perpendicularly to the beam axis and measure z. The readout chain of DSSDs is based on CMOS-integrated circuit placed outside of the tracking volume.

The impact parameter resolution $\sigma_{r\phi}$ and σ_z measured using cosmic rays events are shown in Fig. 3.5. Obtained performance are

$$\begin{aligned}
\sigma_{r\phi}(\mu\text{m}) &= 19.2 \oplus 54.0/\tilde{p} \quad \text{for SVD1,} && (3.4)\\
\sigma_{r\phi}(\mu\text{m}) &= 21.9 \oplus 35.5/\tilde{p} \quad \text{for SVD2,} && (3.5)\\
\sigma_{z}(\mu\text{m}) &= 42.2 \oplus 44.3/\tilde{p} \quad \text{for SVD1,} && (3.6)\\
\sigma_{z}(\mu\text{m}) &= 27.8 \oplus 31.9/\tilde{p} \quad \text{for SVD2,} && (3.7)
\end{aligned}$$

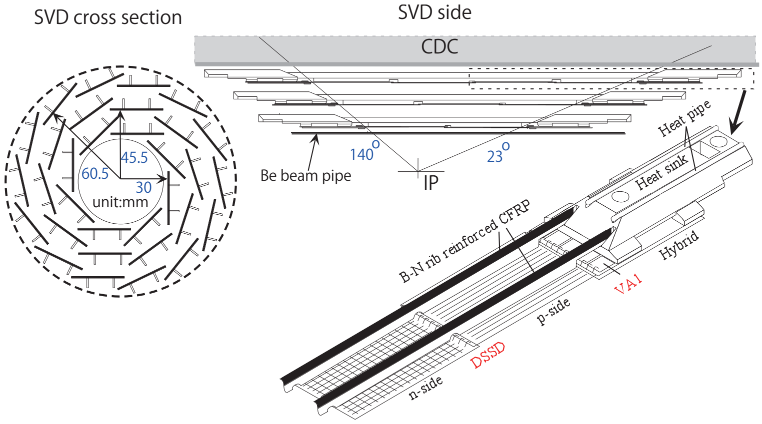
where pseudo momentum of $\tilde{p}$ is defined as $\tilde{p} = p\beta\sin\theta^{3/2}\theta$ for r-ϕ side and $\tilde{p} = p\beta\sin^{5/2}\theta$ for z side. The SVD upgrade significantly improved the impact parameter resolution in both the r-ϕ and z coordinates.

3.2.2 Central Drift Chamber (CDC)

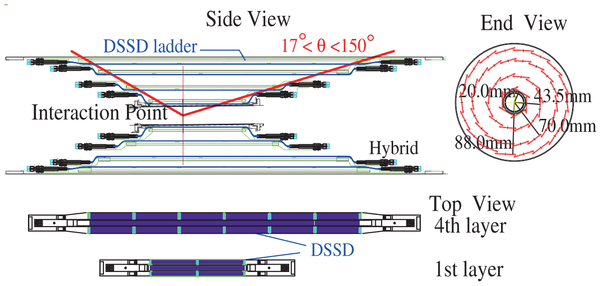
The main role of the Central Drift Chamber (CDC) [34] is detection of charged particle tracks and determination of their momenta from their curvature in the magnetic field of 1.5 T provided by the superconducting solenoid. The CDC also provides particle identification information in the form of dE/dx measurements for charged particles.

Figure 3.6 shows the structure of the CDC. It is asymmetric in the z direction in order to provide an angular coverage of $17° < \theta < 150°$. The CDC has 50 cylindrical layers of anode wires, which consist of 32 axial- and 18 stereo-wire layers, and three cathode strip layers. In summer of 2003, the inner three layers have been replaced by two small-cell layers for making a space of SVD2, maintaining the performance of the trigger. Axial wires are parallel to the z axis, while stereo wires slant to the z axis to provide z position information. A total number of drift cells is 8400(8464) for SVD1(SVD2) configuration.

A mixture of helium (50%) and ethane (50%) gas is filled in the chamber. A charged particle passing through CDC ionizes the gas. A charge avalanche is caused by the ionized gas and drifts to a sense wire with a specific drift velocity, then the measured signal height



(a) SVD1.



(b) SVD2.

Figure 3.4: Detector configuration of SVD.

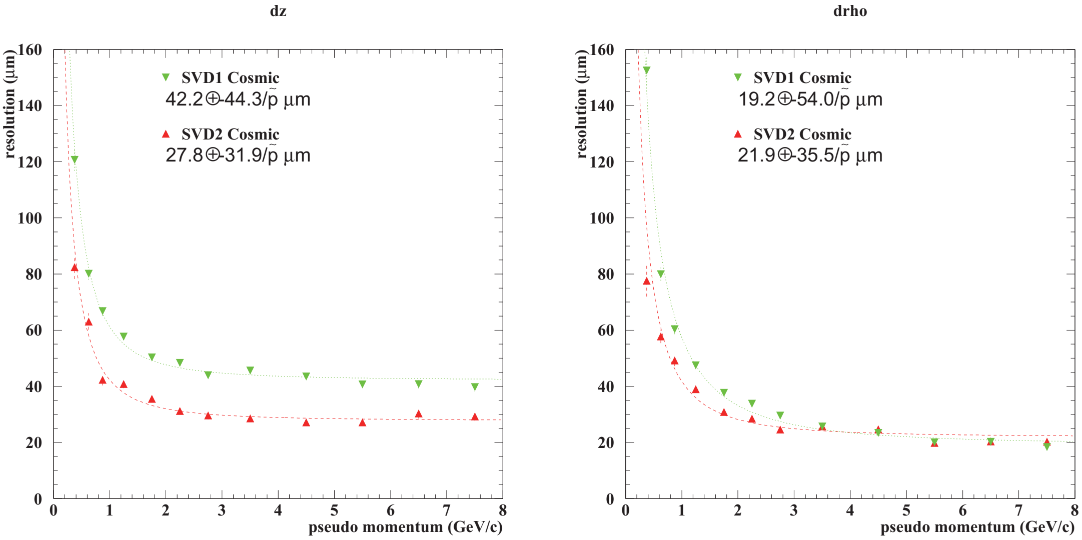


Figure 3.5: Impact parameter resolutions (left) in z and (right) in r-ϕ coordinates for the SVD1 and SVD2.

and drift time provides information of the energy deposit and distance from the sense wire. Even though the gas mixture has a low Z to minimize the multiple-Coulomb scattering, a good dE/dx resolution is provided by the large ethane component. The transverse momentum resolution measured using the cosmic ray events are shown in Fig. 3.7(a). Obtained performance is

$$\frac{\sigma_{p_t}}{p_t}(\%) = 0.19 p_t \oplus 0.30/\beta. \tag{3.8}$$

A scatter plot of measured $< dE/dx >$ and particle momentum is shown in Fig. 3.7(b), together with the expected mean energy losses for different particle species. Populations of pions, kaons, protons, and electrons can be clearly seen. The $< dE/dx >$ resolution was measured to be 7.8% in the momentum range from 0.4 to 0.6 GeV/c. It provides K/π separation up to 0.8 GeV/c and also in the region of relativistic rise (above 2.5 GeV/c).

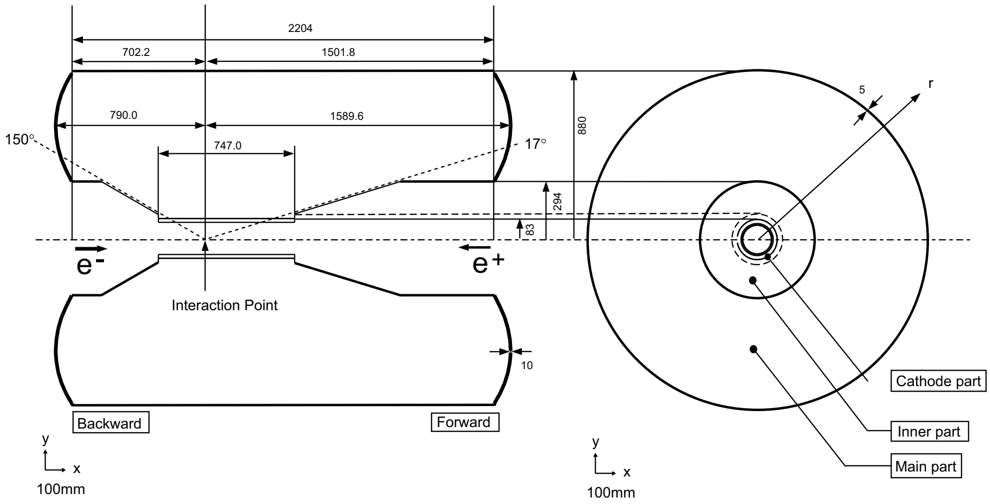


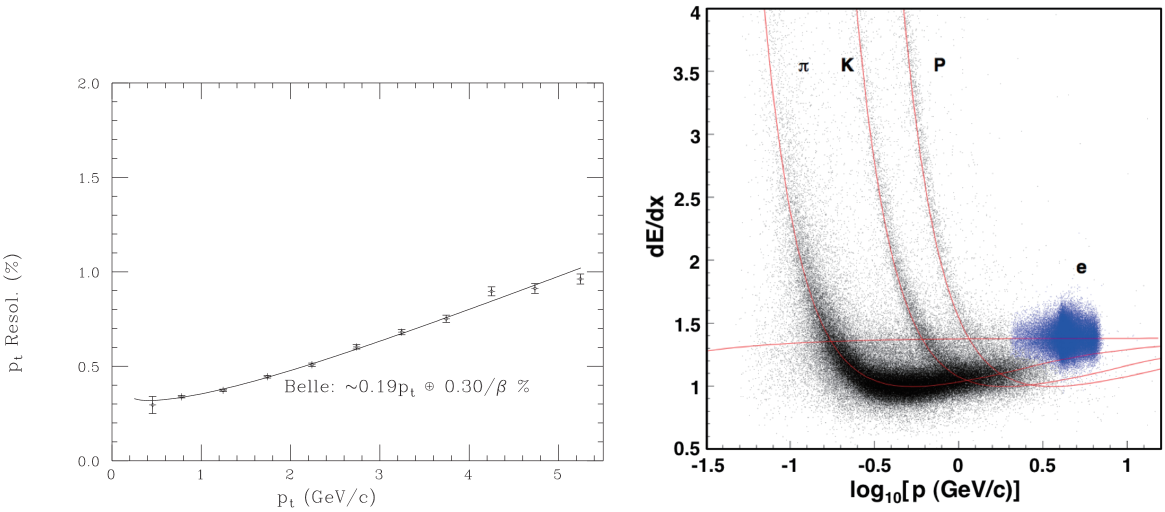
Figure 3.6: Overview of the CDC structure.

3.2.3 Aerogel Cherenkov Counter (ACC)

Aerogel Cherenkov Counter (ACC) [35] [36] provides an information to separate $K^{\pm}$ from $\pi^{\pm}$ in high momentum range (1.2 GeV/$c < p <$ 3.5 GeV/c), which extend the momentum coverage beyond the reach of CDC and TOF. ACC is silica aerogel threshold Cherenkov counters, which detect if a particle emits Cherenkov light or not and distinguishes particle species. Cherenkov radiation is emitted when a charged particle passes through a material medium at a speed greater than the phase velocity of light in that medium. The condition to emit Cherenkov light is given as

$$n > \frac{1}{\beta} = \sqrt{1 + \left(\frac{m}{p}\right)^2}, \tag{3.9}$$

where m and p are the particle mass and the momentum and n is the refractive index of the matter.



(a) Transverse momentum resolution measured by CDC with SVD.

(b) dE/dx versus charged track momentum in collision data.

Figure 3.7: CDC performance.

Figure 3.8 illustrates the configuration of the ACC in the Belle detector. ACC consists of 960 counter modules segmented into 60 cells in the ϕ direction for the barrel part and 228 modules arranged in five concentric layers for the forward end-cap part of the detector. All the modules are arranged in a semi-tower geometry, pointing to the IP. In order to obtain good K/π separation for the whole kinematical range, the refractive indices of aerogels are selected to be between 1.01 and 1.03, depending on their polar angle region. The choice of the refractive index for the barrel ACC is optimized for separation of high momentum pions and kaons from the two-body B decay, such as $B \to \pi\pi$ and $K\pi$. For the end-cap ACC, $n = 1.030$ has been chosen to cover low momentum region, which is necessary for flavor tagging, to cover lack of TOF in the endcap. A typical single ACC module is shown in Fig. 3.9 for the barrel and the end-cap ACC. Five aerogel tiles are stacked in thin (0.2 mm thick) aluminum box of approximate dimensions 12×12×12 cm^2. To detect the Cherenkov lights, two(one) fine-mesh type photomultiplier tubes (FM-PMTs) are attached to each module in the barrel (end-cap) part. These FM-PMTs are designed to operate in strong magnetic fields of 1.5 T.

The performance of ACC is confirmed using the decay chain $D^{*-} \to \bar{D}^0\pi^-$ followed by $\bar{D}^0 \to K^+\pi^-$. The slow π^- from D^{*-} allows to identify the daughter K and π from the D^0 directly by their relative charges with respect to the slow pion. Figure 3.10 shows the distribution of the number of photoelectrons, where the K/π separation is good and consistent with MC.

3.2.4 Time-of-Flight Counter (TOF)

The Time-of-Flight Counter (TOF) [37] gives particle identification to distinguish charged kaon from pions in the momentum region, below 1.2 GeV/c. TOF also provides fast timing signals for the trigger system, together with thin trigger scintillation counters (TSC). The TSC is used for keeping the fast trigger rate below 70 kHz.

The mass of the particle m can be determined from the time-of-flight T measured with the TOF and the momentum p measured with the CDC, according to the following

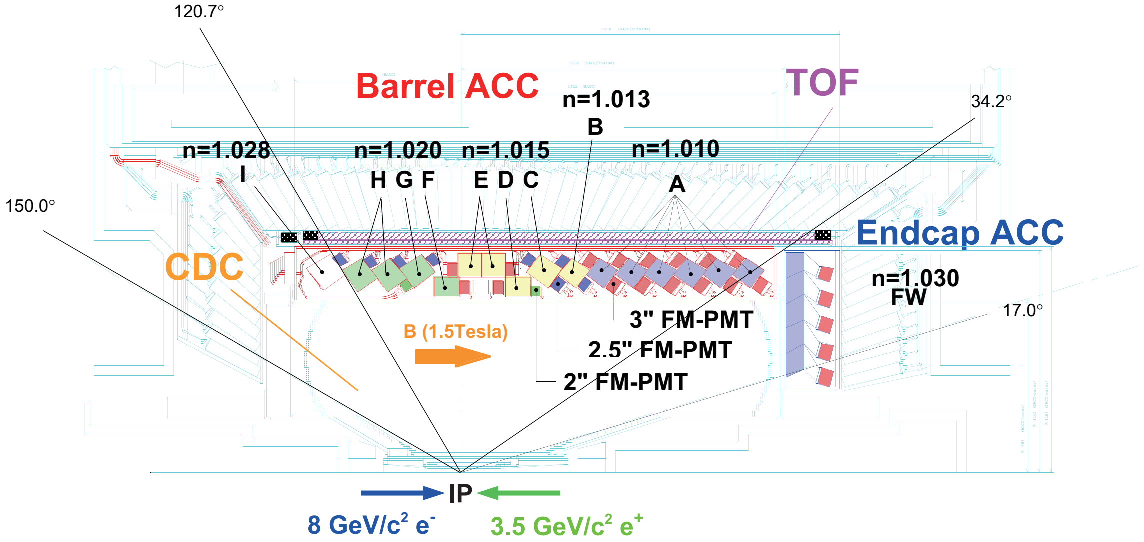


Figure 3.8: Arrangement of the ACC.

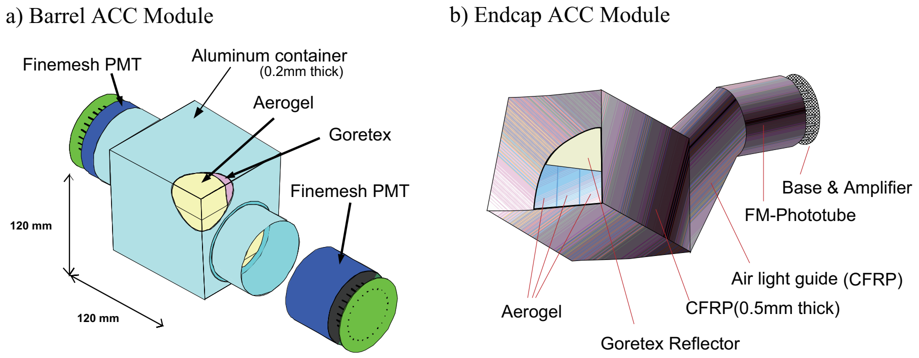


Figure 3.9: Schematic drawing of a typical ACC counter module: (a) barrel and (b) end-cap ACC.

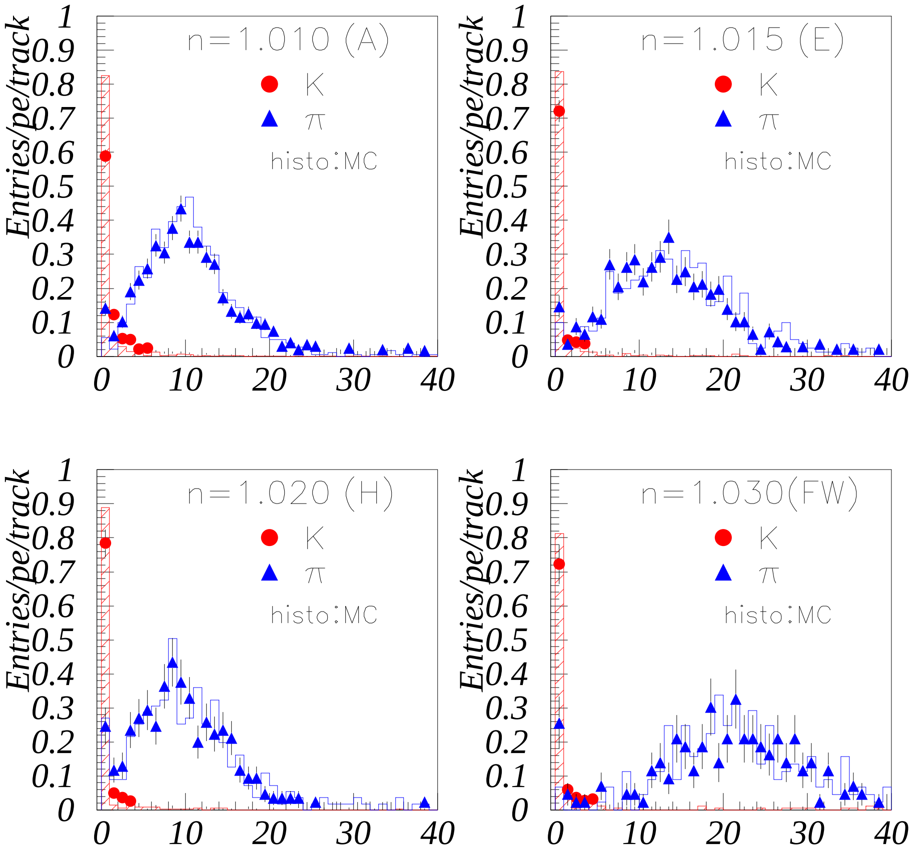


Figure 3.10: Distribution of photo-electron for $K^{\pm}$ and $\pi^{\mp}$ in $D^{*\mp}$ decays. Each plot corresponds to the different set of modules with a different refractive index

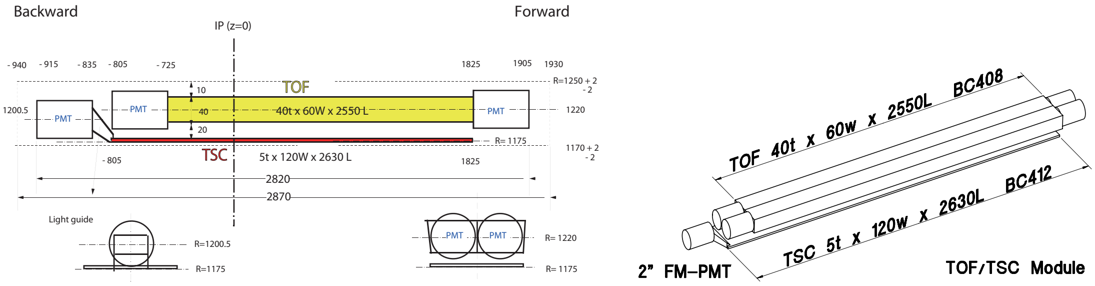
relation:

$$m = p\sqrt{\left(\frac{cT}{L}\right)^2 - 1}, \tag{3.10}$$

where L is a length of the flight.

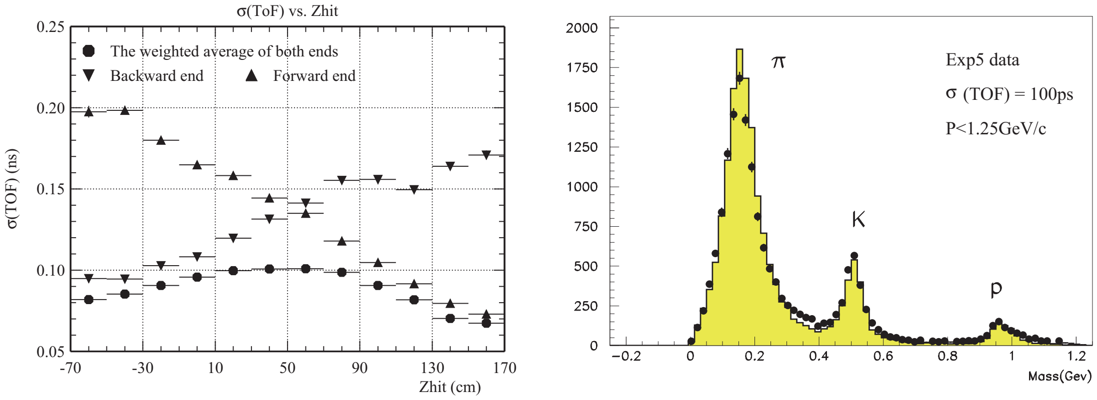
The TOF system consists of 128 TOF counters and 64 TSC counters. Two trapezoidally shaped TOF counters and one TSC counters form one module as shown in Fig. 3.11. In total 64 TOF/TSC modules located at a radius of 1.2 m from the IP cover a polar angle range from 34° to 120°. Each TOF counter is read out by a FM-PMT at each end. Each TSC counter is read out by only one FM-PMT from the backward end.

Figure 3.12(a) shows TOF time resolution for forward and backward PMTs and for the weighted average as a function of z position. The resolution for the weighted average time is about 100 ps with a small z dependence. Figure 3.12(b) shows the mass distribution for each track in hadron events. Clear peaks corresponding to pion, kaon and proton are seen. The data points well agree with a MC prediction (histogram) obtained by assuming resolution of $\sigma_{TOF} = 100$ ps.



(a) (b)

Figure 3.11: Schematic drawing of a TOF/TSC module.



(a) Time resolution for μ-pair events.

(b) Mass distribution from TOF measurements for particle momenta below 1.2 GeV/c. Histogram shows the expectation assuming time resolution of 100 ps. The point with error bars are data.

Figure 3.12: TOF performance.

3.2.5 Electromagnetic Calorimeter (ECL)

The main purpose of the electromagnetic calorimeter (ECL) [38] is the detection of electrons and photons from B meson decays with high efficiency and good resolution in energy and position. The ECL measures energy deposited by electromagnetic showers. High energy electrons and photons entering the calorimeter initiate an electromagnetic shower through subsequent bremsstrahlung and electron pair production processes followed by Coulomb scatterings. As a result, all of the incident energy is absorbed as ionization or excitation (light) in the calorimeter. Other particles only deposit small amounts of energy via dE/dx ionization. The matching of the energy measured by the ECL and the momentum measured by the CDC is used for the electron identification.

The overall configuration of the ECL is shown in Fig. 3.13. ECL contains 8,736 thallium doped CsI crystal counters. The ECL consists of three sections: the forward endcap section consist of 1152 crystals and cover $12.4° < \theta < 31.4°$; the barrel section consist of 6,624 crystals and cover $32.2° < \theta < 128.7°$; and the backward endcap section consist of 960 crystals and cover $130.7° < \theta < 155.1°$; Each crystal has tower shape with about 6 cm × 6 cm cross section and 30 cm length (16.2 radiation lengths). Total weight of the crystals is about 43 tons. The light of each crystal is readout by two PIN photodiodes and a preamplifier mounted at the end of each crystal.

The energy resolution is measured by a beam test [39] to be

$$\frac{\sigma_E}{E}(\%) = \frac{0.066}{E} \oplus \frac{0.81}{\sqrt[4]{E}} \oplus 1.34 \quad (E \text{ in GeV}), \tag{3.11}$$

where the value is affected by the electronic noise (1st term), the shower leakage fluctuation (2nd and 3rd terms), and the systematic effect such as the uncertainty of calibration (3rd term). The spacial resolution is approximately found to be 0.5 cm/$\sqrt{E}$ (E in GeV).

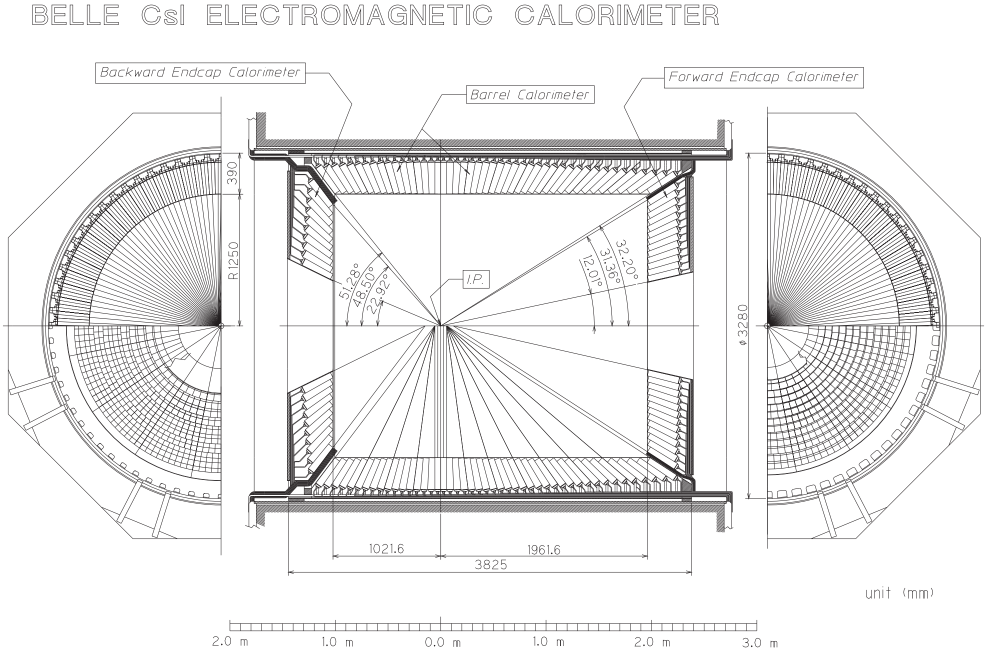


Figure 3.13: Configuration of ECL.

3.2.6 K_L and Muon Detector (KLM)

The purpose of K_L and Muon Detector (KLM) [40] is to identify K_L's and muons with high efficiency over a broad momentum range greater than 600 MeV/c.

The KLM consists of successive layers of charged particle detector (resistive plate counters, RPC [41]) and iron plates (4.7 cm thick). The neutral K_L meson produces a hadronic shower when interacting in the iron, allowing for position detection. However, no useful measurement of its energy is possible because of the fluctuations of this shower. The muons go through all the detector and the hits in the RPC allow for energy and position measurements. Other particles, such as pions and kaons are stopped in the system and can easily be separated from muons.

The KLM contains 15 detector layers and 14 iron layers in the barrel part ($45° < \theta < 125°$), and 14 detector layers. in each of the forward and backward endcaps region ($20° < \theta < 155°$). The iron plates provide 3.9 interaction length of material, in addition to 0.8 interaction length of ECL. The iron layers also seve as a return yoke for the magnetic flux provided by the superconducting solenoid.

KLM layers are grouped in superlayers, as shown in Fig. 3.14. A superlayer is made of θ and ϕ cathode strips surrounding two RPCs. Resistive-plate counters have two parallel-plate electrodes separated by a gas-filled gap. An ionizing particle traveling the gap initiates a streamer in the gas that results in a local discharge. This discharge creates a signal on the external cathode strips which can be used to record the location and time of the ionization. The number of K_L clusters per event is in good agreement with the prediction. Typical muon identification efficiency is 90% with a fake rate around 2%.

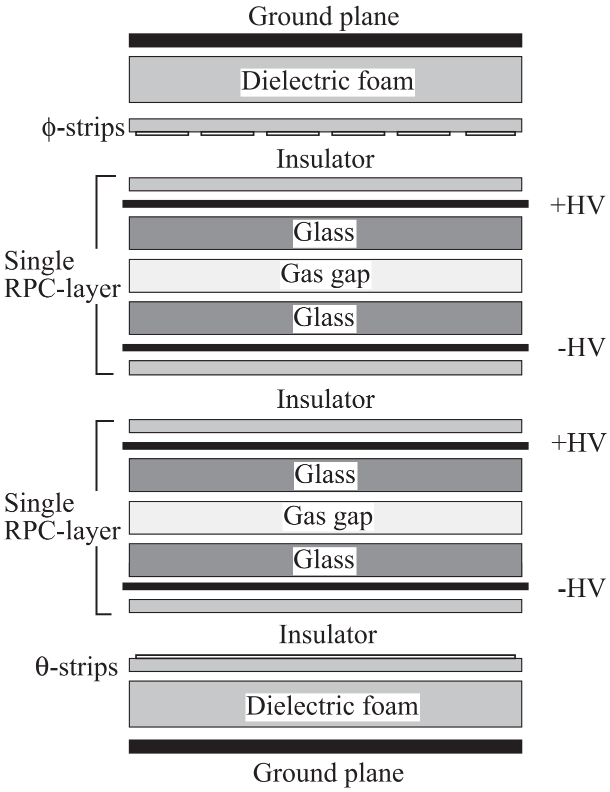


Figure 3.14: Cross section of a KLM super layer.

Table 3.2: Total trigger rates with $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} from various processes at $\Upsilon(4S)$. † The values is pre-scaled by a factor 1/100.

Process	Rate [Hz]
$\Upsilon(4S) \to B\bar{B}$	12
$e^+e^- \to q\bar{q}, (q = u, d, s, c)$	28
$e^+e^- \to \ell^+\ell^-, (\ell = \mu, \tau)$	16
Bhabha ($\theta_{lab} > 17°$)	4.4 †
$\gamma\gamma$ ($\theta_{lab} > 17°$)	0.24 †
two-photon process ($\theta_{lab} > 17°$, $p_t > 0.3$ GeV/c)	35
Total	86

3.2.7 Trigger and Data Acquisition

An important part of the Belle experiment is the trigger and the data acquisition (DAQ) systems. Most of events are not interesting for physics studies, like e^+e^- scattering (Bhabha interaction), beam-gas interaction in the beam pipe, cosmic rays, etc. The purpose of the trigger is to reject uninteresting events as much as possible and to forward interesting ones to the DAQ system. with high efficiency, within a very short decision time. The trigger rates at high luminosity 10^{34} cm^{-2}s^{-1} for various processes of interest are listed in Table 3.2. The Belle trigger system consists of a hardware trigger and a software trigger.

An overview of the hardware trigger system [42] is shown in Fig. 3.15. It consists of the sub-detector trigger systems and the central trigger system called the Global Decision Logic (GDL). The GDL receives sub-detector triggers within 1.85 μs after the collision and issues a decision 2.2 μs after the collision. The sub-detector trigger systems are based on two categories : track triggers and energy triggers. CDC and TOF are used to yield trigger signals for charged particles. The ECL trigger system provides triggers based on total energy deposit and cluster counting of crystal hits. These two categories allow sufficient redundancy. The KLM trigger gives additional information on muons and the ECL triggers are used for tagging two photon events as well as Bhabha events.

When the hardware trigger is issued, the Data Acquisition system (DAQ) collects signal data from sub-detectors and them on the data storage system. Figure 3.16 shows the overview of the DAQ system. The entire system is segmented into 7 subsystems running in parallel, each handling the data from a sub-detector. The signals from most sub-detectors go through a charge-to-time (Q-to-T) converter and are processed by a time-to-digital converter (TDC). The KLM does not have a Q-to-T converter since the pulse does not provide useful information. For the SVD, DSSDs are read out by on-board chips and passed to flash analog-to-digital converters (FADC). The readout sequence starts when the sequence controller (SEQ) receives a final trigger from the GDL and distributes a common stop signal to the TDCs. The event builder converts detector-by-detector parallel data streams to an event-by-event data river and sent the data to an online computer farm.

The online computer formats an event data into an offline event format and performs a background reduction (a hardware trigger) after a fast event reconstruction. The data

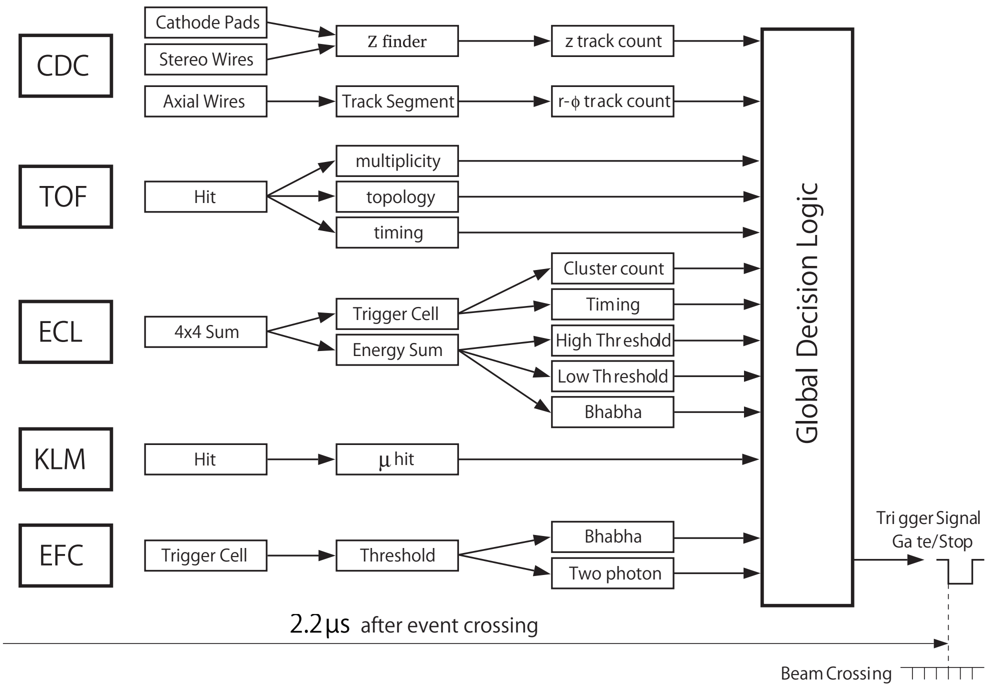


Figure 3.15: Overview of the software trigger.

are then sent to a mass storage system via optical fibers.

3.3 Analysis Tools

3.3.1 Particle Identification (PID) for $K^\pm/\pi^\pm$

The $K^\pm/\pi^\pm$ identification [44] is based on the complementary measurements performed in three sub-detectors:

- dE/dx measurement by CDC
- the Cherenkov light yield in ACC
- the time-of-flight information from TOF

The momentum coverage of kaon over pion separation of each sub-detector is illustrated in Fig. 3.17. A separation of more than 3σ between kaons and pions is realized up to momenta of 3.5 GeV/c. The likelihood functions $\mathcal{L}_K$ and $\mathcal{L}_\pi$ are constructed on the product of the likelihood functions for three discriminants.

$$\mathcal{L}_i = \mathcal{L}_i^{dE/dx} \cdot \mathcal{L}_i^{ACC} \cdot \mathcal{L}_i^{TOF} \quad (i = K, \pi). \tag{3.12}$$

The likelihood ratio $\mathcal{P}_{K/\pi}$ is then calculated as

$$\mathcal{P}_{K/\pi} = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi} \tag{3.13}$$

The performance of kaon identification is checked by measuring the decay chain $D^{*+} \to D^0\pi^+$; $D^0 \to K^-\pi^+$. With $\mathcal{P}_{K/\pi} > 0.6$, the average kaon efficiency and π fake rate over $0.5 < p < 4.0$ GeV/c are about 88% and 8.5%.

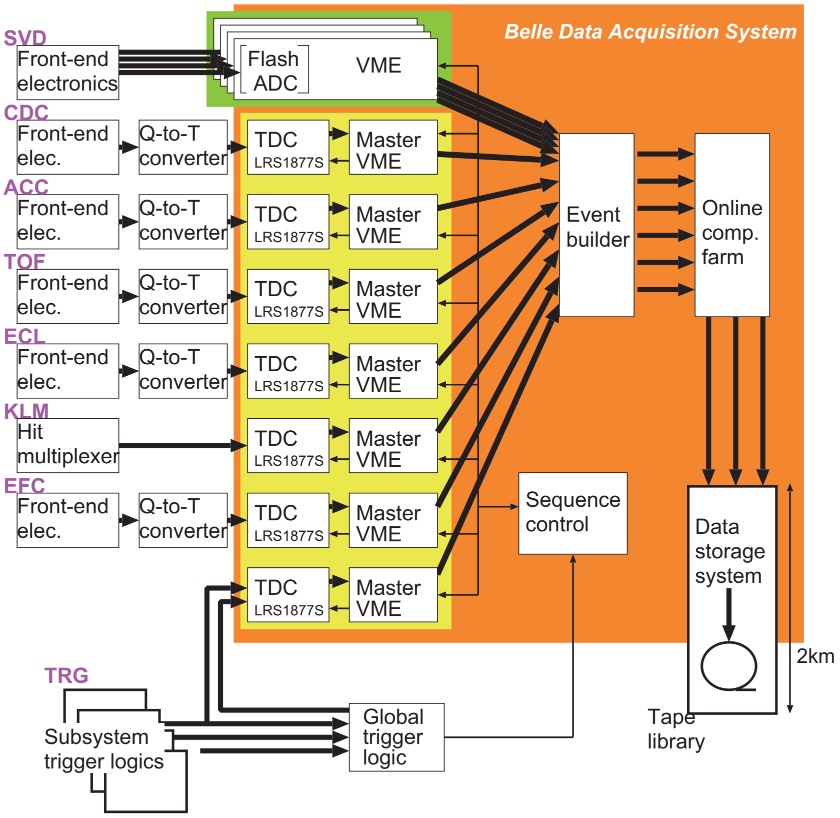


Figure 3.16: Overview of the Belle Data Acquisition (DAQ) system.

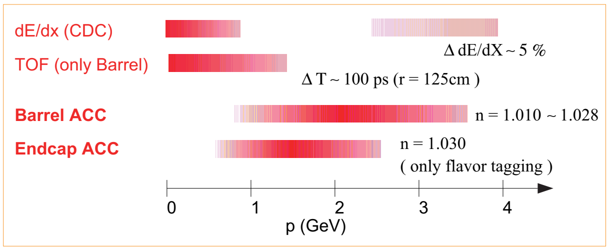


Figure 3.17: Separation power of kaon over pion identification provided by different sub-detectors in Belle.

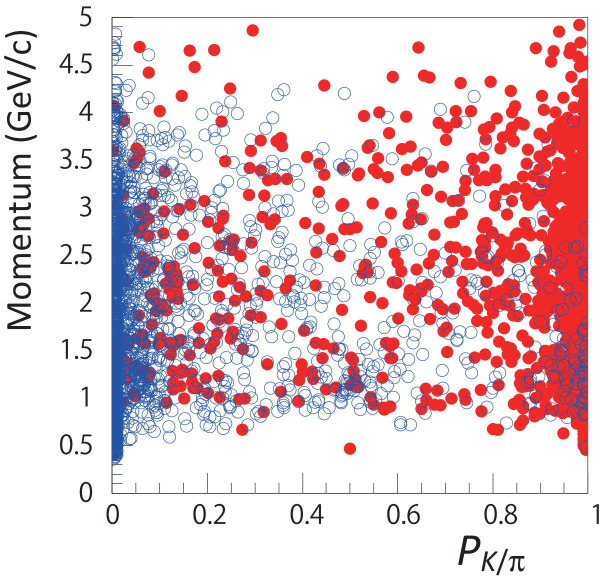


Figure 3.18: Scatter plot of the track momentum and the likelihood ratio $\mathcal{P}_{K/\pi}$ for kaon (closed red circle) and pion (open blue circle) tracks.

3.3.2 Particle Identification (PID) for e

The electron identification [45] is based on differences in the shape of the electromagnetic shower and the velocity of electrons and hadrons with same momentum. The following five discriminants are used in the electron identification.

- the ratio of cluster energy and track momentum
- the value of dE/dx measured by the CDC
- matching between the track and ECL cluster
- cluster shape parameter
- ACC light yields

A likelihood function for the electron $\mathcal{L}_e$ and non-electron $\mathcal{L}_{non\text{-}e}$ hypotheses are constructed by combining the probability density functions from the above five variables. The likelihood ratio $\mathcal{P}_e$ is then calculated as

$$\mathcal{P}_e = \frac{\mathcal{L}_e}{\mathcal{L}_e + \mathcal{L}_{non\text{-}e}}. \tag{3.14}$$

The performance of electron identification is estimated using the dedicated hadronic MC samples. With $\mathcal{P}_e > 0.5$, the average electron efficiency over $1.0 < p < 3.0$ GeV/c are about 92%. The average pion fake rates are determined using inclusive $K_S^0 \to \pi^+\pi^-$ decays and is found to be 0.22% over $0.5 < p < 3.0$ GeV/c.

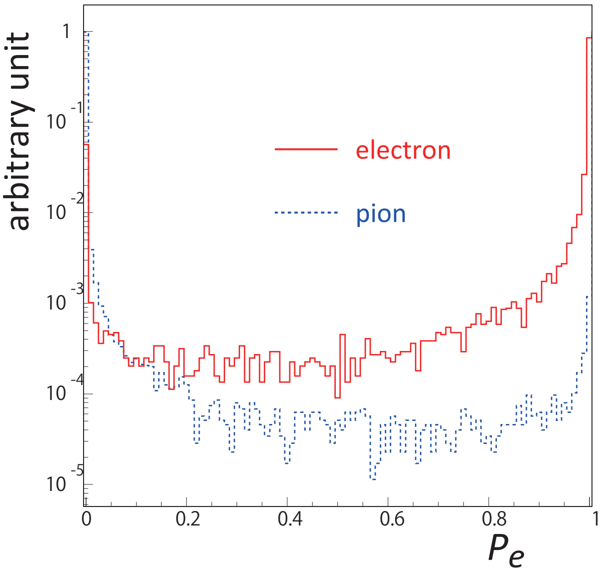


Figure 3.19: Likelihood ratio $\mathcal{P}_e$ for electron (red) and pion (blue) tracks.

3.3.3 Particle Identification (PID) for μ

Muons are heavy charged leptons that lose their energy mainly by multiple scattering in the detector material. If a muon's momentum is greater than 500 MeV, it can penetrate easily to the outermost part of the detector, the KLM. The KLM hits are associated to the reconstructed track, if they are near the extrapolated track from the CDC and SVD to the KLM. The charged track is then refitted with the associated KLM hits, minimizing the χ^2, defined as the deviation of hits from the track, in the units of the corresponding uncertainties. A likelihood function for the muon identification [46] is calculated based on the following two discriminants.

- the difference between the expected and the actual penetration in the KLM
- the distance between KLM hits and the extrapolated track

The likelihood ratio $\mathcal{P}_\mu$ is then calculated as

$$\mathcal{P}_\mu = \frac{\mathcal{L}_\mu}{\mathcal{L}_\mu + \mathcal{L}_K + \mathcal{L}_\pi}. \tag{3.15}$$

The performance of muon identification is checked using two-photon samples $e^+e^- \to e^+e^-\mu^+\mu^-$. The efficiency is measured to be around 89% for $\mathcal{P}_\mu > 0.9$ and 93% for $\mathcal{P}_\mu > 0.1$ over $1.0 < p < 3.0$ GeV/c The average pion fake rates are determined using inclusive $K_S^0 \to \pi^+\pi^-$ decays and is found to be 1.4% for $\mathcal{P}_\mu > 0.9$ and 2.8% for $\mathcal{P}_\mu > 0.1$ over $1.5 < p < 3.0$ GeV/c.

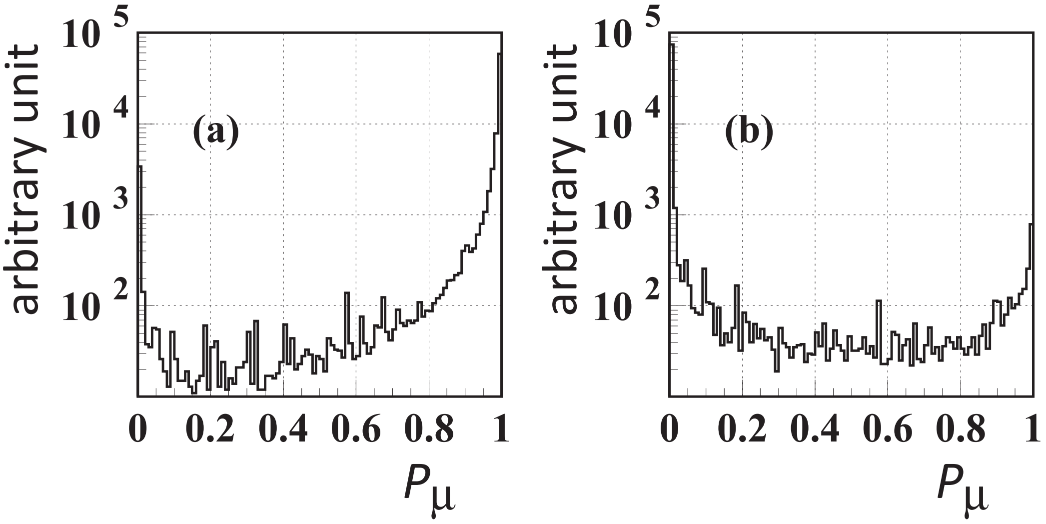


Figure 3.20: Likelihood ratio $\mathcal{P}_\mu$ for (a) muon and (b) pion tracks.

Chapter 4

Signal Reconstruction

4.1 Analysis Overview

We measure the forward-backward asymmetry of the $B \to X_s \ell^+ \ell^-$ decay with semi-inclusive method. In this analysis, we reconstruct $X_s \ell^+ \ell^-$ from an electron pair or a muon pair and one of 18 hadronic states with one $K^{\pm}$ or K_S^0 and one to four pions, where at most one pion can be neutral, as shown in Tab. 4.1. We choose particle candidates using "particle selection" (Sec. 4.3), and then apply "event selection" (Sec. 4.4) to reduce wrong combination. The event selection includes pre-selections for the background suppression. The background events are mainly suppressed using the multivariate analysis package NeuroBayes (Chap. 5). If there are still multiple B candidates left, we select a most-B-like B candidate based on the NeuroBayes output.

The forward-backward asymmetry (A_{FB}) is extracted from the $M_{\rm bc}$ distributions in the region $M_{\rm bc} > 5.22$ GeV$/c^2$ using the extended unbinned maximum likelihood fit. We divide q^2 into six bins, and obtain the A_{FB} for each q^2 bin, where $q^2 = (p_{\ell^+} + p_{\ell^-})^2$ is the dilepton mass-square. The fit is applied simultaneously for forward ($\cos\theta_\ell > 0$) and backward ($\cos\theta_\ell < 0$) events, to obtain the A_{FB} directly. For the measurement of the forward-backward asymmetry, we need flavor information of reconstructed B meson. For B^0 ($\bar{B}^0$), only the self-tagging modes with a $K^+(K^-)$ are used. Furthermore, we do not use $K4\pi$ modes due to too low signal yields. To reduce cross-feed from modes not used for the A_{FB} measurements, we reconstruct B candidates also from these 8 modes and then remove them after the best candidate selection. Finally, we use 10 hadronic states for the measurement of the forward-backward asymmetry.

In this analysis, we define reconstruction efficiency $\epsilon = N_{\rm rec}/N_{\rm gen}$, where $N_{\rm rec}$ is the number of reconstructed events and $N_{\rm gen}$ is the number of generated events, including missing states [1].

---

[1] In case of the reconstruction efficiency for specific mode mentioned in Fig. 4.2 and Fig. 5.7, $N_{\rm gen}$ is the number of generated events of the specific mode.

Table 4.1: The 18 hadronic states. The $X_s\ell^+\ell^-$ is reconstructed from 18 hadronic states, and the A_{FB} is measured using 10 hadronic states (red).

	$B^0/\bar{B}^0$		$B^\pm$	
K		K_S^0	$K^\pm$	
$K\pi$	$K^\pm\pi^\mp$	$K_S^0\pi^0$	$K^\pm\pi^0$	$K_S^0\pi^\pm$
$K2\pi$	$K^\pm\pi^\mp\pi^0$	$K_S^0\pi^\pm\pi^\mp$	$K^\pm\pi^\mp\pi^\pm$	$K_S^0\pi^\pm\pi^0$
$K3\pi$	$K^\pm\pi^\mp\pi^\pm\pi^\mp$	$K_S^0\pi^\pm\pi^\mp\pi^0$	$K^\pm\pi^\mp\pi^\pm\pi^0$	$K_S^0\pi^\pm\pi^\mp\pi^\pm$
$K4\pi$	$K^\pm\pi^\mp\pi^\pm\pi^\mp\pi^0$	$K_S^0\pi^\pm\pi^\mp\pi^\pm\pi^\mp$	$K^\pm\pi^\mp\pi^\pm\pi^\mp\pi^\pm$	$K_S^0\pi^\pm\pi^\mp\pi^\pm\pi^0$

4.2 Data Set

4.2.1 Data Samples

This analysis is based on 711 fb^{-1} of data which corresponds to 772×10^6 $B\bar{B}$ pairs. The data with SVD2 has been reprocessed using improved track finding and photon reconstruction.

4.2.2 Monte Carlo (MC) Samples

We use Monte Carlo (MC) simulation to tune the selection criteria, evaluate the reconstruction efficiency, study background sources and so on. The MC samples are generated using an event generator called EvtGen [47] and hadronization process is implemented by JETSET [48]. Detector response simulation is performed based on GEANT [49].

Signal MC Sample

For the signal MC samples, the $B \to K\ell^+\ell^-$, $B \to K^*\ell^+\ell^-$, and non-resonant $B \to X_s\ell^+\ell^-$ events are produced separately. The X_s mass is required to be more than 1.1 GeV/c^2 in the non-resonant $B \to X_s\ell^+\ell^-$ events, We mixed these three MC samples, using the experimental branching fraction (see Table 2.2). We used the average branching fraction of e^+e^- and $\mu^+\mu^-$ assuming lepton universality. we use the branching fraction, 0.45×10^{-6}, 1.08×10^{-6}, and $2.13 (= 3.66 - 0.45 - 1.08) \times 10^{-6}$ for the $B \to K\ell^+\ell^-$, $B \to K^*\ell^+\ell^-$, and non-resonant $B \to X_s\ell^+\ell^-$, respectively

The fraction of the X_s decay states covered by this semi-inclusive method is approximately 62% (41% and 21% for $K^\pm$ and K_S^0 states respectively) by 18 hadronic states, and 50% (40% and 10% for $K^\pm$ and K_S^0 states respectively) by 10 hadronic states. If the fraction of the states containing a K_L^0 is taken to be equal to that containing a K_S^0, the missing states that remain unaccounted for is about 18% (41%) of the total rate by 18 (10) hadronic states.

Note that the A_{FB} of the non-resonant $B \to X_s\ell^+\ell^-$ MC samples is calculated to match the theoretically predicted A_{FB} in quark level $b \to s\ell^+\ell^-$ transition, where the spin of the X_s is neglected. Namely, the A_{FB} is independent from the final state of the X_s in our non-resonant $B \to X_s\ell^+\ell^-$ MC samples. However, the A_{FB} would actually depend on the X_s mass and the final state of the X_s. We assume the A_{FB} of the quark level $b \to s\ell^+\ell^-$ transition is equal to the A_{FB} of the sum of the $K\ell^+\ell^-$, $K^*\ell^*\ell^-$ and non-resonant $X_s\ell^+\ell^-$:

$$A_{FB}(b \to s\ell^+\ell^-) = A_{FB}\left([B \to K\ell^+\ell^-] + [B \to K^*\ell^+\ell^-] + [B \to \text{non-resonant } X_s\ell^+\ell^-]\right)$$

Under the above assumptions, the A_{FB} measured in this analysis is equal to the theoretically predicted $A_{FB}(b \to s\ell^+\ell^-)$.

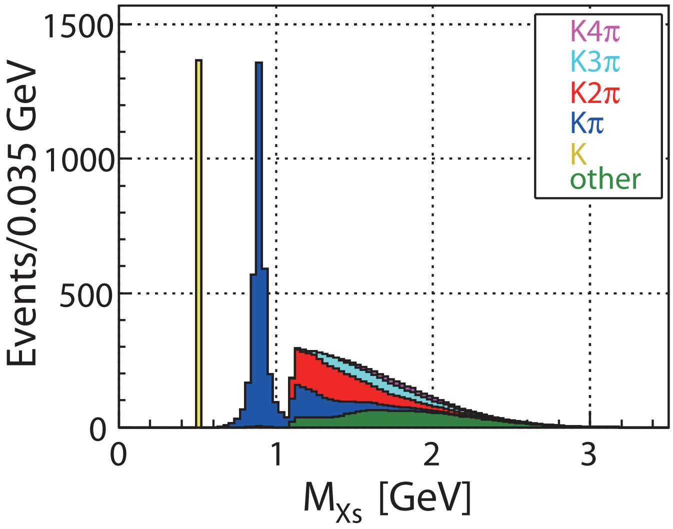


Figure 4.1: X_s mass distribution for generated signal MC samples. The histograms are scaled to the Bell full dataset.

Background MC sample

For the background study, we use continuum ($e^+e^- \to q\bar{q}$; $q = u, d, s, c$) and $B\bar{B}$ ($B^0\bar{B}^0$ and B^+B^-) MC samples which corresponds to six and ten times the amount of data available, respectively. The background MC samples of same amount as data is packed up by the unit of ”stream”.

4.3 Particle Selection

We use $e^\pm, \mu^\pm, K^\pm, \pi^\pm, K_S^0$ and π^0 for the reconstruction of $X_s\ell^+\ell^-$ in this study. All charged tracks other than $K_S^0 \to \pi^+\pi^-$ daughters are required to have a point of closest approach to the beam line within 1.0 cm in the radial direction and within 5.0 cm in the beam-axis direction.

Charged tracks are idendified using the partcile identification (PID) likelihood ratio ($\mathcal{P}_e$, $\mathcal{P}_\mu$, $\mathcal{P}_{K/\pi}$). Details on PID can be found in Sec. 3.3. The $e^\pm$ candidates are required to satisfy $\mathcal{P}_e > 0.80$. The $\mu^\pm$ candidates are required to satisfy $\mathcal{P}_\mu > 0.97$. Electrons and muons are required to have lab-frame momenta greater than 0.40 GeV/c and 0.80 GeV/c, respectively. To recovery bremsstrahlung photon, photons detected within 0.05 radians cone along the original lepton direction are added.

The $K^{\pm}$ candidates are equired $\mathcal{P}_{K/\pi} > 0.6$ for K candidates. After selecting the electron, charged kaon and muon candidate tracks, the remaining tracks are assumed to be charged pions.

The K^0_S candidates are formed by combining two oppositely charged tracks, assuming both are pions. Since the K^0_S's can be selected with low background, we apply a loose mass selection that requires an invariant mass within 15 MeV/c^2 of the K^0 mass. We then impose the following additional requirements: (1) The closest approach of the two pion tracks must have a large distance to the IP in the plane perpendicular to the electron beam line; (2) the pion tracks must intersect at a common vertex that is displaced from the IP; (3) the K^0_S candidate's momentum vector should originate from the IP.

Neutral pions are reconstructed from two photon and required to have a lab-frame energy greater than 400 MeV, photon daughter energies greater than 50 MeV, and a $\gamma\gamma$ invariant mass satisfying $|M_{\gamma\gamma} - M_{\pi^0}| < 10$ MeV/c^2.

The particle selection criteria are summarized in Table 4.2.

Table 4.2: Summary of the particle selection criteria.

Event selection		cut value
Charged tracks	$\lvert dr\rvert$	< 1.0 cm
	$\lvert dz\rvert$	< 5.0 cm
$e^{\pm}$ candidate	P_{lab}	> 0.4 GeV/c
	eid.prob(3,-1,5)	> 0.80
$\mu^{\pm}$ candidate	Not selected as electron candidate	
	P_{lab}	> 0.8 GeV/c
	muid.Likelihood()	> 0.97
	atc_pid probability	< 0.6
$K^{\pm}$ candidate	Not selected as electron, muon candidate	
	atc_pid probability	> 0.6
$\pi^{\pm}$ candidate	Remaining tracks after selecting lepton and $K^{\pm}$ tracks.	
K^0_s candidate	good K_s of K_s finder class	
	$\left\lvert M_{\pi^+\pi^-} - M_{K^0_S}\right\rvert$	< 15 MeV/c^2
π^0 candidate	E_γ	> 50 MeV
	E_{π^0}	> 400 MeV
	$\lvert M_{\gamma\gamma} - M_{\pi^0}\rvert$	< 10 Mev/c^2

4.4 Event selection criteria

We reconstruct 18 hadronic states from the combination of the $K^{\pm}$, $\pi^{\pm}$, K^0_S, and π^0 candidates. The total charge of an X_s candidate is required to be 0 or ± 1. We prohibit charged B decaying to the opposite charged K. For example, the decay of $B^+ \to K^-\pi^+\pi^+\ell^+\ell^-$ is not allowed.

There is a large contamination from charmonium background events $B \to J/\psi\,(\psi(2S))\,X_s$ decays with $J/\psi\,(\psi(2S)) \to \ell^+\ell^-$. To reject such events, we applied charmonium veto.

Veto windows are defined as follow:

$$
\begin{aligned}
-0.40\,\mathrm{GeV}/c^2 &< M_{ee(\gamma)} - M_{J/\psi} < +0.15\,\mathrm{GeV}/c^2 && (4.1)\\
-0.25\,\mathrm{GeV}/c^2 &< M_{\mu\mu} \quad - M_{J/\psi} < +0.10\,\mathrm{GeV}/c^2 && (4.2)\\
-0.25\,\mathrm{GeV}/c^2 &< M_{ee(\gamma)} - M_{\psi(2S)} < +0.10\,\mathrm{GeV}/c^2 && (4.3)\\
-0.15\,\mathrm{GeV}/c^2 &< M_{\mu\mu} \quad - M_{\psi(2S)} < +0.10\,\mathrm{GeV}/c^2. && (4.4)
\end{aligned}
$$

If one of lepton from a J/ψ or $\psi(2S)$ decay erroneously pick up a random photon in the Bremsstrahlung-recovery process, the dilepton mass can increase sufficiently to evade the above selections. Therefore the charmonium veto is applied to the dilepton mass both before and after Bremsstrahlung recovery. Some bremsstrahlung photons, which fail in bremsstrahlung recovery process, form fake π^0 with random photon. Events including such a fake π^0 peak in the $M_{\rm bc}$ distribution. Additional charmonium veto is applied to following four modes : $B \to K^{\pm}\pi^0 e^+e^-$, $K^0_S\pi^0 e^+e^-$, $K^{\pm}\pi^{\mp}\pi^0 e^+e^-$, and $K^0_S\pi^{\pm}\pi^0 e^+e^-$. Veto windows are defined as follow using the invariant mass of dilepton with energetic photon from π^0 :

$$-0.15\,\mathrm{GeV}/c^2 < M_{ee\gamma} - M_{J/\psi} < +0.05\,\mathrm{GeV}/c^2. \qquad (4.5)$$

The di-electron from γ conversion ($\gamma \to e^+e^-$) or π^0 Dalitz decay($\pi^0 \to e^+e^-\gamma$) can be background sources. We also require $M_{ee} > 0.2$ GeV/c^2.

The random combinations of dilepton from semileptonic B or D decays are one of the largest background sources. To remove such events, we applied loose selection for vertex information. Since particles except K^0_S should have a same vertex, we require $|\Delta z_{\ell^+\ell^-}| < 190$ μm, where $|\Delta z_{\ell^+\ell^-}|$ is the distance between the two leptons along the beam direction; the z-coordinate of each lepton is determined at the point of closest approach to the beam axis. We also require confidence level of the B vertex constructed from all the charged daughter particles, excluding the K^0_S daughters is more than 10^{-18}. The random combination of dilepton from semileptonic B or D decays are effectively reduced by the Neural network later.

To remove a large fraction of the combinatorial background, we reject candidates with an X_s invariant mass M_{X_s} greater than 2.0 GeV/c^2. The dilepton candidate and an X_s candidate are combined to form a B meson candidate. Two independent kinematic variables, the beam constrained B meson mass ($M_{\rm bc}$) and the energy difference(ΔE) are calculated in the $\Upsilon(4S)$ rest frame, which are defined as follows

$$
\begin{aligned}
M_{\rm bc} &\equiv \sqrt{E^2_{beam} - |\vec{p}_B|^2} && (4.6)\\
\Delta E &\equiv E_B - E_{beam}, && (4.7)
\end{aligned}
$$

where E_{beam} is the CM-frame beam energy and $(\vec{p}_B, E_B)$ is the reconstructed CM-frame four-momenta of B candidate. The $M_{\rm bc}$ can be calculated more precisely than $M_B = \sqrt{E^2_B - |\vec{p}_B|^2}$, because the spread of beam energy is smaller than that the resolution of E_B. In addition, the information of B energy can be used as another independent variable, energy difference. For truly reconstructed signal events, the $M_{\rm bc}$ is the B meson mass and the ΔE is zero. To remove combinatorial background events, we require -0.10 GeV $< \Delta E <$ 0.05 GeV for the dielectron case (-0.05 GeV $< \Delta E <$ 0.05 GeV for the dimuon case). The fit region for the signal extraction is 5.22 GeV/$c^2 < M_{\rm bc} <$ 5.30 GeV/c^2. The selection criteria are summarized in Table 4.3, and the number of reconstructed events after event selection are shown in Fig. 4.2.

	K	K_S	$K\pi$	$K_S\pi$	$K\pi^0$	$K_S\pi^0$	$K2\pi$	$K_S2\pi$	$K1\pi\pi^0$	$K_S1\pi\pi^0$	$K3\pi$	$K_S3\pi$	$K2\pi\pi^0$	$K_S2\pi\pi^0$	$K4\pi$	$K_S4\pi$	$K3\pi\pi^0$	$K_S3\pi\pi^0$	Total
K	69.4 69.15 20.18	0.06	1.46	0.06	0.25		0.22		0.01		0.01								342.76
K_S	0.16	22 21.91 12.95	0.13	0.94	0.09	0.04	0.02	0.07	0.02	0.01	0.01								169.06
$K\pi$	0.17	0.02	112.8 102.07 13.85	0.69	2.62	0.07	7.05	0.24	0.97	0.03	0.69	0.07	0.23	0.01	0.08		0.03		737.11
$K_S\pi$	0.05	0.03	1.02	35.11 29.47 7.96	0.28	0.7	0.53	1.47	0.25	0.43	0.1	0.3	0.05	0.04	0.02	0.01			370.43
$K\pi^0$	0.08		5.04	0.12	20.67 17.5 4.75	0.09	1.98	0.02	1.3	0.04	0.22		0.33				0.03		368.51
$K_S\pi^0$	0.01		0.22	3.25	0.34	5.65 4.78 2.61	0.09	0.72	0.15	0.79		0.16	0.01	0.1					183.19
$K2\pi$			0.38		0.2		32.58 26.83 10.82	0.16	0.99	0.01	2.13	0.07	0.78	0.01	0.29	0.01	0.08		247.88
$K_S2\pi$			0.01	0.18		0.03	0.42	9.4 7.49 6.04	0.08	0.54	0.23	1.21	0.17	0.21	0.02	0.07	0.04	0.04	124.38
$K1\pi\pi^0$			0.31		0.23		4.68	0.07	14.33 10.56 3.51	0.11	2.02	0.03	2.32	0.03	0.28		0.32		301.54
$K_S1\pi\pi^0$			0.04	0.18		0.05	0.14	1.1	0.19	4.47 2.99 2.01	0.1	0.56	0.18	0.55	0.02	0.03	0.01	0.1	148.22
$K3\pi$							0.34		0.13		6.22 4.14 7.31	0.07	0.56		0.53		0.18		56.75
$K_S3\pi$							0.01	0.07		0.06	0.11	1.93 1.16 4.09	0.04	0.15	0.01	0.07	0.01	0.04	28.11
$K2\pi\pi^0$			0.03				0.95	0.02	0.55	0.02	2.98	0.1	7.33 3.99 2.27	0.06	0.69		0.77	0.01	176.56
$K_S2\pi\pi^0$						0.01	0.04	0.23	0.07	0.31	0.08	1.71	0.22	1.94 0.97 1.12	0.05	0.26	0.08	0.4	88.38
$K4\pi$							0.01				0.2		0.1		0.89 0.45 3.56	0.01	0.07		13.6
$K_S4\pi$							0.01	0.01			0.01	0.09		0.02		0.25 0.14 2.05	0.02	0.04	6.77
$K3\pi\pi^0$							0.08		0.03		0.75	0.04	0.74	0.03	1.51	0.03	2.15 0.79 1.09	0.05	70.9
$K_S3\pi\pi^0$								0.03			0.02	0.29	0.01	0.15		0.27	0.02	0.58 0.15 0.44	35.62
Others	0.21	0.02	1.87	0.64	0.7	0.21	2.72	0.65	2.13	0.94	2.42	1.05	3.82	0.99	1.64	0.34	2.51	1.07	2147.54
Total	70.17	22.23	123.41	41.2	25.44	6.92	51.97	14.28	21.34	7.9	18.34	7.77	16.89	4.36	6.24	1.49	6.39	2.41	5617.24 448.75 5.42
True	69.15 98.58	21.91 98.49	102.07 82.7	29.47 71.56	17.5 68.75	4.78 68.97	26.83 51.62	7.49 52.54	10.56 49.55	2.99 37.62	4.14 22.6	1.16 14.75	3.99 23.68	0.97 22.64	0.45 7.77	0.14 9.32	0.79 12.12	0.15 6.48	304.54 67.87

	K	K_S	$K\pi$	$K_S\pi$	$K\pi^0$	$K_S\pi^0$	$K2\pi$	$K_S2\pi$	$K1\pi\pi^0$	$K_S1\pi\pi^0$	$K3\pi$	$K_S3\pi$	$K2\pi\pi^0$	$K_S2\pi\pi^0$	$K4\pi$	$K_S4\pi$	$K3\pi\pi^0$	$K_S3\pi\pi^0$	Total
K	71.24 71.01 20.58	0.01	0.38	0.06	0.08		0.05												345.15
K_S	0.07	22.24 22.11 12.85	0.08	0.17	0.03	0.02		0.01	0.01										172.2
$K\pi$	0.12	0.01	111.57 103.43 13.9	0.43	1.92	0.05	3.65	0.17	0.55	0.04	0.27	0.01	0.16		0.01		0.02		744.06
$K_S\pi$	0.01		0.74	32.74 28.77 7.68	0.24	0.53	0.38	0.74	0.13	0.28	0.05	0.09	0.01	0.02				0.02	374.73
$K\pi^0$	0.04		3.53	0.08	20.5 17.93 4.76	0.1	1.4	0.02	0.8	0.03	0.11		0.12				0.01		376.04
$K_S\pi^0$			0.12	2.24	0.25	5.4 4.71 2.51	0.08	0.39	0.07	0.39		0.1	0.02	0.01				0.01	187.15
$K2\pi$			0.22		0.12		27.01 23.02 9.12	0.14	0.79	0.01	1.1	0.1	0.43		0.14		0.04		252.18
$K_S2\pi$			0.01	0.1		0.03	0.28	7.43 6.32 4.98	0.06	0.34	0.11	0.68	0.08	0.11	0.01	0.05	0.01		126.81
$K1\pi\pi^0$			0.24		0.2	0.01	3.12	0.03	12 9.45 3.11	0.09	1.28	0.02	1.28	0.04	0.24		0.18		304.09
$K_S1\pi\pi^0$			0.04	0.1		0.01	0.14	0.77	0.17	3.47 2.42 1.61	0.03	0.39	0.12	0.29		0.04		0.03	151.68
$K3\pi$							0.21		0.09		4.52 3.27 5.81	0.01	0.25		0.26		0.09		56.43
$K_S3\pi$							0.01	0.01	0.01	0.04	0.06	1.36 0.89 3.1	0.04	0.09	0.01	0.02		0.02	28.65
$K2\pi\pi^0$			0.01	0.01	0.01		0.66	0.01	0.35	0.03	1.77	0.04	4.86 2.96 1.68	0.01	0.48		0.38	0.01	177.48
$K_S2\pi\pi^0$							0.01	0.13	0.05	0.16	0.13	0.9	0.17	1.34 0.72 0.83	0.02	0.12	0.05	0.21	89.28
$K4\pi$											0.12		0.06		0.69 0.4 3.1		0.03		12.97
$K_S4\pi$												0.01		0.01		0.12 0.1 1.33		0.04	6.69
$K3\pi\pi^0$							0.04		0.02		0.42	0.02	0.44	0.01	0.75		1.16 0.44 0.69	0.03	69.34
$K_S3\pi\pi^0$									0.01		0.02	0.15	0.01	0.06	0.01	0.1	0.03	0.26 0.11 0.32	34.98
Others	0.14	0.04	1.19	0.46	0.54	0.17	1.75	0.41	1.54	0.72	1.64	0.63	2.28	0.61	0.99	0.22	1.37	0.6	2137.33
Total	71.67	22.35	118.22	36.53	23.94	6.34	38.83	10.4	16.86	5.67	11.78	4.63	10.36	2.64	3.69	0.73	3.5	1.28	5647.14 389.42 5.28
True	71.01 99.1	22.11 99.01	103.43 87.51	28.77 78.79	17.93 74.82	4.71 73.96	23.02 59.27	6.32 60.72	9.45 56.07	2.42 43.11	3.27 27.84	0.89 19.15	2.96 28.74	0.72 28.04	0.4 10.9	0.1 12.03	0.44 13.59	0.11 8.87	298.06 76.57

Figure 4.2: Efficiency matrix after the event selection for $X_s e^+e^-$ (left) and $X_s\mu^+\mu^-$ (right). The number of event is scaled to full-data set. Reconstruction efficiency for each mode is shown in diagonal red cells. The total reconstruction efficiency, in which the effect of missing modes is taken into account, is 5.42% and 5.28% for $X_s e^+e^-$ and $X_s\mu^+\mu^-$, respectively. Purity for each mode is shown in red cells at lowest row.

Table 4.3: Summary of the event selection criteria.

Selection on the lepton pairs				
Charmonium veto window	-0.40 GeV/c^2	$<$	$M_{ee(\gamma)} - M_{J/\psi}$	$<$ 0.15 GeV/c^2
	-0.25 GeV/c^2	$<$	$M_{\mu\mu(\gamma)} - M_{J/\psi}$	$<$ 0.10 GeV/c^2
	-0.25 GeV/c^2	$<$	$M_{ee(\gamma)} - M_{\psi(2S)}$	$<$ 0.10 GeV/c^2
	-0.15 GeV/c^2	$<$	$M_{\mu\mu(\gamma)} - M_{\psi(2S)}$	$<$ 0.10 GeV/c^2
Charmonium veto window	-0.15 GeV/c^2	$<$	$M_{ee\gamma_{\pi^0}} - M_{J/\psi}$	$<$ 0.05 GeV/c^2
for fake π^0	for $X_s e^+e^-$ where $X_s = K^\pm\pi^0, K_S\pi^0, K^\pm\pi^\mp\pi^0, K_S\pi^\pm\pi^0$			
γ-conv./π^0 dalitz decay rejection	M_{ee}	$>$ 0.2 GeV/c^2		
Minimum z distance	$\lvert\Delta z_{\ell^+\ell^-}\rvert$	$<$ 190 μm		
Selection on the reconstructed $B \to X_s\ell^+\ell^-$ candidates				
Beam constraint mass	5.22	$<$	$M_{\rm bc}$	$<$ 5.29 GeV/c^2
Energy difference	-0.10	$<$	ΔE	$<$ 0.05 GeV/c^2 for $X_s e^+e^-$
	-0.05	$<$	ΔE	$<$ 0.05 GeV/c^2 for $X_s\mu^+\mu^-$
X_s mass cut			M_{X_s}	$<$ 2.0 GeV/c^2
$X_s\ell^+\ell^-$ vertex cut	10^{-18}	$<$	C.L. of B-vertex	

Chapter 5

Background Study

5.1 Background Sources

In this study, main background source is random combination from semi-leptonic B decay. The B meson decays into semi-leptonic state via the $b \to c \to s, d$ decay chain. Leptons from semi-leptonic B decays are picked up as lepton candidate in $X_s\ell^+\ell^-$, and form large combinatorial background. In these events, the lepton candidates are accompanied by different decay products of B or D mesons. This background also has a significant amount of missing energy due to the neutrinos from the semi-leptonic decays. Another background source is continuum processes $e^+e^- \to q\bar{q}$ $(q = u, d, s, c)$. This background is efficiently suppressed with event-shape variables. There are a few background modes that can peak in the signal window (”peaking background”), besides above backgrounds. Details about the peaking backgrounds will be described in Sec. 5.3.

5.2 Background Suppression with NeuroBayes

To remove background events effectively, we employ the NeuroBayes package [50], which is a highly sophisticated tool for multivariate analysis of correlated data on the basis of Bayesian statistics. An three-layered feed-forward neural network is combined with an automated preprocessing of the input variables. The complex relationships between the input variables are learned by using a provided dataset such as simulated data, and transformed into the output for analyzing the data of interest. The output can be utilized both for classification and shape reconstruction.

The preprocessing enables to find the optimal starting point for the subsequent network training. All variables can be normalized and linearity decorrelated such that the covariance matrix of new set of input variables becomes the unit matrix. Binary or discrete variables are automatically recognized and treated accordingly. A very important option is to handle the variables with a default value or delta-function, which can e.g. occur if the vertex fitting is failed. The preprocessing is very robust and efficient for powerful classification.

In the neural network, it is of vital importance to improve the generalization, which is realized by an advanced regularization and pruning techniques. Employing regularization based on Bayesian statistics practically eliminates the risk of overtraining and enhance the generalization abilities. Insignificant network connections and even entire nodes are removed to ensure that the network learns only the physical features of data. The resulting network represents the minimal topology needed to correctly reproduce the characteristics of the data while being insensitive to statistical fluctuations.

5.2.1 Input variables

To separate signal and background events, we use 23 variables for NeuroBayes inputs.

- Likelihood ratio of energy difference (ΔE)
  Energy difference is defined as $\Delta E \equiv E_B - E_{beam}$, where E_{beam} is the beam energy and E_B is the reconstructed B meson energy at e^+e^- center-of-mass system (CM). For the signal event, the energy difference is around zero. To take account of bremsstrahlung photons and the energy resolution of the ECL, we use the likelihood ratio of energy difference, which is calculated from four signal PDFs (for $X_s e^+e^-$ including π^0, $X_s e^+e^-$ not including π^0, $X_s\mu^+\mu^-$ including π^0, and $X_s\mu^+\mu^-$ not including π^0) and one background PDF, as the NeuroBayes input.

- cosθ_B
  cosθ_B is cosine of the angle between the B-flight and the beam axis. The B meson tends to have perpendicular direction to the beam axis, while the $q\bar{q}$ and $B\bar{B}$ background has random direction which makes the distribution about flat.

- $|\Delta z_{\ell^+\ell^-}|$
  $\Delta z_{\ell^+\ell^-}$ is the distance between the two leptons along the beam direction; the z-coordinate of each lepton is determined at the point of closest approach to the beam axis. For the signal event, the $|\Delta z_{\ell^+\ell^-}|$ is around zero.

- Confidence level of the B vertex construction
  We construct B vertex from all the charged daughter particles, excluding the K_S^0 daughters. To avoid dependence of the number of tracks, we use confidence level instead of χ^2.

- Visible energy
  Visible energy is defined as $E_{vis} \equiv \sum_i E_i$, where E_i are the reconstructed energy at the CM-frame of all tracks (assumed to be pions) and photons in the event. The background from semi-leptonic B decay tend to have lower visible energy than the signals, due to the neutrinos from the semi-leptonic decays.

- Missing mass
  Missing mass is defined as $M_{miss} \equiv \sqrt{(2E_{beam} - \sum E_i)^2 - |\sum \vec{p_i}|^2}$, where E_{beam} is the beam energy and $(\vec{p_i}, E_i)$ is the reconstructed CM-frame four-momenta of all tracks (assumed to be pions) and all photons in the events. The missing mass for the signal event is around zero, while the missing mass for the background from semi-leptonic B decay shift from zero due to the neutrinos from the semi-leptonic decays.

- 17 shape parameters
  we employ 17 shape parameters calculated from modified Fox-Wolfram moments, called as KSFW [51]. The definition of shape parameters are described in Appendix B.

We check consistency the distribution of the NeuroBayes input parameters between the data and MC, using control samples of $K^{(*)}J/\psi$ and $X_se\mu$ samples. Here, the "J/ψ" control samples are the samples rejected by the charmonium (J/ψ) mass veto. The "$e\mu$" samples are samples reconstructed the nominal selection criteria but requiring that the two leptons have different flavor ($e^{\pm}, \mu^{\mp}$). As a result, the large discrepancies are found in the distributions of ΔE, $|\Delta z_{\ell^+\ell^-}|$, and confidence level of the B vertex construction. Then, the MC calibration is performed. The details about the MC calibration are described in Appendix C. The distributions after the MC calibration are shown in Fig. 5.1 and 5.2. The MC after the calibration and data samples are in almost good agreement. There are small discrepancies in several variables. However it is not serious for the measurement of the A_{FB}. We also check the correlation factors to $M_{\rm bc}$ and cosθ_ℓ for all input variables are less than 1%.

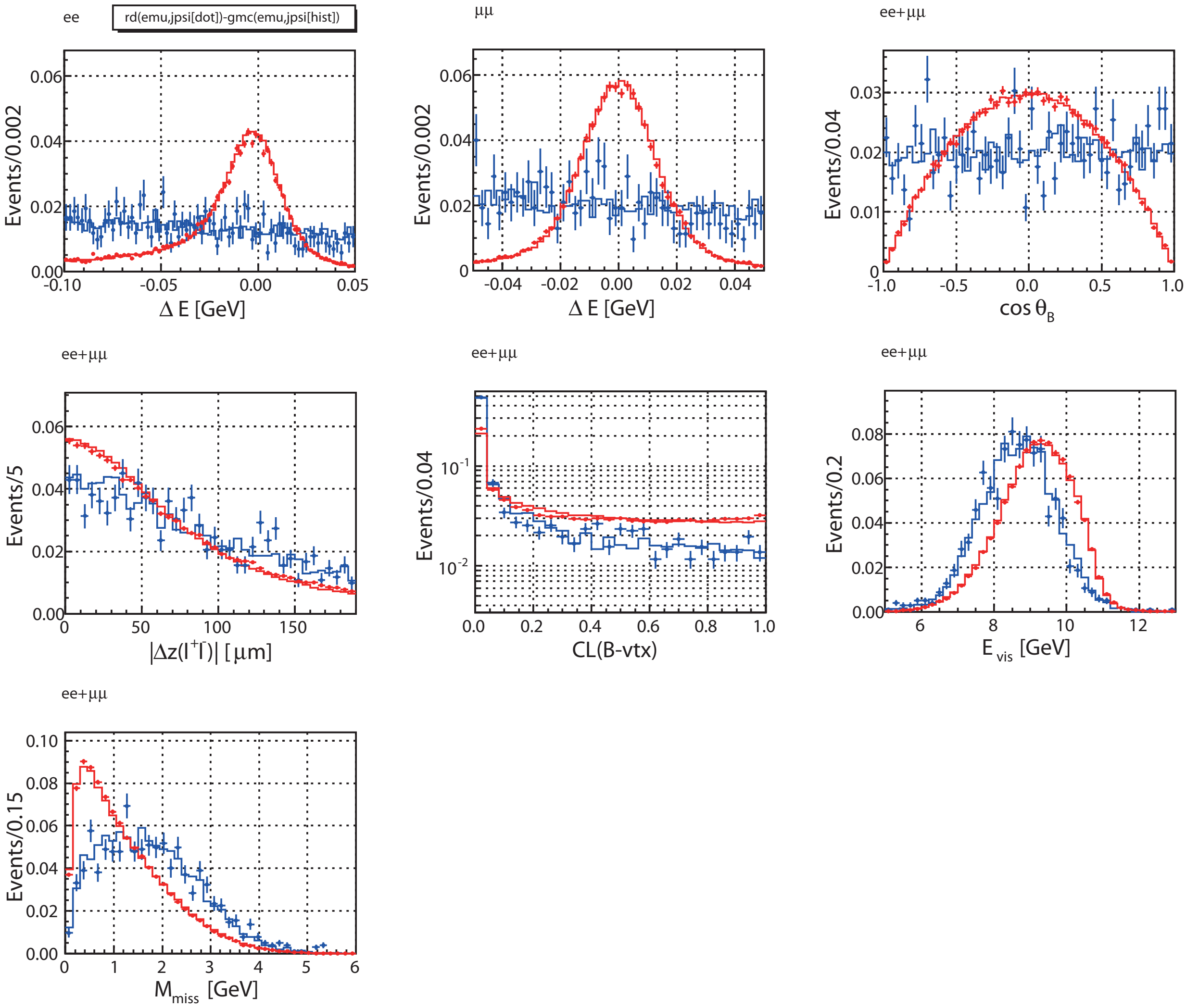


Figure 5.1: Distributions of the NeuroBayes input parameters except shape parameters. The histograms show the distribution for MC samples of $K^{(*)}J/\psi$ (red) and $X_se\mu$ (blue). The dots with error bar show the distributions for data samples.

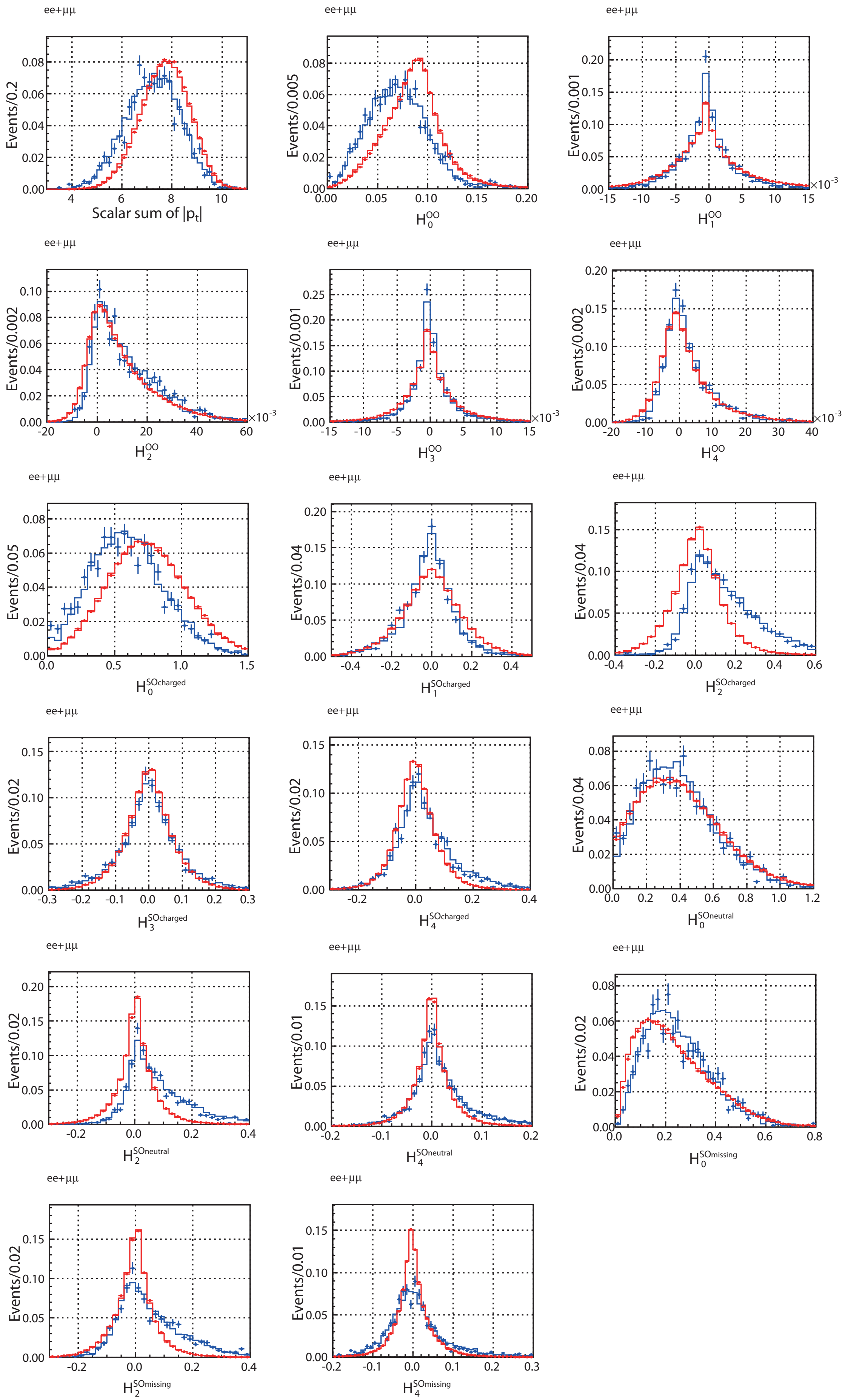


Figure 5.2: Distributions of shape parameters. The histograms show the distribution for MC samples of $K^{(*)}J/\psi$ (red) and $X_s e\mu$ (blue). The dots with error bar show the distributions for data samples. The definition of shape parameters are described in Appendix B.

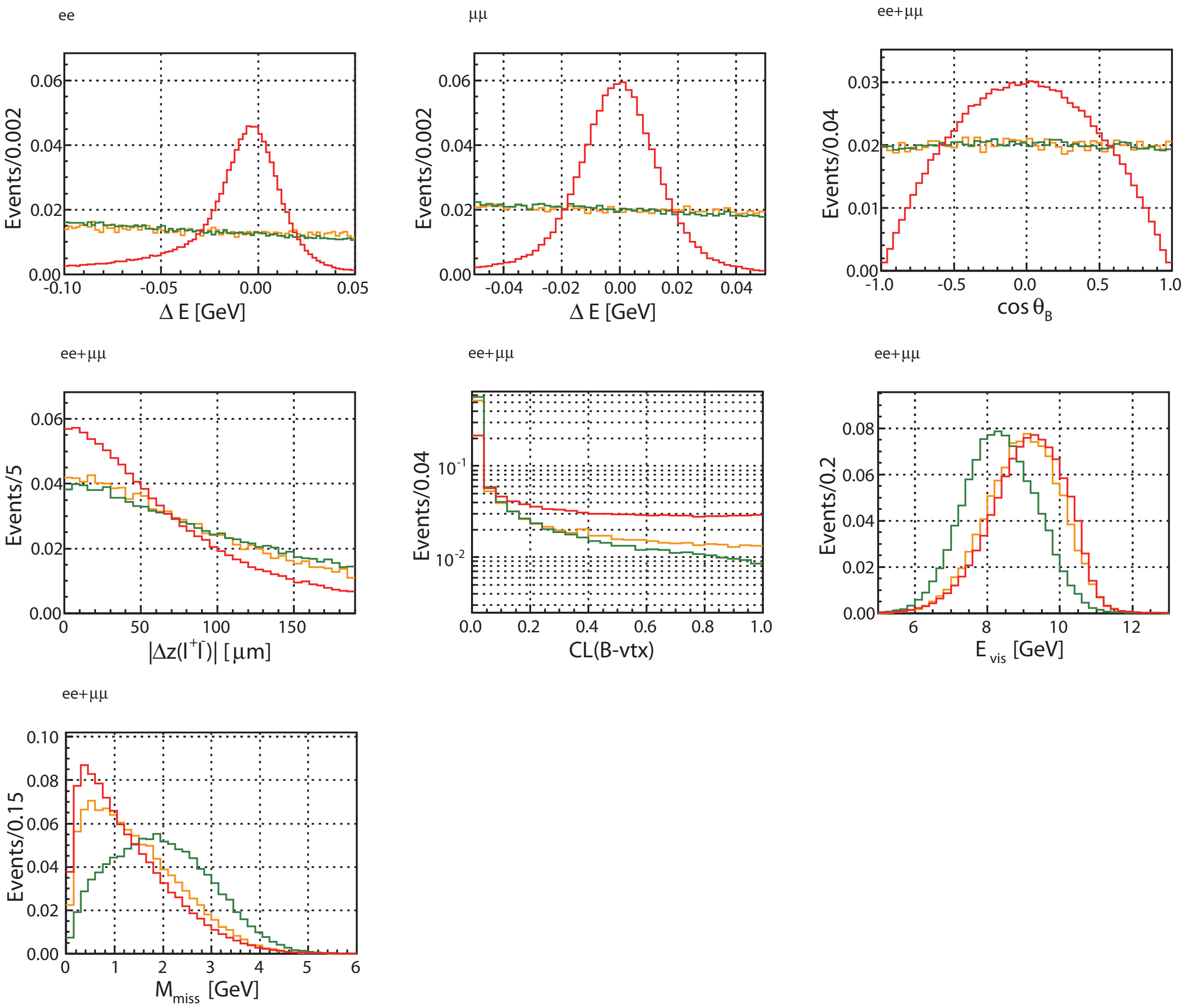


Figure 5.3: Distribution of the NeuroBayes input parameters except shape parameters for signal (red), $q\bar{q}$ (orange), and $B\bar{B}$ (green) MC.

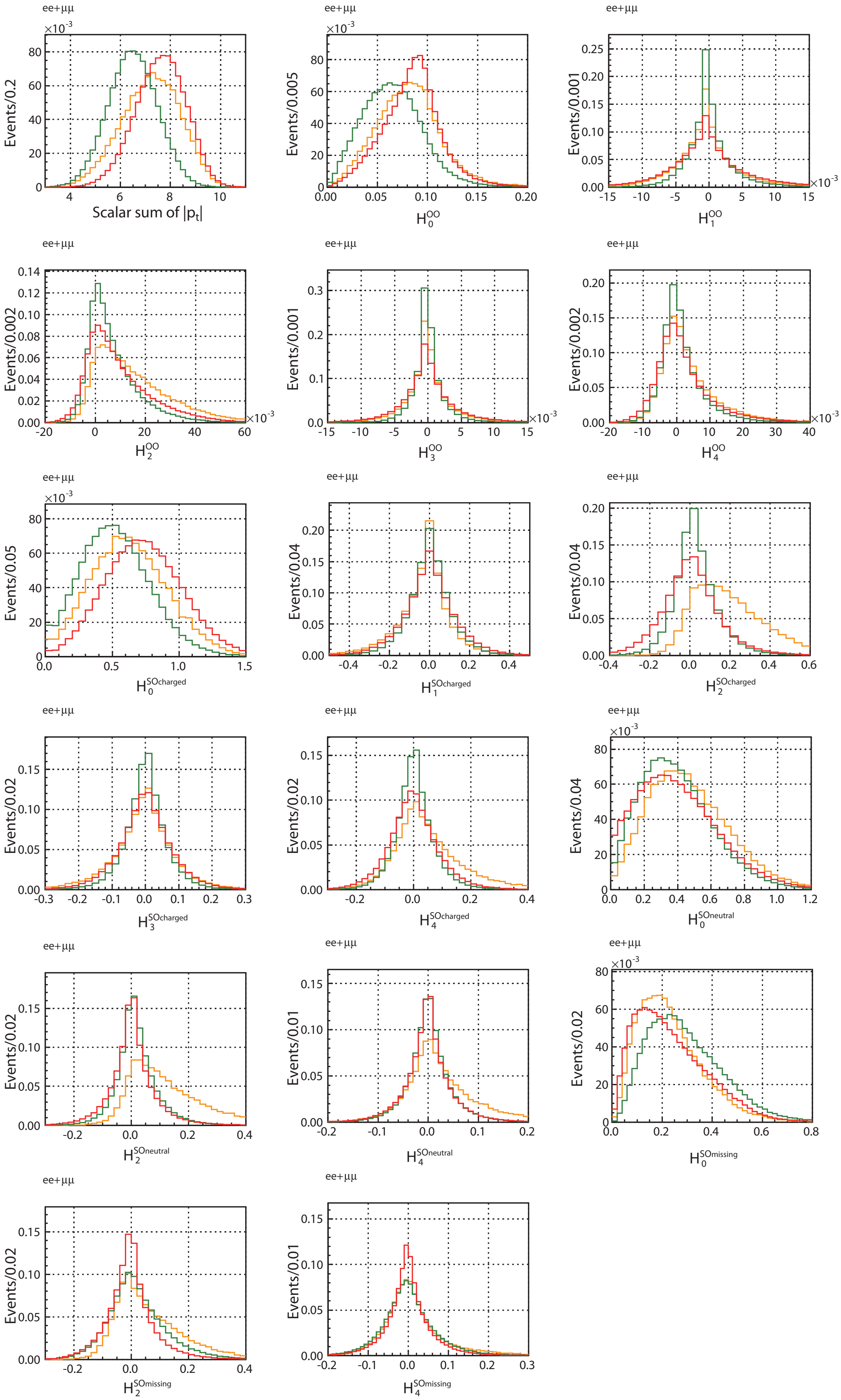


Figure 5.4: Distribution of shape parameters for signal (red), $q\bar{q}$ (orange), and $B\bar{B}$ (green) MC. The definition of shape parameters are described in Appendix B.

5.2.2 Training and Optimization

The NeuroBayes is trained with MC samples. The signal MC samples are divided into $X_s e^+e^-$ and $X_s\mu^+\mu^-$, and the background MC samples are divided into four types of $q\bar{q}$ in $X_s e^+e^-$, $q\bar{q}$ in $X_s\mu^+\mu^-$, $B\bar{B}$ in $X_s e^+e^-$, and $B\bar{B}$ in $X_s\mu^+\mu^-$. Figure 5.3 and 5.4 show the distributions of the NeuroBayes input parameters for signal, $q\bar{q}$, and $B\bar{B}$ MC samples. The NeuroBayes is trained four times separately with different types of MC samples, and obtain four NeuroBayes outputs : $NB^{ee}_{q\bar{q}}$, $NB^{\mu\mu}_{q\bar{q}}$, $NB^{ee}_{B\bar{B}}$, and $NB^{\mu\mu}_{B\bar{B}}$, where the output it denoted as NB. The MC samples used for each output are shown in Tab. 5.1. To check consistency the NB distributions between the data and MC, using $K^{(*)}J/\psi$ and $X_s e\mu$ samples as shown in Fig. 5.5. The MC and data samples are in almost good agreement. There are small discrepancies in a few outputs. However it is not serious for the measurement of the A_{FB}. Figure 5.6 show the NB distributions for signal, $q\bar{q}$, and $B\bar{B}$ MC samples. The final suppression of the combinatorial background is achieved with a cut on the NB, optimized to maximize the statistical significance of the signal. The optimization is performed separately in the resonant X_s region ($M_{X_s} < 1.1$ GeV/c^2) and non-resonant X_s region ($M_{X_s} > 1.1$ GeV/c^2). The optimized cut values are summarized in Table 5.1.

In the cases where there are more than one candidate in an event, we select the best candidate on the basis of the NeuroBayes output for $B\bar{B}$. The best candidate B selection is peripheral in this analysis because the probability to obtain multiple B candidate per event is not large. Average number of B candidates per event after background suppression with NeuroBayes is 1.12 (1.10) for $X_s e^+e^-$ ($X_s\mu^+\mu^-$) [1]. The number of reconstructed events and $M_{\rm bc}$ distributions after background suppression are shown in Fig. 5.7 and Fig. 5.8.

Table 5.1: Optimization of NeuroBayes output (NB)

NeuroBayes output	signal sample	background sample	Low M_{X_s}	High M_{X_s}
$NB^{ee}_{q\bar{q}}$	$X_s e^+e^-$	$q\bar{q}$ in $X_s e^+e^-$	> 0.91	> 0.93
$NB^{\mu\mu}_{q\bar{q}}$	$X_s\mu^+\mu^+$	$q\bar{q}$ in $X_s\mu^+\mu^-$	> 0.86	> 0.92
$NB^{ee}_{B\bar{B}}$	$X_s e^+e^-$	$B\bar{B}$ in $X_s e^+e^-$	> 0.39	> 0.91
$NB^{\mu\mu}_{B\bar{B}}$	$X_s\mu^+\mu^-$	$B\bar{B}$ in $X_s\mu^+\mu^-$	> 0.56	> 0.87

5.3 Peaking Background

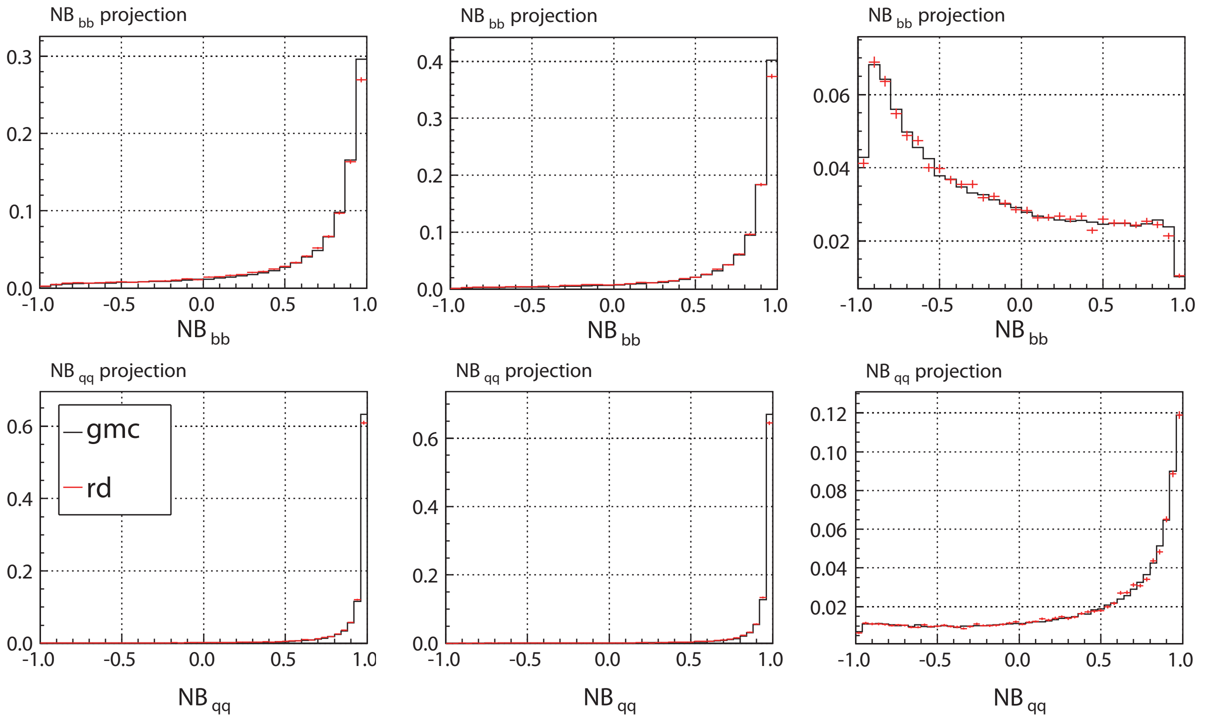
As shown in Fig. 5.8, there are peaking backgrounds. We consider three sources for peaking background events: (i) Double mis-PID events, (ii) Swapped mis-PID events, and (iii) Charmonium events.

5.3.1 Double mis-PID events

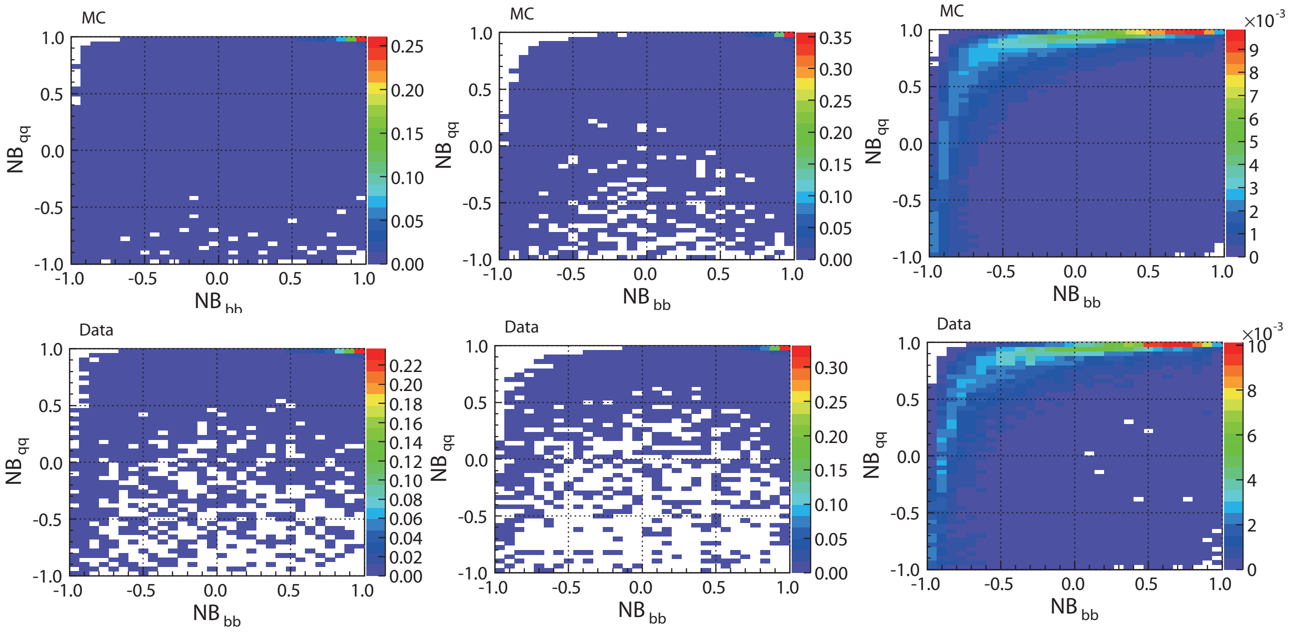
All hadronic B decays, such as $B \to D^{(*)}n\pi$ ($n > 0$), $D \to K\pi$, can be peaking background source, when two pions are misidentified as leptons. The normalization and shape of this

---

[1] Average number of B candidates per event after event selection is 1.57 (1.42) for $X_s e^+e^-$ ($X_s\mu^+\mu^-$).

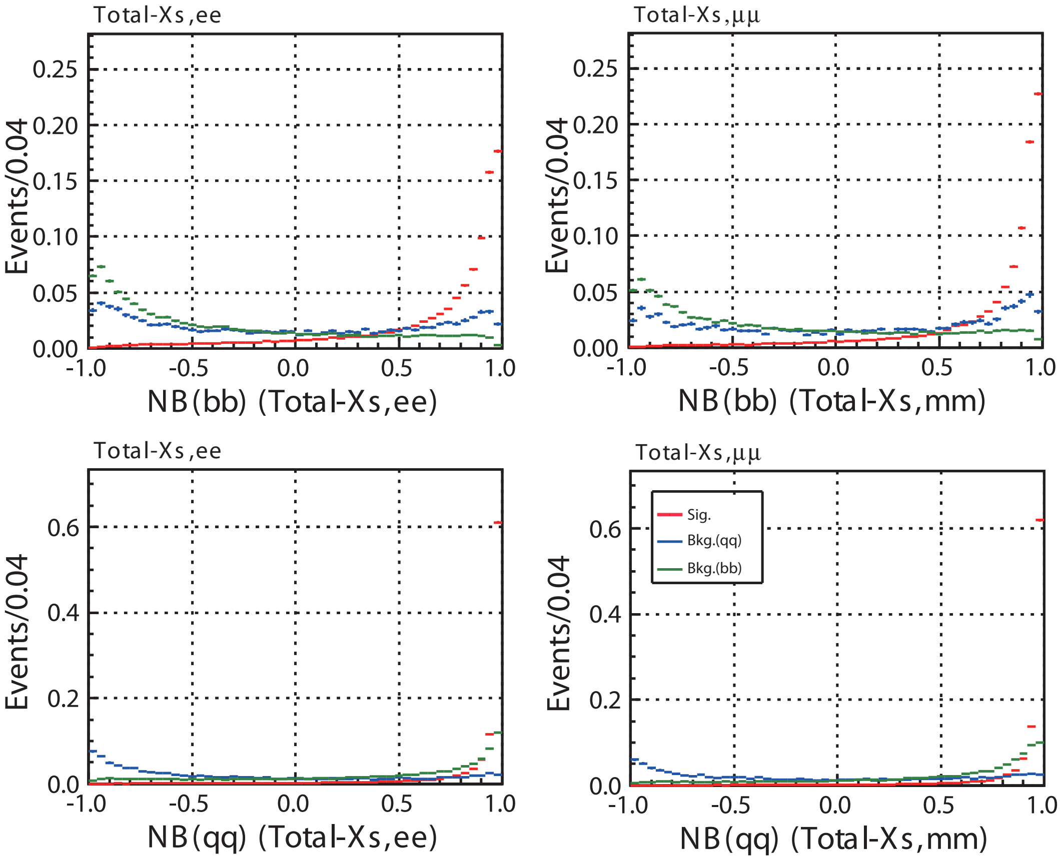


(a) Projections of $NB_{B\bar{B}}$ (top) and $NB_{q\bar{q}}$ (bottom). Black histograms show the distribution for MC samples, and red dots with error bar show the distribution for data samples.

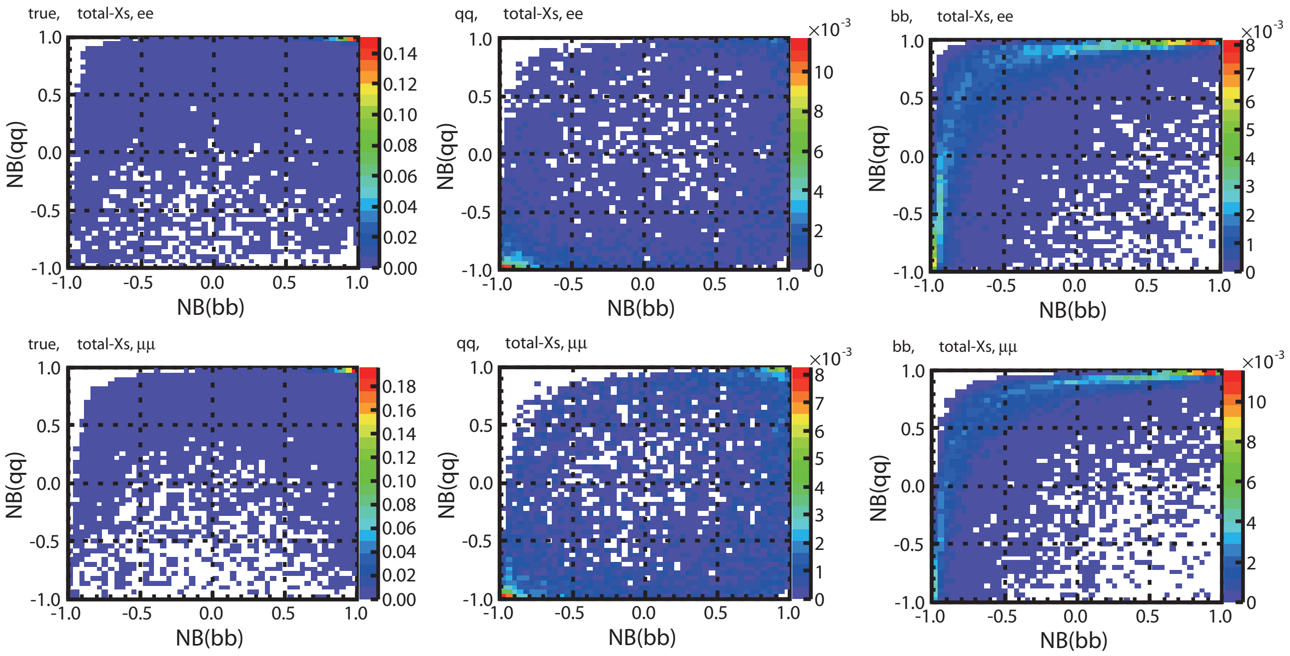


(b) Two-dimensional NB distributions for MC (top) and data (bottom).

Figure 5.5: (a) One and (b) two dimensional NB distributions in $K^{(*)}J/\psi(\rightarrow e^+e^-)$ (left), $K^{(*)}J/\psi(\rightarrow \mu^+\mu^-)$ (center), and $X_s e^\pm \mu^\mp$ (right).

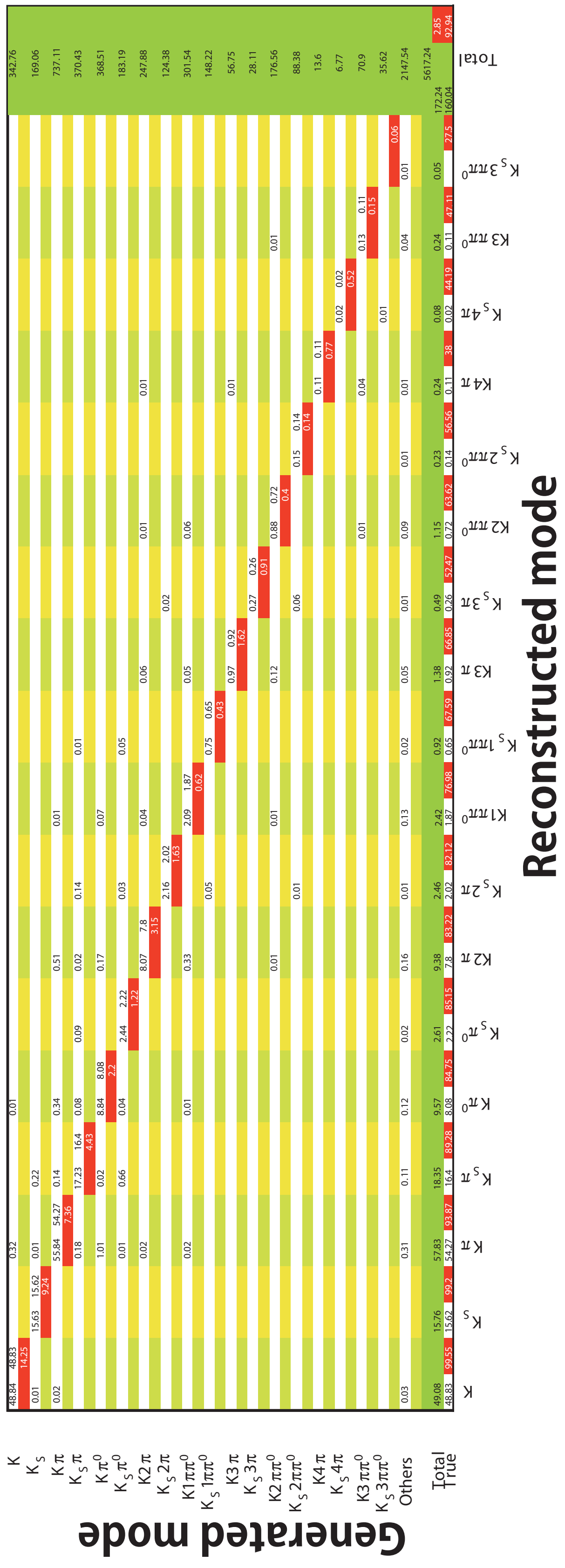


(a) Projections of $NB^{ee}_{B\bar{B}}$ (top left), $NB^{\mu\mu}_{B\bar{B}}$ (top right), $NB^{ee}_{q\bar{q}}$ (bottom left), and $NB^{\mu\mu}_{q\bar{q}}$ (bottom right). The histograms show the MC samples of signal (red), $q\bar{q}$ (blue), and $B\bar{B}$ (green).



(b) Two dimensional NB output distributions for MC samples of signal (left), $q\bar{q}$ (center), and $B\bar{B}$ (right), where the top (bottom) is for $X_s e^+e^-$ ($X_s\mu^+\mu^-$).

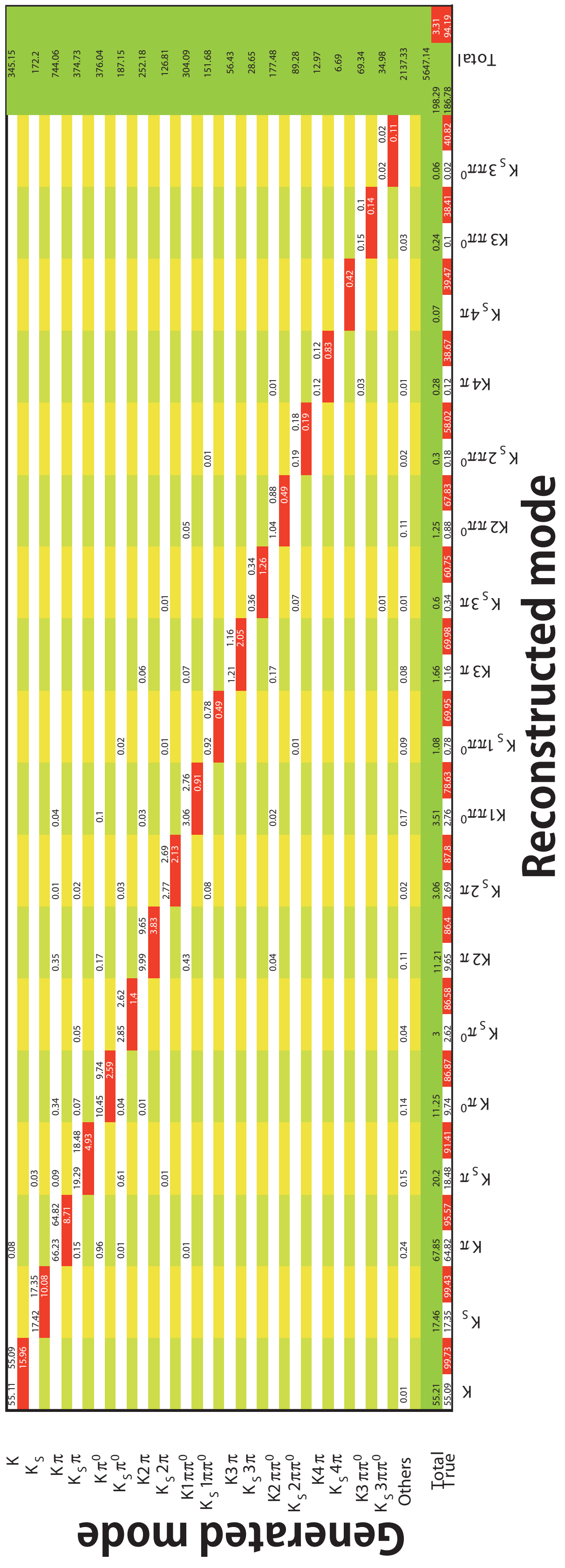
Figure 5.6: NB distributions for signal, $q\bar{q}$, and $B\bar{B}$ MC samples.



Reconstructed mode

Generated mode

	K	K_S	$K\pi$	$K_S\pi$	$K\pi^0$	$K_S\pi^0$	$K2\pi$	$K_S2\pi$	$K1\pi\pi^0$	$K_S1\pi\pi^0$	$K3\pi$	$K_S3\pi$	$K2\pi\pi^0$	$K_S2\pi\pi^0$	$K4\pi$	$K_S4\pi$	$K3\pi\pi^0$	$K_S3\pi\pi^0$	Total
K	48.84 48.83 14.25		0.32		0.01														342.76
K_S	0.01	15.63 15.62 9.24	0.01	0.22															169.06
$K\pi$	0.02		55.84 54.27 7.36	0.14	0.34		0.51		0.01										737.11
$K_S\pi$			0.18	17.23 16.4 4.43	0.08	0.09	0.02	0.14		0.01									370.43
$K\pi^0$			1.01	0.02	8.84 8.08 2.2		0.17		0.07										368.51
$K_S\pi^0$			0.01	0.66	0.04	2.44 2.22 1.22		0.03		0.05									183.19
$K2\pi$			0.02				8.07 7.8 3.15		0.04		0.06		0.01		0.01				247.88
$K_S2\pi$								2.16 2.02 1.63				0.02							124.38
$K1\pi\pi^0$			0.02		0.01		0.33		2.09 1.87 0.62		0.05		0.06						301.54
$K_S1\pi\pi^0$								0.05		0.75 0.65 0.43									148.22
$K3\pi$											0.97 0.92 1.62				0.01				56.75
$K_S3\pi$												0.27 0.26 0.91							28.11
$K2\pi\pi^0$							0.01		0.01		0.12		0.88 0.72 0.4				0.01		176.56
$K_S2\pi\pi^0$								0.01				0.06		0.15 0.14 0.14					88.38
$K4\pi$															0.11 0.11 0.77				13.6
$K_S4\pi$																0.02 0.02 0.52			6.77
$K3\pi\pi^0$													0.01		0.04		0.13 0.11 0.15		70.9
$K_S3\pi\pi^0$																0.01		0.06	35.62
Others	0.03		0.31	0.11	0.12	0.02	0.16	0.01	0.13	0.02	0.05	0.01	0.09	0.01	0.01		0.04	0.01	2147.54
Total	49.08	15.76	57.83	18.35	9.57	2.61	9.38	2.46	2.42	0.92	1.38	0.49	1.15	0.23	0.24	0.08	0.24	0.05	5617.24 172.24 2.85
True	48.83 99.55	15.62 99.2	54.27 93.87	16.4 89.28	8.08 84.75	2.22 85.15	7.8 83.22	2.02 82.12	1.87 76.98	0.65 67.59	0.92 66.85	0.26 52.47	0.72 63.62	0.14 56.56	0.11 38	0.02 44.19	0.11 47.11	27.5	160.04 92.94



Reconstructed mode

Generated mode

	K	K_S	$K\pi$	$K_S\pi$	$K\pi^0$	$K_S\pi^0$	$K2\pi$	$K_S2\pi$	$K1\pi\pi^0$	$K_S1\pi\pi^0$	$K3\pi$	$K_S3\pi$	$K2\pi\pi^0$	$K_S2\pi\pi^0$	$K4\pi$	$K_S4\pi$	$K3\pi\pi^0$	$K_S3\pi\pi^0$	Total
K	55.11 55.09 15.96		0.08																345.15
K_S		17.42 17.35 10.08		0.03															172.2
$K\pi$			66.23 64.82 8.71	0.09	0.34		0.35	0.01	0.04										744.06
$K_S\pi$			0.15	19.29 18.48 4.93	0.07	0.05		0.02											374.73
$K\pi^0$			0.96		10.45 9.74 2.59		0.17		0.1										376.04
$K_S\pi^0$			0.01	0.61	0.04	2.85 2.62 1.4		0.03		0.02									187.15
$K2\pi$					0.01		9.99 9.65 3.83		0.03		0.06								252.18
$K_S2\pi$				0.01				2.77 2.69 2.13		0.01		0.01							126.81
$K1\pi\pi^0$			0.01				0.43		3.06 2.76 0.91		0.07		0.05						304.09
$K_S1\pi\pi^0$								0.08		0.92 0.78 0.49				0.01					151.68
$K3\pi$											1.21 1.16 2.05								56.43
$K_S3\pi$												0.36 0.34 1.26							28.65
$K2\pi\pi^0$							0.04		0.02		0.17		1.04 0.88 0.49		0.01				177.48
$K_S2\pi\pi^0$										0.01		0.07		0.19 0.18 0.19					89.28
$K4\pi$															0.12 0.12 0.83				12.97
$K_S4\pi$																0.42			6.69
$K3\pi\pi^0$															0.03		0.15 0.1 0.14		69.34
$K_S3\pi\pi^0$												0.01						0.02 0.02 0.11	34.98
Others	0.01		0.24	0.15	0.14	0.04	0.11	0.02	0.17	0.09	0.08	0.01	0.11	0.02	0.01		0.03		2137.33
Total	55.21	17.46	67.85	20.2	11.25	3	11.21	3.06	3.51	1.08	1.66	0.6	1.25	0.3	0.28	0.07	0.24	0.06	5647.14 198.29 3.31
True	55.09 99.73	17.35 99.43	64.82 95.57	18.48 91.41	9.74 86.87	2.62 86.58	9.65 86.4	2.69 87.8	2.76 78.63	0.78 69.95	1.16 69.98	0.34 60.75	0.88 67.83	0.18 58.02	0.12 38.67	39.47	0.1 38.41	0.02 40.82	186.78 94.19

Figure 5.7: Efficiency matrix after the background suppression for $X_se^+e^-$ (left) and $X_s\mu^+\mu^-$ (right). The number of event is scaled to full-data set. Reconstruction efficiency for each mode is shown in diagonal red cells. The total reconstruction efficiency, in which the effect of missing modes is taken into account, is 2.85% and 3.31% for $X_se^+e^-$ and $X_s\mu^+\mu^-$, respectively. Purity for each mode is shown in red cell at lowest row.

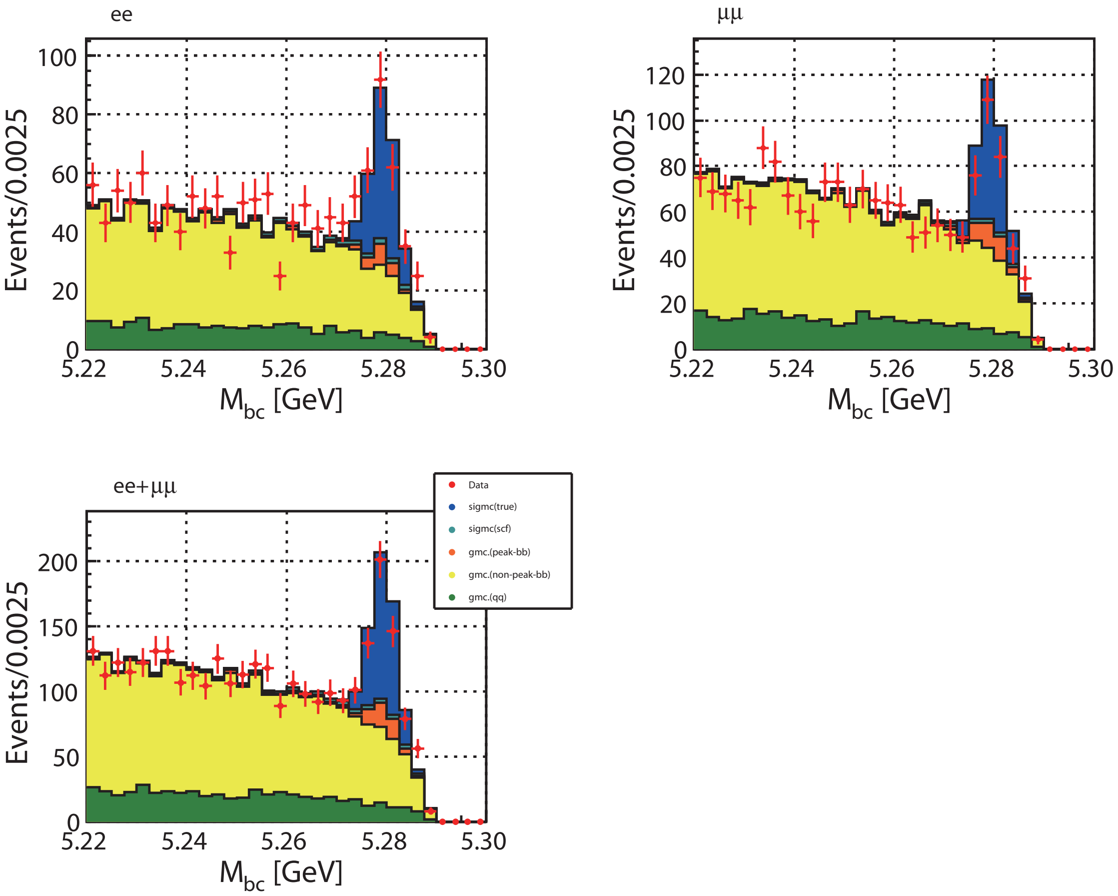


Figure 5.8: $M_{\rm bc}$ distributions after the background suppression for $X_s e^+e^-$ (top left), $X_s\mu^+\mu^-$ (top right), and $X_s\ell^+\ell^-$ (bottom left). The histograms show distributions for MC samples of $q\bar{q}$ (green), non-peaking $B\bar{B}$ (yellow), peaking $B\bar{B}$ (orange), falsely reconstructed signals (light blue), and truly reconstructed signals (blue). The dots with error bar show the distribution for data samples. The events from higher ψ resonances will be considered as signals, as explained later. However, in these plots, the events from higher ψ resonances are included in the peaking background (orange histograms).

peaking background are estimated directly from the data. We repeat the same event selection in which the lepton identification requirements are flipped, and then pick up two hadrons (mostly pion in this case) instead of dilepton. We then weigh each event with the "lepton fake rate (f_ℓ)", which is the probability to mis-identify a hadron as an electron or muon as a function of hadron momentum and direction. The details about the lepton fake rate will be described in Appendix. The weight factor w is

$$w = f_{\ell^+} \cdot f_{\ell^-}. \tag{5.1}$$

To validate the method to estimate the double mis-PID events, we compare the expected and estimated number of events in background MC samples. Figure 5.9 shows the weighted $M_{\rm bc}$ distributions estimated using the nominal event selection with flipped lepton identification requirement. The estimated number of the double mis-PID events are 0.2, 6.7, and 6.9 events for $X_s e^+e^-$, $X_s\mu^+\mu^-$, and $X_s\ell^+\ell^-$, respectively. The expected number of events are 0.3 ± 0.2, 7.2 ± 1.1, and 7.5 ± 1.1 events for $X_s e^+e^-$, $X_s\mu^+\mu^-$, and $X_s\ell^+\ell^-$, respectively, which is counted from Fig. 5.8 using generator information. The estimated and expected numbers of events are consistent within statistical errors.

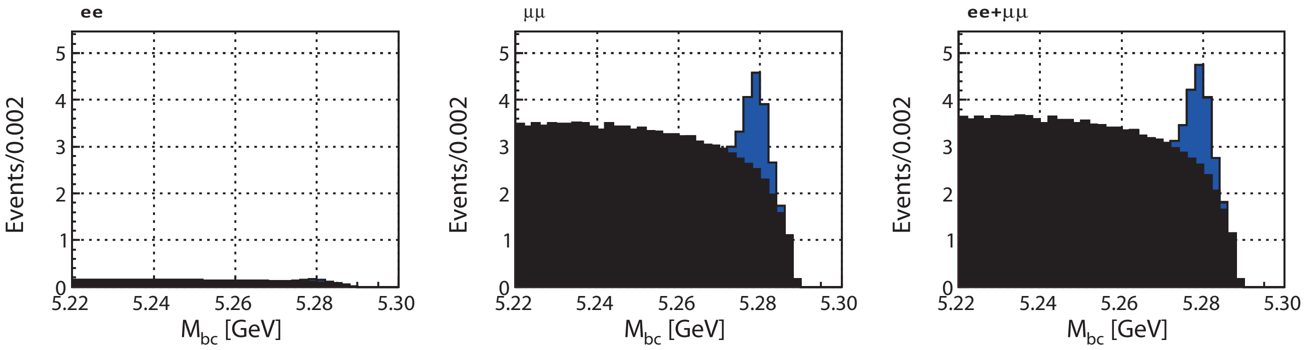


Figure 5.9: Weighted $M_{\rm bc}$ distributions estimated using the nominal event selection with flipped lepton identification requirement in $X_s e^+e^-$ (left), $X_s\mu^+\mu^-$ (center), and $X_s\ell^+\ell^-$ (right). The histograms are shown for non-peaking events (black) and double mis-PID events (blue), which correspond to the full data-set.

5.3.2 Swapped mis-PID events

$J/\psi X_s$ events can easily evade the charmonium veto, when a lepton are misidentified as a pion and a pion are misidentified as a lepton. The normalization and shape of this peaking background are estimated directly from the data. To pick up $X_s J/\psi$ and $X_s\psi(2S)$ samples, we flip the charmonium veto in the event selection. At that time, the selections for $M_{\ell^+\ell^-}$, M_{X_s}, $\Delta z_{\ell^+\ell^-}$, and NB are skipped. For the $X_s J/\psi$ and $X_s\psi(2S)$ samples, we swap lepton candidate and a hadron candidate, and calculate the $M_{\ell^+\ell^-}$, M_{X_s}, $\Delta z_{\ell^+\ell^-}$, and NB again. We require the event selection for the re-calculated variables. And then, we weigh each event with the lepton fake rate and the efficiency of the PID. The weight factor w is

$$w = \frac{f_\ell}{\epsilon_\pi} \cdot \frac{(1-\epsilon_\ell)}{\epsilon_\ell}, \tag{5.2}$$

where the ϵ_π and ϵ_ℓ are efficiency for the K/π-PID and lepton-PID, respectively.

To validate the method to estimate the swapped mis-PID events, we compare the expected and estimated number of events in background MC samples in the same way as the double mis-PID events. Figure 5.10 shows the weighted $M_{\rm bc}$ distributions estimated using the swapped X_sJ/ψ and $X_s\psi(2S)$ events. The estimated number of the swapped mis-PID events are 0.2, 6.7, and 6.9 events for $X_se^+e^-$, $X_s\mu^+\mu^-$, and $X_s\ell^+\ell^-$, respectively. The expected number of events are 0.2 ± 0.2, 7.5 ± 1.1, and 7.7 ± 1.1 events for $X_se^+e^-$, $X_s\mu^+\mu^-$, and $X_s\ell^+\ell^-$, respectively, which is counted from Fig. 5.8 using generator information. The estimated and expected numbers of events are consistent within statistical errors.

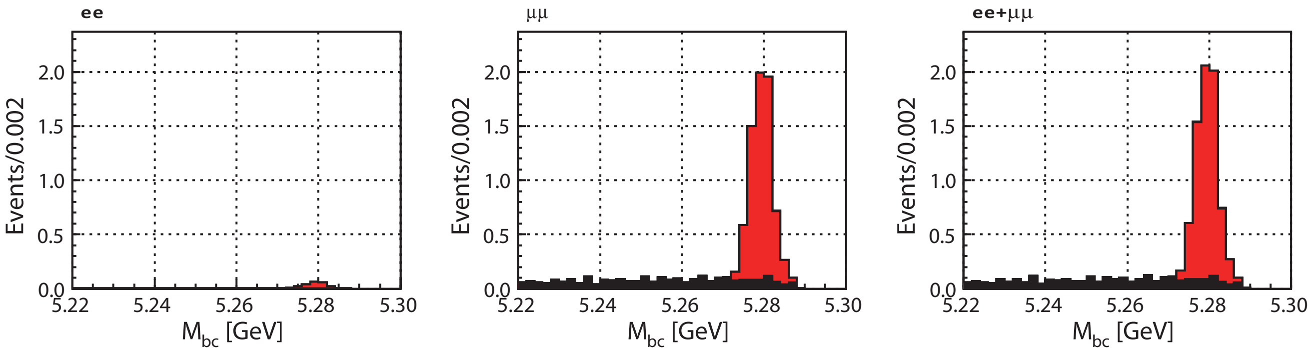


Figure 5.10: Weighted $M_{\rm bc}$ distributions estimated for the swapped mis-PID events in $X_se^+e^-$ (left), $X_s\mu^+\mu^-$ (center), and $X_s\ell^+\ell^-$ (right). The histograms are shown for non-peaking events (black) and swapped mis-PID events (red), which correspond to the full data-set.

5.3.3 Charmonium events

To veto charmonium events such as $J/\psi X_s$ and $\psi(2S)$, we applied several selections (Tab. 4.3). However, some of them would pass the selections. To estimate this peaking background, we use inclusive J/ψ and $\psi(2S)$ MC samples which corresponds to one hundred times the amount of data available. The normalization of the PDF is determined from the comparison of X_sJ/ψ samples between the MC and data samples. Figure 5.11 shows the $M_{\rm bc}$ distributions constructed using the inclusive J/ψ and $\psi(2S)$ MC and generic background MC samples. We found the difference that the peak in the generic background MC is larger than that in J/ψ and $\psi(2S)$ MC, especially in $X_se^+e^-$. The difference originates from the events in which a bremsstrahlung photon and random photon form fake π^0. However, it is difficult to estimate which MC samples describe the data, because the control sample modes including π^0, such as $K^\pm\pi^0J/\psi$, do not have enough yields. Therefore, the difference is included in systematic uncertainties. Originally, the charmonium veto for π^0 (see Tab. 4.3) is not adopt. After finding of this difference, we introduce the charmonium veto for π^0 to reduce the difference.

Events originating from higher ψ resonances, such as $\psi(4040)$ and $\psi(4160)$ can also peak. These higher resonance events have interferences with $X_s\ell^+\ell^-$, and it is difficult to separate them. Therefore, in this analysis, the higher resonance events are included in the signal, and the interpretation is left to theorists.

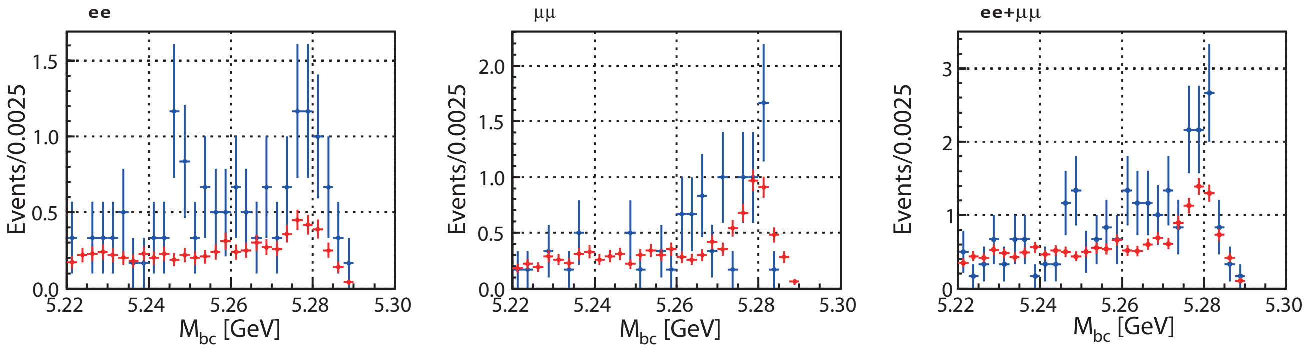


Figure 5.11: $M_{\rm bc}$ distributions estimated for the charmonium peaking events in $X_s e^+e^-$ (left), $X_s\mu^+\mu^-$ (center), and $X_s\ell^+\ell^-$ (right). The dots with error bars are constructed from inclusive J/ψ and $\psi(2S)$ MC (red) and generic background MC samples (blue). The events originating from higher ψ resonance are not included. These histograms correspond to the full data-set.

Chapter 6

Maximum Likelihood Fit

6.1 Method

The forward-backward asymmetry (A_{FB}) is extracted from the $M_{\rm bc}$ distributions in the region $M_{\rm bc} > 5.22$ GeV/c^2 using the extended unbinned maximum likelihood fit [1] . We divide q^2 into six bins, and obtain the A_{FB} for each q^2 bin, where $q^2 = (p_{\ell^+} + p_{\ell^-})^2$ is the dilepton mass-square. The binning is indicated in Fig. 6.1. The lowest q^2 bin is named bin number 1 and highest q^2 bin is named bin number 6 in sequence. The 3rd and 5th q^2 bins correspond to the J/ψ and $\psi(2S)$ veto regions, respectively. The boundary between 1st and 2nd q^2 bin is 4.3 GeV$^2/c^4$. The fit is applied simultaneously for forward ($\cos\theta_\ell > 0$) and backward ($\cos\theta_\ell < 0$) events, to obtain the A_{FB} directly.

The fit is applied simultaneously for $X_x e^+e^-$ and $X_s\mu^+\mu^-$ samples, in addition to the simultaneous fitting of the forward and backward events. Figure 6.1 shows the reconstruction efficiency of $X_s\ell^+\ell^-$ which depends on the q^2 and $\cos\theta_\ell$. The reconstruction efficiency of events with low q^2 and $\cos\theta_\ell$ close to ±1 drops by lepton's momentum selection ($P_e > 0.4$ GeV/c and $P_\mu > 0.8$ GeV/c). The observed A_{FB}, (hereafter denoted as A_{FB}^{raw}) deviates from true A_{FB} by this dependence. We derive the relation between the $A_{FB}^{raw.}$ and A_{FB}, which called "correction function" in this note. To construct the correction function, we generate MC sample, varying A_7, A_9 and A_{10} from the SM values, where A_i is a leading term of the Wilson coefficients, C_i. The values of A_i used in the EVTGEN are listed in Tab. 6.1. The A_7 is fixed at the SM value ($+A_7^{SM}$) or the sign-flipped value ($-A_7^{SM}$). The A_9 and A_{10} are varied in the range of $-2\cdot A_{9,10}^{SM}$ to $+2\cdot A_{9,10}^{SM}$, respectively. The true A_{FB} is derived from the generator information of MC samples. Then, we employ the reconstruction efficiency as the detector simulation, instead of the GSIM, and obtain the A_{FB}^{raw}. Figure 6.2 shows the correction functions with $A_7 = +A_7^{SM}$. We find the correction function especially for low q^{-2} bins and $X_s\mu^+\mu^-$ are deviated from

---

[1] The likelihood function $\mathcal{L}$ for the extended maximum likelihood fit is expressed as:

$$\mathcal{L} \;=\; \frac{\exp\left(-\sum\limits_j N_j\right)}{N!}\prod_{i=1}^{N}\left(\sum_j N_j P_j^i\right), \qquad (6.1)$$

where i runs over all events, j runs over the possible event categories (e.g. signal, background), N_j is the number of the events in each category, and P_j is the corresponding PDF.

the ideal line, as expected. These correction functions allow to reduce two parameters of $A_{FB}^{raw(ee)}$ and $A_{FB}^{raw(\mu\mu)}$ to single parameter of A_{FB} in the simultaneous fitting of $X_se^+e^-$ and $X_s\mu^+\mu^-$.

Finally, for each q^2 bin, we simultaneously fit four $M_{\rm bc}$ distributions : $X_se^+e^-/X_s\mu^+\mu^-$ and forward/backward events.

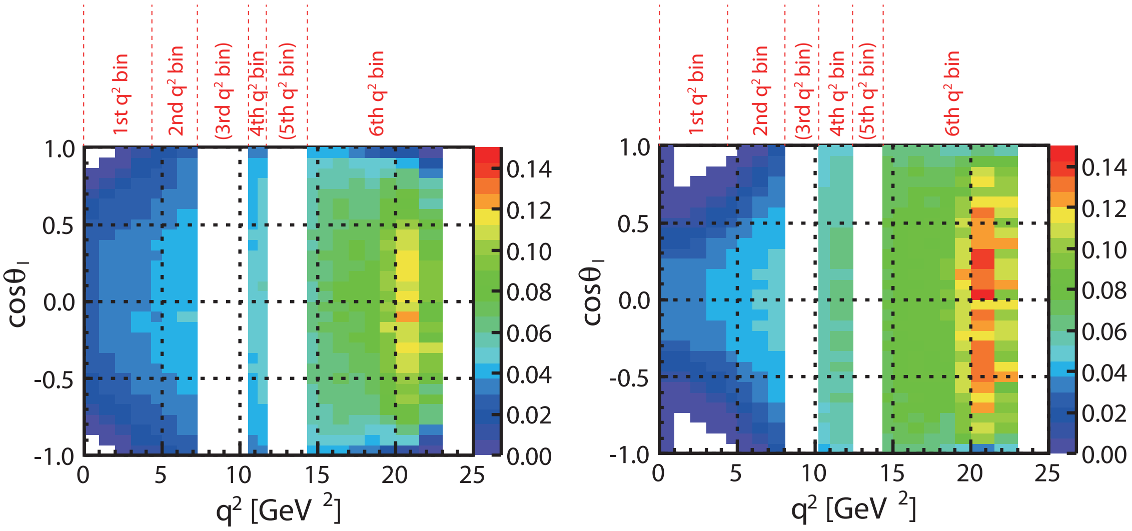


Figure 6.1: Reconstruction efficiency on q^2-cosθ_ℓ plane for $X_se^+e^-$ (left) and $X_s\mu^+\mu^-$ (right).

Table 6.1: The Wilson coefficient values used in the EVTGEN [8]. The A_i is the leading term of Wilson coefficient, C_i, with i = 7, 9, 10. In the EVTGEN, the values estimated at scale $\mu = 2.5$ (5.0) GeV are used for $q^2/m_b^2 > 0.25$ ($q^2/m_b^2 < 0.25$). The superscript (0) refers to the lowest order quantities while the superscript (1) denotes the correction terms of order α_s, i.e., $A_i = A_i^{(0)} + A_i^{(1)}$.

Parameter	Scale	
	$\mu = 2.5$ GeV	$\mu = 5$ GeV
$(A_7^{(0)}, A_7^{(1)})$	$(-0.353, +0.023)$	$(-0.312, +0.008)$
$(A_9^{(0)}, A_9^{(1)})$	$(+4.287, -0.218)$	$(+4.177, -0.035)$
$(A_{10}^{(0)}, A_{10}^{(1)})$	$(-4.592, +0.379)$	$(-4.592, +0.379)$

6.2 Probability Density Function (PDF)

In the likelihood function, we consider four components: signal, self cross-feed, non-peaking background, and peaking background. The parameters are summarized in Tab. 6.2.

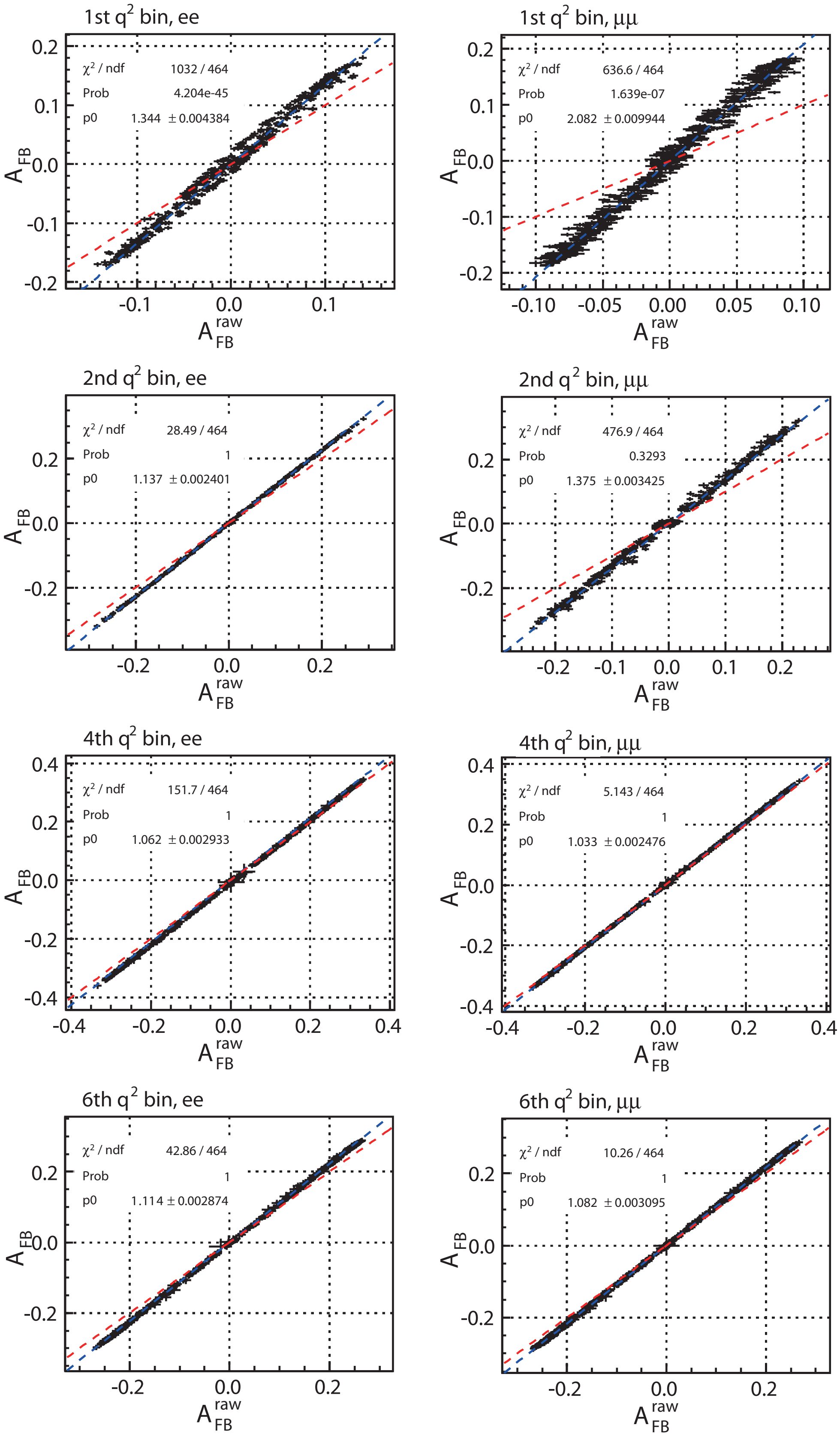


Figure 6.2: Correction functions for $X_s e^+e^-$ (left) and $X_s\mu^+\mu^-$ (right) with $C_7 = +C_7^{SM}$. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom. The distribution is fitted by linear function (blue). The ideal line ($A_{FB} = A_{FB}^{raw}$) is also shown (red). Each dot with error bar is obtained from the MC samples which have different A_i parameters.

Table 6.2: Summary of the fitting method. For each q^2 bin, we simultaneously fit four $M_{\rm bc}$ distribution : $X_s e^+e^-/X_s\mu^+\mu^-$ and forward/backward events. Floated parameters are $N_{sig(ee)}$, $N_{sig(\mu\mu)}$, A_{FB}, and four shape and yields for the ARGUS function.

PDF component	Used function	Parameter (fix or float)
Signal(sig)	Gaussian	$N_{sig(ee)}$, $N_{sig(\mu\mu)}$: float A_{FB} : float all shapes : fix
Self cross-feed(scf)	histogram-PDF	all N_{scf}/N_{sig} : fix
Non-peaking background(bkg)	ARGUS	four yields : float four shapes : float four endpoints : fix
Peaking background(pkg)	histogram-PDF	all N_{pkg} : fix

- Signal
  The signal PDF is modeled by a Gaussian:

$$f_{\rm Gaussian}(M_{\rm bc}) \propto \exp\left(-\frac{(M_{\rm bc}-\mu)^2}{2\sigma^2}\right) \tag{6.2}$$

  Mean (μ) and sigma (σ) of the Gaussian are fixed using the ARGUS+Gaussian fitting of $X_s J/\psi$ events in data samples, as shown in Fig. 6.3. The obtained parameters for Gaussian shapes are listed in Tab. 6.3. The floated parameters are signal yields for $X_s e^+e^-$ ($N_{sig(ee)}$) and $X_s\mu^+\mu^-$ ($N_{sig(\mu\mu)}$) and the A_{FB}. Note that four degree of freedom related to the signal yield, is reduced to three by the correction function.

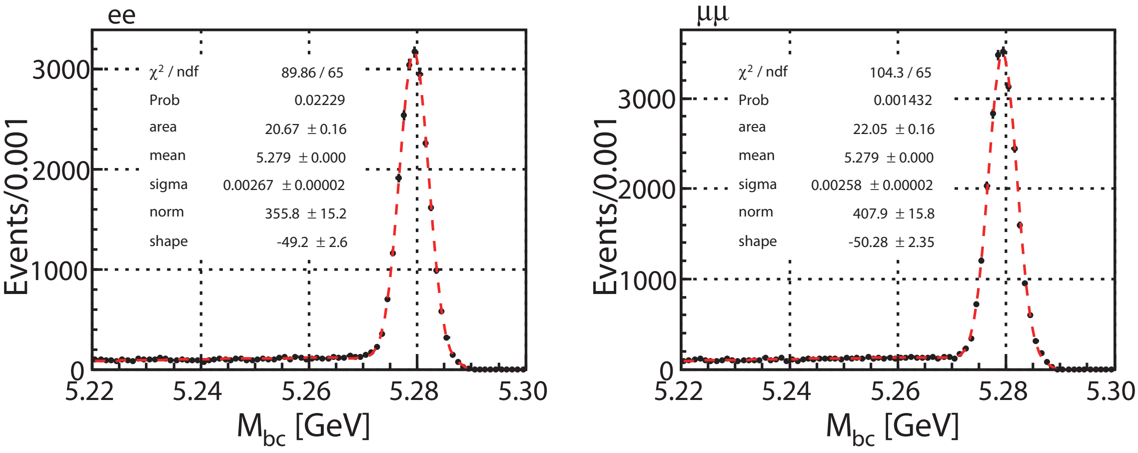


Figure 6.3: $M_{\rm bc}$ distribution for the $X_s J/\psi$ data samples for e^+e^- (left) and $\mu^+\mu^-$ (right).

- Self cross-feed
  The Self cross-feed is categorized into following three types:

Table 6.3: The fixed parameter using the $X_s J/\psi$ in data samples.

Parameter	Mode	
	$X_s e^+e^-$	$X_s \mu^+\mu^-$
μ (mean) [GeV]	5.27936 ± 0.00002	5.27932 ± 0.00002
σ (width) [GeV]	0.00267 ± 0.00002	0.00258 ± 0.00002

(i) events with right q^2 and right b-flavor
(ii) events with right q^2 and wrong b-flavor
(iii) events with wrong q^2.

For self cross-feed of (i) and (ii), we construct the histogram-PDF from signal MC samples as shown in Fig. 6.4 and 6.5, respectively. The yield of self cross-feed should be proportional to the signal, therefore, the ratio of the self cross-feed of (i)/(ii) to the signal in same/opposite cosθ bin is fixed to the values obtained from MC samples. The self cross-feed of (iii), whose ratio to the signal is less than 0.3%, is included in the systematic uncertainties.

- Non-peaking background
  The non-peaking background PDF is modeled by an ARGUS function [54]:

$$f_{\text{ARGUS}}(M_{\text{bc}}) \propto M_{\text{bc}} \cdot \sqrt{t} \cdot \exp(-at), \tag{6.3}$$

$$t \equiv 1 - \left(\frac{M_{\text{bc}}}{E_{beam}}\right)^2. \tag{6.4}$$

  The endpoint of the ARGUS function is fixed to the beam energy in the CM-frame, $E_{beam} = 5.289$ GeV. Shape parameters (a) and yields are independently floated parameters for $X_s e^+e^-/X_s\mu^+\mu^-$ and forward/backward events.

- Peaking background
  We consider three peaking background sources as explained in Sec. 5.3 :

(i) charmonium events
(ii) double mis-PID events
(iii) swapped mis-PID events.

We construct the histogram-PDF for each peaking backgrounds as shown in Fig. 6.6, 6.7, and 6.8, respectively. The data samples are used for double mis-PID and swapped mis-PID events. The inclusive J/ψ and $\psi(2S)$ MC samples are used for charmonium events. The details of method to estimate the peaking backgrounds are described in Sec. 5.3.

6.3 Fitter Check

To confirm that our fitter works correctly, we performed two kinds of tests: a toy MC test and an ensemble test.

6.3.1 Toy MC test

In the toy MC test, we check the linearity of the signal yields (N_{sig}) and the forward-backward asymmetry (A_{FB}). We generate test samples for signal, self cross-feed, non-peaking background, and peaking background events from corresponding PDFs. The number of the peaking and non-peaking background events generated are sampled from a Poisson distribution around the expected event yields in data for each component.

For the linearity check of the N_{sig}, we see the relation of the input and output N_{sig}, varying the input N_{sig}. The signal and self cross-feed events generated are sampled from a Poisson distribution, whose mean is changed from 10% to 300% of the expected yields, in steps of 5%. For each input N_{sig}, 10 MC samples are generated, and fitted. Figure 6.9 shows the relation between the input and output N_{sig}. We find the output N_{sig} is obtained correctly.

For the linearity check of the A_{FB}, we see the relation of the input and output A_{FB}, varying the input A_{FB} from -1 to $+1$ in steps of 0.05. The signal and self cross-feed events are generated by a Poisson distribution. The mean values of the Poisson distribution for forward and backward events ($\mu^{f,b}$) are determined as follows:

$$\mu^{f} = \frac{N_{sig}^{f+b}}{2} \cdot (1 + A_{FB}^{raw}) \tag{6.5}$$

$$\mu^{b} = \frac{N_{sig}^{f+b}}{2} \cdot (1 - A_{FB}^{raw})\,, \tag{6.6}$$

where superscript (f) refers to the forward events while (b) denotes the backward events. The A_{FB}^{raw} for $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$ are calculated from the input A_{FB} using the correction functions. The sum of the yields for forward and backward events (N_{sig}^{f+b}) are set to the expected yields in the data. For each input A_{FB}, 40 MC samples are generated, and fitted. Figure 6.10 shows the relation between the input and output A_{FB}. We find the output A_{FB} is obtained correctly.

From both results, we confirmed good linearity for the signal yield and the A_{FB} and no bias in the fitter.

6.3.2 Ensemble test

Since the background MC samples are limited in statistics, the ensemble test is performed only ten times. The background MC samples are used for the test sample of non-peaking and peaking background events. Since we have only six streams of continuum background MC samples, we used six streams of continuum background MC samples for the first six ensemble tests and re-used them for the rest four ensemble tests. The dominant background is $B\bar{B}$ events, the re-used is reasonable. The test sample for signal and self cross-feed events are picked up from a large amount of signal MC samples. Figure 6.11 and 6.12 show the result of ensemble test. We confirm the N_{sig} and A_{FB} are obtained correctly within errors.

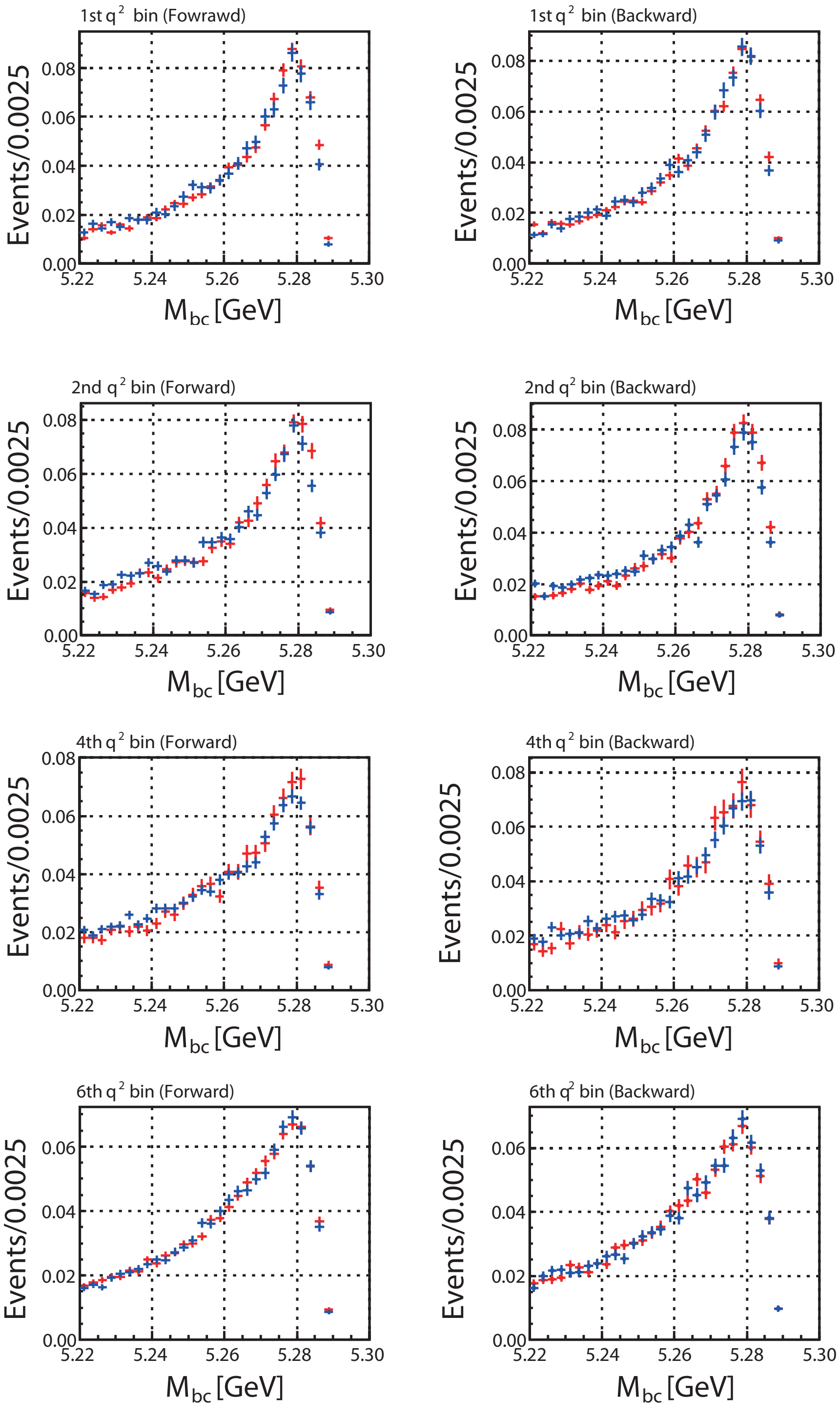


Figure 6.4: PDF of the self cross-feed events with right q^2 and right b-flavor for $X_s e^+ e^-$ (red) and $X_s \mu^+ \mu^-$ (blue). Each column corresponds to the forward (left) and backward (right) events. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

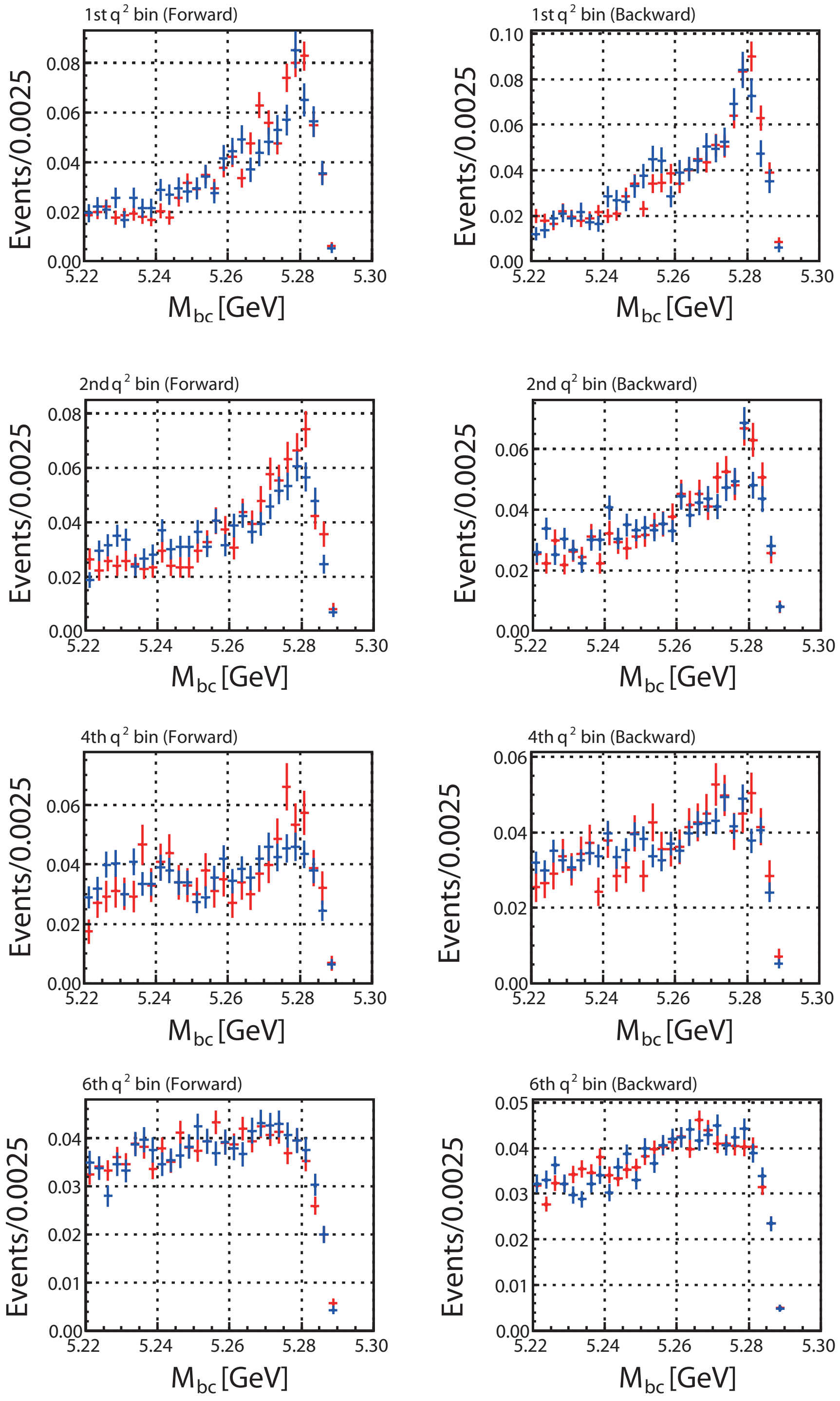


Figure 6.5: PDF of the self cross-feed events with right q^2 and wrong b-flavor for $X_s e^+ e^-$ (red) and $X_s \mu^+ \mu^-$ (blue). Each column corresponds to the forward (left) and backward (right) events. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

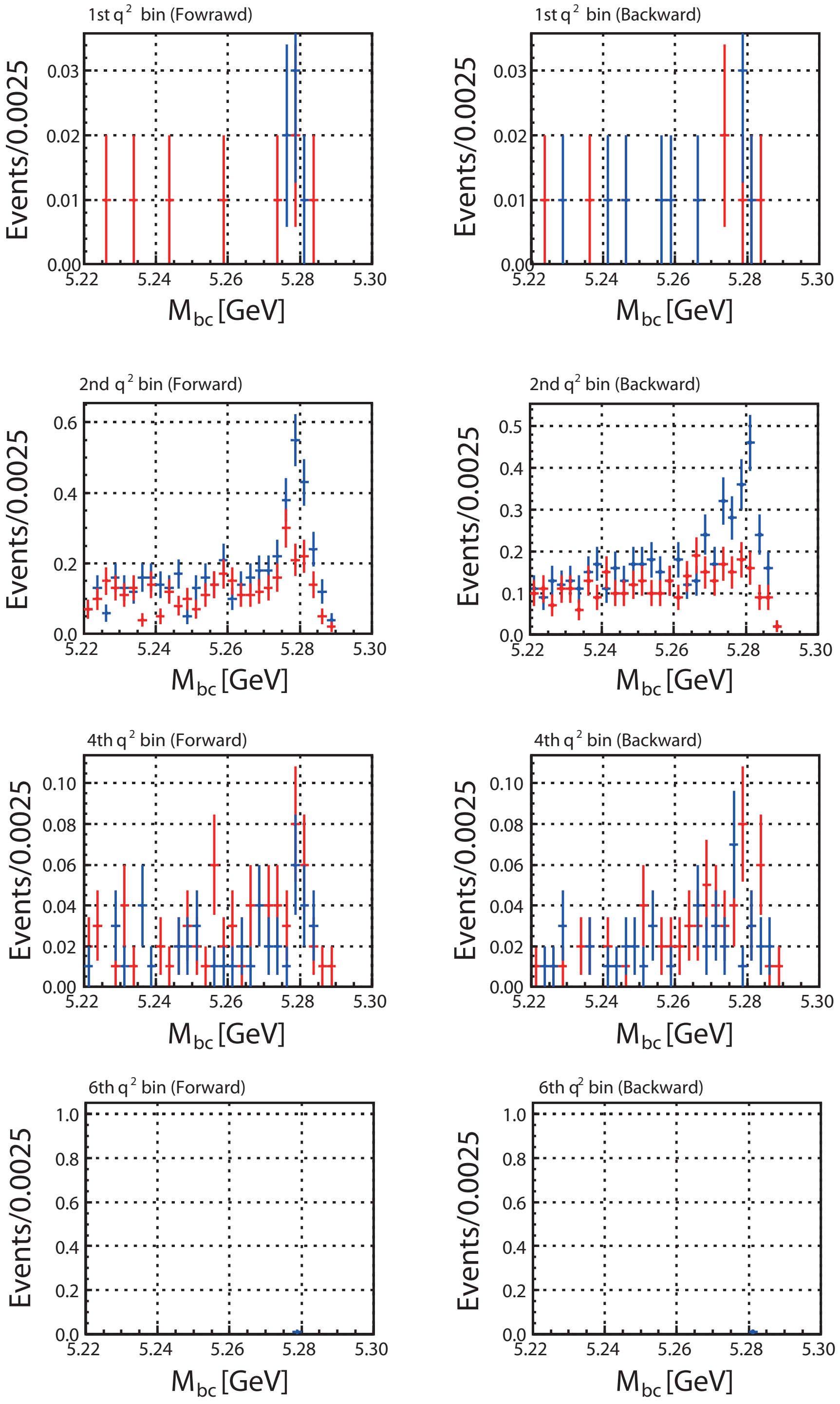


Figure 6.6: PDF of the charmonium peaking events for $X_s e^+ e^-$ (red) and $X_s \mu^+ \mu^-$ (blue). Each column corresponds to the forward (left) and backward (right) events. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

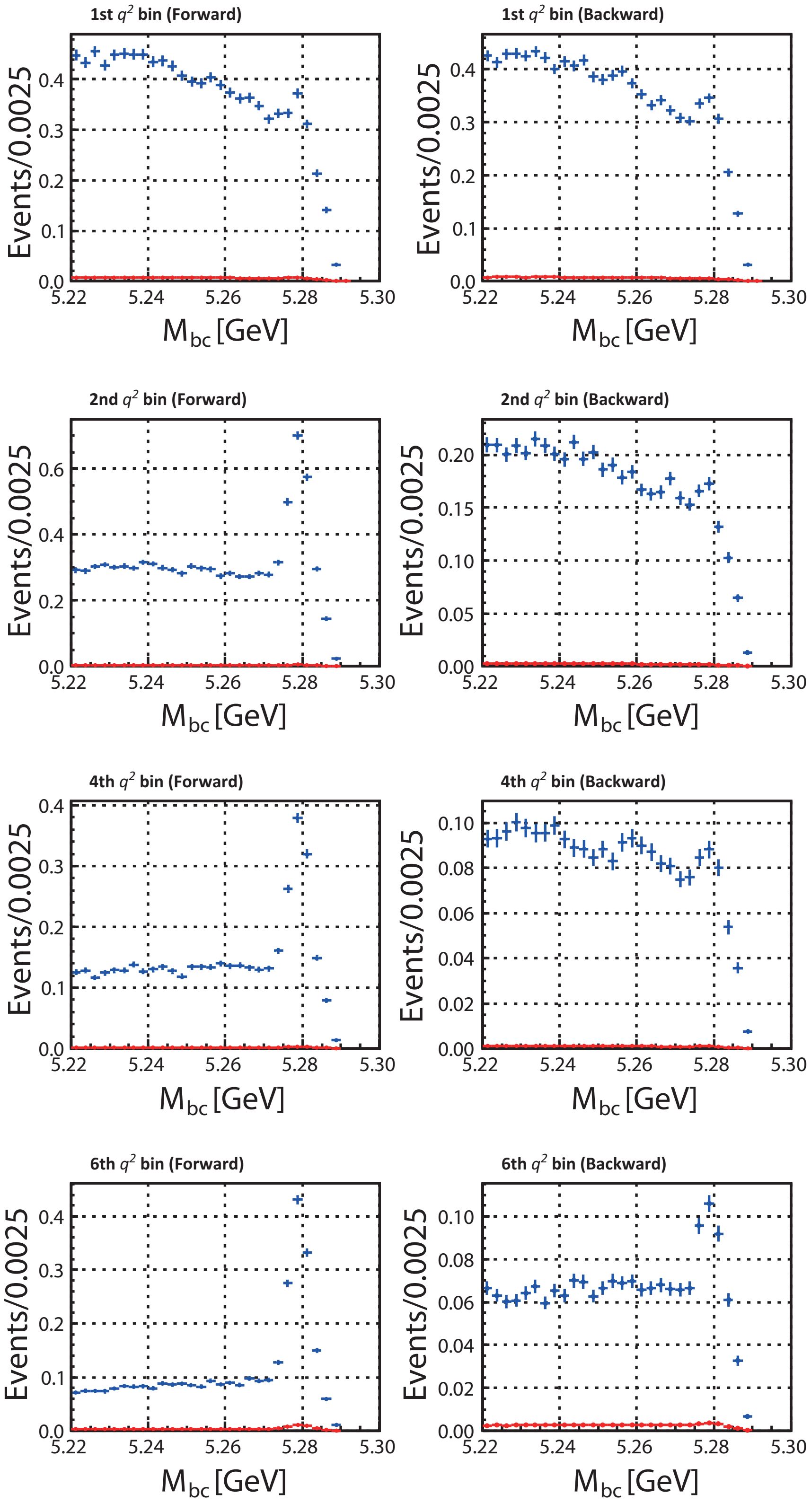


Figure 6.7: PDF of the double mis-PID events for $X_s e^+ e^-$ (red) and $X_s \mu^+ \mu^-$ (blue). Each column corresponds to the forward (left) and backward (right) events. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

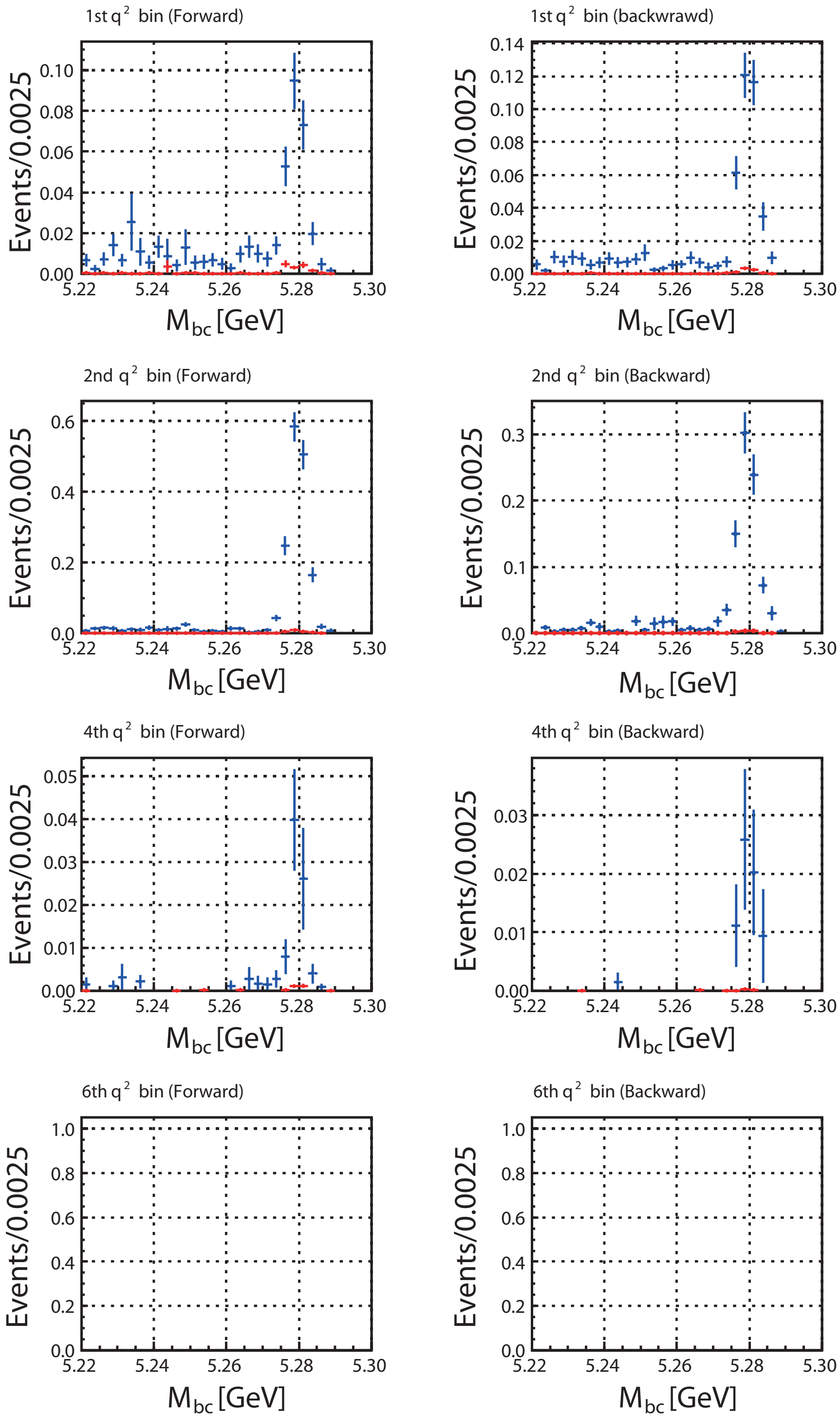


Figure 6.8: PDF of the swapped mis-PID events for $X_s e^+ e^-$ (red) and $X_s \mu^+ \mu^-$ (blue). Each column corresponds to the forward (left) and backward (right) events. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

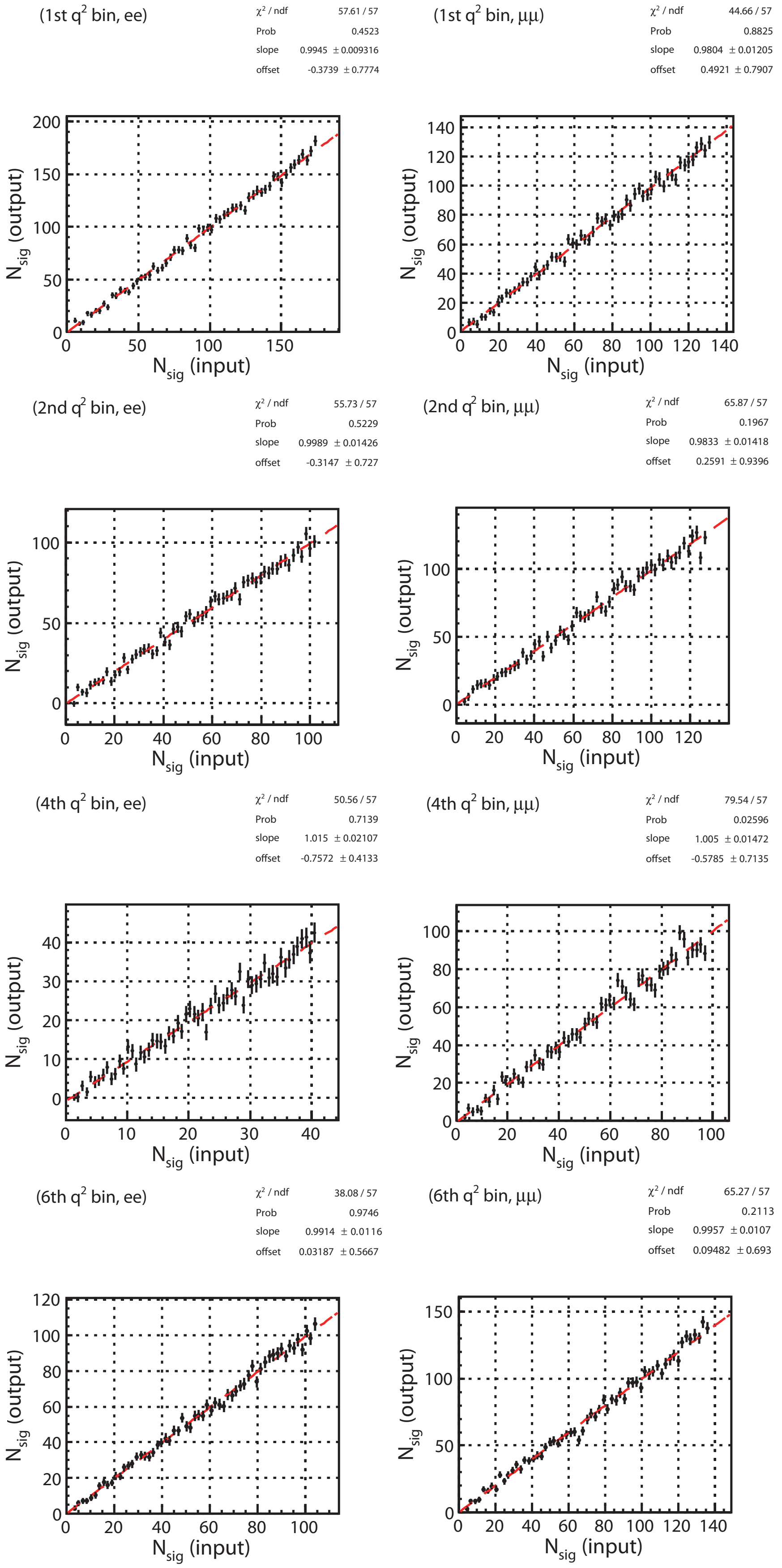


Figure 6.9: Linearity test of the N_{sig} for each q^2 bin in $X_s e^+ e^-$ (left) and $X_s \mu^+ \mu^-$ (right) events. Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

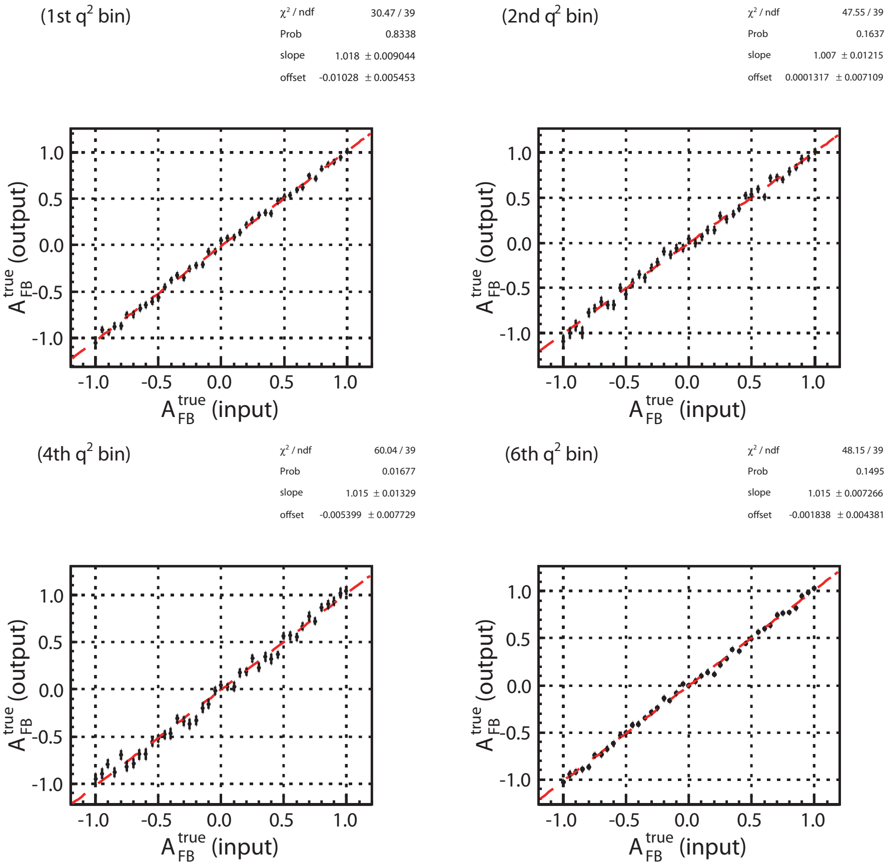


Figure 6.10: Linearity test of the A_{FB} for 1st (top left), 2nd (top right), 4th (bottom left), and 6th (bottom right) q^2 bin.

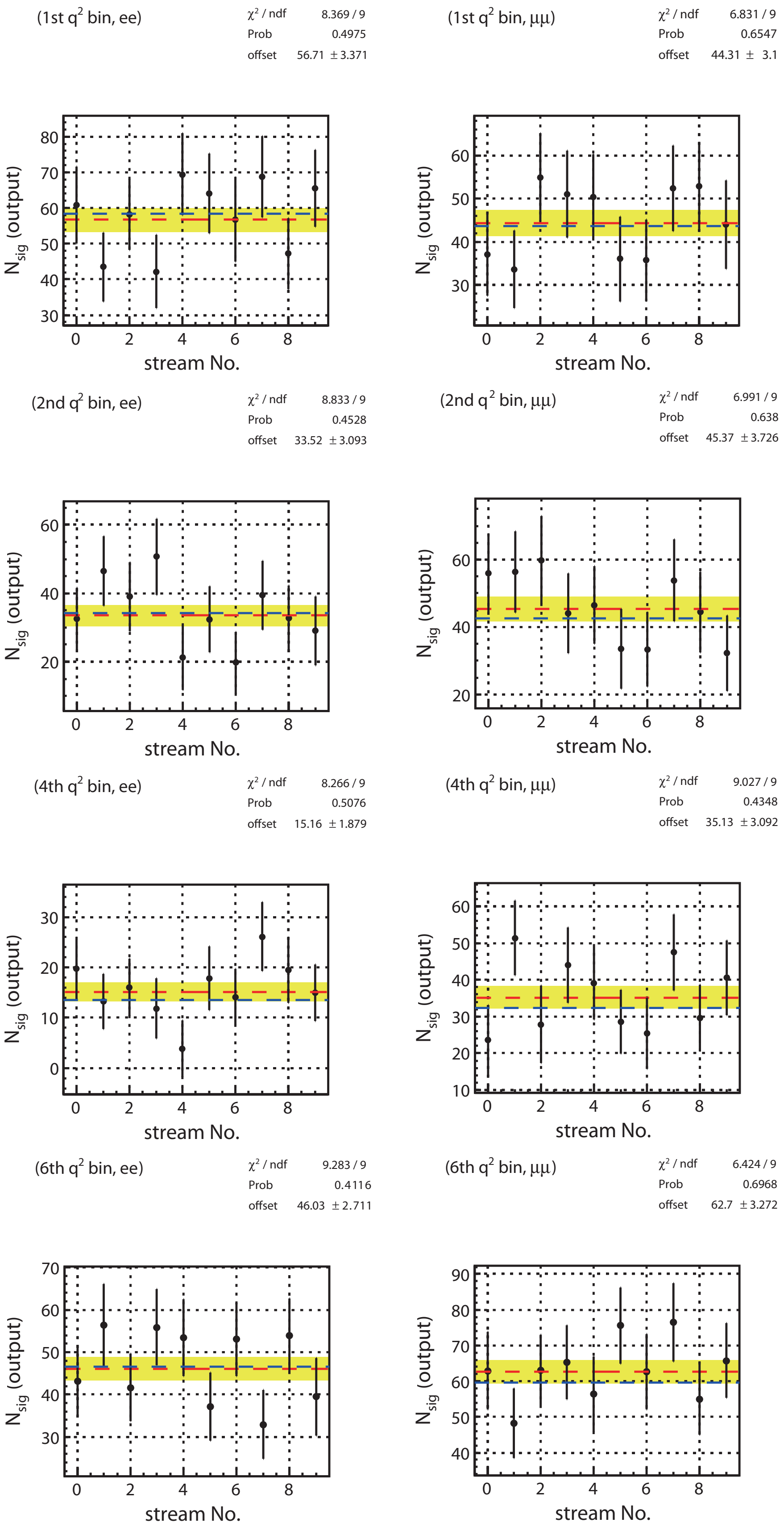


Figure 6.11: Ensemble test of the N_{sig} for $X_s e^+e^-$ (left) and $X_s\mu^+\mu^-$ (right). Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom. Red line shows the average value of total streams with 1σ error band (yellow band). Blue line shows the expected value, which is obtained from the counting using generator information.

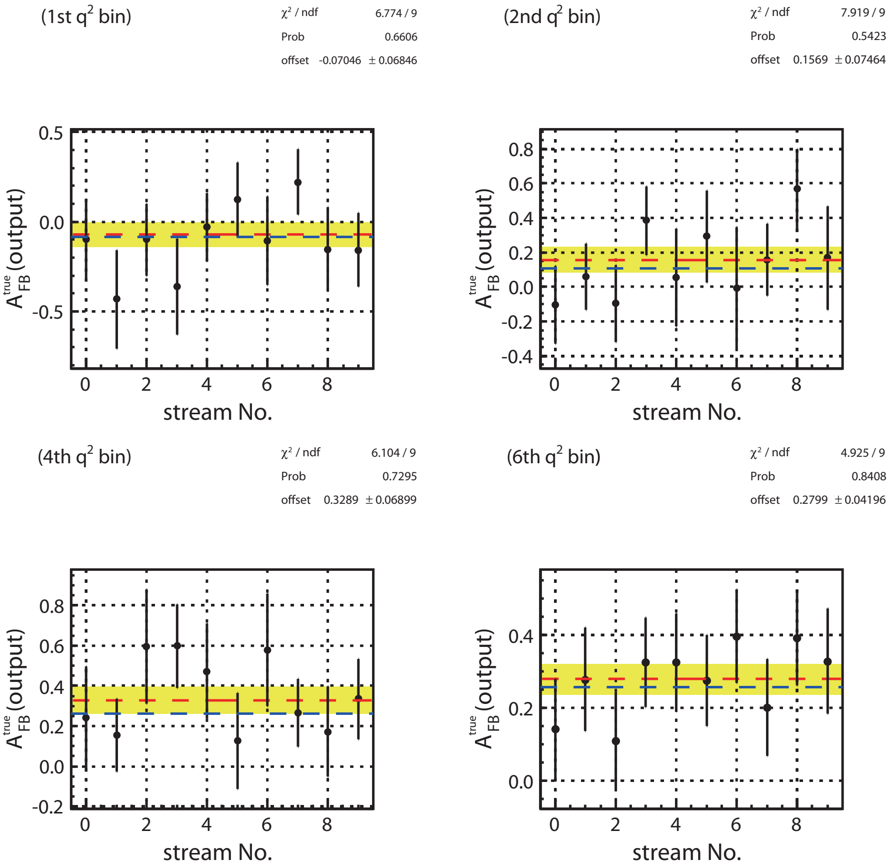


Figure 6.12: Ensemble test of the A_{FB} for 1st (top left), 2nd (top right), 4th (bottom left), and 6th (bottom right) q^2 bin. Red line shows the average value of total streams with 1σ error band (yellow band). Blue line shows the expected value, which is obtained from the counting using generator information.

Chapter 7

Systematic Study

In this section, we discuss the systematic errors. As for the systematic error of A_{FB}, following sources are considered.

- Fit parameterization
- Correction function

Note that we assume only operators in the SM, when the correction function is constructed. In our estimation of systematic uncertainties, the uncertainties from operators other than SM are not included.

7.1 Fit Parameterization

We estimate the systematic error due to the fixed parameter in the fitting. Basically, we varied the fixed parameter by reasonable range and the difference of obtained A_{FB} value is taken as the systematic uncertainties.

- Signal
  The signal PDF was modeled by a Gaussian and the shape parameters (mean and width) are fixed using the $X_s J/\psi$ data samples. The systematic uncertainties due to fixed shape parameters are estimated by varying them by $\pm 1\sigma$.

- Self Cross-Feed
  Two self cross-feed components were considered in the fitter: events with right q^2 and right/wrong b-flavor. The entries in the bins in the histogram-PDFs are varied according to the statistical error of the MC samples. We repeat the final fit to extract the A_{FB} 100 times with varied PDF and take the RMS of the A_{FB} distribution as the systematic error.

  The self cross-feed component with wrong q^2 was not considered in the fitter. The systematic uncertainties due to this self cross-feed is estimated by the fitting including the PDF of this self cross-feed component.

- Peaking Background
  The double and swapped mis-PID events were estimated using the lepton fake rate and PID efficiencies. We calculate the error of PDF normalization from the error

of the lepton fake rate and PID efficiencies, and then vary the normalization by calculated error of $\pm 1\sigma$. The errors for double miss-PID are 55% and 7% in $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$, respectively. The errors for swapped miss-PID are 28% and 4% in $X_s e^+ e^-$ and $X_s \mu^+ \mu^-$, respectively.

The charmonium peaking background was estimated using the inclusive J/ψ and $\psi(2S)$ MC samples. However, we found the difference between the inclusive J/ψ and $\psi(2S)$ MC and the generic MC samples. We perform a fitting using the PDF constructed from the generic MC, and the difference of the obtained A_{FB} value is taken as the systematic error. The normalization of charmonium peaking PDF is determined by the MC. Therefore, we varied the normalization of charmonium peaking PDF by ±100%, conservatively.

7.2 Correction Function

In the fitting, we use the correction function, which transfers from the A_{FB}^{raw} to the true A_{FB}. The reconstruction efficiency and signal modeling would affect the correction function and make systematic uncertainties.

7.2.1 Reconstruction and PID Efficiency

We constructed the correction function using the reconstruction efficiency which is estimated by the MC samples. However, the efficiencies of K_S^0 and π^0 reconstruction and the particle identification (PID) estimated by MC samples are slightly different from those in the data samples. The difference between the MC and data samples are estimated as a function of momentum and direction by the collaborators. Using them, we calculate the correction factor of K_S^0 and π^0 reconstruction and PID $(e^{\pm}, \mu^{\pm}, K^{\pm}/\pi^{\pm})$ efficiencies for each q^2-cosθ_ℓ bin. The correction factors are shown in Tab. 7.1. We apply these correction factors to the reconstruction efficiencies and construct the correction functions. However, the correction function with and without efficiency correction are almost same, because there are almost no differences between correction factors for forward and backward events. Therefore, the systematic uncertainties due to the efficiency correction is negligible.

7.2.2 Signal Modeling

When we generate non-resonant $X_s \ell^+ \ell^-$ MC samples, we input the Fermi motion parameter and b-quark mass. The hadronization process is implemented by PHYTIA. We require $X_s > 1.1$ GeV/c^2 for non-resonant $X_s \ell^* \ell^-$ MC samples and mixed them with $K\ell^+ K\ell^-$ and $K^* \ell^+ \ell^-$ MC samples. The correction functions would be influenced by the signal modeling as described above. We construct the correction functions with varied signal model, and perform the fitting to extract the A_{FB}. The difference of the obtained A_{FB} value is taken as the systematic uncertainties.

Fermi Motion Model

In our signal Monte Carlo model, we assume the Fermi motion parameter p_F to be 410 MeV/c. The parameter is varied in accordance with measurement of hadronic moments

Table 7.1: Correction ($\epsilon_{Data}/\epsilon_{MC}$) of PID and reconstruction efficiency, where ϵ_{Data} and ϵ_{MC} are the efficiencies in the data and MC samples, respectively.

$X_s e^+e^-$		e-ID	K/π-ID	Find K_S	π^0
1st q^2 bin	Forward	0.955 ± 0.036	0.970 ± 0.014	0.998 ± 0.006	0.993 ± 0.002
	Backward	0.955 ± 0.036	0.970 ± 0.014	0.998 ± 0.006	0.993 ± 0.002
2nd q^2 bin	Forward	0.963 ± 0.034	0.981 ± 0.014	0.998 ± 0.006	0.995 ± 0.002
	Backward	0.963 ± 0.034	0.981 ± 0.014	0.998 ± 0.006	0.995 ± 0.002
4th q^2 bin	Forward	0.966 ± 0.030	0.994 ± 0.015	0.997 ± 0.006	0.995 ± 0.002
	Backward	0.964 ± 0.030	0.993 ± 0.014	0.997 ± 0.006	0.995 ± 0.002
6th q^2 bin	Forward	0.964 ± 0.030	1.008 ± 0.015	0.998 ± 0.006	0.998 ± 0.001
	Backward	0.964 ± 0.030	1.008 ± 0.015	0.998 ± 0.006	0.998 ± 0.001
Total		0.960 ± 0.033	0.983 ± 0.014	0.998 ± 0.006	0.995 ± 0.002

$X_s \mu^+\mu^-$		μ-ID	K/π-ID	Find K_S	π^0
1st q^2 bin	Forward	0.930 ± 0.047	0.970 ± 0.014	0.998 ± 0.006	0.993 ± 0.003
	Backward	0.929 ± 0.047	0.970 ± 0.014	0.998 ± 0.006	0.992 ± 0.003
2nd q^2 bin	Forward	0.913 ± 0.049	0.982 ± 0.014	0.998 ± 0.006	0.994 ± 0.002
	Backward	0.912 ± 0.049	0.981 ± 0.014	0.998 ± 0.006	0.994 ± 0.002
4th q^2 bin	Forward	0.913 ± 0.045	0.994 ± 0.015	0.997 ± 0.006	0.994 ± 0.002
	Backward	0.912 ± 0.046	0.994 ± 0.015	0.997 ± 0.006	0.995 ± 0.002
6th q^2 bin	Forward	0.897 ± 0.046	1.008 ± 0.016	0.998 ± 0.006	0.998 ± 0.001
	Backward	0.897 ± 0.046	1.008 ± 0.016	0.998 ± 0.006	0.998 ± 0.001
Total		0.913 ± 0.047	0.986 ± 0.016	0.998 ± 0.006	0.995 ± 0.002

in semileptonic B decays [56] and the photon spectrum in inclusive $B \to X_s\gamma$ decays [57]. The varied range of the Fermi motion parameter is 200 MeV/$c < p_F <$ 480 MeV/c. M_{X_s} and $M_{\ell^+\ell^-}$ distributions for varied Fermi motion parameter is shown in Fig 7.1.

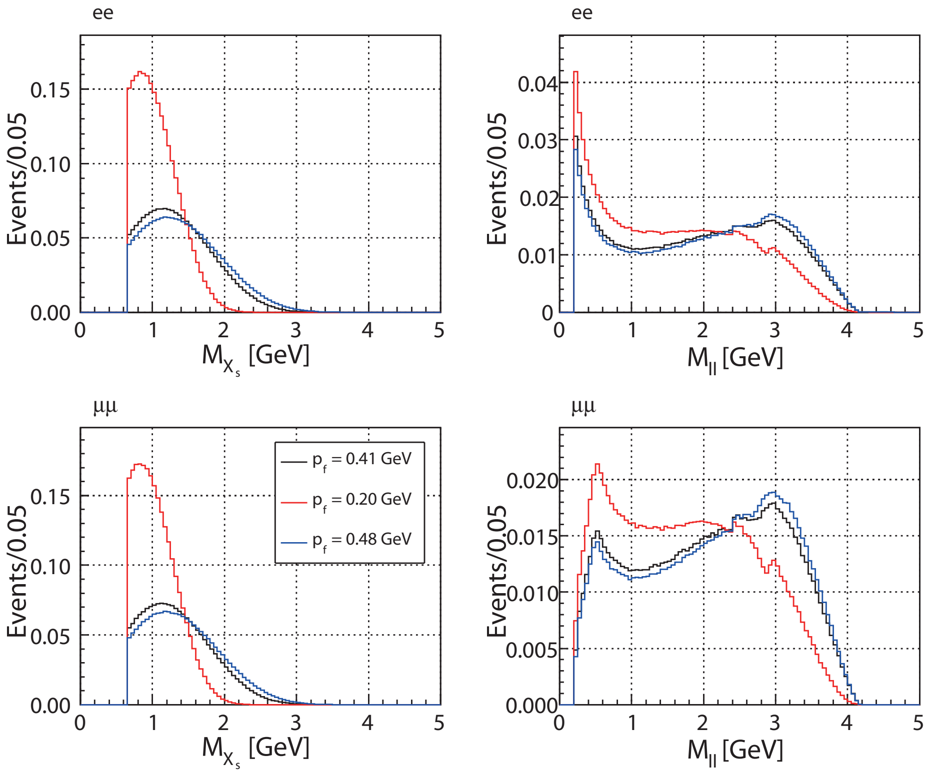


Figure 7.1: M_{X_s} (left) and $M_{\ell^+\ell^-}$ (right) distributions for the various Fermi motion parameters. Upper figure are for e^+e^- mode and lower figure are for $\mu^+\mu^-$ mode. The black, red and blue line represents the distribution for $p_F = 410, 200$, and 480 MeV/c, respectively. M_{X_s} distributions are normalized to have the same entries in 1.1 GeV/$c^2 < M_{X_s} <$ 2.0 GeV/c^2.

b-quark Mass

In our signal Monte Carlo model, we assume the b-quark mass M_b to be 4.8 GeV/c^2. The b-quark mass is varied in the range of ±0.15 GeV/c^2. M_{X_s} and $M_{\ell^+\ell^-}$ distributions for varied Fermi motion parameter is shown in Fig 7.2.

X_s-K^* Transition

We required $M_{X_s} >$ 1.1 GeV/c^2 to the non-resonant $X_s\ell^+\ell^-$ MC samples. The transition point is varied by ±0.1 GeV/c^2.

Hadronization

The non-resonant Monte Carlo event generator relies on JETSET to fragment and hadronize the system consisting of a final state s quark and a spectator quark from the B meson.

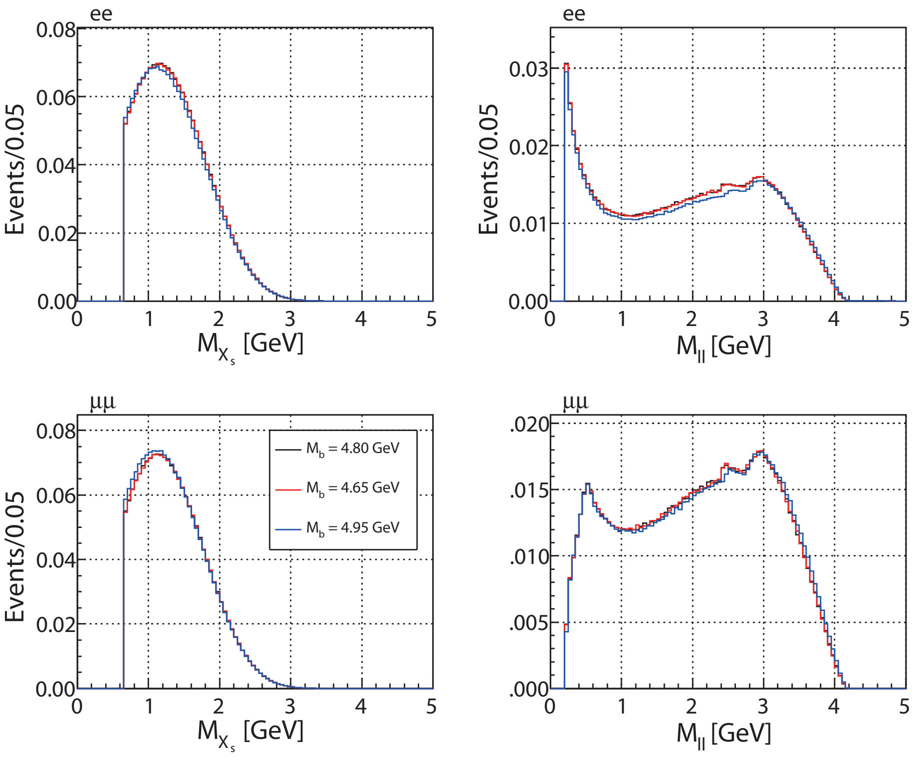


Figure 7.2: M_{X_s} (left) and $M_{\ell^+\ell^-}$ (right) distributions for the various b-quark masses. Upper figure are for e^+e^- mode and lower figure are for $\mu^+\mu^-$ mode. The black, red and blue line represents the distribution for $M_b = 4.80, 4.65$, and 4.95 GeV/c^2, respectively. M_{X_s} distributions are normalized to have the same entries in 1.1 GeV/c^2 $< M_{X_s} <$ 2.0 GeV/c^2.

The signal efficiencies depend strongly on the particle content of the final state. We estimate the scale factor for each X_s decay mode by comparing the yields in X_sJ/ψ data and MC samples. Since we find no yield in $K4\pi$ modes, the scale factor for the $K4\pi$ modes is set to 1. To keep the fraction of $K^{(*)}\ell^+\ell^-$ and $X_s\ell^*\ell^-$, the scale factor for each X_s decay modes are multiplied by the normalization factor (=0.986). Table 7.2 shows the scale factors multiplied by the normalization factor. After applying the scaling factor, we calculate reconstruction efficiency and construct correction function. The difference of the A_{FB} is taken as the systematic uncertainties. We also estimate the scale factor for the modes with/without K_S^0 and π^0 as shown in Tab. 7.3.

Table 7.2: Scale factors (Data/MC) for each X_s mode.

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

Table 7.3: Scale factors (Data/MC) for modes with/without K_S^0 and π^0.

Mode	Factor
modes including K_S^0	0.91 ± 0.03
modes not including K_S^0	1.04 ± 0.02
modes including π^0	1.01 ± 0.04
modes not including π^0	0.98 ± 0.02

Fraction of $K\ell^+\ell^-$, $K^*\ell^+\ell^-$ and $X_s\ell^+\ell^-$

The signal Monte Carlo samples of $K\ell^+\ell^-$, $K^*\ell^*\ell^-$ and non-resonant $X_s\ell^+\ell^-$ are mixed assuming the experimental branching fraction. The fraction of these three MC samples are varied according to experimental uncertainties.

7.2.3 Correction Function

Figure 6.2 shows the correction functions with SM A_7 value. We also construct the correction functions with sign-flipped SM A_7 value. The systematics uncertainties arise from the difference of correction functions between SM and sign-flipped SM A_7 value.

The correction function fitted by linear function, is incorporated into the fitter. However there is spread in the relation between the A_{FB}^{true} and $A_{FB}^{meas.}$ as shown in Fig. 6.2. To estimate the width of the correction function, Fig. 6.2 are projected to the axis perpendicular to the fitted line. Figure 7.3 are the projected distributions fitted by the Gaussian. The mean is deviated from zero. This deviation would arise from statistical fluctuation of MC samples used for the determination of reconstruction efficiencies. When we construct correction functions, we used the reconstruction efficiency as the detector simulation. Therefore the statistical fluctuation in the reconstruction efficiency make offset of correction functions. The error bars plotted in Fig. 6.2 include the error of reconstruction efficiency. All the markers with error bars are clearly consistent with the fitted linear line assuming null offset. Conservatively, we add the deviation of the mean to the width of the correction functions. To estimate this uncertainty, we repeat the final fit to extract the A_{FB} 100 times with varied correction function and the take the RMS of the A_{FB} distribution as the systematic error.

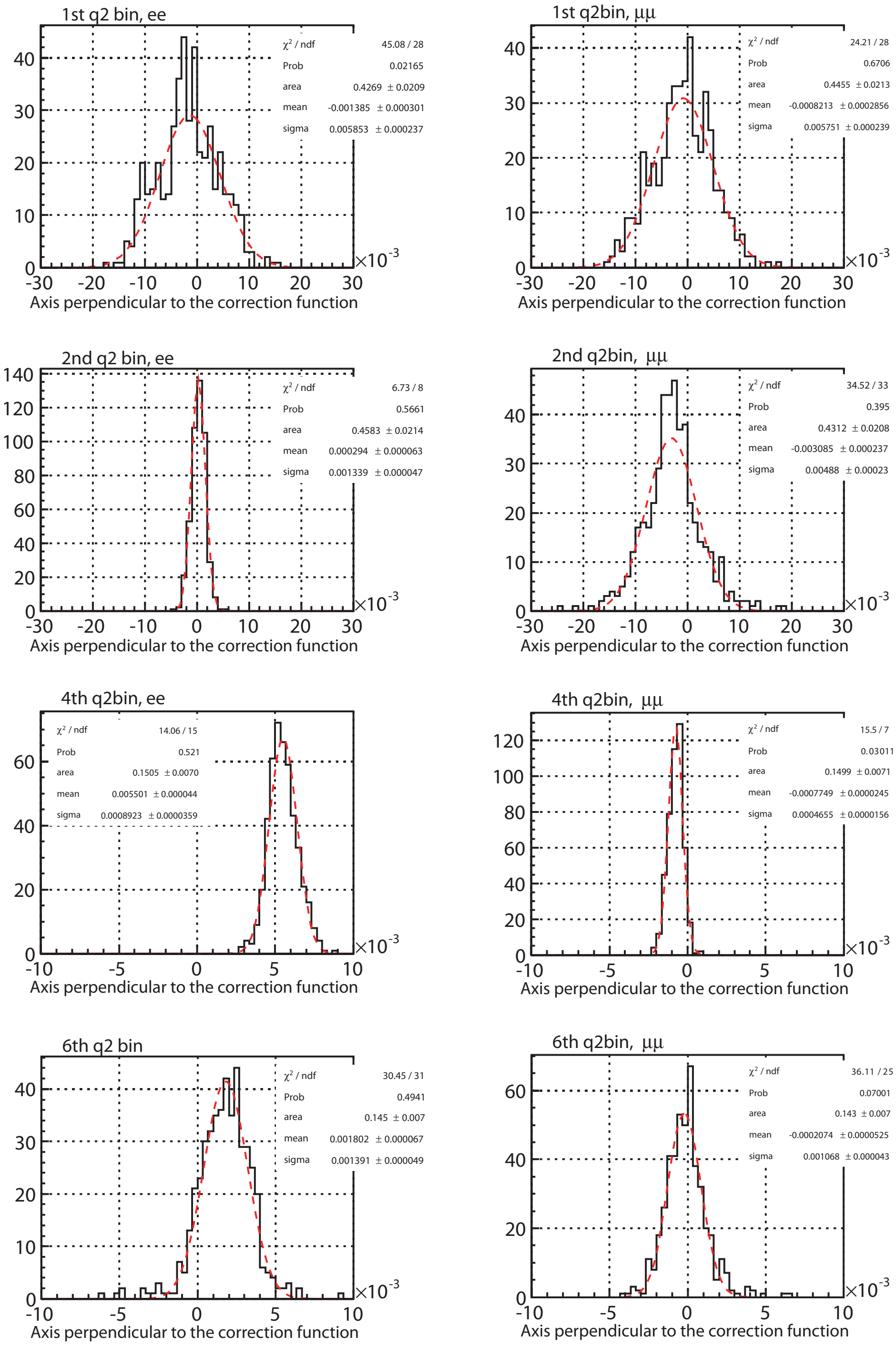


Figure 7.3: Width of correction functions for $X_se^+e^-$ (left) and $X_s\mu^+\mu^-$ (right). Each row corresponds to 1st, 2nd, 4th, and 6th q^2 bin from top to bottom.

Chapter 8

Results and Discussion

8.1 Results

Figure 8.1 shows the fitting results. We obtain:

	Our results	SM prediction [15] [8]
A_{FB}(1st q^2 bin)	0.34 ± 0.24(stat) ± 0.02(syst)	-0.11
A_{FB}(2nd q^2 bin)	0.04 ± 0.31(stat) ± 0.05(syst)	0.12
A_{FB}(4th q^2 bin)	0.28 ± 0.21(stat) ± 0.01(syst)	0.32
A_{FB}(6th q^2 bin)	0.28 ± 0.15(stat) ± 0.01(syst)	0.40

According to the method described in the Chap. 7, the systematic errors are estimated (See Table 8.1). The total uncertainties are estimated by the quadratic sum of them.

Table 8.1: Summary of systematic errors.

Sources	$\sigma_{A_{FB}}$			
	1st q^2 bin	2nd q^2 bin	4th q^2 bin	6th q^2 bin
Signal shape and Self cross-feed	0.002	0.002	0.002	0.002
Peaking background	0.003	0.050	0.004	0.001
Reconstruction and PID efficiency	0.000	0.000	0.000	0.000
Fermi motion	0.002	0.001	0.004	0.004
b-quark mass	0.002	0.001	0.003	0.000
X_s-K^* transition	0.001	0.001	0.002	0.003
Hadronization	0.001	0.000	0.001	0.000
Fraction of $[K/K^*/X_s]\ell^+\ell^-$	0.001	0.001	0.002	0.003
Sign flipped SM A_7	0.015	0.001	0.000	0.001
Width of the correction function	0.012	0.004	0.006	0.003
Total	0.020	0.050	0.009	0.007

The signal yields are obtained to be 140, 161, and 301 events for $X_s e^+ e^+$, $X_s \mu^+ \mu^-$, and $X_s \ell^+ \ell^-$, respectively. The signal yields for each q^2 bin are summarized in Tab. 8.2. Figure 8.2 shows the $-2\ln(\mathcal{L}/\mathcal{L}_{max})$ as a function of the A_{FB}, where $\mathcal{L}_{max}$ and $\mathcal{L}$ are

Table 8.2: Signal yields for each q^2 bin. Errors are only statistical uncertainties.

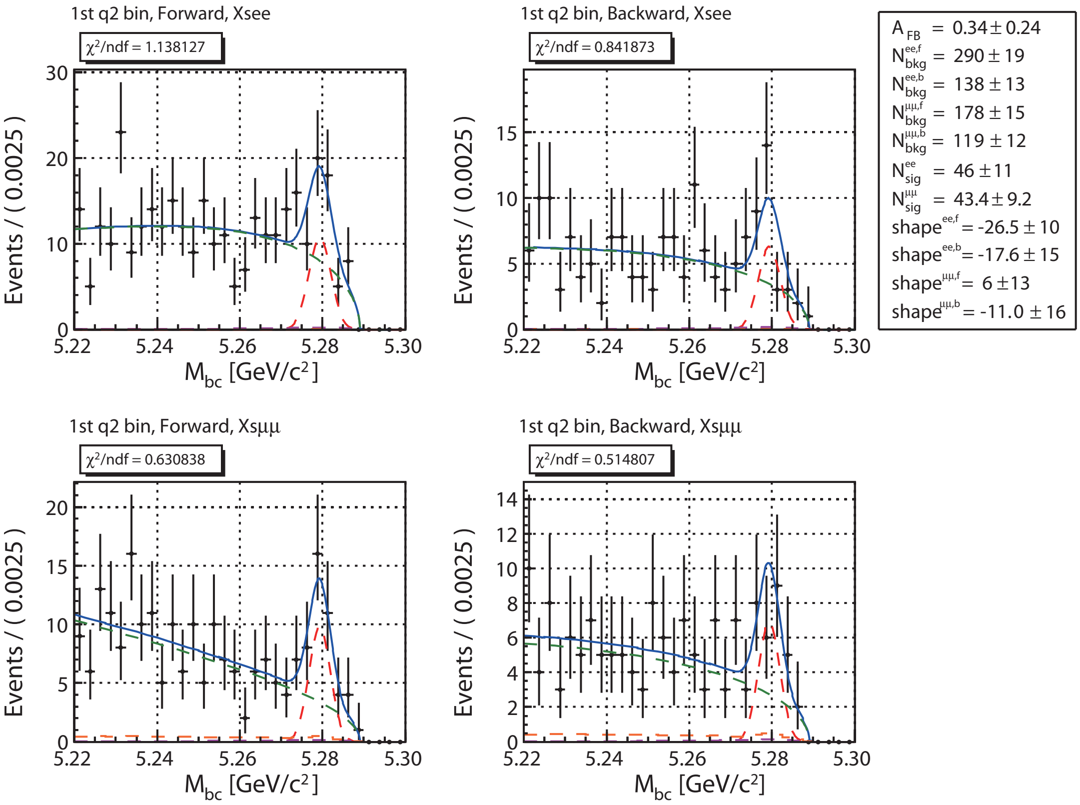
bin	$N_{sig}^{e^+e^-}$	$N_{sig}^{\mu^+\mu^-}$	$N_{sig}^{\ell^+\ell^-}$
1st q^2	45.7 ± 10.9	43.4 ± 9.2	89.1 ± 14.3
2nd q^2	30.1 ± 9.2	23.9 ± 10.5	53.9 ± 14.0
4th q^2	25.0 ± 7.0	30.7 ± 9.9	55.7 ± 12.1
6th q^2	39.2 ± 9.6	62.9 ± 10.4	102.0 ± 14.1
Total	139.9 ± 18.6	160.8 ± 20.0	300.7 ± 27.3

the maximum likelihood of the nominal fitting and the fitting when the A_{FB} is fixed. In Fig. 8.2, the systematic uncertainties are not included.

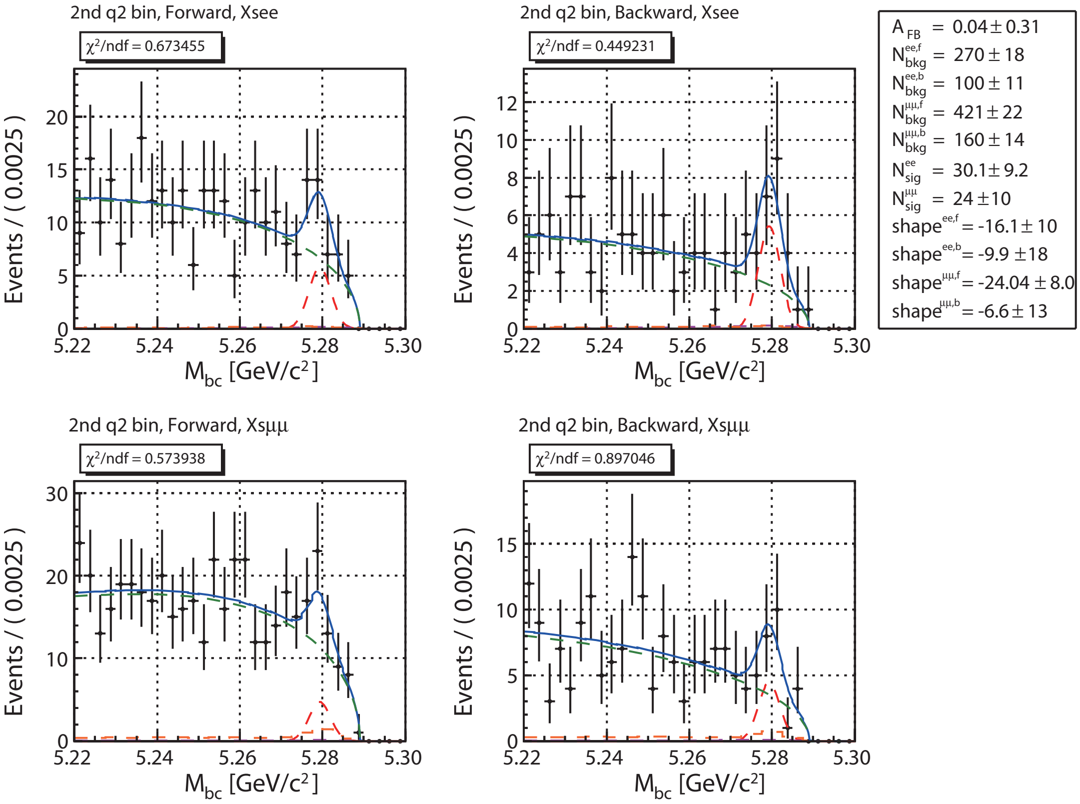
8.2 Discussion

Figure 8.3 shows our result of forward-backward asymmetry and its theoretical prediction in the SM. Equations for the SM prediction of the matrix element of the $B \to X_s\ell^+\ell^-$ transition are referred from Ref. [15] and the expression for the Wilson coefficients are taken from Ref. [8]. The forward-backward asymmetry in the high q^2 region is consistent with the SM prediction. Our result of 4th and 6th q^2 bin favors $C_{10}{\cdot}\mathrm{Re}(C_9)$ term in Eq. 2.21 to be negative with 2.3σ (97.9% C.L.). The forward-backward asymmetry in the lowest q^2 bin is 1.8σ (6.6% C.L.) away from the SM prediction and favors the relative sign of C_7 to C_{10} to be negative with 1.4σ (84.1% C.L.).

We compare our result with resuts using exclusive $B \to K^*\ell^+\ell^-$ decay (See Fig. 2.6). In the high q^2 region, all results using the exclusive decay favor $C_{10}{\cdot}\mathrm{Re}(C_9)$ term to be negative, which is consistent with our results. In the low q^2 region ($q^2 < 4.3$ GeV$^2/c^4$), results using the exclusive decay by lepton colliders, Belle and BaBar, favor the relative sign of C_7 to C_{10} to be negative, which is the flipped sign to the SM case. On the other hand, the results by hadron colliders, LHCb and CDF, favor it to be positive. Our result favors it to be negative and 1.7σ (8.6% C.L.) away from the result of LHCb, which is the most precise. Our result is useful to constrain the new physics beyond the SM.

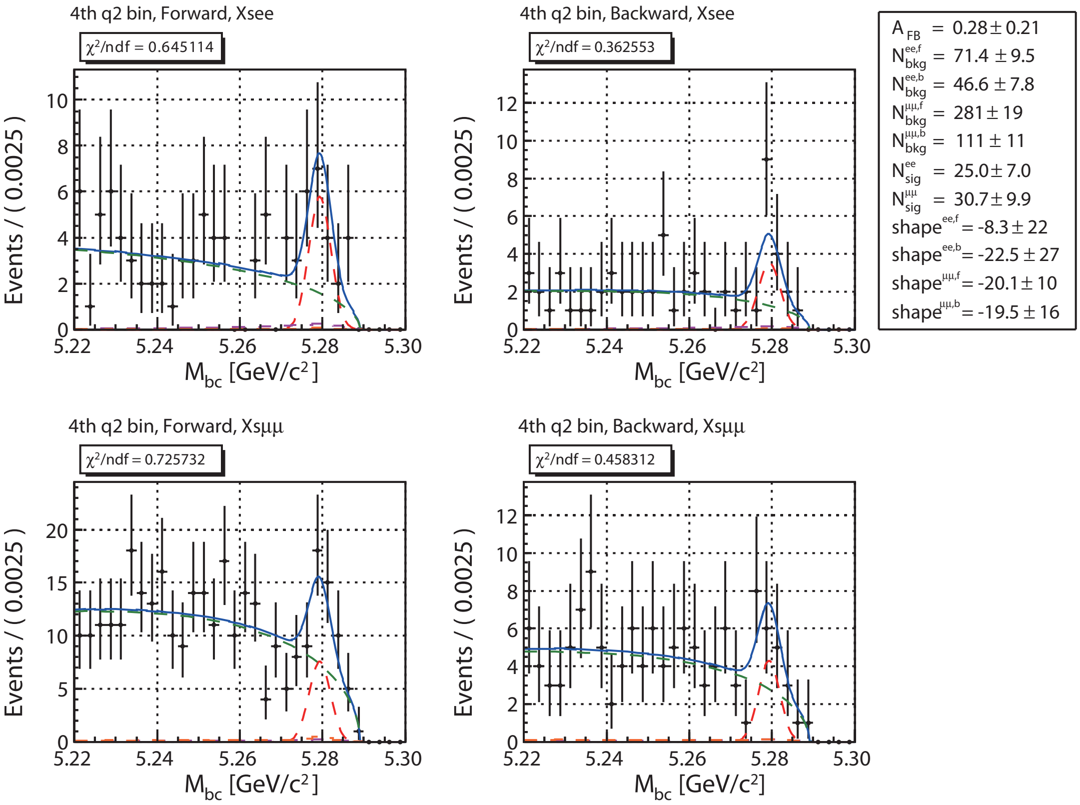


(a) 1st q^2 bin.

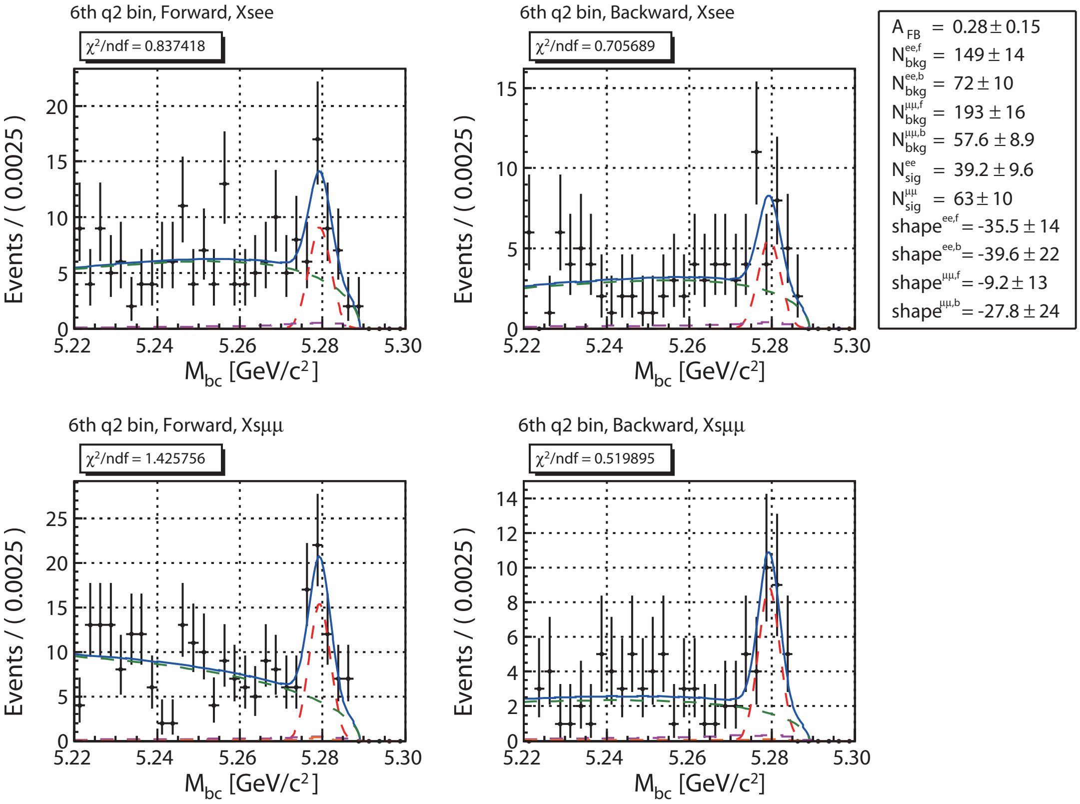


(b) 2nd q^2 bin.

Figure 8.1: Fitted $M_{\rm bc}$ distributions for forward in $X_s e^+e^-$ (top left), backward in $X_s e^+e^-$ (top right), forward in $X_s\mu^+\mu^-$ (bottom left), and backward in $X_s\mu^+\mu^-$ (bottom right). The curves shows total (solid blue), signal (dashed red), self cross-feed (dashed purple), non-peaking background (dashed green), and peaking background (dashed orange) components, respectively.

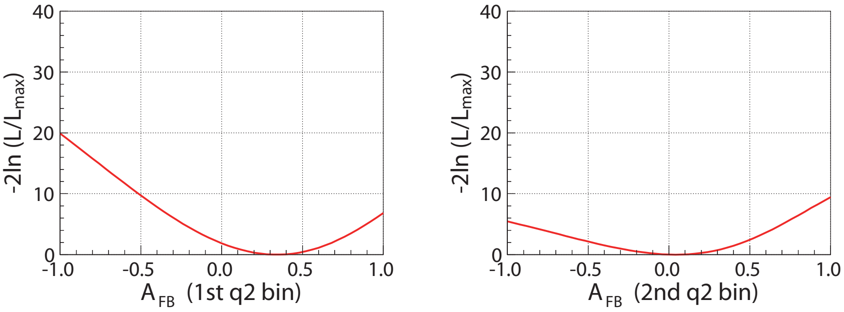


(a) 4th q^2 bin.



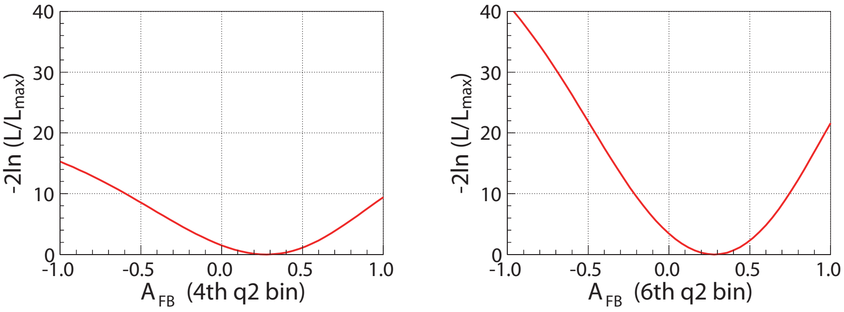
(b) 6th q^2 bin.

Figure 8.1: Continued.



(a) 1st q^2 bin.

(b) 2nd q^2 bin.



(c) 4th q^2 bin.

(d) 6th q^2 bin.

Figure 8.2: $-2\ln(\mathcal{L}/\mathcal{L}_{max})$. The systematic uncertainties are not included.

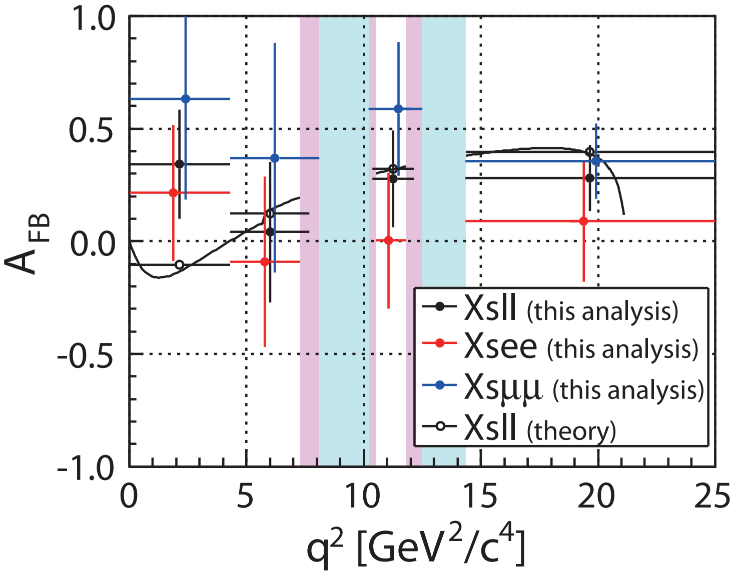


Figure 8.3: Result of the A_{FB}. Theoretical SM prediction is shown by the black curve and open circle. The closed black circles show the result by the simultaneous fitting of $X_se^+e^-$ and $X_s\mu^+\mu^-$. The closed red (blue) circles show the results by the fitting of only $X_se^+e^-$ ($X_s\mu^+\mu^-$). The light red and light blue bands correspond to the J/ψ and $\psi(2S)$ veto regions for $X_se^+e^-$ and $X_s\mu^+\mu^-$.

Chapter 9

Conclusion

The $B \to X_s \ell^+ \ell^-$ decay is a flavor changing neutral current process, highly suppressed in the standard model and therefore a probe for searching new physics beyond the standard model. We have measured the forward-backward asymmetry in $B \to X_s \ell^+ \ell^-$ decays with semi-inclusive reconstruction method for the first time.

We use a data sample of 711 fb^{-1} recorded on the $\Upsilon(4S)$ resonance that containing 772×10^6 $B\bar{B}$ pairs collected with the Belle detector at the KEKB asymmetric e^+e^- collider. We extract the forward-backward asymmetry A_{FB} with an unbinned maximum likelihood fit. We obtain:

$$\begin{aligned}
A_{FB}(\text{1st } q^2 \text{ bin}) &= 0.34 \pm 0.24(\text{stat}) \pm 0.02(\text{syst}), \quad (9.1)\\
A_{FB}(\text{2nd } q^2 \text{ bin}) &= 0.04 \pm 0.31(\text{stat}) \pm 0.05(\text{syst}), \quad (9.2)\\
A_{FB}(\text{4th } q^2 \text{ bin}) &= 0.28 \pm 0.21(\text{stat}) \pm 0.01(\text{syst}), \quad (9.3)\\
A_{FB}(\text{6th } q^2 \text{ bin}) &= 0.28 \pm 0.15(\text{stat}) \pm 0.01(\text{syst}). \quad (9.4)
\end{aligned}$$

Our result of 4th and 6th q^2 bin favors $C_{10}\cdot\mathrm{Re}(C_9)$ term in Eq. 2.21 to be negative with 2.3σ (97.9% C.L.). The forward-backward asymmetry in the lowest q^2 bin is 1.8σ (6.6% C.L.) away from the theoretical SM prediction. Our result is useful to constrain the new physics beyond the Standard Model. The uncertainty of our measurement is dominated by statistical errors, and thus it will be reduced with more data.

Appendix A

Neural Network/NeuroBayes

In data analysis, one often has to deal with classification problems. Generally speaking one wants to classify each event in a set of data to be either signal or background. For classification problems, the use of multivariate methods has many advantages compared to classical methods.

In this analysis, we employ the NeuroBayes package as a multivariate analysis tool, which was originally developed at the University of Karlsruhe and is now maintained and further developed by physicists in the company <phi-t>® (Physics Information Technologies), where it is also applied to non-physical problems.

A.1 Neural Network

The basic idea of neural network is based on the real biological neuronal networks in the brain. A neuronal circuit like the central nervous system consists of a multitude of interconnected neurons. Each neuron has several inputs (dendrites) and output (axons) structure. The neurons in the brain communicate among each other by exchanging electrical impulses. They are fired by the axon of one cell and received by the dendrites of other neurons. If the power of the electrical signal exceeds a certain threshold, the neurons are triggered and fire output signals themselves along their axons. The junctions between axons and dendrites of different cells are called synapses. These are able to increase or decrease the electrical impulse, when received at a dendrite, using biochemical process.

The layout of an artificial neuron is shown in Fig. A.1. The input structure consists of an input vector $\vec{x}_i$. The weighted sum of the input is calculated, where the weights w_i are in analogy to the synapses. Additionally there is a hidden input variable x_0 called bias, which can provide a constant shift to the total neuronal input *net*. This sum is then passed to an activation function ϕ. The most common choice for classification is to use the sigmoid function:

$$\phi(net) = \frac{2}{1 + e^{-net}} - 1. \tag{A.1}$$

This symmetric function maps the interval $(-\infty, +\infty)$ to the output interval $[-1, +1]$. In total, the numerical output for a given input vector $\mathbf{x}$ is $o = \phi(net)$.

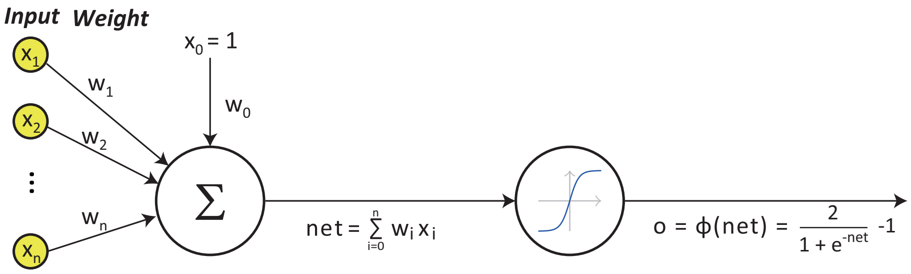


Figure A.1: Layout of an artificial neuron.

A.1.1 Feed-Forward Networks

In the human brain, consisting of up to 30 billion neurons, each neuron has up to 10,000 synapses to other neurons. This structure is far too complicated for the purpose of data analysis. For classification problems, the three layer feed-forward network with a single output node is a sufficient choice. The information of the n input nodes x_i is transferred via m nodes y_i in a hidden layer to a single output node o. The term feed-forward arises from the fact that information, in contrast to the brain, is only transferred in one direction.

The choice of the number of neurons in the hidden layer is slightly arbitrary. In practice using the same number of hidden neurons as there are input neurons will often result in good performance. Figure A.2 shows the topology of a three layer feed-forward network.

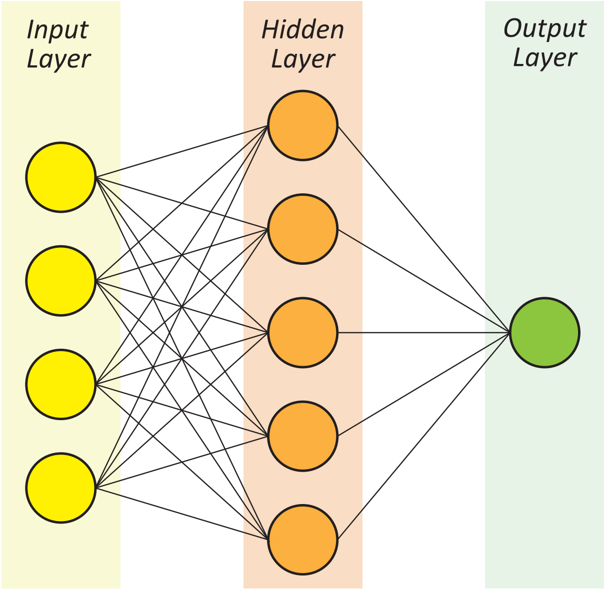


Figure A.2: Topology of a three layer feed-forward network.

A.1.2 Network Training

To use a neural network for a classification problem, it has to be trained first. For such training, one needs a set of data for which the truth is known. In general, this can be historical data, but in particle physics, simulations are usually used. Using a data set for signal and background where the truth of each entry or event is known, the weights of each individual connection in the neural network are determined in a way that the network output for each event is as close to the known truth as possible. For a network with N input neurons, M hidden neurons and one output neuron, the total number of weights is given by:

$$N_W = N \cdot M + M + N. \tag{A.2}$$

To solve this problem, the weights are adjusted iteratively. The weights are usually adjusted using the whole training data set. For each event i the network output is computed with the current weights and compared to the truth t_i. This is usually expressed as a cost function. One possibility of a cost function is the sum over the squared difference between network output and truth or target value

$$\chi^2 = \sum_i^N (o_i(\vec{w}) - t_i)^2 \tag{A.3}$$

where N is the total number of training events and $\vec{w}$ is the vector of all weights. Another possible cost function is the entropy function

$$E = \sum_i^N \ln\left(\frac{1}{2}(1 + o_i(\vec{w}) \cdot t_i)\right) \tag{A.4}$$

In general, training a neural network is a non-trivial minimization problem in high dimensional space. To be more precise, we minimize the cost function by adjusting each weight. The gradient descent method can be used for this purpose and if the minimum has not been reached

$$\frac{\partial E}{\partial w_{ij}} \neq 0, \tag{A.5}$$

the change of weights Δw_{ij} is proportional to the respective gradient

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial w_{ij}}, \tag{A.6}$$

where η is the proportionality constant. Starting with random weights, they are adjusted in each iteration of the training until the minimum of the cost function is reached. This kind of neural network training algorithm is called Backpropagation-Algorithm.

Appendix B

Event Shape Parameters

The KSFW [51] is a Fisher discriminant [52] extended from the Fox-Wolfram variables [53] using information such as missing mass calculated from the daughter particles of the signal candidates and all the other particles in this event. The KSFW is constructed from 17 variables. We use these variables as the NeuroBayes input parameters.

B.1 Fox-Wolfram Momenta

In many analyses of the B meson decays, large backgrounds arise from the continuum processes $e^+e^- \to q\bar{q}(q = u, d, s, c)$.

The event topology is characterized on the basis of the Fox-Wolfram momenta. The l-th moment is defined in CM the frame as

$$H_l \equiv \sum_{i,j} |\vec{p}_i|\,|\vec{p}_j|\,P_l(\mathrm{cos}\theta_{ij}), \tag{B.1}$$

where P_l is the l-th Legendre polynomial, $\vec{p}_i$ and $\vec{p}_j$ are the momenta of the i-th and j-th particles, respectively, θ_{ij} is the angle between the two momentum vectors. The sum is over the particles in the final state. Note that the overall constant is ignored here for simplicity.

B.2 SFW

SFW is devised by modifying the Fox-Wolfram moment H_l. In the SFW, the H_l is divided into three components and categorize the particles to the two type; B signal candidate particles and the remaining particles.

$$\begin{aligned} H_l &= H_l^{SS} + H_l^{SO} + H_l^{OO} & \text{(B.2)} \\ H_l^{SS} &= \sum_{ij} |p_i||p_j|P_l(\cos\theta_{ij}) & \text{(B.3)} \\ H_l^{SO} &= \sum_{jk} |p_j||p_k|P_l(\cos\theta_{jk}) & \text{(B.4)} \\ H_l^{OO} &= \sum_{kl} |p_k||p_l|P_l(\cos\theta_{kl}) & \text{(B.5)} \end{aligned}$$

where i and j iterate over B signal candidates particles (denoted by S for the signal) and the indices k and l iterate over the remaining particles (denoted by O for other) in the event. We define a Fisher discriminants named SFW (Super Fox-Wolfram) by the divided Fox-Wolfram moments as

$$SFW \equiv \sum_l \alpha_l \left(\frac{H_l^{SO}}{H_0^{SO}}\right) + \sum_l \beta_l \left(\frac{H_l^{OO}}{H_0^{OO}}\right), \tag{B.6}$$

where α_l and β_l are Fisher coefficients.

B.3 KSFW

In order to improve the discrimination, the SFW is modified to derive so-called KSFW (Kakuno Super Fox-Wolfram) by considering the charges of the particles, the missing mass of the event, and normalization factors. The KSFW is defined as

$$KSFW \equiv \sum_{l=0}^{4} R_l^{SO} + \sum_{l=0}^{4} R_l^{OO} + \gamma \sum_{n=1}^{N_t} |p_{t,n}|, \tag{B.7}$$

where N_t is the number of the particles and γ is a Fisher coefficients to be optimized. We define these terms in the Eq. B.7 in turn.

- R_l^{SO}

  We add the missing pseudo-particle as one particle that has the event's missing energy and momentum (p_{miss} and furthermore divide the remaining particles in the event to three categories; "charged", "neutral" and "missing".

  $$R_l^{SO} = \frac{\alpha_l^{charged} H_l^{SO_{charged}} + \alpha_l^{neutral} H_l^{SO_{neutral}} + \alpha_l^{missing} H_l^{SO_{missing}}}{E_{beam} - \Delta E}, \tag{B.8}$$

  where α_l^i $(i = charged, neutral, missing)$ are Fisher coefficients.

  For signal and other remaining charged particles,

  $$H_l^{SO_{charged}} = \begin{cases} \sum_i \sum_j |p_j| P_l(\cos\theta_{ij}) & \text{if } l \text{ is even} \\ \sum_i \sum_j |p_j| Q_i Q_j P_l(\cos\theta_{ij}) & \text{if } l \text{ is odd} \end{cases} \tag{B.9}$$

  where the index i iterates over the particles in the B signal candidates and the index j iterates over all other remaining charged particles. The $Q_{i,j}$ are the charge of the particle i and j.

  For signal and other remaining neutral particles,

  $$H_l^{SO_{neutral}} = \begin{cases} \sum_i \sum_j |p_j| P_l(\cos\theta_{ij}) & \text{if } l \text{ is even} \\ 0 & \text{if } l \text{ is odd} \end{cases} \tag{B.10}$$

  where the index i iterates over the particles in the B signal candidates and the index j iterates over all other remaining neutral particles.

For signal and missing particles,

$$H_l^{SO_{missing}} = \begin{cases} \sum_i |p_i| P_l(\cos\theta_{iM}) & \text{if } l \text{ is even} \\ 0 & \text{if } l \text{ is odd} \end{cases} \tag{B.11}$$

where the index i iterates over the particles in the B signal candidates and the θ_{iM} is the opening angle between p_i and p_{miss}.

- R_l^{OO}

$$R_l^{OO} = \frac{\beta_l H_l^O O}{(E_{beam} - \Delta E)^2} \tag{B.12}$$

$$H_l^{OO} = \begin{cases} \sum_j \sum_k |p_j||p_k| P_l(\cos\theta_{jk}) & \text{if } l \text{ is even} \\ \sum_j \sum_k |p_j||p_k| Q_j Q_k P_l(\cos\theta_{jk}) & \text{if } l \text{ is odd} \end{cases} \tag{B.13}$$

where the indices j and k iterates over all other remaining particles.

- $\sum_{n=1}^{N_t} |p_{t,n}|$

$\sum_{n=1}^{N_t} |p_{t,n}|$ is the scalar sum of the transverse momenta (p_t) of all the particles in the signal candidates and all the other remaining particles.

The KSFW has totally 17 (=11+5+1) Fisher coefficients.

Appendix C

MC calibration

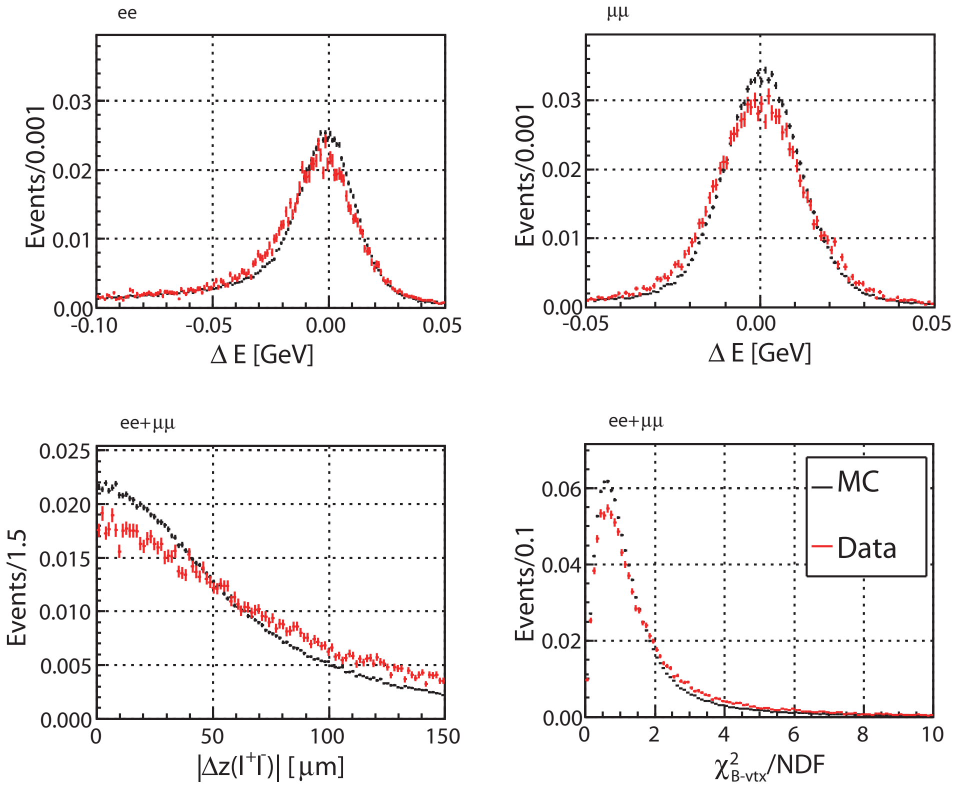
We compare the distribution of the NeuroBayes input parameters between MC and data, using $X_s J/\psi$ samples. As a result, we found large discrepancies in the ΔE, $\Delta z_{\ell^+\ell^-}$, and confidence level of the B vertex construction. Therefore, the MC samples were tuned to adequately reproduce the observed data distribution. The following tuning was performed.

- Tuning the shape of the signal peaks in the ΔE distribution
  The signal peak for the ΔE distribution in $K^{(*)}J/\psi$ samples are fitted by a Crystal Ball function [55]. The calibration factor for mean and width are determined by the fitting results.

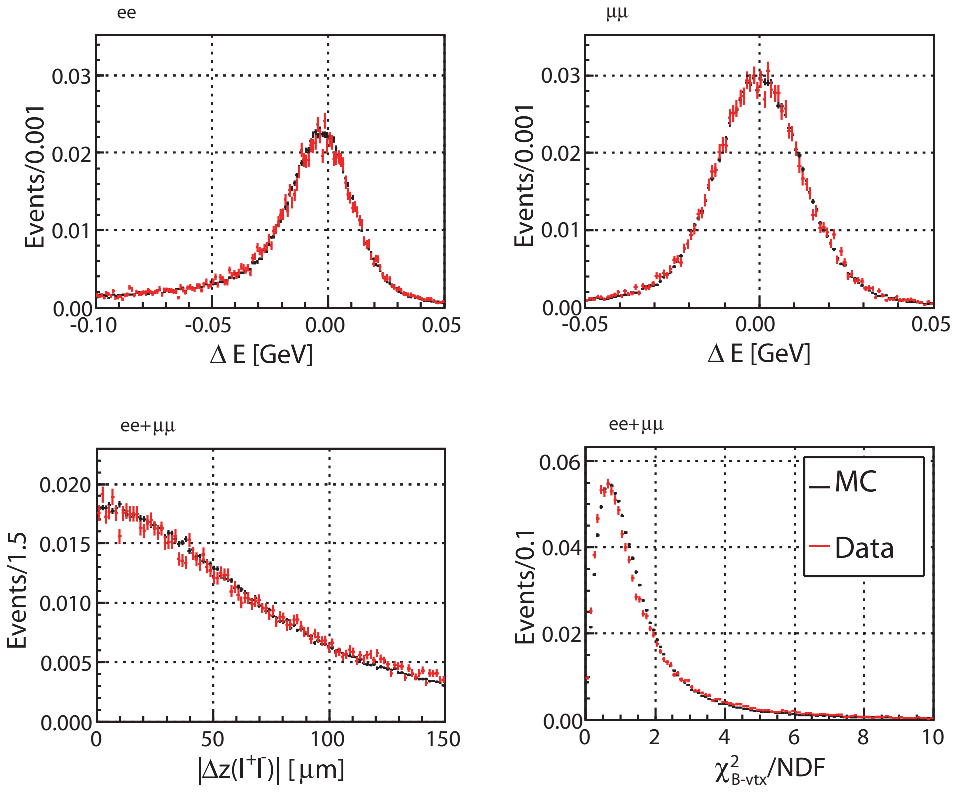
- Tuning the width in the $\Delta z_{\ell^+\ell^-}$ distribution
  The $\Delta z_{\ell^+\ell^-}$ distribution were fitted by the double Gaussian. The calibration factors for signal and background were determined from the fitting results of $K^{(*)}J/\psi$ samples and $X_s e^\pm \mu^\mp$ samples, respectively.

- Tuning of the confidence level of the B vertex construction
  The χ^2 of B vertex construction ($\chi^2_{B\text{-}vtx}$) is calibrated for the calibration of the confidence level. The confidence level is recalculated from the calibrated $\chi^2_{B\text{-}vtx}$. The $\chi^2_{B\text{-}vtx}$ distribution in MC samples are widened to minimize the χ^2 which is calculated from the difference between data and MC. The calibration factor are determined using the $K^{(*)}J/\psi$ samples and $X_s e^\pm \mu^\mp$ samples, respectively.

The calibration factors are summarized in Tab. C.1. The distributions for before and after MC calibration are shown in Fig. C.1.



(a) Before MC calibration



(b) After MC calibration

Figure C.1: Distributions of ΔE in $X_s e^+e^-$ (top left), ΔE in $X_s\mu^+\mu^-$ (top right), $\Delta z_{\ell^+\ell^-}$ (bottom left), and $\chi^2_{B\text{-}vtx}$ (bottom right). The $K^{(*)}J/\psi$ MC (black) and data (red) samples are shown.

Table C.1: Summary of the calibration factor.

Variables	Mode	Parameter	Calibration factor
ΔE	Signal ($X_s e^+ e^-$)	mean	+1.9 MeV
		width	×1.13
	Signal ($X_s \mu^+ \mu^-$)	mean	+0.9 MeV
		width	×1.15
$\Delta z_{\ell^+\ell^-}$	Signal	width	×1.28
	Background	width	×1.17
χ^2_{B-vtx}	Signal	width	×1.14
	Background	width	×1.04

Appendix D

Estimation of Lepton Fake Rate

When we estimate the peaking background of double and swapped mis-PID events, we use the lepton fake rate, which is the probability to mis-identify a hadron as an electron or muon. In this appendix, we show the method to estimate the lepton fake rate using the decay $D^{*+} \to D^0\pi^+$ followed by $D^0 \to K^-\pi^+$.

The lepton fake rate is determined by the ratio of the yields with and without the lepton PID requirement in the D mass fitting. The requirement of e-ID and μ-ID is same with the event selection for $X_s\ell^+\ell^-$: eid.prob$(3,-1,5) > 0.80$ and muid.Likelihood()$>$ 0.97. To estimate the lepton fake rate as a function of momentum(p) and direction($\cos\theta$), the momentum and $\cos\theta$ is divided into bins of 0.60 GeV/c and 0.10, respectively.

D.1 Selection Criteria

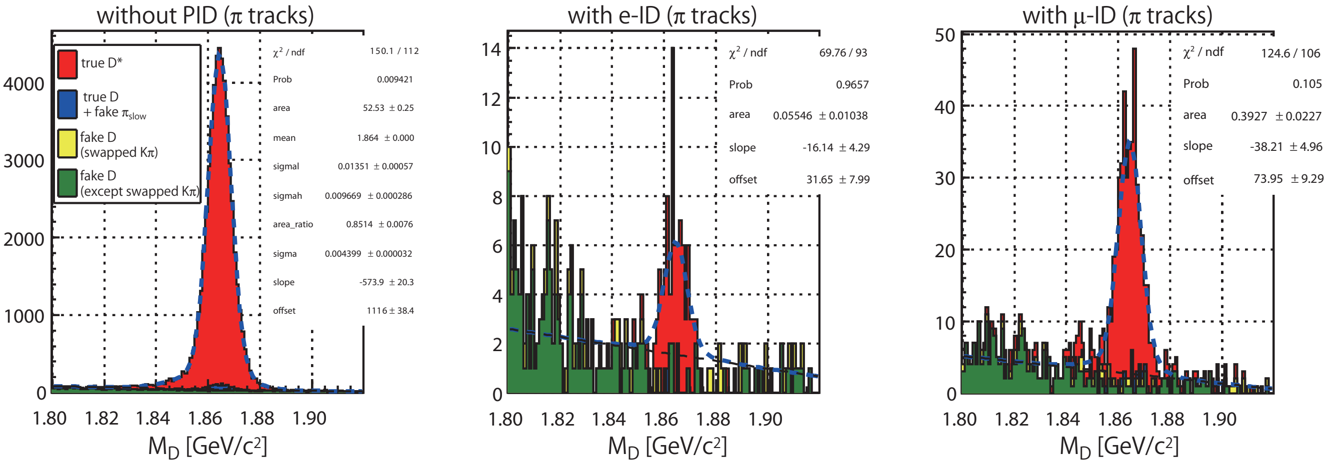
D^{*+} is reconstructed from $D^{*+} \to D^0\pi^+_{\mathrm{slow}} \to K^-\pi^+\pi^+_{\mathrm{slow}}$. We require following selection criteria.

- $|\Delta M - (M_{D^*} - M_D - M_\pi)| < 1.5$ MeV/c^2, where ΔM is the mass difference between reconstructed D^* and D.
- $|M_D^{\mathrm{rev}} - M_D| > 30$ MeV/c^2, where M_D^{rev} is the $K\pi$ invariant mass with reversed mass assignment.
- $p_{D^*}/E_{beam} > 0.50$, where p_{D^*} and E_{beam} are the CM-momentum of D^* and CM-energy of beam, respectively.
- $\cos\theta_D < 0.8$, where $\cos\theta_D$ is the D decay angle, defined as the angle of K momentum int the D rest frame with respect to the D momentum in the CM-frame.

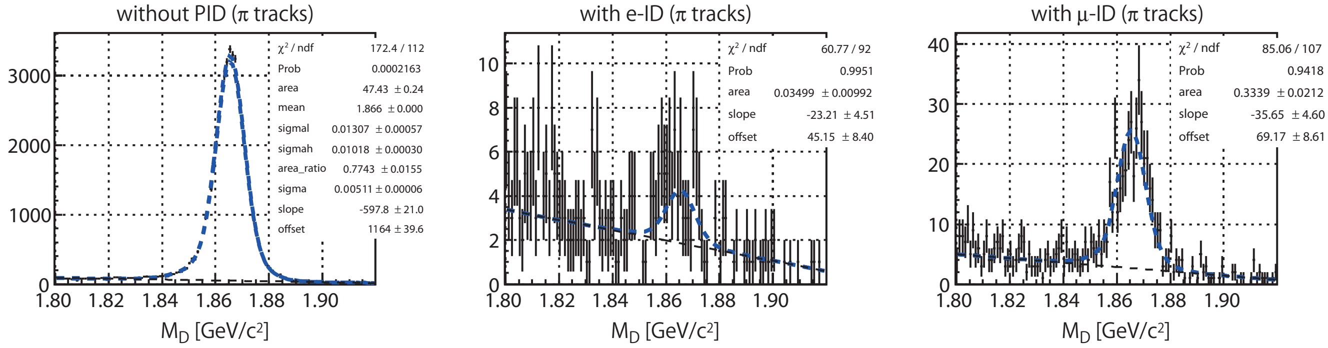
The vertex fitting to find D decay vertex is performed using K and π. Then, the slow pion vertex is refitted to the fixed D vertex. We can obtain the clean D^* samples with purity more than 99% using the selection criteria.

D.2 D Mass Fitting

In order to derive the lepton fake rate, we fit the D mass distribution in each momentum-$\cos\theta$ bin. We model the signal shape by the sum of the Gaussian and bifurcated Gaussian, which have common mean. In the fitting without PID, the all signal shape parameters are floated. In the fitting with PID, the all signal shape parameters are fixed by the result of fitting without PID. The background shape is modeled by the linear function. The typical D mass fitting are shown in Fig. D.1(a) and D.1(b) for MC and data samples. The lepton fake rate is derived from the ratio of the yields with and without the lepton PID requirement as shown in Fig. D.2.



(a) MC samples for truly reconstructed D^* (red), truly reconstructed D with fake π_{slow} (blue), falsely reconstructed D by the swapped K and π (yellow), and falsely reconstructed D not by the swapped K and π (green).



(b) Data samples.

Figure D.1: Examples of D mass fitting without PID (left), with e-ID (center) and with μ-ID (right).

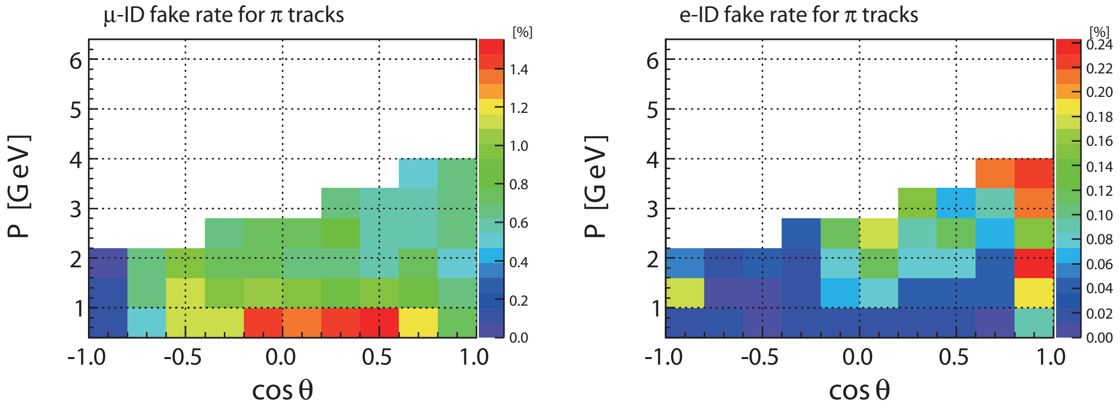


Figure D.2: Lepton fake rate of e-ID (left) and μ-ID (right).

Bibliography

[1] ATLAS Collaboration, Phys. Lett. B **716**, 1, (2012).

[2] CMS Collaboration, Phys. Lett. B **716**, 30, (2012).

[3] ATLAS Collaboration, ATLAS-CONF-2012-104 (2012).

[4] ATLAS Collaboration, ATLAS-CONF-2012-041 (2012).

[5] E. Lunghi, A. Masiero, I. Scimemi, L. Silvestrini, Nucl. Phys. B **568**, 120 (2000).

[6] J. R. Ellis, M. K. Gaillard, D. V. Nanopoulos and S. Rudaz, Nucl. Phys. B **131**, 285 (1977) [Erratum-ibid. B **132**, 541 (1978)].

[7] Shifman MA. arXiv:hep-ph/9510397 (1995)

[8] A. Ali, E. Lunghi, C. Greub, and G. Hiller, Phys. Rev. D **66**, 034002 (2002).

[9] R. Ammar *et al.* (CLEO Collaboration), Phys. Rev. Lett. **71**, 674 (1993).

[10] N. G. Deshpande, OITS-530.

[11] M. Misiak *et al.*, Phys. Rev. Lett. **98**, 022002 (2007).

[12] Ulrich Haisch and Andreas Weiler, Phys. Rev. D **76**, 034014 (2007).

[13] D. Atwood, M. Gronau, and A. Soni, Phys. Rev. Lett. **79**, 185 (1997). D. Atwood, T. Gershon, M. Hazumi, and A. Soni, Phys. Rev. D **71**, 076003 (2005).

[14] K. Abe *et al.* (Belle Collaboration), Phys. Rev. Lett. **88**, 021801 (2001).

[15] S. Fukae, C. S. Kim, T. Morozumi, and T. Yoshikawa, Phys. Rev. D **59**, 074013 (1999).

[16] Keith S. M. Lee, Zoltan Ligeti, Iain W. Stewart, and Frank J. Tackmann, Phys. Rev. D **75**, 034016 (2007).

[17] J.-T. Wei *et al.*, Phys. Rev. Lett. **103**, 171801 (2009).

[18] B. Aubert *et al.*, Phys. Rev. Lett. **102**, 091803 (2009).

[19] talk by T. Iijima presented at Lepton Photon (2009).

[20] B. Aubert *et al.*, Phys. Rev. Lett. **93**, 081802 (2002).

[21] Gambino P, Haisch U, Misiak M., Phys. Rev. Lett. **94**, 061803 (2005).

[22] A. Ali, hep-ph/0210183, CERN-TH/2002-284 (2002).

[23] Heavy Flavor Averaging Group Web Page : http://www.slac.stanford.edu/xorg/hfag/

[24] A. Ali *et al.*, Phys. Rev. D **61**, 074024 (2000).

[25] F. Mahmoudi, hep-ph/12053099 .

[26] Buras. Isidori, hep-ph/12080934 .

[27] De Bruyn *et al.*, hep-ph/12041737 .

[28] The LHCb collaboration, hep-ex/12112674 .

[29] T. Hurth, F. Mahmoudi, Nucl. Phys. B **865**, 461 (2012).

[30] S. Kurokawa and E. Kikutani, Nucl. Instrum. Meth. A **499**, 1 (2003), and other papers included in this Volume.

[31] T. Abe *et al.*, arXiv:hep-ex/07063248 (2011)

[32] Belle Collaboration, A. Abashian *et al.*, Nucl. Instrum. Meth. A **479**, 117 (2002).

[33] Z. Natkaniec *et al.*, Nucl. Instrum. Meth. A **560**, 1 (2006).

[34] H. Hirano *et al.*, Nucl. Instrum. Meth. A **455**, 294, (2000).

[35] T. Iijima *et al.*, Nucl. Instrum. Meth. A **453** 321 (2000).

[36] T. Iijima *et al.*, Nucl. Instrum. Meth. A **379** 457 (1996).

[37] H. Kichimi *et al.*, Nucl. Instrum. Meth. A **453** 315 (2000).

[38] K. Miyabayashi *et al*, Nucl. Instrum. Meth. A **494**, 298, (2002).

[39] H. Ikeda *et al.*, Nucl. Instrum. Meth. A **441**, 401 (2000).

[40] A. Abashian *et al.*, Nucl. Instrum. Meth. A **491**, 69 (2002).

[41] K. Abe *et al.*, Nucl. Instrum. Meth. A **455**, 397 (2000).

[42] Y. Ushiroda *et al.*, Nucl. Instrum. Meth. A **438**, 460 (1999).

[43] T. Ziegler *et al.*, IEEE Trans. Nucl. Sci., **51**, 1852 (2004).

[44] E. Nakano, Nucl. Instrum. Meth. A **494**, 402 (2002).

[45] K. Hanagaki, *et al.*, Nucl. Instrum. Meth. A **485**, 490 (2002).

[46] A. Abashian, *et al.*, Nucl. Instrum. Meth. A **491**, 69 (2002).

[47] D. J. Lange, Nucl. Instrum. Meth. A **462**, 152 (2001).

[48] T Sjöstrand, Comput. Phys. Commun. **82**. 74 (1994)

[49] R. Brun *et al.*, GEANT 3 Manual, CERN Program Library Long Writeup W5013, 1994.

[50] M. Feindt and U. Kerzel, Nucl. Instrum. Meth. A **559**, 190 (2006).

[51] S. H. Lee *et al.* (Belle Collaboration), Phys. Rev. Lett. **91**, 261801 (2003)

[52] R. A. Fisher, Annals of Human genetics **7**, 179 (1936).

[53] G. C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978).

[54] H. Albrecht *et al.* (ARGUS Collaboration), Phys. Lett. B **241**, 278 (1990).

[55] T. Skwarnicki, Ph.D. Thesis, Institute for Nuclear Physics, Krakow 1986. DESY Internal Report, DESY F31-86-02 (1986).

[56] D. Cronin-Hennessy *et al.* (CLEO Collaboration), Phys. Rev. Lett. **87**, 251808 (2001).

[57] S. Chen *et al.* CLEO Collaboration, Phys. Rev. Lett. **87**, 251807 (2001).

[58] T. Abe *et al.* KEK Report 2010-1, arXiv:1011.035 (2010).