

Precision Measurement of the Electroweak
Flavor-Changing Neutral Current Decays of B Mesons

Hiroyuki NAKAYAMA
Department of Physics, The University of Tokyo

December 2009

Abstract

We report the precise measurement of the branching fraction for the electroweak penguin process $B \rightarrow X_s \ell^+ \ell^-$, where ℓ is an electron or a muon and X_s is a hadronic system containing an s -quark, reconstructed from one K^\pm or K_S^0 and up to four pions, where at most one pion can be neutral.

$B \rightarrow X_s \ell^+ \ell^-$ decay is a Flavor Changing Neutral Current (FCNC) process, highly suppressed in the standard model and therefore a probe for searching new physics beyond the standard model.

We use a data sample that contains $657 \times 10^6 B\bar{B}$ pairs collected on the $\Upsilon(4S)$ resonance, observed by the Belle detector at the KEKB asymmetric energy e^+e^- collider.

The measured branching fractions for $M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$ are:

$$\begin{aligned}\mathcal{B}(B \rightarrow X_s e^+ e^-) &= \left(4.59 \pm 1.15(\text{stat.})_{-0.51}^{+0.56}(\text{syst.}) \right) \times 10^{-6}, \\ \mathcal{B}(B \rightarrow X_s \mu^+ \mu^-) &= \left(1.91 \pm 1.02(\text{stat.})_{-0.14}^{+0.15}(\text{syst.}) \right) \times 10^{-6}, \\ \mathcal{B}(B \rightarrow X_s \ell^+ \ell^-) &= \left(3.22 \pm 0.79(\text{stat.})_{-0.25}^{+0.28}(\text{syst.}) \right) \times 10^{-6}.\end{aligned}$$

Here the first and second errors are statistical and systematic errors. These results are consistent with Standard Model predictions within errors.

The differential branching fractions as functions of invariant mass of X_s and $\ell^+\ell^-$ systems are also measured, and are consistent with the Standard Model.

Acknowledgments

First, I would like to express my great appreciation to my supervisor, Prof. H. Aihara, for giving me the opportunity to work on this research and his great supervision.

I greatly appreciate Prof. M. Iwasaki (Abe) and Dr. H. Kakuno for their powerful support and advice on my research works.

I am grateful to the members, students, and alumni of our group at University of Tokyo: Dr. T. Abe, Dr. N. Hastings, Dr. T. Uchida, Dr. Y. Nakahama, Mr. H. Miyatake, Mr. H. Fujimori, Mr. S. Mineo, Mr. C. Ng, Mr. S. Sugihara; Dr. T. Nakadaira, Mr. H. Kawai, Dr. A. Kusaka, Mr. K. Tanabe, Mr. R. Ishida, Mr. K. Ito and Mr. K. Yamada. They have been good teachers, capable colleagues, and nice drinking companies.

I would like to express my gratitude to all the members of the EWP, DCPV and DAQ groups in Belle; especially Prof. M. Nakao for the great supervision on both analysis and DAQ work, over four years at KEK. I am grateful to Dr. A. Ishikawa for his instructions on this analysis. I would like to express my appreciation to Prof. R. Ito, Dr. T. Higuchi and Dr. S. Y. Suzuki for the supervision on my DAQ work.

I also wish to give many thanks to my friends in Tsukuba Hall; Dr. J. Dalseno, Dr. M. Iwabuchi, Dr. M. Fujikawa and Mr. D. Heffernan. We share many good memories from life at KEK.

I appreciate all the members of the Belle collaboration and the KEKB accelerator group for keeping the good performance of the experimental apparatus, great contributions to the development of analysis frameworks, and the high luminosity from the excellent operation.

Finally, I am particularly indebted to my family; I am grateful to my wife for her encouragement and overall support. I also wish to express my obligation to my parents for their continuous understanding.

Contents

1	Introduction	13
2	Flavor-Changing Neutral Current B meson decays	15
2.1	Effective Hamiltonian approach	17
2.2	Branching fraction of $b \rightarrow s\ell^+\ell^-$, $b \rightarrow s\gamma$	18
2.2.1	$b \rightarrow s\ell^+\ell^-$	18
2.2.2	$b \rightarrow s\gamma$	21
2.3	Dilepton mass distribution of $b \rightarrow s\ell^+\ell^-$	23
2.4	Forward-backward asymmetry of $b \rightarrow s\ell^+\ell^-$	25
2.5	New physics impact on Wilson Coefficients	27
3	The Experimental Apparatus	29
3.1	The KEKB accelerator	29
3.2	Belle Detector	33
3.2.1	Silicon Vertex Detector (SVD)	36
3.2.2	Central Drift Chamber (CDC)	39
3.2.3	Aerogel Cerenkov Counter System (ACC)	42
3.2.4	Time-of-Flight Counter (TOF)	44
3.2.5	Electromagnetic Calorimeter (ECL)	46
3.2.6	K_L^0 and Muon Detection System (KLM)	49
3.2.7	Extreme Forward Calorimeter (EFC)	50
3.2.8	Trigger system	51
3.2.9	Data Acquisition system	52
4	Analysis overview	54
4.1	Semi-inclusive Approach	54
4.2	Particle Selections	54
4.3	Event Selections	56
4.4	Background Suppression	56
4.5	Best Candidate Selection and Likelihood ratio cut	57
4.6	Data Samples	63
4.6.1	Real data sample	63
4.6.2	Signal Monte Carlo sample and efficiency	63
4.6.3	Background Monte Carlo sample	63

5	Maximum Likelihood Fit	65
5.1	Fitter method	65
5.1.1	Signal component (sig)	66
5.1.2	Background Component (bkg)	67
5.1.3	Peaking Background Component (pkg,psi)	67
5.1.4	Self cross-feed component (scf)	68
5.1.5	PDFs for each of M_{X_s} or $M_{\ell^+\ell^-}$ bins	71
5.2	Fitter validation test	72
5.2.1	Toy MC check	72
5.2.2	Ensemble check	74
6	Results with full (605 fb^{-1}) data	78
7	Systematic uncertainties	83
7.1	Signal yield uncertainty	83
7.2	Detector Modeling Uncertainty	85
7.3	Signal Model Uncertainty	89
7.4	Uncertainty in number of B mesons	90
8	Final results	94
8.1	A simple method	94
8.2	The sum-up method	96
8.3	Extrapolation for entire M_{X_s} region (over $M_{X_s} > 2.0 \text{ GeV}/c^2$)	96
8.4	MC normalization from measured branching fraction	97
9	Conclusion	99
9.1	Summary of the results	99
9.2	Discussion	99
A	The new peaking background sources	104
A.1	Charmonium higher resonances	104
A.2	Hadronic with neutrino	104
A.3	Cross check with generic MC	106
B	PDF for each of M_{X_s} or $M_{\ell^+\ell^-}$ bins	108
C	Cross check with the previous 140 fb^{-1} analysis	125
C.1	Summary of the previous results	125
C.2	140 fb^{-1} results by simultaneous fit	127
C.3	140 fb^{-1} results by non-simultaneous fit	132

List of Figures

1.1	The loop diagram (left) and the box diagram (right) for $b \rightarrow sl^+\ell^-$ process.	14
2.1	The loop diagram (left) and the box diagram (right) for $b \rightarrow sl^+\ell^-$ process.	15
2.2	The loop diagram for $b \rightarrow s\gamma$ process.	16
2.3	Penguin drawings (courtesy A. Lenz)[10].	16
2.4	Model-independent constraints on additive new physics contributions to $C_{9,10}$ at 90% C.L for the SM-like (upper plot) and opposite (lower plot) sign of C_7 . The three lines correspond to three different values of $\mathcal{B}(B \rightarrow X_s\gamma)$. The dot at $(C_9, C_{10}) = (0, 0)$ indicates the SM case for $C_{9,10}$	20
2.5	The measurements and theoretical calculations for $B \rightarrow X_s\gamma$ branching fraction.	21
2.6	The constraints on the charged Higgs mass from the $B \rightarrow X_s\gamma$ branching fraction [65].	22
2.7	The dilepton invariant mass distribution of $B \rightarrow X_s\ell^+\ell^-$ calculated in partial (dashed lines) or full (solid lines) NNLO computation [2]. The left (right) figure is for the e^+e^- ($\mu^+\mu^-$) modes. In the left figure, the three lines correspond to the scale $\mu = 2.5, 5, 10$ GeV from upper to lower. In the right figure, the three lines correspond to the scale $\mu = 10, 5, 2.5$ GeV from upper to lower.	23
2.8	The differential forward-backward asymmetry for $B \rightarrow X_s\ell^+\ell^-$ decay theoretically calculated assuming various value of Wilson Coefficients. Line SM, 1, 2, 3 correspond to the points in Figure 2.10, $(\Delta C_9, \Delta C_{10}) = (0, 0), (1, 8.5), (-9, 9), (-8.5, -1)$, respectively.	25
2.9	The A_{FB} measurements for $B \rightarrow K^*\ell^+\ell^-$ decay with 605 fb^{-1} data in Belle. Points represent the measured A_{FB} and solid (dotted) curve represents the SM ($C_7 = -C_7^{SM}$) prediction. The two shaded regions are veto windows to reject $J/\psi K^*$ events.	26
2.10	The experimental constraints on additive new physics contributions to $C_{9,10}$ at 90% C.L for the SM-like (upper plot) and opposite (lower plot) sign of C_7 . The dot at $(C_9, C_{10}) = (0, 0)$ indicates the SM case for $C_{9,10}$. The other dots represent the possible $C_{9,10}$ values in case of EMFV theory.	27

2.11	The upper figure shows the m_H dependence of the real part of the effective Wilson coefficient $C_{\tau\gamma}(m_b)$ in the SM (solid line) and T2HDM for $\sigma = 0^\circ$, and $\tan\theta = 10$ (dots curve), 30 (dot-dashed curve) and 50 (dashed curve), respectively. The lower figure shows the m_H dependence of the real part of the effective Wilson coefficient $C_{9V}(m_b)$ in the SM (solid line) and T2HDM for $\sigma = 0^\circ$, $\hat{s} = 0.2$, and $y = 10$ (dots curve), 30 (dot-dashed curve) and 50 (dashed curve), respectively.	28
3.1	Schematic view of the layout of KEKB.	30
3.2	Drawing of KEKB crab cavity.	31
3.3	Bunch crossing schemes without (top) and with (bottom) use of crab cavities.	31
3.4	The KEKB luminosity history.	32
3.5	Overview of the Belle Detector.	33
3.6	Detector configuration of SVD1.	36
3.7	Detector configuration of SVD2.	37
3.8	Schematic drawing of the DSSD.	38
3.9	Comparison of the impact parameter resolutions in the directions of r - ϕ (left) and z (right) measured with cosmic ray data.	38
3.10	Overview of the CDC structure.	39
3.11	p_t resolution studied using cosmic rays.	40
3.12	Charged track momenta vs. dE/dx observed in collision data.	41
3.13	Arrangement of the ACC in Belle detector.	42
3.14	ACC number of photo-electron distribution for π^\pm and K^\pm from $D^{*\pm}$ decays. Each plot corresponds to the different set of modules with a different refractive index.	43
3.15	Dimensions of a TOF/TSC module.	44
3.16	Time resolution of the TOF for $e^+e^- \rightarrow \mu^+\mu^-$ events.	45
3.17	Mass distribution from TOF measurements for particles with momenta below 1.2 GeV/c. The histogram corresponds to MC distribution.	45
3.18	Configuration of ECL.	46
3.19	Energy dependence of the average position resolution. The solid curve is the result of the fit to the MC.	47
3.20	Energy resolutions calibrated and measured with $e^+e^- \rightarrow e^+e^-$ (Bhabha) events. The plots correspond to the overall average (top left) and each of the barrel (top right), forward end-cap (bottom left), and backward end-cap (bottom right) sections.	48
3.21	Cross section of a KLM superlayer.	49
3.22	Distribution of the difference between directions of the neutral cluster detected by KLM and the missing momentum.	50
3.23	Schematic figure of the Level-1 trigger system for Belle detector.	51
3.24	An overview of the Belle DAQ system.	52
3.25	The picture of a COPPER module.	53
3.26	The dead time before and after the CDC COPPER installation At the typical data size condition (~ 100 hits per TDC), the dead time decreased from ~ 30 ns to ~ 3 ns.	53

4.1	Distributions of three input parameters of $\mathcal{F}_{\text{miss}}$: a) E_{vis} (in GeV), b) M_{miss} (in GeV/ c^2) and c) ΔE^{ROE} (in GeV); and d) the Fisher output. Black (red) histogram represents the distribution of signal (background) Monte Carlo sample.	57
4.2	Distributions of six input parameters for \mathcal{F}_{FW} : a) R_2^{so} , b) R_4^{so} , c) R_1^{oo} , d) R_2^{oo} , e) R_3^{oo} , and f) R_4^{oo} ; and g) the \mathcal{F}_{FW} . Black(red) histogram represents the distribution of signal(background) Monte Carlo sample.	58
4.3	The distributions of discriminating variables: ΔE (e^+e^- modes), ΔE ($\mu^+\mu^-$ modes), $\cos\theta_B$, \mathcal{F}_{FW} , $\mathcal{F}_{\text{miss}}$, $ \Delta z_{\ell^+\ell^-} $ and $\chi_{\text{vtx}}^2/\text{NDF}$, for the signal Monte Carlo sample (black) and the real $J/\psi X_s$ sample(green). The signal PDFs are determined by the fitting to the real $J/\psi X_s$ distribution. The obtained PDFs are also shown in thick black lines. Fitting to the ΔE distribution of $X_s e^+e^-$ modes and $X_s \mu^+\mu^-$ modes are separately done.	59
4.4	The distributions of discriminating variables: ΔE (e^+e^- modes), ΔE ($\mu^+\mu^-$ modes), $\cos\theta_B$, \mathcal{F}_{FW} , $\mathcal{F}_{\text{miss}}$, $ \Delta z_{\ell^+\ell^-} $ and $\chi_{\text{vtx}}^2/\text{NDF}$, for the background Monte Carlo events. The background PDFs are determined by the fitting to the background Monte Carlo distribution. The obtained PDFs are also shown in thick black lines. Fitting to the ΔE distribution of $X_s e^+e^-$ modes and $X_s \mu^+\mu^-$ modes of are separately done.	60
4.5	Likelihood ratio check for a) signal events and b) background events. For signal events, we compare signal MC (black) with real $J/\psi X_s$ events (green). For background events, we compare generic background MC (black) and real $B \rightarrow X_s e^\pm \mu^\mp$ events (red). The differences between distribution of data and MC are taken as the systematic uncertainty.	61
5.1	The M_{bc} distributions for the real $J/\psi X_s$ sample for a) e^+e^- , b) $\mu^+\mu^-$ and c) $\ell^+\ell^-$ modes. The obtained fitted values are used for the $B \rightarrow X_s \ell^+\ell^-$ signal Gaussian fitting parameters, as described in the text. These histograms correspond to 605 fb^{-1}	69
5.2	The estimated M_{bc} distributions for J/ψ (upper) and $\psi(2S)$ (lower) charmonium peaking background events, for dielectron(left), dimuon(center) and dilepton(right) cases. These histograms correspond to 605 fb^{-1}	69
5.3	The estimated M_{bc} distributions for hadronic peaking background events, for dielectron(left) and dimuon(right) cases. These histograms include both of two hadronic peaking background events. These histograms correspond to 605 fb^{-1}	70
5.4	The estimated M_{bc} distributions for self cross-feed events, for dielectron(left), dimuon(center) and dilepton(right) cases. These histograms correspond to 605 fb^{-1}	70
5.5	Toy MC test result for the N_{sig} . For 140 fb^{-1} , estimated number of signal is 69. a) is the fitter output as a function of the input, fitted with a linear function. b) is the residual from the fitted linear function. c) is the mean of the pull distribution. d) is the resolution of the pull distribution. We found no problem in the linearity and the pull distribution.	72

5.6	Toy MC test result for the ARGUS shape parameter. We found the linearity and the pull distribution is good. a) is the fitter output as a function of the input, fitted with linear function. b) is the residual from the fitted linear function. c) is the mean of the pull distribution. d) is the resolution of the pull distribution. We found no problem in the linearity and the pull distribution.	73
5.7	Ensemble test result for the fitter check. Signal, background and self cross feed components are randomly picked up from the corresponding MC sample. N_{sig} is estimated to be 69 for 140 fb^{-1} according to the MC. The peaking background component is generated from PDF. a) is the fitter output as a function of the input, fitted with a linear function. b) is the residual from the fitted linear function. c) is the mean of the pull distribution. d) is the resolution of the pull distribution. We found a excess in the N_{sig} output.	74
5.8	a) is N_{sig} excess as a function of the N_{bkg} input in the ensemble test, fitted with a linear function. b) is the residual from the fitted linear function. N_{bkg} is estimated to be ~ 1500 for 140 fb^{-1} according to the MC. We found the N_{sig} excess is proportional to the N_{bkg} input, which implies excess is purely from the background component, not from others. OB	75
5.9	Fitting M_{bc} distribution of background MC with Gaussian + ARGUS function, for a) the $B\bar{B}$ background MC and b) the continuum background MC. Gaussian shape parameter is fixed to be the value determined from real $J/\psi X_s$ sample, in the same way of the data fitting. ARGUS shape parameter is floated.	76
5.10	Fitting M_{bc} distribution of background MC with Gaussian + ARGUS function, for a) the estimated peaking background, b) the generic background MC and c) sum of a) and b). Gaussian shape parameter is fixed to be the same value as the real data fitting. ARGUS shape parameter is floated. The histogram for a) is the one estimated in Section 5. . . .	77
6.1	M_{bc} distribution of the 605 fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample(upper) and 605 fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample(lower). Each column corresponds to $e^+ e^-$ (left), $\mu^+ \mu^-$ (center) and $\ell^+ \ell^-$ (right) cases. Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in the black line.	79
6.2	M_{bc} distribution of 605 fb^{-1} data for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to $e^+ e^-$ (left), $\mu^+ \mu^-$ (center) and $\ell^+ \ell^-$ (right) cases. Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in the black line.	80

6.3	M_{bc} distribution of 605 fb^{-1} data for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- (left), $\mu^+\mu^-$ (center) and $\ell^+\ell^-$ (right) cases. Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in the black line.	81
7.1	Tracking efficiency errors for several charged particles, as a function of momentum [60].	85
7.2	Typical electron identification efficiency for data (left) and MC (right), as a function of electron momentum [61].	86
7.3	Typical muon identification efficiency for data (left) and MC (right), as a function of muon momentum [61].	87
7.4	The typical kaon identification efficiency as a function of kaon momentum, for real data and Monte Carlo sample [62]. This figure is for $0.017 < \cos \theta < 0.029$	87
7.5	The typical pion identification efficiency as a function of pion momentum, for real data and Monte Carlo sample [62]. This figure is for $0.017 < \cos \theta < 0.029$	88
7.6	M_{X_s} (left) and $M_{\ell^+\ell^-}$ (right) distribution for the various Fermi motion parameter. Upper figures are for e^+e^- mode and lower figures are for $\mu^+\mu^-$ mode. The black, red and green line represents the distribution for $p_F = 410, 200, \text{ and } 480 \text{ MeV}/c$, respectively. M_{X_s} distributions are normalized to have the same entries in $1.1 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$	89
8.1	Differential branching fractions for M_{X_s} bins, by simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 605 fb^{-1} results and the histogram represents the signal MC simulation. The signal MC normalization for non-resonant $X_s\ell^+\ell^-$ sample is based on the measurements in this 605 fb^{-1} analysis, not the HFAG values.	98
8.2	Differential branching fractions for q^2 bins, by simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 605 fb^{-1} results and the histogram represents the signal MC simulation. The signal MC normalization for non-resonant $X_s\ell^+\ell^-$ sample is based on the measurements in this 605 fb^{-1} analysis, not the HFAG values.	98
9.1	The latest HFAG averages for Electroweak radiative B decays (2009 summer).	100
9.2	Model-independent constraints on additive new physics contributions to $C_{9,10}$ at 95% C.L for the SM-like (upper plot) and opposite (lower plot) sign of C_7 . The dot at $(C_9, C_{10}) = (0, 0)$ indicates the SM case for $C_{9,10}$	102

9.3	The constraints from this $B \rightarrow X_s \ell^+ \ell^-$ measurement on new physics parameters in the Extended MFV model, in case of SM-like (upper) and opposite (lower) sign of C_7 . Red dots represents the possible $C_{9,10}$ values in the Extended MFV model (shown in Figure 2.10).	103
A.1	M_{bc} distributions for hadronic peaking background events with neutrino, for electron(left) and muon(right) cases. These histograms correspond to 605 fb^{-1}	105
A.2	M_{bc} distributions for generic MC sample (10 streams). The black histogram corresponds to the events after all analysis cut. The blue histogram corresponds to the events with an additional requirement that all X_s daughters are coming from the same B meson. The green histogram corresponds to the events with same requirements except for π^0 in X_s daughter.	106
A.3	M_{bc} distributions for generic MC sample (10 streams). The black histogram corresponds to the events after all analysis cut. The blue histogram corresponds to the events with an additional requirement that all X_s daughters are coming from the same B meson. The green histogram corresponds to the events with same requirements except for the π^0 in X_s daughter.	107
B.1	M_{bc} distribution of real $J/\psi X_s$ sample for each M_{X_s} bin, fitted with Gaussian + ARGUS function. Obtained Gaussian shape parameters are used for signal PDF $\mathcal{P}_{\text{sig}}^i$. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	109
B.2	M_{bc} distribution of real $J/\psi X_s$ sample for each M_{X_s} bin, fitted with Gaussian + ARGUS function. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} . Because the real $J/\psi X_s$ sample has all entries in $q^2 = [6.0, 14.4]$, and no entry in other q^2 regions, we use the value obtained from the full sample for $\mathcal{P}_{\text{sig}}^i$ for each q^2 region.	110
B.3	M_{bc} distribution of peaking background for each M_{X_s} bin, estimated from the charmonium MC and real hadronic candidates. Histogram shapes are used for peaking background PDF $\mathcal{P}_{\text{pkg}}^i$. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	111

B.4	M_{bc} distribution of peaking background for each q^2 bin, estimated from the charmonium MC and real hadronic candidates. Histogram shapes are used for peaking background PDF $\mathcal{P}_{\text{pkg}}^i$. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	112
B.5	M_{bc} distribution of $J/\psi X_s + \psi(2S)X_s$ peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	113
B.6	M_{bc} distribution of $J/\psi X_s + \psi(2S)X_s$ peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	114
B.7	M_{bc} distribution of $B \rightarrow X_s hh$ peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	115
B.8	M_{bc} distribution of $B \rightarrow X_s hh$ peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	116
B.9	M_{bc} distribution of $B \rightarrow X_s h \ell \nu + B \rightarrow X_s hh$ peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	117

- B.10 M_{bc} distribution of $B \rightarrow X_s h \ell \nu + B \rightarrow X_s h h$ peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} 118
- B.11 M_{bc} distribution of generic MC charmonium peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson. 119
- B.12 M_{bc} distribution of generic MC charmonium peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson. 120
- B.13 M_{bc} distribution of generic MC hadronic peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson. 121
- B.14 M_{bc} distribution of generic MC hadronic peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson. 122
- B.15 M_{bc} distribution of self cross-feed candidates in signal MC for each M_{X_s} bin. Histogram shapes are used for self cross-feed PDF $\mathcal{P}_{\text{scf}}^i$. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} 123

B.16	M_{bc} distribution of self cross-feed candidates in signal MC for each M_{X_s} bin. Histogram shapes are used for self cross-feed PDF $\mathcal{P}_{\text{scf}}^i$. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1}	124
C.1	M_{bc} distributions of selected (a) $B \rightarrow X_s e^+e^-$, (b) $B \rightarrow X_s \mu^+\mu^-$, (c) $B \rightarrow X_s \ell^+\ell^-$ ($\ell = e, \mu$), and (d) $B \rightarrow X_s e^\pm \mu^\mp$ candidates. The solid lines represent the result of the fits, and dashed lines represent the sum of all background components under the signal peaks, respectively. . . .	126
C.2	M_{bc} fit to the 140 fb^{-1} of $B \rightarrow X_s \ell^+\ell^-$ sample(upper) and 605 fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample(lower). Each column corresponds to e^+e^- (left), $\mu^+\mu^-$ (center) and $\ell^+\ell^-$ case (right). Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in black line.	128
C.3	Differential branching fraction in each M_{X_s} bin, by the simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 140 fb^{-1} results and the histogram represents the signal MC simulation.	131
C.4	Differential branching fraction in each q^2 bin, by the simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 140 fb^{-1} results and the histogram represents the signal MC simulation.	131
C.5	M_{bc} fit with 140 fb^{-1} data, for a) $B \rightarrow X_s e^+e^-$, b) $B \rightarrow X_s \mu^+\mu^-$ and c) $B \rightarrow X_s \ell^+\ell^-$. Background, peaking background, self cross feed components are shown in yellow, green, blue area, respectively. Signal component is shown in black line.	132
C.6	M_{bc} fit with 140 fb^{-1} data for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right.	133
C.7	M_{bc} fit with 140 fb^{-1} data for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right.	134

Chapter 1

Introduction

The final goal of particle physics is to find the ultimate framework which explains the fundamental constituents of matter and their interactions. The most established framework we have at present is the Standard Model (SM). The SM is a gauge theory of the electroweak and strong interactions with the gauge group $SU(3) \otimes SU(2) \otimes U(1)$, and it has been very successful in explaining many elementary particles and phenomena in the past decades. However, we consider it could not be the ultimate theory.

To search new physics beyond the SM, there exist several direct and indirect approaches. One possible approach is to find the deviation from the SM prediction by precise measurements with high statistics which can only be possible with very high luminosity experiments such as those at B factories.

Two B factories, the Belle experiment at KEK in Japan and the BaBar experiment at SLAC in USA, have enjoyed successful operation and performed independent tests of the Standard Model description of CP violation as well as stringent tests of the unitarity of the Cabibbo-Kobayashi-Maskawa (CKM) matrix with a rich CKM phenomenology.

The B factories also provide a possibility to measure very rare decays such as $b \rightarrow s\ell^+\ell^-$ and $b \rightarrow s\gamma$. These processes are called ‘‘Flavor Changing Neutral Currents’’ (FCNC’s). In the SM, FCNC processes are forbidden at the tree level and strongly suppressed, and can only be generated via a loop or a box diagram. For example, the diagrams for $b \rightarrow s\ell^+\ell^-$ decay are shown in the Figure 1.1.

If non-SM particles exist, they can appear in the loop and these additional processes will create a deviation in the branching fraction. Thus we can find new physics by measuring the branching fractions of FCNC processes and comparing them with their SM predictions.

In this thesis, we present a measurement of $b \rightarrow s\ell^+\ell^-$ FCNC decay. There are two approaches for the analysis of this decay, exclusive and inclusive reconstructions. The exclusive $B \rightarrow K^{(*)}\ell^+\ell^-$ process can be reconstructed with low background, but has a large theoretical uncertainty associated with the form-factor of the hadronic B to K meson transition process. In case of the inclusive $B \rightarrow X_s\ell^+\ell^-$ process, form-factor uncertainty is not as severe as the exclusive case, but the background rejection is challenging.

In this thesis, we measure the inclusive $B \rightarrow X_s\ell^+\ell^-$ decay with 605 fb^{-1} data sam-

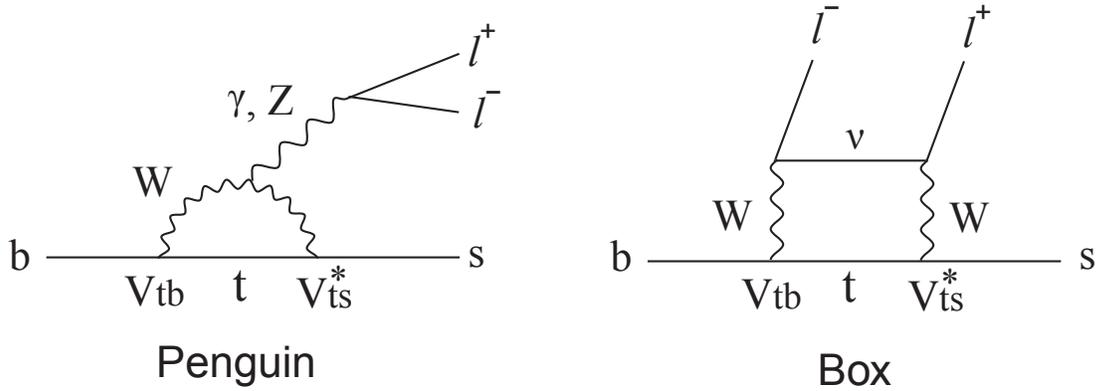


Figure 1.1: The loop diagram (left) and the box diagram (right) for $b \rightarrow s \ell^+ \ell^-$ process.

ples collected with the Belle detector at the KEK e^+e^- asymmetric collider. This measurement updates and supersedes our previous result based on the 140 fb^{-1} data [20]. We also perform the model-independent measurement of the mass spectra of X_s and dilepton systems. The measurement in this thesis is the most precise in the world.

The outline of this thesis is as follows. At first, the details of the $b \rightarrow s$ FCNC process are described in Chapter 2. In Chapter 3, an overview of the experimental apparatus is presented. The analysis procedure and the fitting method are explained in Chapter 4 and 5, respectively. The results with the full 605 fb^{-1} data sample are given in Chapter 6. The study of systematic uncertainties is explained in Chapter 7. In Chapter 8, the final results are summarized. Finally, the discussions and conclusions of this thesis are shown in Chapter 9.

Chapter 2

Flavor-Changing Neutral Current B meson decays

In the SM, Flavor Changing Neutral Currents (FCNC) processes are forbidden at the tree level and strongly suppressed, and can only be generated via a “penguin”¹ diagram or a box diagram. Figures 2.1 and 2.2 show the diagrams for $b \rightarrow s\ell^+\ell^-$ and $b \rightarrow s\gamma$ FCNC processes, respectively.

If non-SM particles exist, they can appear in the loop and these additional processes result in a deviation in the branching fraction. Thus we can find new physics by measuring the branching fractions of FCNC processes and comparing them with their SM predictions.

Therefore $b \rightarrow s\ell^+\ell^-$ (or $b \rightarrow s\gamma$) transition provides a deep insight into the effective Hamiltonian that describes FCNC processes and are sensitive to the effects of non-SM physics that may enter the loop [2, 3].

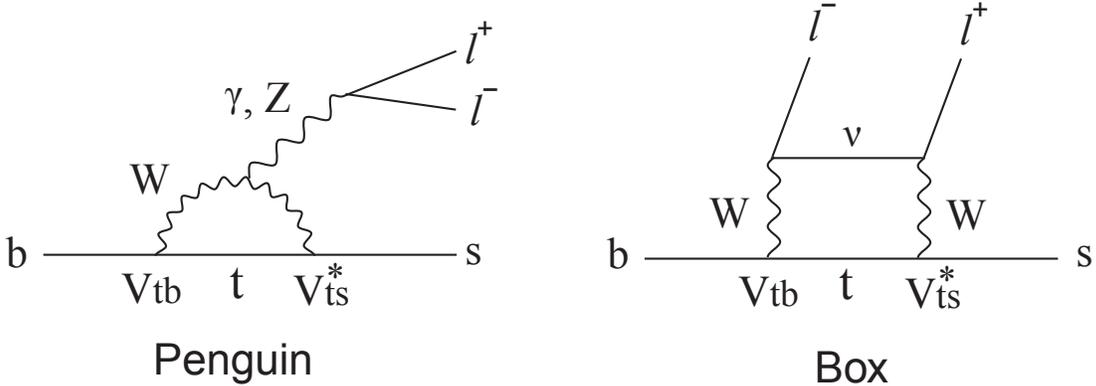


Figure 2.1: The loop diagram (left) and the box diagram (right) for $b \rightarrow s\ell^+\ell^-$ process.

¹These loop diagrams are called “penguins”, first named by [10]. The drawing in Figure 2.3 helps to demonstrate how they might be viewed as penguins.

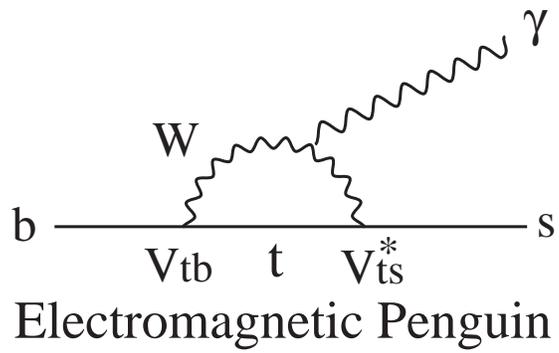


Figure 2.2: The loop diagram for $b \rightarrow s\gamma$ process.

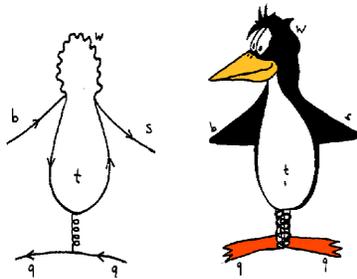


Figure 2.3: Penguin drawings (courtesy A. Lenz)[10].

2.1 Effective Hamiltonian approach

The FCNC process can be generated through an electromagnetic or weak penguin diagram and a box diagram. These Feynman diagrams containing W , Z^0 and a top quark represent a picture at a high energy scale $\mathcal{O}(m_W)$ with the theory of all six quarks. The true picture at the bottom quark scale $\mathcal{O}(m_W)$ is more correctly described by effective operators with the effective five-quark theories. The usual procedure is to start at the high energy scale $\mathcal{O}(m_W)$ and to consecutively integrate out the heavy degrees of freedom, such as W , Z^0 and top quarks, from explicitly appearing in the theory.

The effective Hamiltonian of one-loop processes is given by the operator product expansion (OPE):

$$H_{\text{eff}} = -\frac{4G_F}{\sqrt{2}}V_{ts}^*V_{tb}\sum_{i=1}^{10}C_i(\mu)\mathcal{O}_i(\mu), \quad (2.1)$$

where G_F is the Fermi coupling constant, V_{ij} are elements of the CKM matrix, $\mathcal{O}_i(\mu)$ are relevant operators at the scale μ and $C_i(\mu)$ represents effective strength of short distance interactions, called ‘‘Wilson Coefficients’’.

\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 vector component of electroweak penguin operator, respectively. The local operators are described as:

$$\mathcal{O}_1 = (\bar{s}_\alpha\gamma_\mu Lc_\beta)(\bar{c}_\beta\gamma^\mu Lb_\alpha), \quad (2.2)$$

$$\mathcal{O}_2 = (\bar{s}_\alpha\gamma_\mu Lc_\alpha)(\bar{c}_\beta\gamma^\mu Lb_\beta), \quad (2.3)$$

$$\mathcal{O}_3 = (\bar{s}_\alpha\gamma_\mu Lb_\alpha)\sum_{q=u,d,s,c,b}(\bar{q}_\beta\gamma^\mu Lq_\beta), \quad (2.4)$$

$$\mathcal{O}_4 = (\bar{s}_\alpha\gamma_\mu Lb_\beta)\sum_{q=u,d,s,c,b}(\bar{q}_\beta\gamma^\mu Lq_\alpha), \quad (2.5)$$

$$\mathcal{O}_5 = (\bar{s}_\alpha\gamma_\mu Lb_\alpha)\sum_{q=u,d,s,c,b}(\bar{q}_\beta\gamma^\mu Rq_\beta), \quad (2.6)$$

$$\mathcal{O}_6 = (\bar{s}_\alpha\gamma_\mu Lb_\beta)\sum_{q=u,d,s,c,b}(\bar{q}_\beta\gamma^\mu Rq_\alpha), \quad (2.7)$$

$$\mathcal{O}_7 = \frac{e}{16\pi^2}\bar{s}_\alpha\sigma_{\mu\nu}(m_sL + m_bR)b_\alpha F^{\mu\nu}, \quad (2.8)$$

$$\mathcal{O}_8 = \frac{g}{16\pi^2}\bar{s}_\alpha\sigma_{\mu\nu}(m_sL + m_bR)T_{\alpha\beta}^a b_\beta G^{a\mu\nu}, \quad (2.9)$$

$$\mathcal{O}_9 = \frac{e}{16\pi^2}(\bar{s}_\alpha\gamma^\mu Lb_\alpha)(\bar{\ell}\gamma_\mu\ell), \quad (2.10)$$

$$\mathcal{O}_{10} = \frac{e}{16\pi^2}(\bar{s}_\alpha\gamma^\mu Lb_\alpha)(\bar{\ell}\gamma_\mu\gamma_5\ell), \quad (2.11)$$

where e and g represent the electromagnetic and strong coupling constants, respectively, α and β are color indices, a labels the SU(3) generators, L and R are the projection operators $(1 - \gamma_5)/2$ or $(1 + \gamma_5)/2$, and $F_{\mu\nu}$ and $G^{\mu\nu}$ denote QED and QCD field strength tensors, respectively.

The $C_i(\mu)$ are evaluated perturbatively 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 equations (RGE). The RGE guarantees that the μ dependence of $C_i(\mu)$ is canceled by the μ dependence of $\mathcal{O}_i(\mu)$, 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. New physics appears as the shift of those Wilson coefficients:

$$C_i(\mu) \rightarrow C_i(\mu) + C_i^{NP}(\mu), \quad (2.12)$$

where C^{NP} represents the effect of new physics. In other words, we can perform searches for new physics by measuring the Wilson Coefficients in experiments and comparing them with the predictions of the SM.

In the following sections, we show several observables in FCNC processes and how they are described using the Wilson Coefficients.

2.2 Branching fraction of $b \rightarrow s\ell^+\ell^-$, $b \rightarrow s\gamma$

The amplitude A for the B meson decays to the final state F is calculated as:

$$A(B \rightarrow F) = \langle F | \mathcal{H}_{\text{eff}} | B \rangle = \frac{G_F}{\sqrt{2}} V_{ts}^* V_{tb} \sum_{i=1}^{10} C_i(\mu) \langle F | \mathcal{O}_i(\mu) | B \rangle. \quad (2.13)$$

If $\langle F |$ is an inclusive final state, the leading order term of $\langle F | \mathcal{O}_i(\mu) | B \rangle$ can be represented as $\langle s | \mathcal{O}_i(\mu) | b \rangle$, where $\langle s | \mathcal{O}_i(\mu) | b \rangle$ is a matrix element for a free quark transition which can be easily calculated. If $\langle F |$ is an exclusive state, it is not so easy to calculate.

Although $b \rightarrow s\ell^+\ell^-$ decay rate is expected to be nearly two orders of magnitude lower than that of $b \rightarrow s\gamma$, $b \rightarrow s\ell^+\ell^-$ process has received considerable attention because it is sensitive to C_7 , C_9 and C_{10} , while $b \rightarrow s\gamma$ is only sensitive to the absolute value of C_7 . $b \rightarrow s\ell^+\ell^-$ process has an advantage that we can measure not only the branching fraction but also the dilepton mass distribution and forward-backward charge asymmetry of dilepton, which are also sensitive to new physics.

2.2.1 $b \rightarrow s\ell^+\ell^-$

The decay amplitude of $b \rightarrow s\ell^+\ell^-$ transition can be expressed as follows [43]:

$$\begin{aligned} M(b \rightarrow s\ell^+\ell^-) = & \frac{G_F \alpha_{EM}}{\sqrt{2}\pi} V_{ts}^* V_{tb} \\ & \times [(C_9 - C_{10})(\bar{s}\gamma_\mu L b)(\bar{\ell}\gamma^\mu L \ell) + (C_9 + C_{10})(\bar{s}\gamma_\mu L b)(\bar{\ell}\gamma^\mu R \ell) \\ & - 2C_7(\bar{s}i\sigma_{\mu\nu})\frac{q^\nu}{s^2}(m_s L + m_b R)b(\bar{\ell}\gamma^\mu \ell)] + (\text{corr.}). \end{aligned} \quad (2.14)$$

As you can see, the decay amplitude (and therefore the branching fraction) has a clear dependence on C_7 , C_9 and C_{10} .

Table 2.1 shows the theoretical calculations of the branching fractions for the exclusive decay process of $B \rightarrow K\ell^+\ell^-$ and $B \rightarrow K^*\ell^+\ell^-$ ($\ell = e, \mu$) and inclusive $B \rightarrow X_s\ell^+\ell^-$ process [2, 4]. As for the experimental results, both the Belle and BaBar collaborations have observed exclusive $B \rightarrow K\ell^+\ell^-$ and $B \rightarrow K^*\ell^+\ell^-$ decays [14, 18] and have measured inclusive $B \rightarrow X_s\ell^+\ell^-$ decay [20, 21]. The HFAG world averages in 2009 winter are shown in Table 2.2.

Table 2.1: Recent theoretical branching fraction calculations for $B \rightarrow X_s\ell^+\ell^-$ decays [2, 4].

Mode	e^+e^- mode [$\times 10^{-6}$]	$\mu^+\mu^-$ mode [$\times 10^{-6}$]
$B \rightarrow K\ell^+\ell^-$	0.35 ± 0.12	0.35 ± 0.12
$B \rightarrow K^*\ell^+\ell^-$	1.58 ± 0.49	1.19 ± 0.39
$B \rightarrow X_s\ell^+\ell^-$	6.9 ± 1.0 4.2 ± 0.7 for $M_{e^+e^-} > 0.2 \text{ GeV}/c^2$	4.2 ± 0.7

Table 2.2: Recent experimentally measured branching fractions for $B \rightarrow X_s\ell^+\ell^-$ decays (HFAG Winter08) [14, 18]. Note that the values of $B \rightarrow X_se^+e^-$ and $B \rightarrow X_s\ell^+\ell^-$ are for $M_{e^+e^-} > 0.2 \text{ GeV}/c^2$.

Mode	e^+e^- mode [$\times 10^{-6}$]	$\mu^+\mu^-$ mode [$\times 10^{-6}$]	$\ell^+\ell^-$ mode [$\times 10^{-6}$]
$B \rightarrow K\ell^+\ell^-$	0.42 ± 0.06	0.47 ± 0.06	0.43 ± 0.04
$B \rightarrow K^*\ell^+\ell^-$	$1.24^{+0.19}_{-0.17}$	$1.08^{+0.16}_{-0.14}$	1.00 ± 0.11
$B \rightarrow X_s\ell^+\ell^-$	4.7 ± 1.3	$4.3^{+1.3}_{-1.2}$	$4.50^{+1.03}_{-1.01}$

The constraints for the C_9 and C_{10} from the experimental measurements are shown in Figure 2.4 [1]. The experimental measurements show that the SM-like sign of C_7 is more favored than the opposite sign.

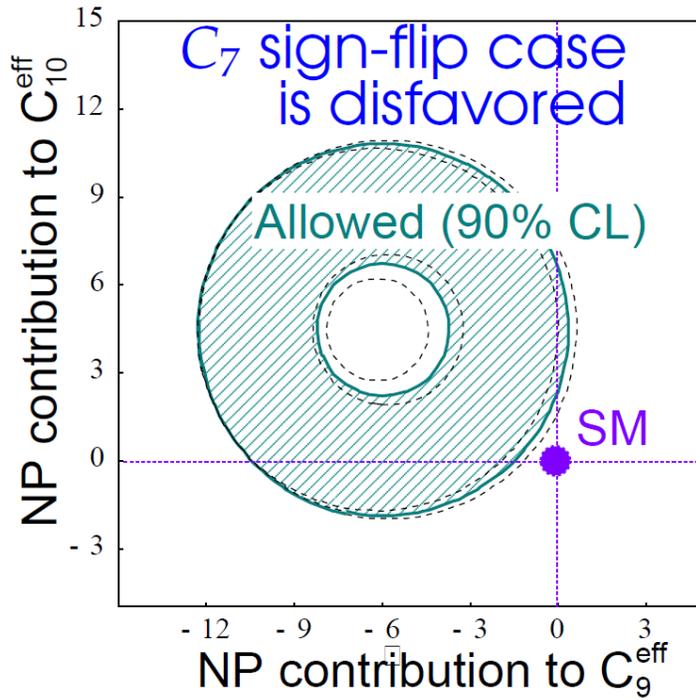
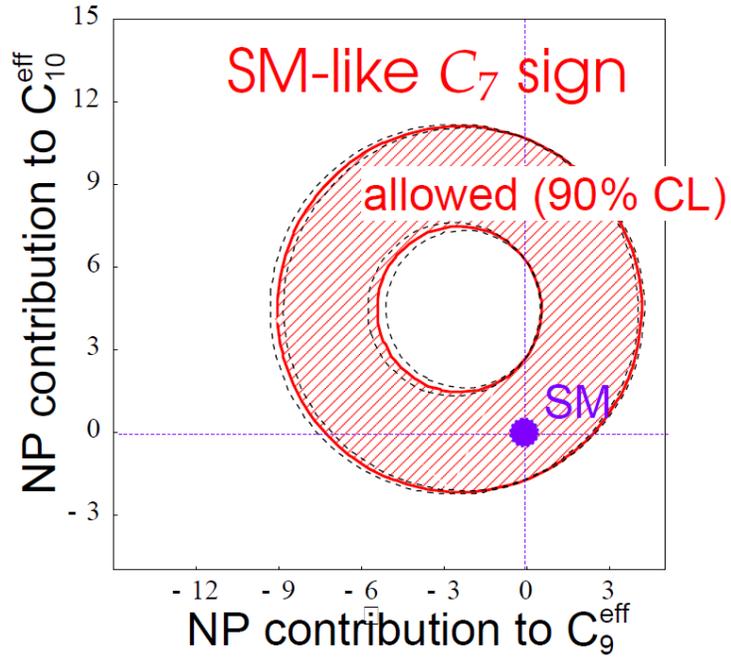


Figure 2.4: Model-independent constraints on additive new physics contributions to $C_{9,10}$ at 90% C.L for the SM-like (upper plot) and opposite (lower plot) sign of C_7 . The three lines correspond to three different values of $\mathcal{B}(B \rightarrow X_s \gamma)$. The dot at $(C_9, C_{10}) = (0, 0)$ indicates the SM case for $C_{9,10}$.

2.2.2 $b \rightarrow s\gamma$

This process is mainly generated through the electromagnetic penguin diagram, therefore this process has a sensitivity for the Wilson Coefficients C_7 . The branching fraction of $b \rightarrow s\gamma$ can be expressed as follows:

$$\mathcal{B}(b \rightarrow s\gamma) = \frac{G_F^2 \alpha_{EM} m_b^5 |V_{ts}^* V_{tb}|^2}{32\pi^4} |C_7|^2 + \text{corr.} \quad (2.15)$$

As one can see, the branching fraction of $b \rightarrow s\gamma$ has a dependence on $|C_7|$.

The measurements and theoretical calculation of $B \rightarrow X_s\gamma$ branching fraction are summarized in Figure 2.5. The constraints on the charged Higgs mass are described in Figure 2.6.

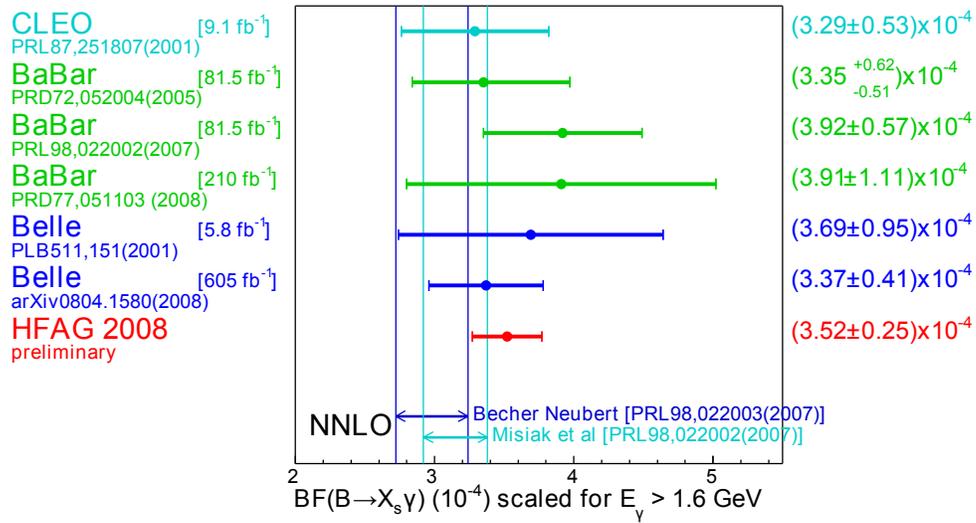


Figure 2.5: The measurements and theoretical calculations for $B \rightarrow X_s\gamma$ branching fraction.

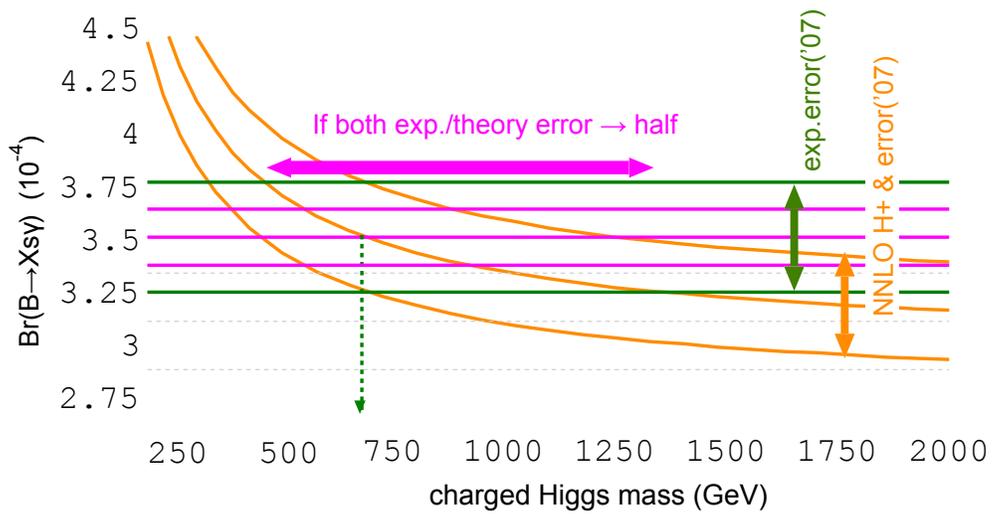


Figure 2.6: The constraints on the charged Higgs mass from the $B \rightarrow X_s \gamma$ branching fraction [65].

2.3 Dilepton mass distribution of $b \rightarrow s\ell^+\ell^-$

The dilepton mass distribution also depends on the Wilson Coefficients. The calculation including the QCD corrections at next-to-next-to leading logarithmic (NNLO) accuracy is written as follows [43]:

$$\frac{d\Gamma(b \rightarrow s\ell^+\ell^-)}{d\hat{s}} = \left(\frac{\alpha_{EM}}{4\pi}\right)^2 \frac{G_F^2 m_b^5 |V_{ts}^* V_{tb}|^2}{48\pi^3} (1 - \hat{s})^2 \times \left[(1 + \hat{s})(|C_9|^2 + |C_{10}|^2) + 4(1 + 2/\hat{s})|C_7|^2 + 12\text{Re}(C_7 C_9^*) \right], \quad (2.16)$$

where \hat{s} is defined as $\hat{s} = M_{\ell^+\ell^-}^2/m_{b,pole}^2$.

Figure 2.7 shows theoretical calculations of dilepton invariant mass distribution in $B \rightarrow X_s \ell^+\ell^-$, with the scale $\mu = 2.5, 5, 10$ GeV.

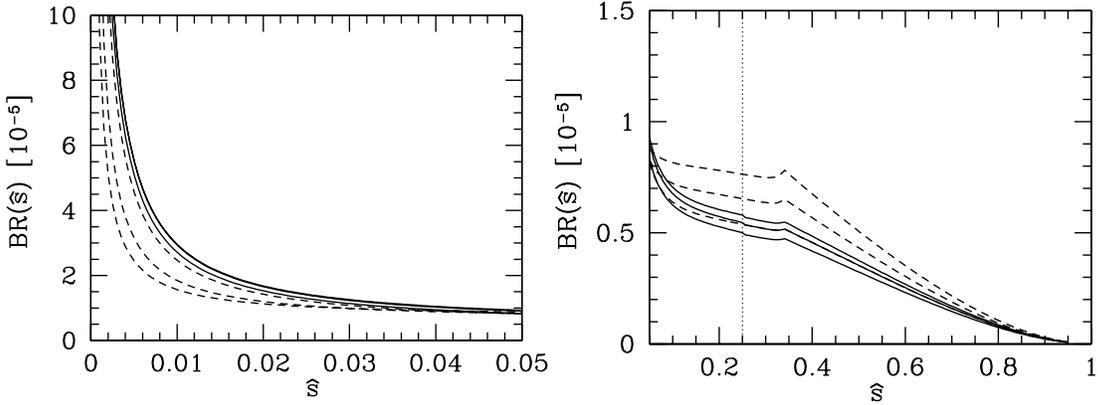


Figure 2.7: The dilepton invariant mass distribution of $B \rightarrow X_s \ell^+\ell^-$ calculated in partial (dashed lines) or full (solid lines) NNLO computation [2]. The left (right) figure is for the e^+e^- ($\mu^+\mu^-$) modes. In the left figure, the three lines correspond to the scale $\mu = 2.5, 5, 10$ GeV from upper to lower. In the right figure, the three lines correspond to the scale $\mu = 10, 5, 2.5$ GeV from upper to lower.

A branching fraction in $1 < M_{\ell^+\ell^-}^2 < 6 (\text{GeV}/c^2)^2$ is of particular interest, since a clean theoretical prediction is available [1]. Tables 2.3 and 2.4 show the experimental measurements and theoretical predictions of the $B \rightarrow X_s \ell^+ \ell^-$ branching fraction, respectively. The experimental measurements show that the SM-like sign of C_7 is more favored than the opposite sign.

Table 2.3: Experimental measurements of the branching fraction in $M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$ and $1 < M_{\ell^+\ell^-}^2 < 6 (\text{GeV}/c^2)^2$.

$M_{\ell^+\ell^-}$ range	BELLE	BaBar	World Average
$M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$	4.11 ± 1.1	5.6 ± 2.0	4.5 ± 1.1
$1 < M_{\ell^+\ell^-}^2 < 6 (\text{GeV}/c^2)^2$	1.5 ± 0.6	1.8 ± 0.9	1.60 ± 0.5

Table 2.4: Theoretical predictions of the branching fraction in $M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$ and $1 < M_{\ell^+\ell^-}^2 < 6 (\text{GeV}/c^2)^2$. with SM parameters and with the opposite sign of C_7 .

$M_{\ell^+\ell^-}$ range	SM	$C_7 = -C_7^{SM}$
$M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$	4.4 ± 0.7	8.8 ± 0.7
$1 < M_{\ell^+\ell^-}^2 < 6 (\text{GeV}/c^2)^2$	1.57 ± 0.16	3.30 ± 0.25

2.4 Forward-backward asymmetry of $b \rightarrow s\ell^+\ell^-$

The forward-backward (FB) asymmetry dA_{FB}/ds is also used to determine the Wilson Coefficients. dA_{FB}/ds is defined as

$$dA_{\text{FB}}/ds = \left(\int_0^1 du \frac{d^2\Gamma}{duds} - \int_{-1}^0 du \frac{d^2\Gamma}{duds} \right) / \left(\int_0^1 du \frac{d^2\Gamma}{duds} + \int_{-1}^0 du \frac{d^2\Gamma}{duds} \right), \quad (2.17)$$

where s is defined as $s = M_{\ell^+\ell^-}^2$ and $u = \cos\theta$ is the cosine of the angle between momentum direction of B meson and positively charged lepton at the center of mass system. Figure 2.8 shows the shape of the FB asymmetry spectrum for the SM and other three cases with non-SM parameters. The distinctive features are the presence or not of a zero and global sign of the asymmetry. A rough indication of the FB asymmetry behavior is thus enough to rule out a large part of the parameter space that the current branching ratios cannot explore.

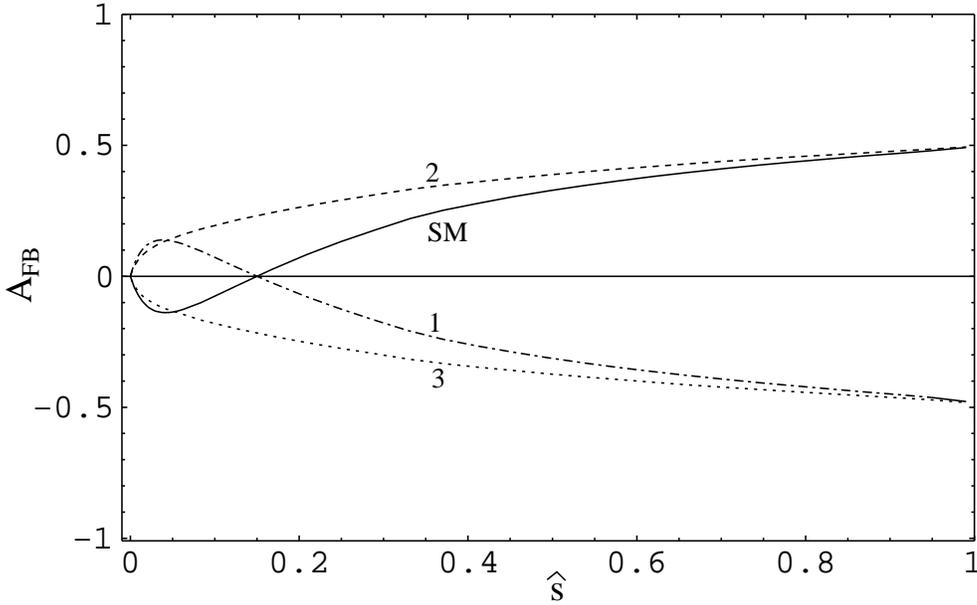


Figure 2.8: The differential forward-backward asymmetry for $B \rightarrow X_s\ell^+\ell^-$ decay theoretically calculated assuming various value of Wilson Coefficients. Line SM, 1, 2, 3 correspond to the points in Figure 2.10, $(\Delta C_9, \Delta C_{10}) = (0, 0), (1, 8.5), (-9, 9), (-8.5, -1)$, respectively.

Figure 2.9 shows the recent measurements of FB asymmetry in $B \rightarrow K^* \ell^+ \ell^-$ with 605 fb^{-1} data in Belle. The solid curve represents the SM prediction and the dotted curve represents the prediction with $C_7 = -C_7^{SM}$. This result shows the implication to favor $C_7 = -C_7^{SM}$, while the $B \rightarrow X_s \ell^+ \ell^-$ branching fraction measurements show the implication to favor the SM (see Section 2.3).

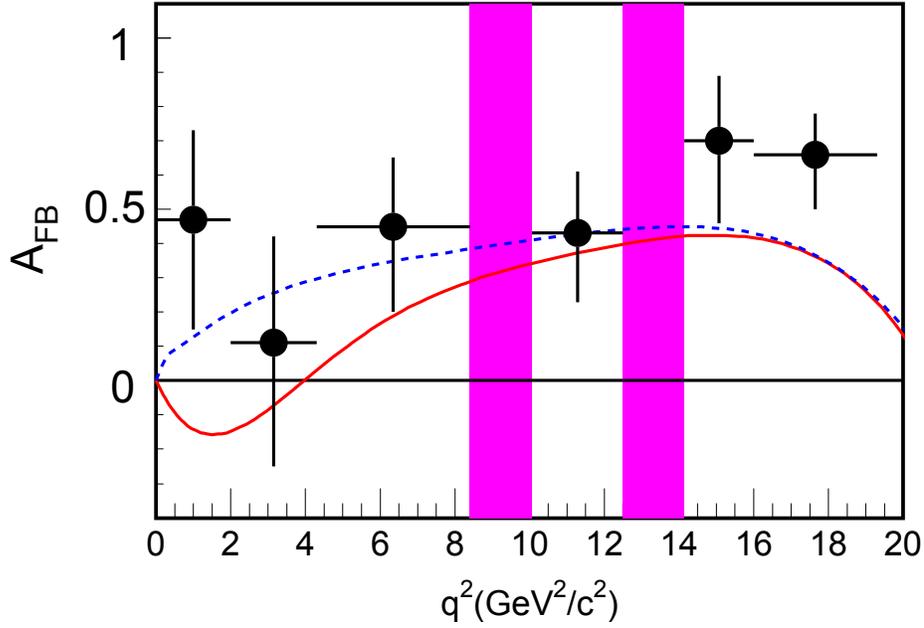


Figure 2.9: The A_{FB} measurements for $B \rightarrow K^* \ell^+ \ell^-$ decay with 605 fb^{-1} data in Belle. Points represent the measured A_{FB} and solid (dotted) curve represents the SM ($C_7 = -C_7^{SM}$) prediction. The two shaded regions are veto windows to reject $J/\psi K^*$ events.

2.5 New physics impact on Wilson Coefficients

The model-independent measurements of Wilson coefficients are a powerful test for possible candidates of new physics sources. Theorists have elaborately calculated the impact from new physics on the Wilson coefficients in various new physics models (See for instance [5]). We will show some examples of new physics model calculations in the following:

Minimal Flavor Violation (MFV) or gluino contributions on C_9 and C_{10} are considered to be small and hard to find, unfortunately [5].

The Extended MFV (EMFV) is estimated to give large contributions to C_9 and especially to C_{10} [5]. Possible C_9 and C_{10} values in case of EMFV are plotted in Figure 2.10 [2]. Note that the experimental measurements used in Figure 2.10 are older than those used in Figure 2.4.

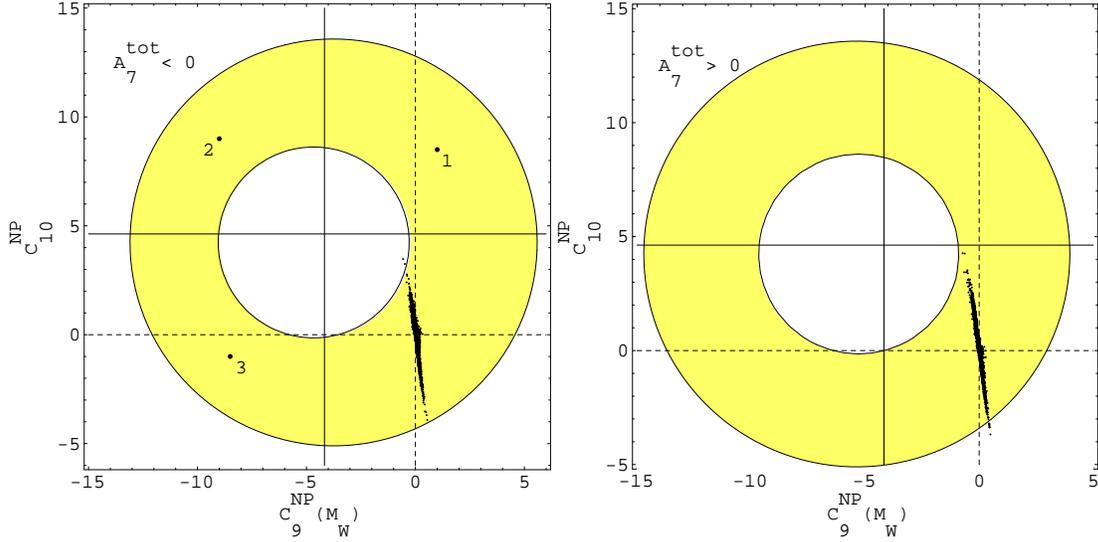


Figure 2.10: The experimental constraints on additive new physics contributions to $C_{9,10}$ at 90% C.L for the SM-like (upper plot) and opposite (lower plot) sign of C_7 . The dot at $(C_9, C_{10}) = (0, 0)$ indicates the SM case for $C_{9,10}$. The other dots represent the possible $C_{9,10}$ values in case of EMFV theory.

The contributions from two-Higgs doublet model (THDM) [8] and top quark two Higgs doublet model (T2HDM) are currently estimated [7]. Figure 2.11 shows the relation between the Wilson coefficients $C_{7,9}$ and the mass of charged-Higgs boson m_H calculated in T2HDM. This shows that our Wilson coefficients measurements can constrain m_H in T2HDM.

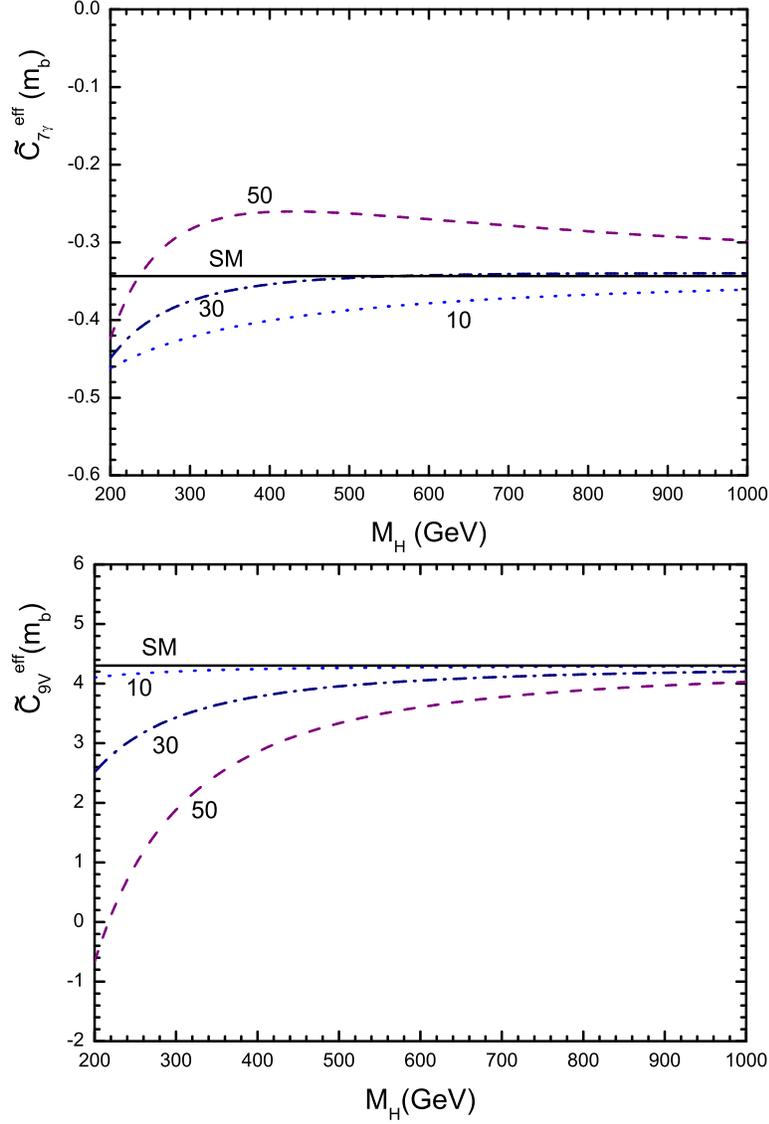


Figure 2.11: The upper figure shows the m_H dependence of the real part of the effective Wilson coefficient $C_{7\gamma}(m_b)$ in the SM (solid line) and T2HDM for $\sigma = 0^\circ$, and $\tan\theta = 10$ (dots curve), 30 (dot-dashed curve) and 50 (dashed curve), respectively. The lower figure shows the m_H dependence of the real part of the effective Wilson coefficient $C_{9V}(m_b)$ in the SM (solid line) and T2HDM for $\sigma = 0^\circ$, $\hat{s} = 0.2$, and $y = 10$ (dots curve), 30 (dot-dashed curve) and 50 (dashed curve), respectively.

Chapter 3

The Experimental Apparatus

In this chapter, 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 the High Energy Accelerator Research Organization (KEK), Tsukuba, Japan. It is one of two major B physics experiments in the world, together with the BaBar experiment, which is performed at the Stanford Linear Accelerator Center (SLAC), California, USA.

3.1 The KEKB accelerator

KEKB [24] 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. It consists of two 3 km-long storage rings, an 8 GeV electron ring (HER) and a 3.5 GeV positron ring (LER), and an injection linear accelerator. The two rings cross at one point, called the interaction point (IP). The linear accelerator accelerates electron and positron up to the required energy and injects them to HER and LER respectively.

The Belle detector is located around IP and detects and measures particles produced by the collisions. The center-of-mass energy is designed to be $\sqrt{s} = 10.58$ GeV, which corresponds to the mass of the $\Upsilon(4S)$ resonance. Due to the energy asymmetry between the electron and positron, the $\Upsilon(4S)$ is produced with a Lorentz boost of $(\beta\gamma)_{\Upsilon} = 0.425$.

The design luminosity of KEKB is $1.0 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, which corresponds to an approximate production rate of 10 pairs of B and anti- B pairs per second. The collider achieved its design luminosity in May 2003 and the luminosity has been continuously increasing since then.

In early 2004, a new method of operation at KEKB called “continuous injection mode” was successfully introduced which removes the dead time of the ordinary injection method. With this new method, data taking is no longer forced to stop every hour to replenish the beams in the storage rings. The best records up to Oct. 2006 were $1.6517 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ for the peak luminosity and 1.2315 fb^{-1} per day.

In early 2007, a new instrument called a “crab cavity” was installed. A diagram of the crab cavity is shown in Figure 3.2. In the original design of KEKB, the two beams do not collide head-on, but with a small crossing angle of ± 11 mrad. This

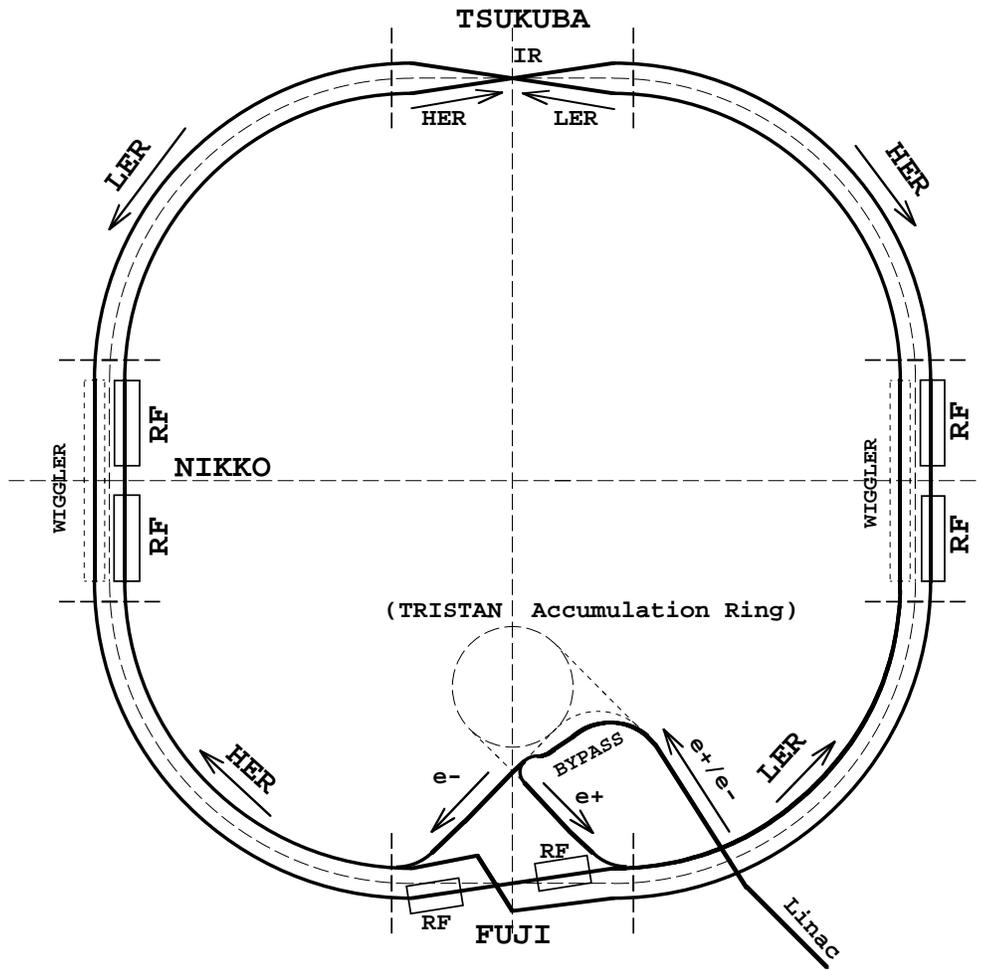


Figure 3.1: Schematic view of the layout of KEKB.

design has the advantage, at some cost on the luminosity, of simplifying the design of the interaction region and reducing the background in the detector. To cope with this luminosity loss, the bunches are tilted by a crab cavity installed in each ring, to collide the bunches with a maximum overlap as shown in Figure 3.3. At first the obtained peak luminosity was smaller than its 2006 values recorded without the crab cavity.

Now KEKB has more than doubled its design luminosity. The best records up to December 2009 are $2.1083 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ for the peak luminosity and 1.4794 fb^{-1} per day, which is the current world record.

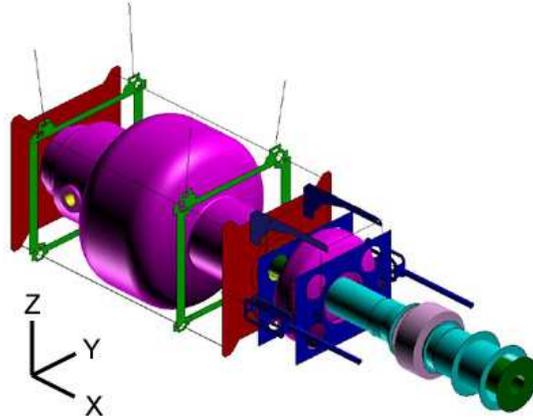


Figure 3.2: Drawing of KEKB crab cavity.

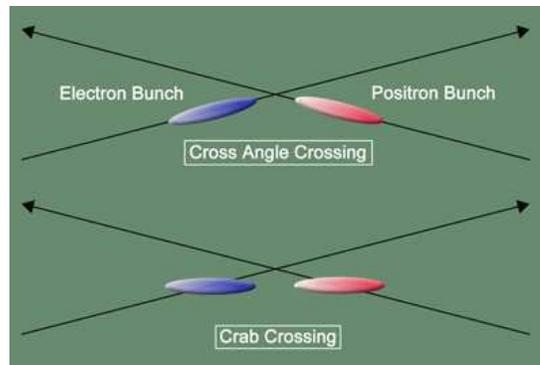


Figure 3.3: Bunch crossing schemes without (top) and with (bottom) use of crab cavities.

Figure 3.4 shows how the luminosity progressed over the ten years of KEKB operation. So far KEKB has provided Belle more than 1000 fb^{-1} of integrated luminosity.

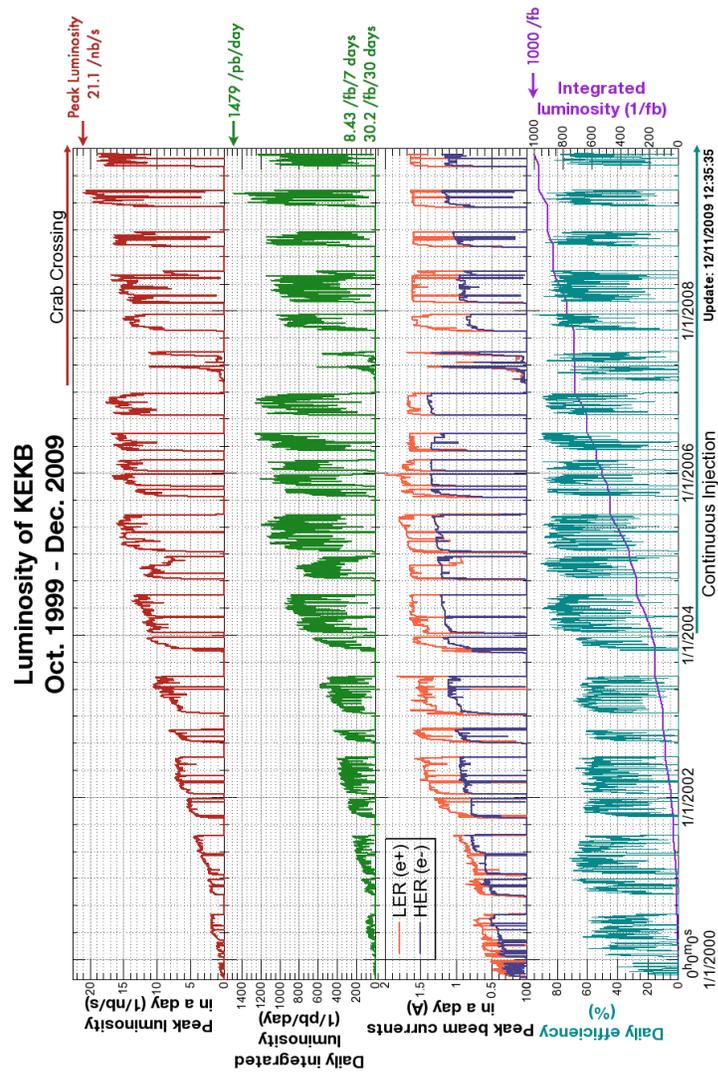


Figure 3.4: The KEKB luminosity history.

3.2 Belle Detector

The Belle detector [25] is a general-purpose 4π hermetic detector surrounding the IP. It consists of barrel, forward, and backward components. Figure 3.5 shows the configuration of the Belle detector.

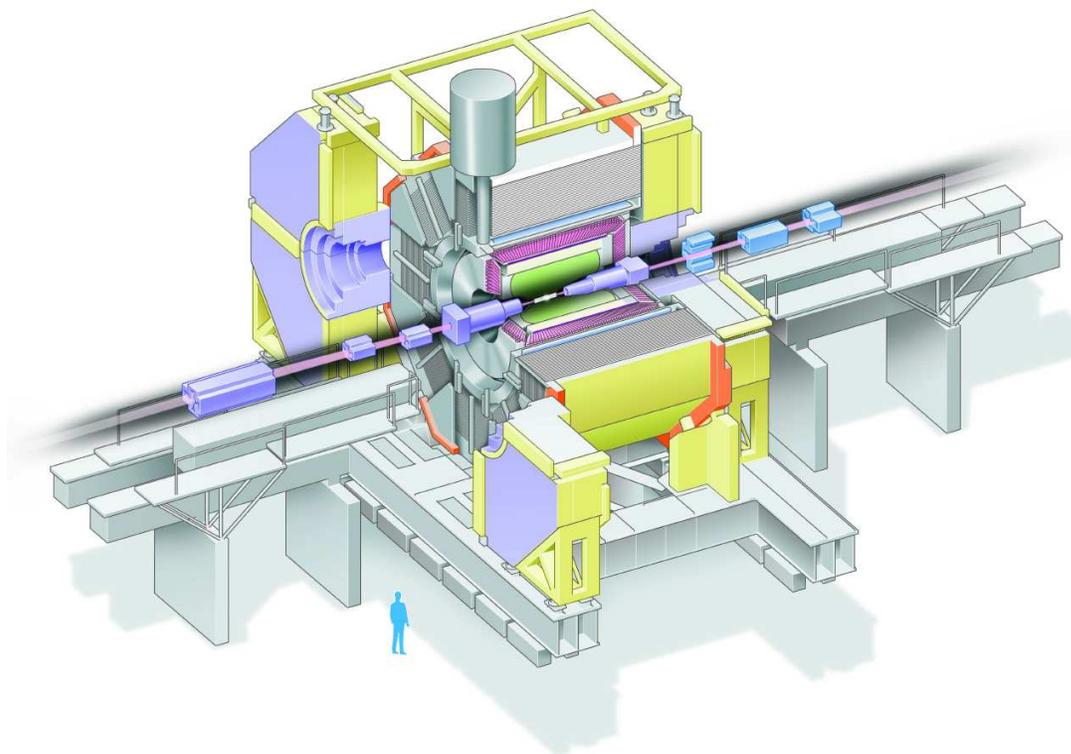


Figure 3.5: Overview of the Belle Detector.

The Belle Detector consists of a silicon vertex detector (SVD), a central drift chamber (CDC), a set of aerogel Cerenkov counters (ACC), a set of time-of-flight counters (TOF), an array of CsI(Tl) crystal calorimeters (ECL), a K_L and muon detector (KLM), and a pair of BGO crystal arrays called the extreme forward calorimeter (EFC).

In the Belle experiment, the final state particles of interest are $K^\pm, \pi^\pm, e^\pm, \mu^\pm, p^\pm$ (charged) and K_L^0, γ (neutral).

Precision tracking and vertex measurements of charge particles are provided by the CDC and SVD. The identifications of charged pions and kaons are based on the information from three sub-detectors: TOF, ACC, and the dE/dx measurement by CDC. Electromagnetic particles are detected in the ECL. Electron identification is based on a combination of the dE/dx measurements by the CDC, the response of the ACC, and the information of position, shape, and energy of the electromagnetic shower in the ECL. The SVD, CDC, TOF, ACC and ECL are located inside a superconducting solenoid of 1.7 m radius that maintains 1.5 T magnetic field. The KLM is the outermost

detector subsystem of the Belle detector. The EFC is placed on the surface of the cryostat of the final focusing quadrupole magnet (QCS) and provides coverage at small angles not covered by the other detectors.

Two inner detector configurations are used. 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} (DS-I) 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. A data sample corresponding to the integrated luminosity of 570 fb^{-1} (DS-II) was collected with this configuration. Performance parameters of the detectors are summarized in Table 3.1.

Table 3.1: Performance parameters for the Belle detector.

Detector	Type	Configuration	Readout	Performance
Beam-pipe (DS-I)	Beryllium double wall	Cylindrical, $r = 20$ mm 0.5/2.5/0.5 mm = Be/He/Be He gas cooled		
Beam-pipe (DS-II)	Beryllium double wall	Cylindrical, $r = 15$ mm 0.5/2.5/0.5 mm = Be/Paraffin/Be Paraffin gas cooled		
SVD (DS-I)	Double-sided Si strip	3-layers: 8/10/14 ladders Strip pitch: $25(p)/50(n)$ μm	$\phi : 40.96$ k, $z : 40.96$ k	$\sigma(\Delta z) \sim 80$ μm for $B_z \rightarrow (\rho\pi)^0$
SVD (DS-II)	Double-sided Si strip	4-layers: 6/12/18/18 ladders Strip pitch: $75(p)/50(n)$ μm (layer 1-3), $73(p)/65(n)$ μm (layer 4)	$\phi : 55.296$ k, $z : 55.296$ k	$\sigma(\Delta z) \sim 70$ μm for $B_z \rightarrow (\rho\pi)^0$
CDC (DS-I)	Small cell drift chamber	Anode: 50 layers, Cathode: 3 layers $r = 8.3 - 87.4$ cm, $-78.72 \leq z \leq 158.77$ cm	Anode: 8.4 k Cathode: 1.8 k	$\sigma_{r\phi} = 130$ μm $\sigma_z = 200 \sim 1400$ μm $\sigma_{p_t}/p_t = 0.3\% \sqrt{p_t^2 + 1}$ $\sigma_{dE/dx} = 8\%$
CDC (DS-II)	Small cell drift chamber	Anode: 49 layers, Cathode: no layer $r = 10.4 - 87.4$ cm, $-78.72 \leq z \leq 158.77$ cm	Anode: 8.5 k	
ACC	Silica aerogel ($n = 1.01 - 1.03$)	$\sim 12 \times 12 \times 12$ cm^3 blocks 960 barrel / 228 endcap FM-PMT readout	1788	$N_{p.e.} \geq 6$ K/π separation at $1.2 < p < 3.5$ GeV
TOF/TSC	Plastic scintillator	128/64 ϕ segmentation $r = 120$ cm, 3-m long	128 \times 2 / 64	$\sigma_t = 100$ ps K/π separation up to 1.2 GeV
ECL	CsI	$r = 125 - 162$ cm (Barrel), $z = -102$ cm (Backward Endcap), $z = +196$ cm (Forward Endcap)	6624 (B), 960 (BE), 1152 (FE)	$\sigma_E/E = 1.3\%/\sqrt{E}$ $\sigma_{\text{pos}} = 0.5$ cm/ \sqrt{E} E in GeV
Magnet	Super-conducting	Inner radius = 170 cm		$B = 1.5$ T
KLM	Resistive plate counters	14 layers (5 cm Fe + 4 cm gap), 2 RPCs for each gap θ and ϕ strips	θ : 16 k, ϕ : 16 k	$\Delta\phi = \Delta\theta = 30$ mrad for K_L $\sigma_t = 1$ ns $\sim 1\%$ hadron fakes
EFC	BGO	$\sim 12 \times 12 \times 12$ cm^3 blocks Photodiode readout Segmentation: 32 (ϕ), 5 (θ)	100 \times 2	Energy resolution (rms): 7.3% at 8 GeV, 5.8% at 8 GeV

3.2.1 Silicon Vertex Detector (SVD)

The SVD is placed at the most inner part of the BELLE detector (just outside of the beam pipe) and provides very precise position measurements for B meson vertex reconstruction, which is crucially important for the time-evolution study of B mesons. In time-dependent CP asymmetry analysis, it is essential to obtain information of the difference between the flight lengths of the two B mesons in the z direction, where z axis is defined as the opposite of the positron beam direction. The SVD also provides the information to reconstruct D and τ decay vertices.

Figure 3.6 shows the side and end views of the SVD for DS-I (SVD1 [26]). SVD1 consists of three concentric cylindrical layers arranged to cover $23^\circ < \theta < 139^\circ$, where θ is defined as the polar angle from the z axis. Its coverage corresponds to 86% of full solid angle. The radii of three layers are 30.0, 45.5, and 60.5 mm (the radius of beam pipe is 2.0 cm) and the three layers are made up of 8, 10 and 14 ladders, from innermost to outermost. Each ladder consists of double-sided silicon strip detectors (DSSDs). SVD1 consists of 102 DSSDs in total.

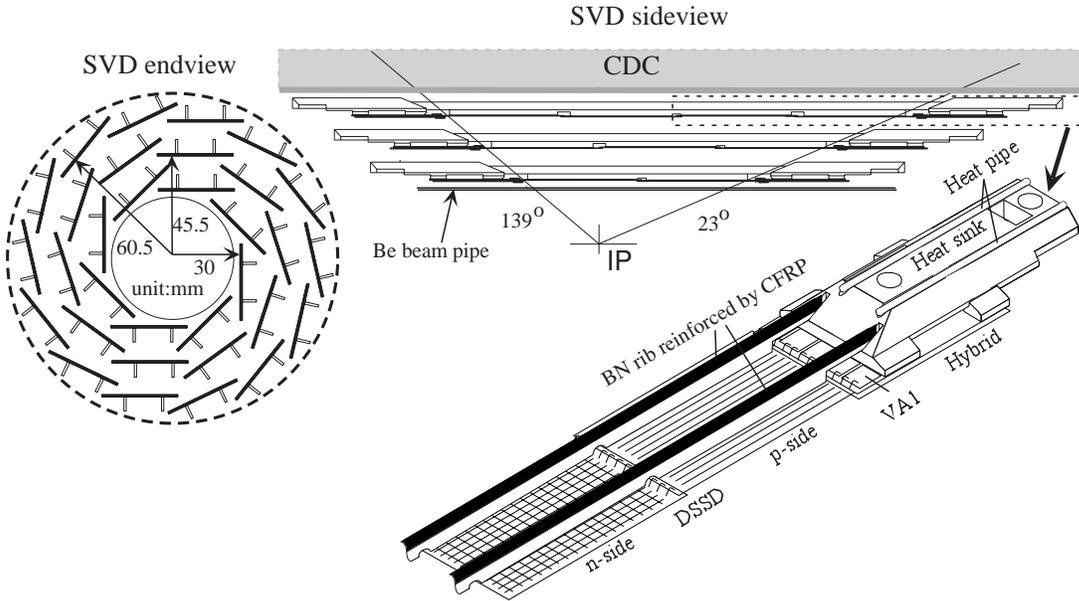


Figure 3.6: Detector configuration of SVD1.

In the summer of 2003, a new vertex detector (SVD2 [27]) was installed. Figure 3.7 schematically shows the configuration of SVD2. SVD2 consists of four concentric cylindrical layers and the polar angle acceptance is improved to cover $17^\circ < \theta < 150^\circ$, which is the same as CDC and corresponds to the 92% of the full solid angle. The radii of four layers are 20.0, 43.5, 70.0 and 88.0 mm (the radius of beam pipe is 1.5 cm) and the four layers are made up of 6, 12, 18 and 18 ladders, from innermost to outermost. SVD2 consists of 138 DSSDs in total.

A DSSD is basically a depleted pn junction 3.8. A charged particle passing through

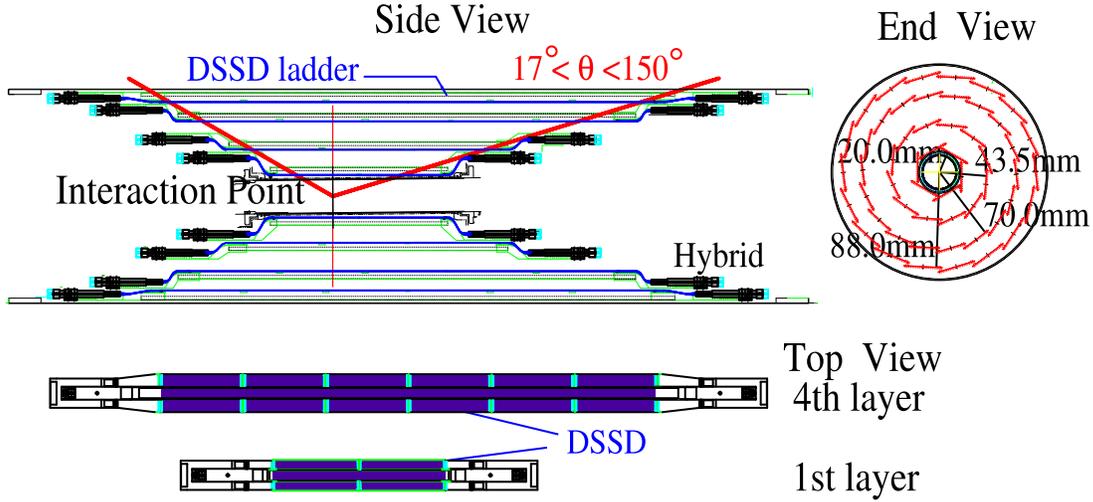


Figure 3.7: Detector configuration of SVD2.

the junction liberates electrons from the valence band into the conduction band creating electron-hole (e^-h^+) 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. The n^+ strips are aligned perpendicularly to the beam axis and measure z . DSSDs are originally designed for the DELPHI microvertex detector and fabricated by Hamamatsu Photonics.

The impact parameter resolution for reconstructed tracks is measured with SVD1 and SVD2, as a function of the track momentum p (measured in GeV/c), the polar angle θ . As shown in Figure 3.9, the impact parameter resolutions of SVD2 is better than those of SVD1, mainly owing to the smaller radius of the innermost layer. The impact parameter resolutions measured with cosmic ray events are

$$\sigma_{r\phi}(\mu\text{m}) = 19.2 \oplus 54.0/\tilde{p}, \sigma_z(\mu\text{m}) = 42.2 \oplus 44.3/\tilde{p}$$

for SVD1 and

$$\sigma_{r\phi}(\mu\text{m}) = 21.9 \oplus 35.5/\tilde{p}, \sigma_z(\mu\text{m}) = 27.8 \oplus 31.9/\tilde{p}$$

for SVD2, where \tilde{p} is defined as $\tilde{p} = p\beta \sin^{3/2} \theta$ for r - ϕ side and $\tilde{p} = p\beta \sin^{5/2} \theta$ for z side.

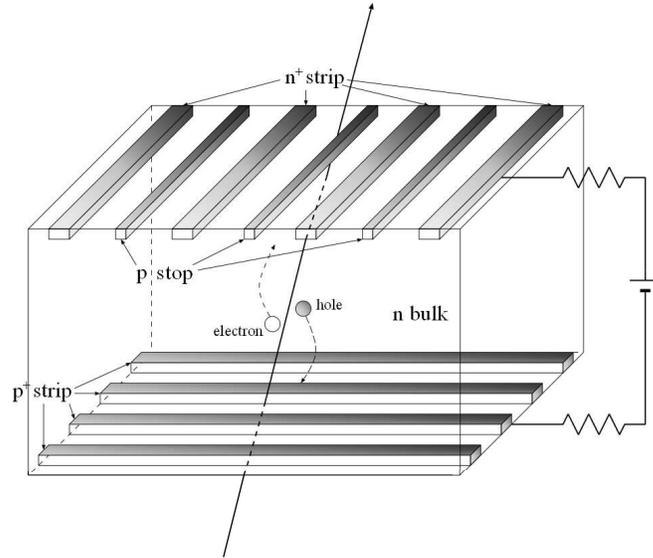


Figure 3.8: Schematic drawing of the DSSD.

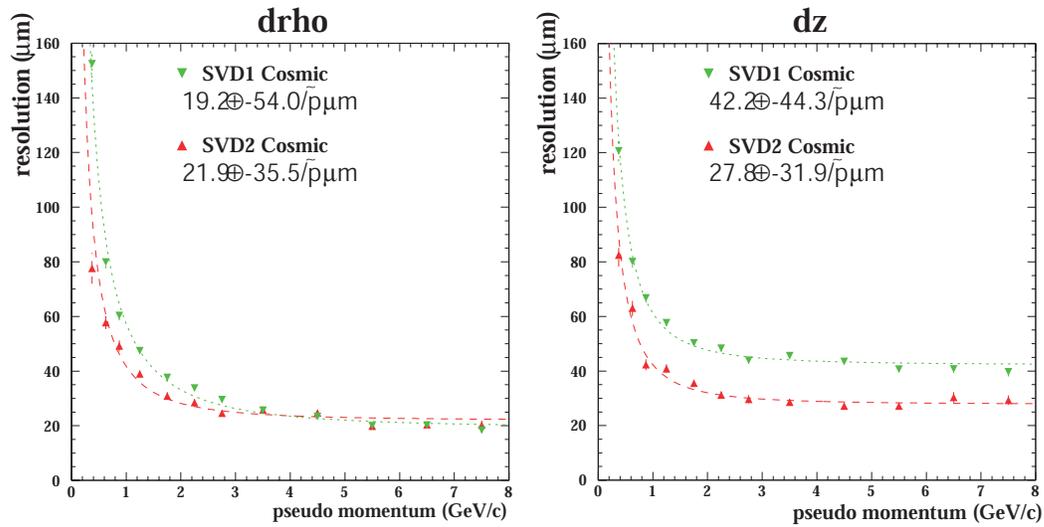


Figure 3.9: Comparison of the impact parameter resolutions in the directions of r - ϕ (left) and z (right) measured with cosmic ray data.

3.2.2 Central Drift Chamber (CDC)

CDC[28] is one of the most important sub-detectors in the Belle detectors and plays a core role in the tracking of the charged particles. CDC is placed in a 1.5 T magnetic field produced by the solenoidal coil, therefore a charged track follows a helicoidal trajectory in CDC. CDC provides following three measurements: it reconstructs effectively the charged particle tracks, it measures their momenta precisely and it helps with particle identification by measuring the energy loss (dE/dx). It also provides important information for the trigger system.

The structure of the CDC used to collect DS-I is shown in Figure 3.10. It is asymmetric in the z , providing an angular coverage of $17^\circ \leq \theta \leq 150^\circ$, which corresponds to 92% of the full solid angle. The longest wires are 2400 mm long. The outer radius is 874 mm and the inner one is extended down to 83 mm. The CDC is a small-cell cylindrical drift chamber with 50 layers of anode wires, which consist of 32 axial- and 18 stereo-wire layers, and three cathode strip layers. Axial wires are parallel to the z axis, while stereo wires slant to the z axis to provide z position information. Stereo wires also provide a highly efficient fast z -trigger combined with the cathode strips. The CDC has a total of 8400 drift cells. At the inner layers of the CDC, three cathode-strip layers are made for higher precision z measurement at the position where the particles enter the CDC, which is especially beneficial for the trigger function.

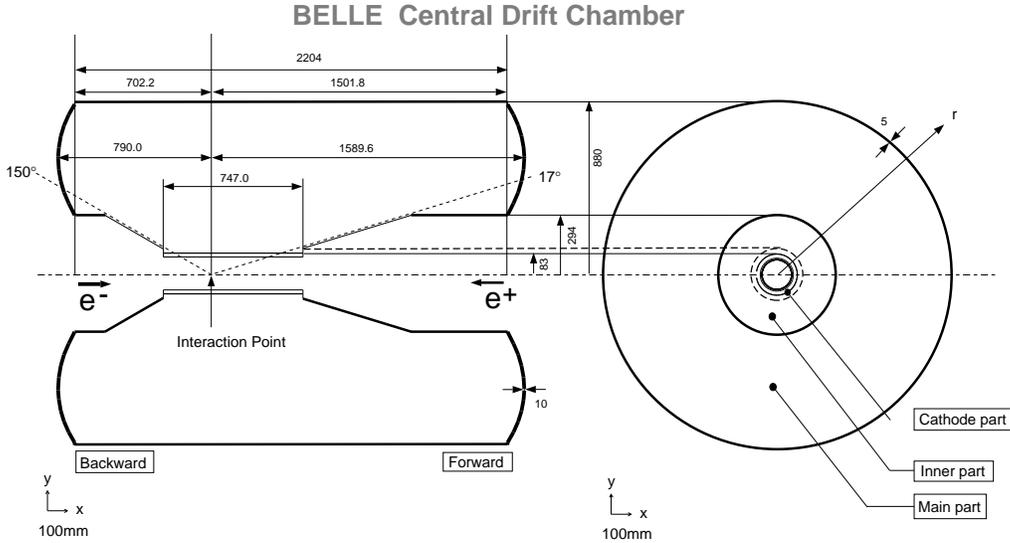


Figure 3.10: Overview of the CDC structure.

In the summer of 2003, the inner part structure of the CDC was jointly modified with the upgrade of the SVD. The three inner layers with cathode strips were removed to make the space for the upgraded SVD with larger radius. Instead, we have installed two layers of smaller cells, which we call small-cell CDC (sCDC). The inner radius after the modification is 104 mm, while the other geometry is unchanged. The sCDC maintains the performance of the Level-1 trigger by keeping the number of inner layers

used for the trigger to be five, which was six before the modification. In addition, we exploit the small drift time due to the smaller cell to provide additional information for the Level-0 trigger logic required by the SVD, which was provided by the information from the TOF alone before the upgrade.

A low- Z gas mixture, consisting of 50% He and 50% ethane (C_2H_6), is used to minimize multiple Coulomb scattering and achieve a good momentum resolution, especially for low momentum particles. Since low- Z gases have a smaller photo-electric cross section than argon-based gases, they have the additional advantage of reduced background from synchrotron radiation. Even though the gas mixture has a low- Z , a good dE/dx resolution is obtained by the large ethane component.

The measured spatial resolution in the r - ϕ direction¹ is $\sim 120 - 150 \mu\text{m}$ with a dependence on the incident angles and layers. The p_t resolution obtained by the study using cosmic ray is

$$\frac{\sigma_{p_t}}{p_t} (\%) = \sqrt{(0.28p_t)^2 + (0.35/\beta)^2} \quad (p_t \text{ in GeV}/c) \quad (3.1)$$

without the SVD information, and

$$\frac{\sigma_{p_t}}{p_t} (\%) = \sqrt{(0.19p_t)^2 + (0.30/\beta)^2} \quad (p_t \text{ in GeV}/c) \quad (3.2)$$

with the SVD information. (Fig 3.11)

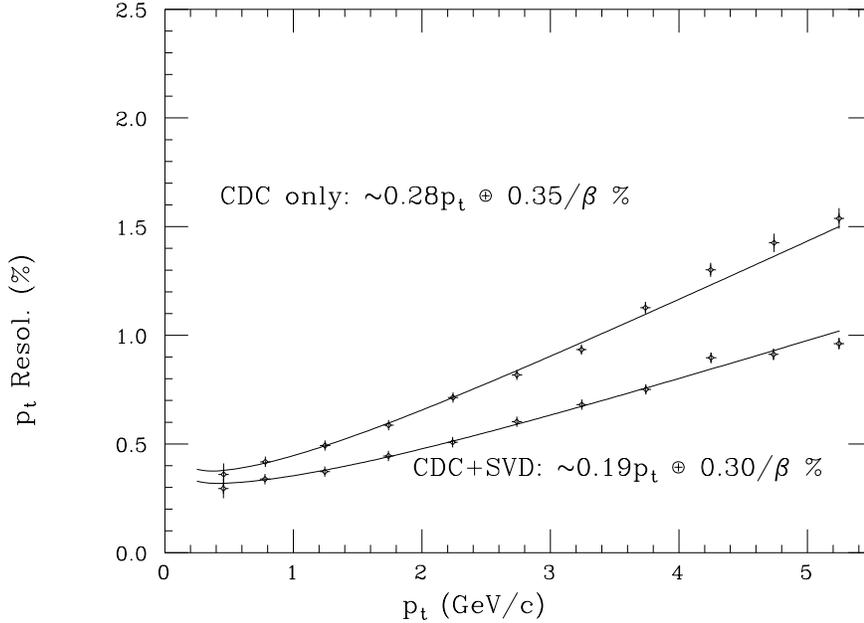


Figure 3.11: p_t resolution studied using cosmic rays.

¹We define the “ r - ϕ direction” as the axis that is perpendicular to the z direction on the plane of each DSSD.

The dE/dx measurement in the CDC provides information which can be used to distinguish particle species, since dE/dx , described by the Bethe-Bloch formula, mainly depends on the velocity β and particles with equal momenta but different masses can be distinguished. A scatter plot of the measured dE/dx and particle momentum is shown in Figure 3.12, 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 is measured to be 7.8% in the momentum range from 0.4 to 0.6 GeV/c.

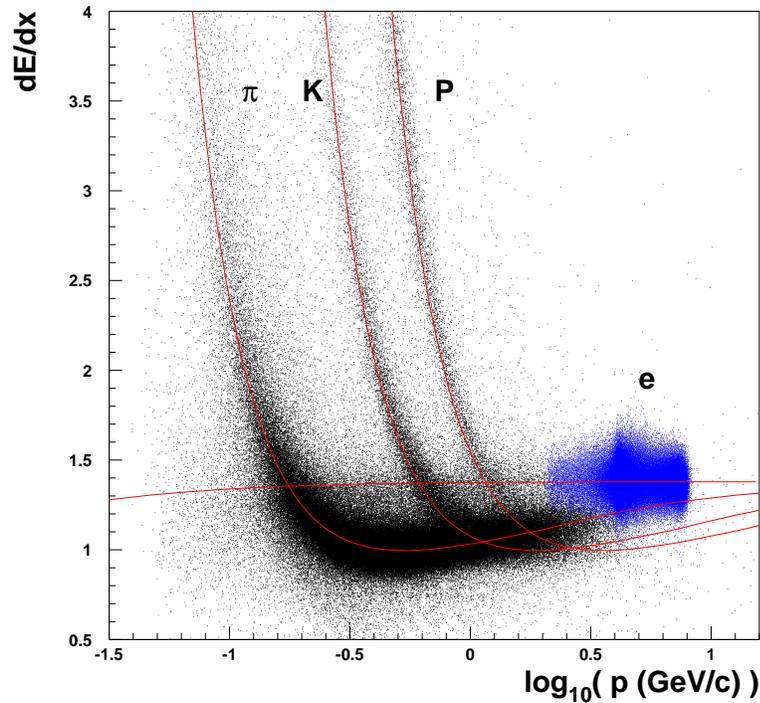


Figure 3.12: Charged track momenta vs. dE/dx observed in collision data.

3.2.3 Aerogel Cerenkov Counter System (ACC)

Particle identification, particularly the identification of π^\pm against K^\pm , plays an important role in many measurements of B decays. An array of silica-aerogel threshold Cerenkov counters is selected as a part of the Belle particle identification system. It covers the momentum range between 1.5 GeV/c and 3.5 GeV/c with respect to the K^\pm/π^\pm separation, extending the coverage beyond the reach of dE/dx measurements by CDC and time-of-flight measurements by TOF.

The Cerenkov radiations are emitted in case of $n > \frac{1}{\beta} = \sqrt{1 + (\frac{m}{p})^2}$, where β , m and p are the velocity, mass and momentum of the charged particle, respectively, and n is the refractive index of the matter through which the particle is passing. Since $m_{K^\pm} > m_{\pi^\pm}$, there is a momentum region where pions emit Cerenkov light but kaons and heavier particles do not. Thus, one can identify pions against kaons by choosing the proper refractive index n for the momentum region of interest.

The configuration of the ACC [30] is shown in Figure 3.13. The 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. A typical ACC module consists of five aerogel tiles stacked in a thin (0.2 mm thick) aluminum box with an approximate size of $12 \times 12 \times 12 \text{ cm}^3$.

To detect the Cerenkov 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 [31].

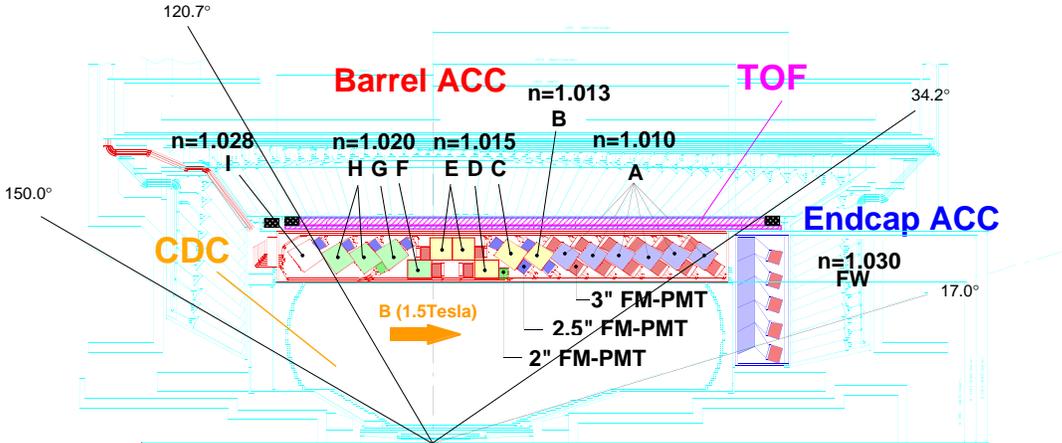


Figure 3.13: Arrangement of the ACC in Belle detector.

In order to obtain a good K^\pm/π^\pm separation for the required kinematic range, the refractive indices of aerogels are selected to be between 1.01 and 1.03, depending on their polar angle region. In barrel part, they are optimized for the momentum corresponding to the daughter particles of B meson two-body decays. In end-cap

part, they are optimized for the momentum of K^\pm from B cascade decays, which is advantageous in B flavor tagging.

The performance of the ACC is confirmed using the decay process of $D^{*+} \rightarrow \pi^+ D^0 (\rightarrow K^- \pi^+)$, where the identification of the charged particles from the D^0 decay can be determined without using the ACC information, by the charge of π from the D^{*+} decay. Figure 3.14 shows the number of photo-electron distribution of π^\pm and K^\pm in this decay process, where π^\pm is well separated from K^\pm , being consistent with MC.

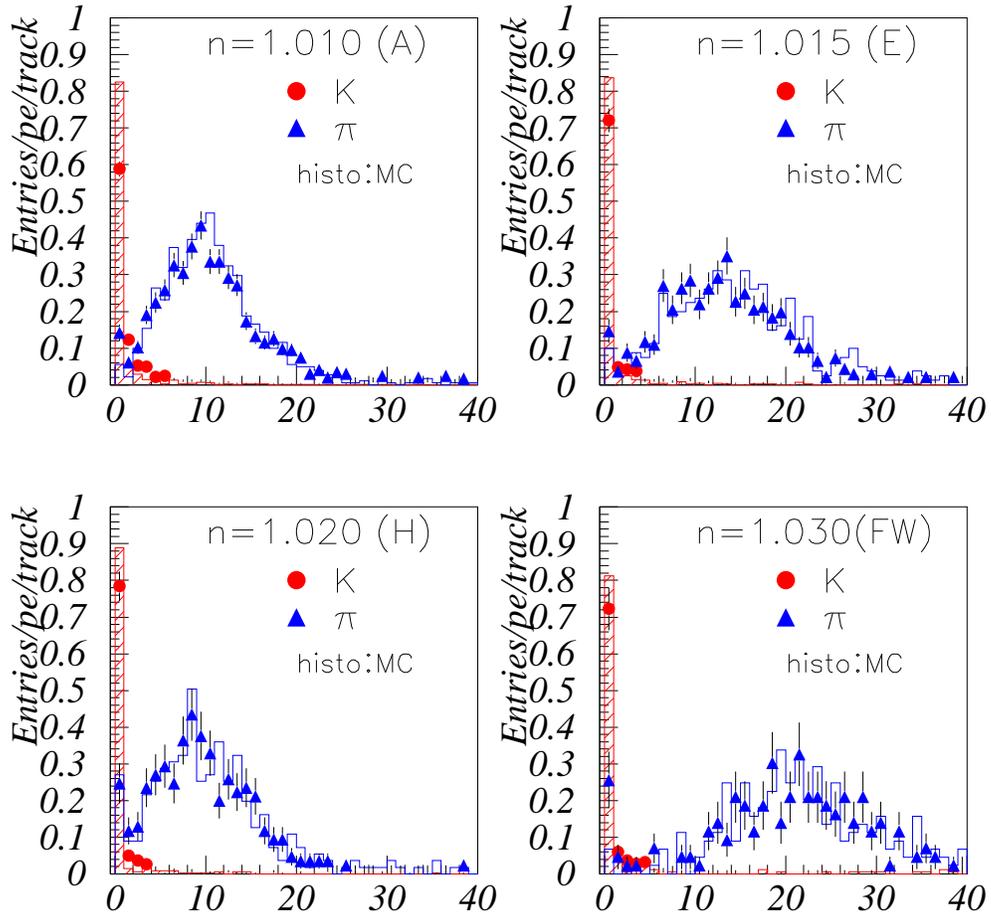


Figure 3.14: ACC number of photo-electron distribution for π^\pm and K^\pm from D^{*+} decays. Each plot corresponds to the different set of modules with a different refractive index.

3.2.4 Time-of-Flight Counter (TOF)

The Time-of-Flight Counter (TOF) gives particle identification information to distinguish charged kaons from pions in the low momentum region, below 1.2 GeV/c. The TOF also provides fast timing signals for the trigger system, together with thin trigger scintillation counters (TSC).

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 relation:

$$T = \frac{L}{c\beta} = \frac{L}{c} \sqrt{1 + \left(\frac{m}{p}\right)^2},$$

where L is a length of the flight.

For example, when $L = 120$ cm and $p = 1.2$ GeV/c, $T = 4.0$ ns for a pion ($m_{\pi^\pm} = 140$ MeV/c²), while $T = 4.3$ for a kaon ($m_{K^\pm} = 494$ MeV/c²). Thus, the difference of T between pions and kaons is ~ 300 ps and K^\pm/π^\pm separation with 3σ significance is obtained with the time resolution of 100 ps.

The Belle TOF system [32] consists of 64 modules and each module consists of two trapezoidally shaped TOF counters and one TSC counter (128 TOFs and 64 TSCs in total). The dimensions of a module are given in Figure 3.15. The TOF/TSC modules are located at a radius of 1.2 m from the IP covering a polar angle range from 34° to 120°. Fine-mesh photomultiplier tubes (FM-PMTs) are attached to both ends of the TOF counter with air gaps of 0.1 mm. As for the TSC counters, the FM-PMTs are glued to the light guides at the backward ends.

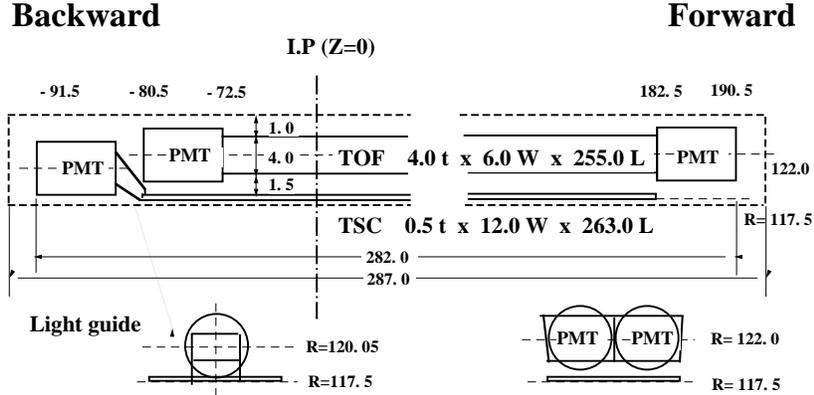


Figure 3.15: Dimensions of a TOF/TSC module.

Figure 3.16 shows time resolutions as a function of z for forward and backward PMTs and for the weighted average. The resolution for the weighted average is about 100 ps with small z dependence. Figure 3.17 shows the mass distribution for each track in hadron events, calculated using the momentum of the particle determined from the CDC track fit assuming muon mass. Clear peaks corresponding to pions, kaons and protons can be seen. The data points are in good agreement with an MC expectation (histogram) obtained assuming the time resolution of TOF $\sigma_{\text{TOF}} = 100$ ps.

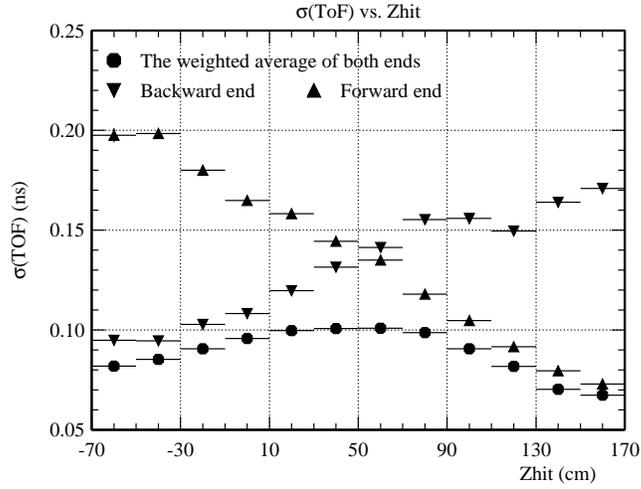


Figure 3.16: Time resolution of the TOF for $e^+e^- \rightarrow \mu^+\mu^-$ events.

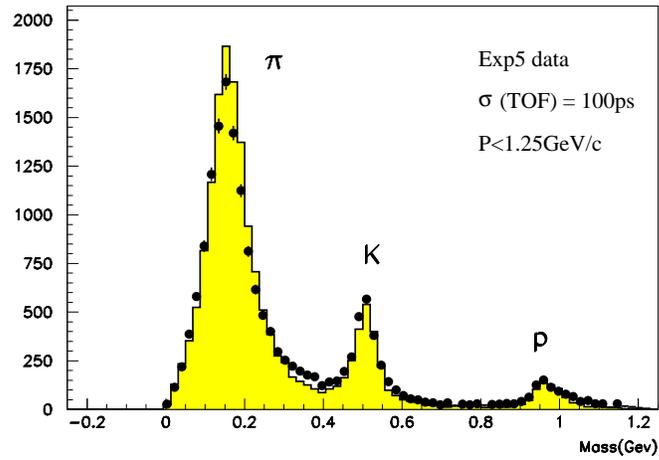


Figure 3.17: Mass distribution from TOF measurements for particles with momenta below $1.2 \text{ GeV}/c$. The histogram corresponds to MC distribution.

3.2.5 Electromagnetic Calorimeter (ECL)

The main purpose of the ECL is to detect and measure photons with high efficiency and good resolutions in energy and position. Photons of interest in Belle are mostly end products of cascade decays and have relatively low energies (below 500 MeV). However, photons in two-body decays such as $B \rightarrow K^*\gamma$ and $B_s \rightarrow \pi^0\pi^0$, are also important and have energies up to 4 GeV. Therefore the ECL is required to cover wide energy range of photon detection.

The ECL is also a main sub-detector for electron identification. When an electron or photon hits a crystal in ECL, its energy is deposited in electromagnetic showers produced by Bremsstrahlung and pair production. While other charged particles only deposit a small amount of energy by dE/dx ionization. Therefore, the ratio of the cluster energy measured by the ECL to the momentum of the charged track momentum as measured by the CDC, E/p , is close to unity for electrons and smaller than unity for other particles.

The overall configuration of the Belle calorimeter system, the ECL [33], is shown in Figure 3.18. The ECL is an array of 8736 tower-shaped CsI(Tl) crystals in total. The ECL consists of three sections: the forward endcap section consists of 1152 crystals and covers $12.4^\circ \leq \theta \leq 31.4^\circ$; the barrel section consists of 6642 crystals and covers $32.2^\circ \leq \theta \leq 128.7^\circ$; and the backward endcap section consists of 960 crystals and covers $130.7^\circ \leq \theta \leq 155.1^\circ$.

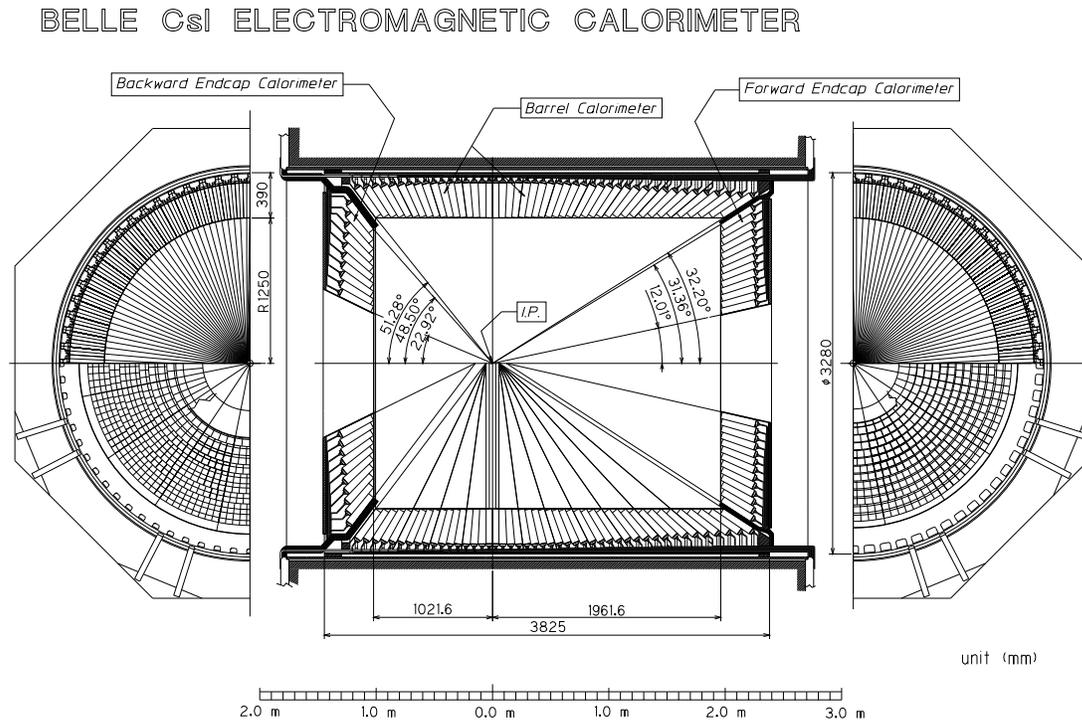


Figure 3.18: Configuration of ECL.

The size of a crystal in the θ - ϕ direction is determined so that a crystal contains approximately 80% of the total energy deposit by a photon injected at the center of its front face. The typical dimension of a crystal is 55 mm \times 55 mm at front face and 65 mm \times 65 mm at rear face for the barrel part. The thickness in r direction is 30 cm, which corresponds to 16.2 X_0 (radiation length). This length is long enough to avoid deterioration of the energy resolution at high energies due to the shower leakage from the rear of the counter.

The energy dependence of the average position resolution estimated by MC and can be approximated by

$$\sigma \text{ (mm)} = 0.27 + \frac{3.4}{\sqrt{E}} + \frac{1.8}{\sqrt[4]{E}} \quad (E \text{ in GeV}),$$

which is shown in Figure 3.19. As can be seen in the figure, the estimation is well consistent with the result of the beam test [33] in the measured energy region.

The energy resolution given by the beam test is

$$\frac{\sigma_E}{E} (\%) = \sqrt{\left(\frac{0.066}{E}\right)^2 + \left(\frac{0.81}{\sqrt[4]{E}}\right)^2} + 1.34^2 \quad (E \text{ in GeV}). \quad (3.3)$$

This is consistent with the collision data calibrated by $e^+e^- \rightarrow e^+e^-$ (Bhabha) events, where the energy resolutions are 1.5%, 1.9%, and 2.5% for the barrel, forward, and backward ECL, respectively. (Figure 3.20)

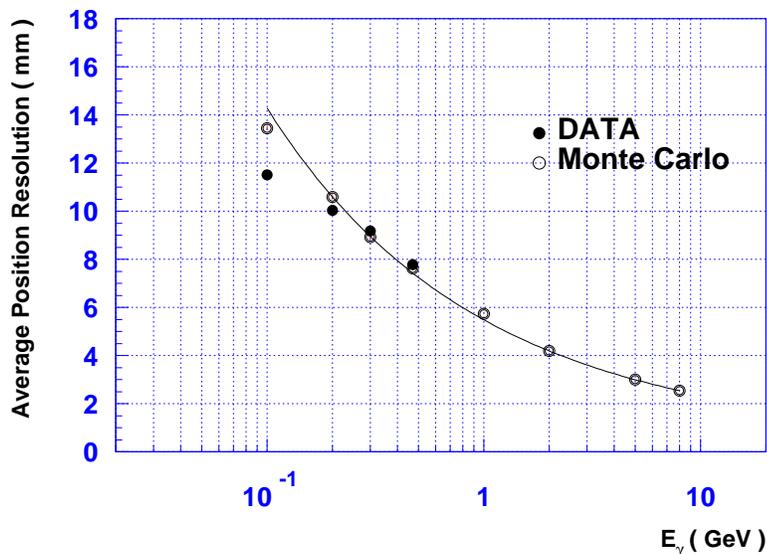


Figure 3.19: Energy dependence of the average position resolution. The solid curve is the result of the fit to the MC.

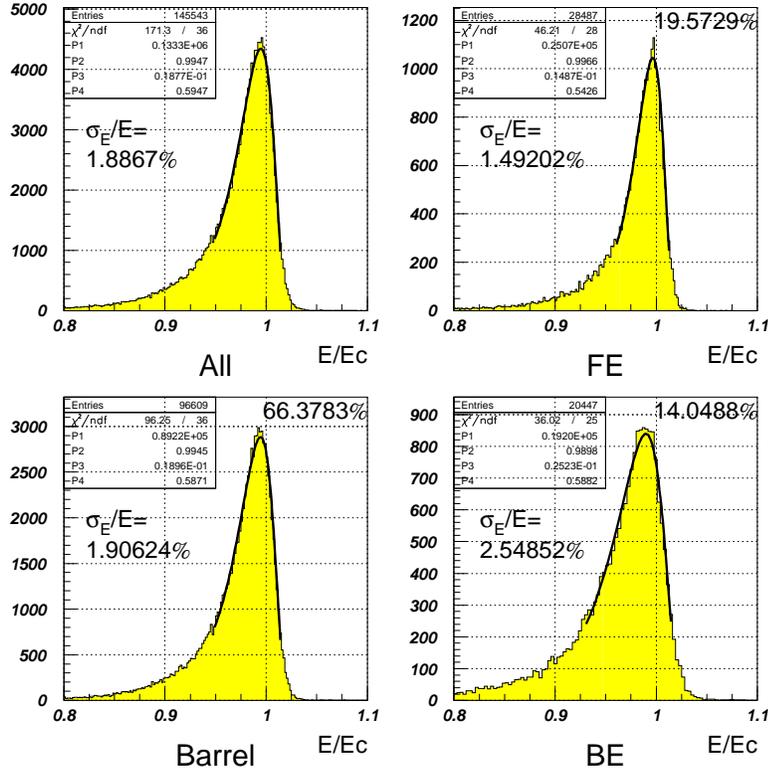


Figure 3.20: Energy resolutions calibrated and measured with $e^+e^- \rightarrow e^+e^-$ (Bhabha) events. The plots correspond to the overall average (top left) and each of the barrel (top right), forward end-cap (bottom left), and backward end-cap (bottom right) sections.

3.2.6 K_L^0 and Muon Detection System (KLM)

The KLM [34] is designed to identify K_L^0 and muon with high efficiency over a broad momentum range greater than 600 MeV/c.

Figure 3.21 shows the cross-section of a superlayer for the barrel region. The KLM consists of alternating layers of glass electrode resistive-plate counters (RPCs) and 4.7 cm-thick iron plates. The barrel section of KLM covers $45^\circ < \theta < 125^\circ$ in the polar angle and the endcap sections in the forward and backward directions extend this range to $20^\circ < \theta < 155^\circ$. There are 15 detector layers and 14 iron layers in the octagonal barrel region and 14 detector layers in each of the forward and backward end-caps. The iron layers also serve as a return yoke for the magnetic flux provided by the superconducting solenoid.

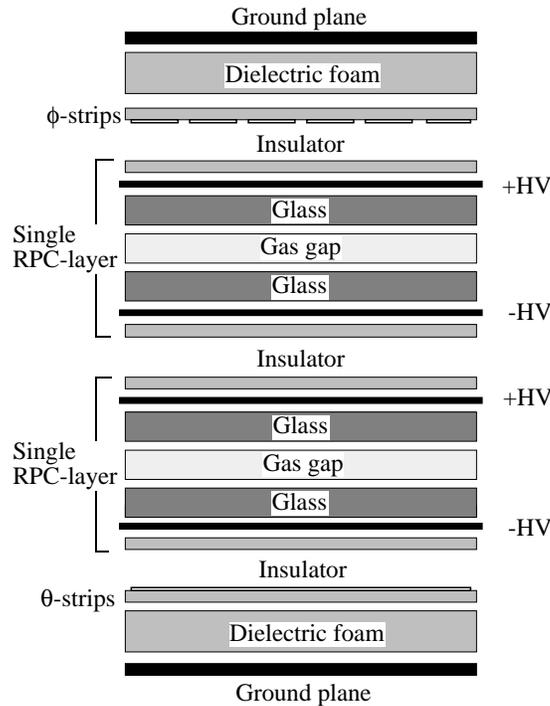


Figure 3.21: Cross section of a KLM superlayer.

Hadrons interacting with the iron plates produce a shower of ionizing particles that are detected by the RPC layers. The result is a cluster of hits deposited in the KLM. A K_L^0 candidate can be distinguished from another charged hadron because it will not leave an associated track in the CDC. A muon, on the other hand, does leave a charged track in the CDC. However, muons can still be distinguished from charged and neutral hadrons because they are not affected by the strong interaction. Hadrons are more rapidly absorbed and deflected by strong interactions with iron, so wide clusters are observed. They are also stopped within a few layers of iron. Muons only experience electromagnetic multiple scattering and energy loss, so their clusters tend to be thinner and they have far greater penetration depth.

Figure 3.22 shows a histogram of the difference between the direction of the neutral cluster (K_L^0 candidate) detected by KLM and the missing momentum direction in data. The missing momentum vector is calculated using all the other measured particles in the event. The histogram shows a clear peak where the direction of the neutral cluster measured in KLM is consistent with the missing momentum in the event, indicating correct detection of K_L^0 . The non-peaking flat-distributed component in the histogram is mainly due to undetected neutrinos and particles escaping the detector acceptance.

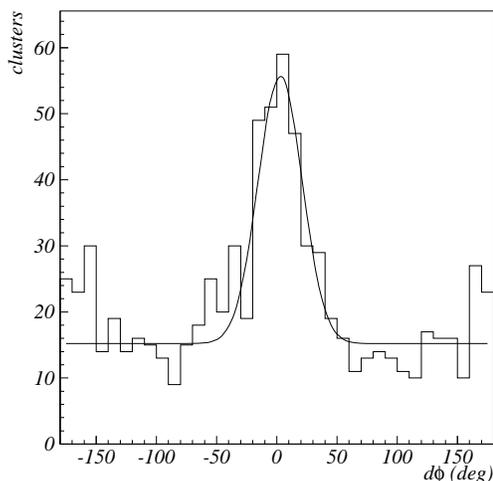


Figure 3.22: Distribution of the difference between directions of the neutral cluster detected by KLM and the missing momentum.

3.2.7 Extreme Forward Calorimeter (EFC)

The EFC [35] extends the polar angle coverage by the ECL, which is $17^\circ < \theta < 150^\circ$. The EFC covers the angular range from 6.4° to 11.5° in the forward direction and 163.3° to 171.2° in the backward direction. The EFC is also required to function as a beam mask to reduce backgrounds for the CDC. In addition, the EFC is used for a beam monitor for the KEKB control and a luminosity monitor for the Belle experiment. Since the EFC is placed in the very high radiation level area around the beam pipe near IP, it is required to be radiation hard. Thus, a radiation-hard BGO (Bismuth Germanate, $\text{Bi}_4\text{Ge}_3\text{O}_{12}$) crystal calorimeter is used for the EFC. The detector is segmented into 32 in ϕ and 5 in θ for both the forward and backward detectors. The radiation lengths of the forward and backward crystals are 12 and 11, respectively.

The energy sum spectra for $e^+e^- \rightarrow e^+e^-$ (Bhabha) events show a correlation between the forward and backward EFC detectors as expected. A clear peak at 8 GeV (3.5 GeV) with a resolution of 7.3% (5.8%) in rms is seen for the forward (backward) EFC.

3.2.8 Trigger system

The Belle trigger system [36] consists of the Level-1 hardware trigger and the Level-3 software trigger. Figure 3.24 shows the schematic view of the Belle Level-1 trigger system. The Level-1.5 hardware trigger is a trigger used in conjunction with SVD2 (see [37] for details). The Level-3 trigger is designed to be implemented in the on-line computer farm.

The trigger system provides the trigger signal with the fixed time of $2.2 \mu\text{s}$ after the event occurrence. The Belle trigger system, including most of the sub-detector trigger systems, is operated in a pipelined manner with clocks synchronized to the KEKB accelerator RF signal. The trigger efficiency for the hadronic events is more than 99%. The event rates for physics processes and background's are $\sim 100 \text{ Hz}$ and $\sim 300 \text{ Hz}$, respectively, at a luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. The current trigger rate is dominated by the beam background.

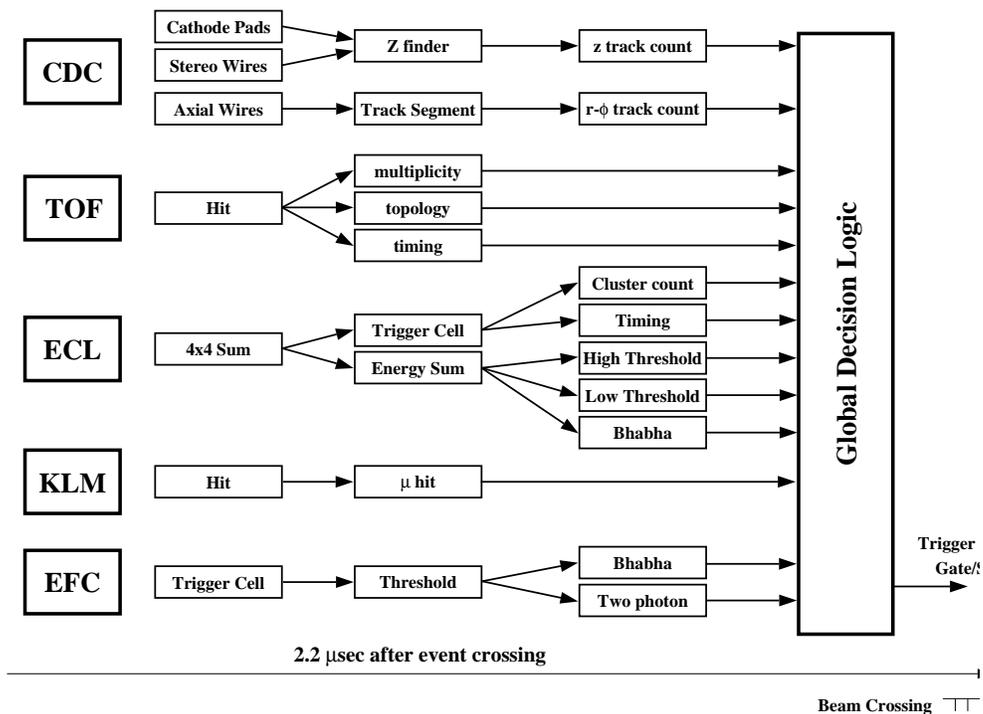


Figure 3.23: Schematic figure of the Level-1 trigger system for Belle detector.

3.2.9 Data Acquisition system

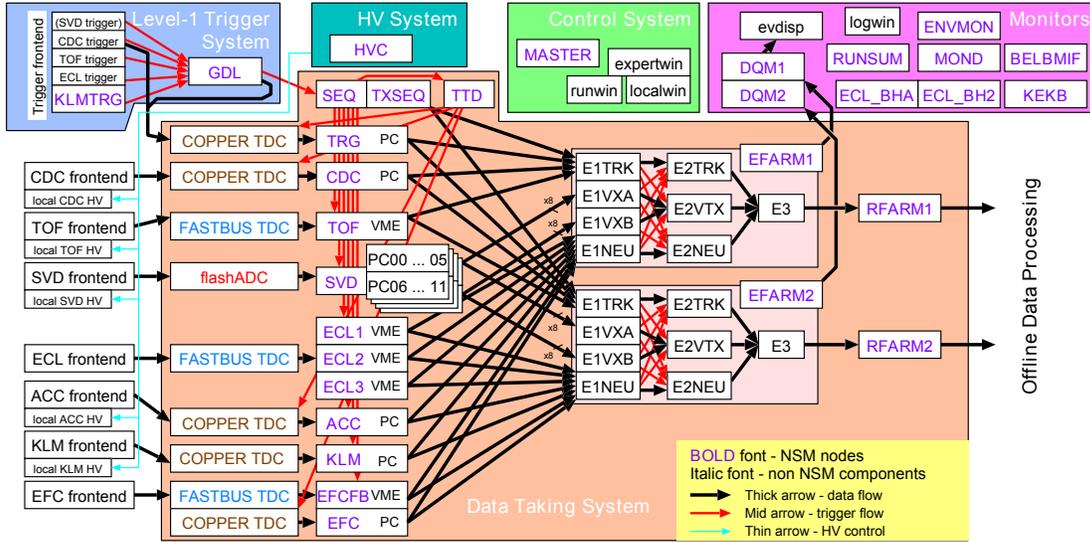


Figure 3.24: An overview of the Belle DAQ system.

The global scheme of the Belle data acquisition system is shown in Figure 3.24. The entire system is segmented into seven subsystems running in parallel, each handling the data from a sub-detector. Data from each subsystem are combined into a single event record by an event builder, which converts “detector-by-detector” parallel data streams to an “event-by-event” data river. The event builder output is transferred to an online computer farm, where another level of event filtering is done after the fast event reconstruction. The data are then sent to a mass storage system located at the computer center via optical fibers.

In the original system, charge-to-time converters (QTCs) and multi-hit FASTBUS time-to-digital converters (TDCs) are used for all the sub-detectors except for the SVD, in which an analog-to-digital converter (FADC) is used.

At a luminosity of $\sim 3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, the level-1 trigger rate is estimated to be higher than 1 to 2 kHz, and the dead time fraction is estimated to be quite bad, above 10%. This dead time is irreducible due to the specification of the FASTBUS TDCs that are currently used. Hence we have started to gradually replace the FASTBUS TDCs with a pipeline readout electronics consisting of COPPER (COmmon Pipelined Platform for Electronics Readout) board [38]. Figure 3.25 shows the picture view of a COPPER module.

In early 2007, we installed the COUNTERS for the first time in large scale (~ 100 COUNTERS) to the Belle CDC readout system [39]. Figure 3.26 shows how the dead time in the CDC decreased after the COUNTER installation. The dead time has decreased down to $\sim 10\%$ after the COUNTER installation, around the typical number of hits (100 \sim 150) in a TDC module.

Based on this experience, COUNTERS are also implemented to the readout of the ACC, KLM and trigger, and has been working successfully for several years.

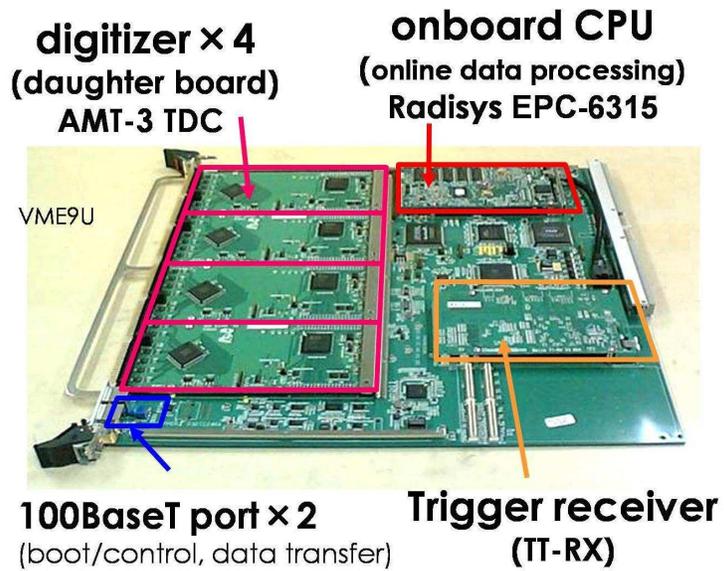


Figure 3.25: The picture of a COPPER module.

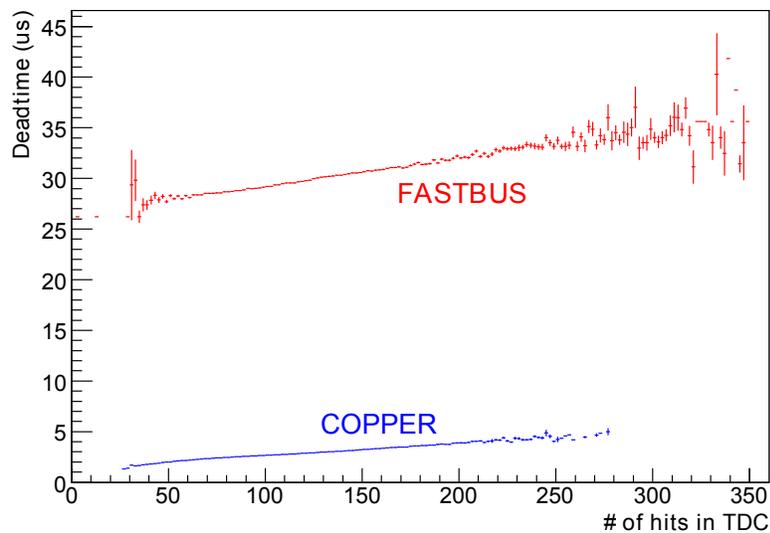


Figure 3.26: The dead time before and after the CDC COPPER installation. At the typical data size condition (~ 100 hits per TDC), the dead time decreased from ~ 30 ns to ~ 3 ns.

Chapter 4

Analysis overview

4.1 Semi-inclusive Approach

We reconstruct inclusive $B \rightarrow X_s \ell^+ \ell^-$ decays with a semi-inclusive reconstruction technique from a dilepton pair $\ell^+ \ell^- (= e^+ e^- \text{ or } \mu^+ \mu^-)$ and one of 18 reconstructed hadronic states. Here the hadronic system consists of one K^\pm or K_S^0 and up to four pions (at most one pion can be neutral). Possible 18 hadronic topologies are: K^\pm , $K^\pm \pi^0$, $K^\pm \pi^\mp$, $K^\pm \pi^\mp \pi^0$, $K^\pm \pi^\mp \pi^\pm$, $K^\pm \pi^\mp \pi^\pm \pi^0$, $K^\pm \pi^\mp \pi^\pm \pi^\mp$, $K^\pm \pi^\mp \pi^\pm \pi^\mp \pi^0$, $K^\pm \pi^\mp \pi^\pm \pi^\mp \pi^\pm$, K_S^0 , $K_S^0 \pi^0$, $K_S^0 \pi^\pm$, $K_S^0 \pi^\pm \pi^0$, $K_S^0 \pi^\pm \pi^\mp$, $K_S^0 \pi^\pm \pi^\mp \pi^0$, $K_S^0 \pi^\pm \pi^\mp \pi^\mp$, $K_S^0 \pi^\pm \pi^\mp \pi^\mp \pi^0$, and $K_S^0 \pi^\pm \pi^\mp \pi^\mp \pi^\pm$.

Compared to a fully inclusive approach, this method has the advantage of having strong kinematic discrimination against background by using the beam-energy constrained mass $M_{bc} = \sqrt{E_{\text{beam}}^2 - p_B^2}$ and the energy difference $\Delta E = E_B - E_{\text{beam}}$, where E_{beam} is the beam energy and E_B (p_B) is the reconstructed B meson energy (3-momentum). All quantities are evaluated in the $e^+ e^-$ center-of-mass system (CM).

From 2-body kinematics, the energy of reconstructed B meson must be half the total energy of the $e^+ e^-$ system in center-of-mass system. We use the beam-energy constrained mass M_{bc} in stead of reconstructed B mass, since M_{bc} has much better resolution than the reconstructed B mass and provides better separation of signal and background. ΔE should peak around zero for correctly reconstructed B mesons.

According to the signal Monte Carlo simulation, 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). 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% of the total rate. In this analysis, we define efficiency $\epsilon = N_{\text{rec}}/N_{\text{gen}}$, where N_{rec} is the number of reconstructed events and N_{gen} is the number of generated events in $M_{X_s} < 2.0 \text{ GeV}/c^2$, including the K_L^0 states or other missing states.

4.2 Particle Selections

Particle identification for e^\pm , μ^\pm , K^\pm , K_S^0 , π^\pm and π^0 is important for this analysis.

We require all charged tracks, except those used in the K_S^0 reconstruction, to have impact parameters with respect to the nominal interaction point of less than 1.0 cm in

the radial direction and 5.0 cm along the beam direction.

Electron identification is based on the ratio of the cluster energy to the track momentum (E/p), the specific energy loss (dE/dx) in the CDC, the position and shower shape of the cluster in the ECL and the response from the ACC. Muon identification is based on the hit positions and the depth of penetration into the ECL and KLM. Electrons and muons are required to have lab-frame momenta greater than 0.4 GeV/c and 0.8 GeV/c, respectively. To select good muon candidates, we apply a kaon veto. Charged kaon candidates are selected by `atc_pid` probability using information from the ACC, TOF and CDC. It is defined in such a way that zero indicates a pion-like track and one indicates a kaon-like track. After selecting the electron, muon and charged kaon candidate tracks, the remaining tracks are assumed to be charged pions. Bremsstrahlung photons from electrons are recovered by combining each electron with photons within a small angular region around the electron direction.

K_S^0 candidates are reconstructed from pairs of oppositely-charged tracks with $|m(\pi^+\pi^-) - m(K_S^0)| < 15 \text{ MeV}/c^2$. We impose additional K_S^0 selection criteria based on the distance and the direction of the K_S^0 vertex, and on the impact parameters of the daughter tracks. Neutral pions are 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 $|m(\gamma\gamma) - m(\pi^0)| < 10 \text{ MeV}/c^2$. The particle selection criteria are summarized in Table 4.1.

Table 4.1: Summary of the particle selection criteria.

Particle	Selection criteria
Charged track	$ dr < 1.0 \text{ cm}$ $ dz < 5.0 \text{ cm}$
Electron	$P_{lab} > 0.4 \text{ GeV}/c$ electron probability > 0.8
Muon	Not selected as electron $P_{lab} > 0.8 \text{ GeV}/c$ muon likelihood > 0.97
Kaon	Not selected as electron, muon <code>atc_pid</code> probability < 0.6 (muon-like)
Pion	Not selected as electron, muon <code>atc_pid</code> probability > 0.6 (kaon-like)
K_S^0	Remaining tracks after selecting the lepton and K^\pm tracks. K_S^0 -like vertex, impact parameters $ m(\pi^+\pi^-) - m(K_S^0) < 15 \text{ MeV}/c^2$
π^0	$E_\gamma > 50 \text{ MeV}$ $E_{\pi^0} > 400 \text{ MeV}$ $ m(\gamma\gamma) - m(\pi^0) < 10 \text{ MeV}/c^2$

4.3 Event Selections

Events are required to have a well determined primary vertex and be tagged as multi-hadronic events.

We reduce the combinatorial background requiring $-0.10(-0.05) \text{ GeV}/c < \Delta E < 0.05 \text{ GeV}/c$ for the dielectron (dimuon) case.

We have large contamination from charmonium background in which B decays to $J/\psi X_s$ or $\psi(2S)X_s$, then J/ψ or $\psi(2S)$ decays to $\ell^+\ell^-$. The charmonium backgrounds are reduced by removing B candidates with a dilepton mass in the ranges of $-0.40 \text{ GeV}/c^2 < M_{ee(\gamma)} - M_{J/\psi} < 0.15 \text{ GeV}/c^2$, $-0.25 \text{ GeV}/c^2 < M_{\mu\mu} - M_{J/\psi} < 0.10 \text{ GeV}/c^2$, $-0.25 \text{ GeV}/c^2 < M_{ee(\gamma)} - M_{\psi(2S)} < 0.10 \text{ GeV}/c^2$, and $-0.15 \text{ GeV}/c^2 < M_{\mu\mu} - M_{\psi(2S)} < 0.10 \text{ GeV}/c^2$. If one of the electrons from a J/ψ or $\psi(2S)$ decay erroneously picks up a random photon in the Bremsstrahlung-recovery process, the dilepton mass can increase sufficiently to evade the above cuts. Therefore the charmonium veto is applied to the dilepton mass both before and after Bremsstrahlung recovery.

The resulting veto sample provides a large and good control sample of decays with a signature identical to that of the signal. We call this control sample as “real $J/\psi X_s$ sample”.

We also require $M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$ to remove background from the conversion of the photon from radiative $B \rightarrow X_s \gamma$ decays and from π^0 Dalitz decays.

Both leptons are required to originate from the same vertex, i.e. $|\Delta z_{\ell^+\ell^-}| < 0.015 \text{ cm}$. Here $|\Delta z_{\ell^+\ell^-}|$ 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 reject candidates with an invariant mass of X_s , M_{X_s} , greater than $2.0 \text{ GeV}/c^2$ to remove a large fraction of the combinatorial background.

We also require $\chi_{\text{vtx}}^2/\text{NDF} < 10$, where χ_{vtx}^2 is the χ^2 of the B vertex constructed from all the charged daughter particles, excluding the K_S^0 daughters.

4.4 Background Suppression

To suppress background events effectively, we adopt a Fisher discriminant method [46]. The Fisher discriminant \mathcal{F} is constructed from the input parameters $P_i^{\text{sig,bg}}$: $\mathcal{F} = \sum_i \alpha_i \mathcal{P}_i$, where coefficients α_i are chosen by training the separation power between signal and background sample.

The largest background source is random combinations from semileptonic B decays. In this case, the leptons which form the dilepton pair in $X_s \ell^+ \ell^-$ are picked up from decay products of different B mesons, where B mesons decay into semileptonic state via the $b \rightarrow c \rightarrow s, d$ decay chain. This background has a significant amount of missing energy due to the neutrinos from the semileptonic decays. We reject this background by placing a cut on the Fisher discriminant variable $\mathcal{F}_{\text{miss}}$ constructed from 3 input parameters: E_{vis} , M_{miss} and ΔE^{ROE} . Here $E_{\text{vis}} = \sum_i E_i$ is the total visible energy and $M_{\text{miss}} = \sqrt{(2E_{\text{beam}} - \sum_i E_i)^2 - |\sum_i \vec{p}_i|^2}$ is the missing mass, where (\vec{p}_i, E_i) are the reconstructed CM-frame four-momenta of all tracks (assumed to be pions) and

all photons in the event, $\Delta E^{\text{ROE}} = E_{\text{ROE}} - E_{\text{beam}}$ is energy difference for Rest Of Event, calculated by summing the energies of all charged tracks and neutral calorimeter clusters not included in the B candidate. Distribution of Fisher input parameters and output are shown in Figure 4.1. We require the selection cut of $\mathcal{F}_{\text{miss}} > -1.856$. The cut value is chosen to have a same efficiency with the previous analysis for signal MC.

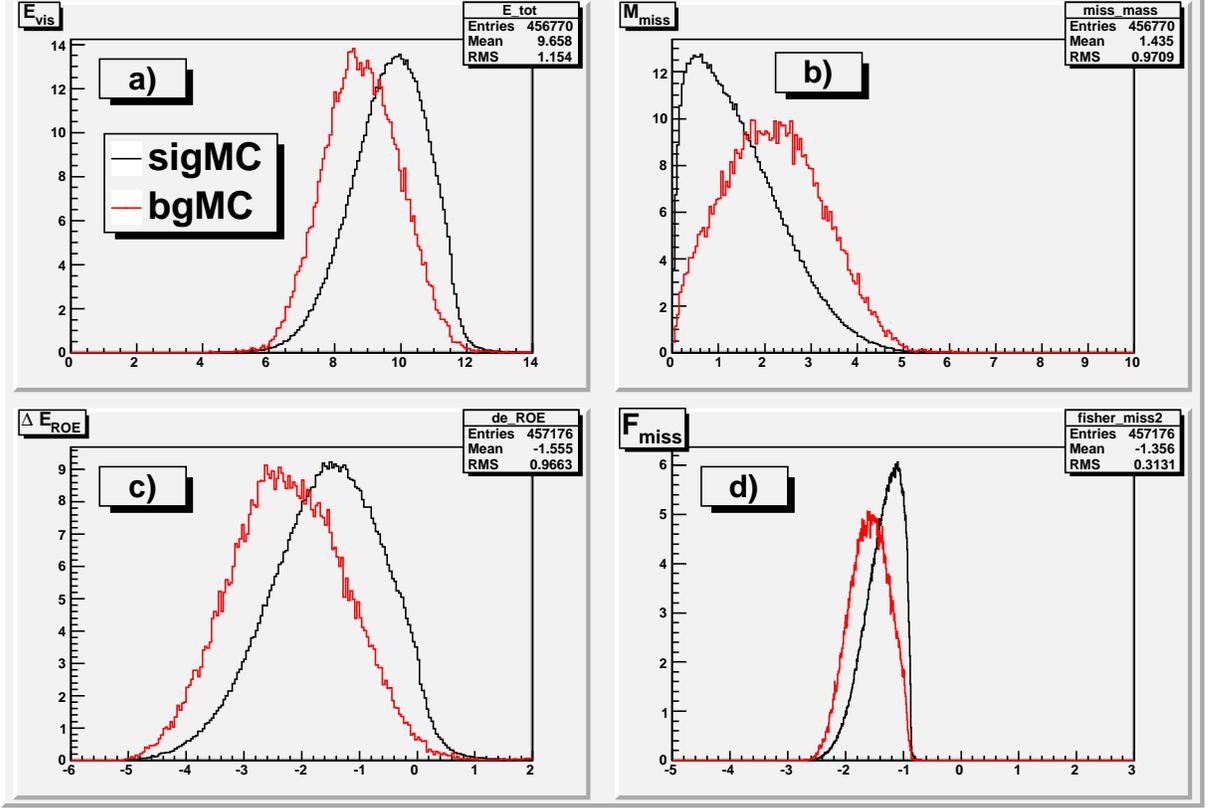


Figure 4.1: Distributions of three input parameters of $\mathcal{F}_{\text{miss}}$: a) E_{vis} (in GeV), b) M_{miss} (in GeV/c^2) and c) ΔE^{ROE} (in GeV); and d) the Fisher output. Black (red) histogram represents the distribution of signal (background) Monte Carlo sample.

We also reject another combinatorial background, which comes from continuum $q\bar{q}$ ($q = u, d, s, c$) background, by cutting on the Fisher discriminant (\mathcal{F}_{FW}) based on a modified set of Super Fox-Wolfram moments [47] that characterizes the event topology by applying a cut $\mathcal{F}_{\text{FW}} > -0.36$. We use six Super Fox-Wolfram moments: $R_{2,4}^{\text{so}}$ and $R_{1,2,3,4}^{\text{oo}}$. Distribution of input parameters and output are shown in Figure 4.2.

4.5 Best Candidate Selection and Likelihood ratio cut

At this stage, there is an average of 1.4 B candidates per event, according to the signal Monte Carlo simulation. In order to select the most signal-like B candidate, we construct a likelihood ratio based on the following six discriminant variables: ΔE ,

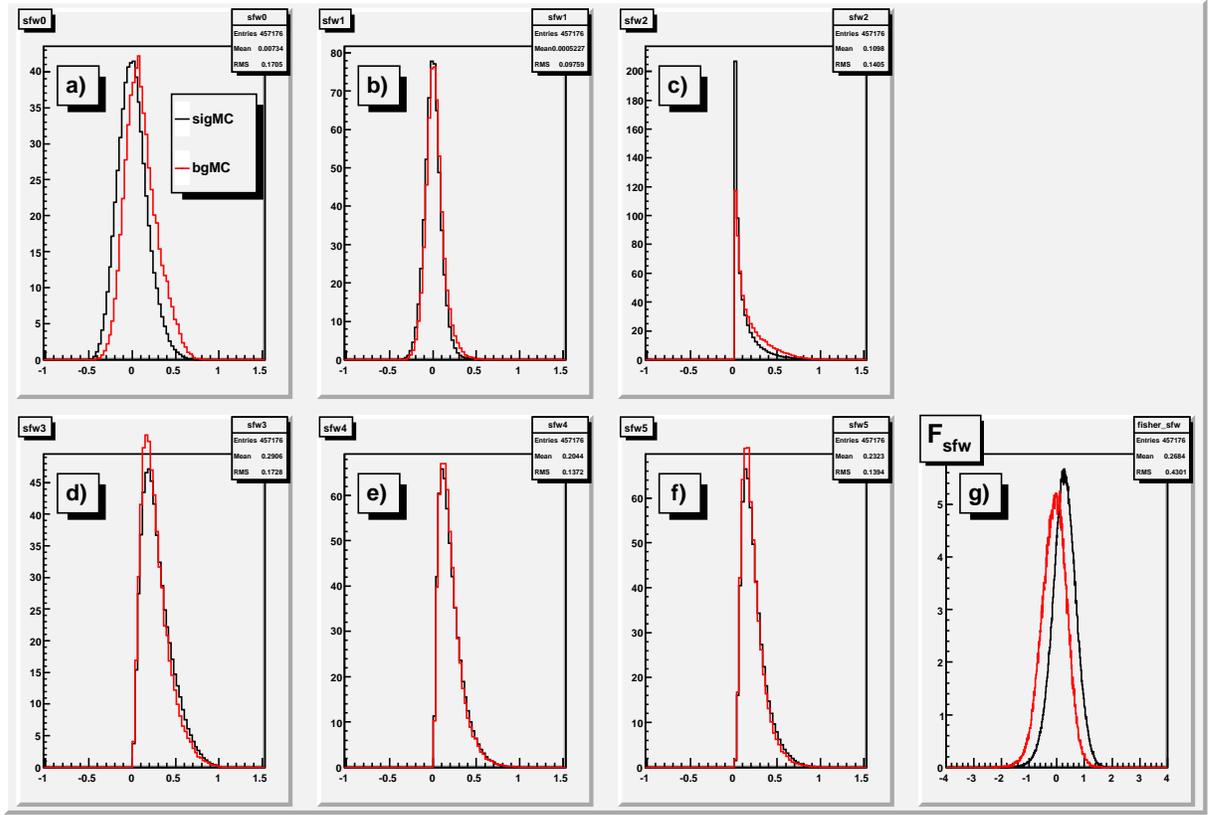


Figure 4.2: Distributions of six input parameters for \mathcal{F}_{FW} : a) R_2^{so} , b) R_4^{so} , c) R_1^{oo} , d) R_2^{oo} , e) R_3^{oo} , and f) R_4^{oo} ; and g) the \mathcal{F}_{FW} . Black(red) histogram represents the distribution of signal(background) Monte Carlo sample.

$\cos\theta_B$, \mathcal{F}_{FW} , $\mathcal{F}_{\text{miss}}$, $\chi_{\text{vtx}}^2/\text{NDF}$ and $|\Delta z_{\ell+\ell-}|$. Here $\cos\theta_B$ is the cosine of the B flight direction with respect to the e^- beam direction in the CM frame. Other discriminants are already defined above.

The variables ΔE and $\mathcal{F}_{\text{miss}}$ are effective at rejecting $B\bar{B}$ background, especially for semileptonic B decays, which have large missing energy. For continuum suppression, the event-shape variable \mathcal{F}_{FW} and $\cos\theta_B$ are useful. $\chi_{\text{vtx}}^2/\text{NDF}$ is effective to reject the random combinatorial background in the high multiplicity modes.

Figure 4.3 shows the distributions of six parameters for signal MC and real $J/\psi X_s$ sample. The probability density function (PDF) fitted to the real $J/\psi X_s$ sample is also shown. ΔE distributions are fitted separately for dielectron and dimuon modes. For the real $J/\psi X_s$ distributions, we subtract the background components estimated from the M_{bc} side-band region. The normalization of the subtracted background events is determined by the number of background events in the M_{bc} signal region, which is estimated by the fitting to the M_{bc} distribution with the Gaussian (signal) and ARGUS (background) functions. Figure 4.4 shows the distributions of six parameters for background MC, together with the fitted PDF function.

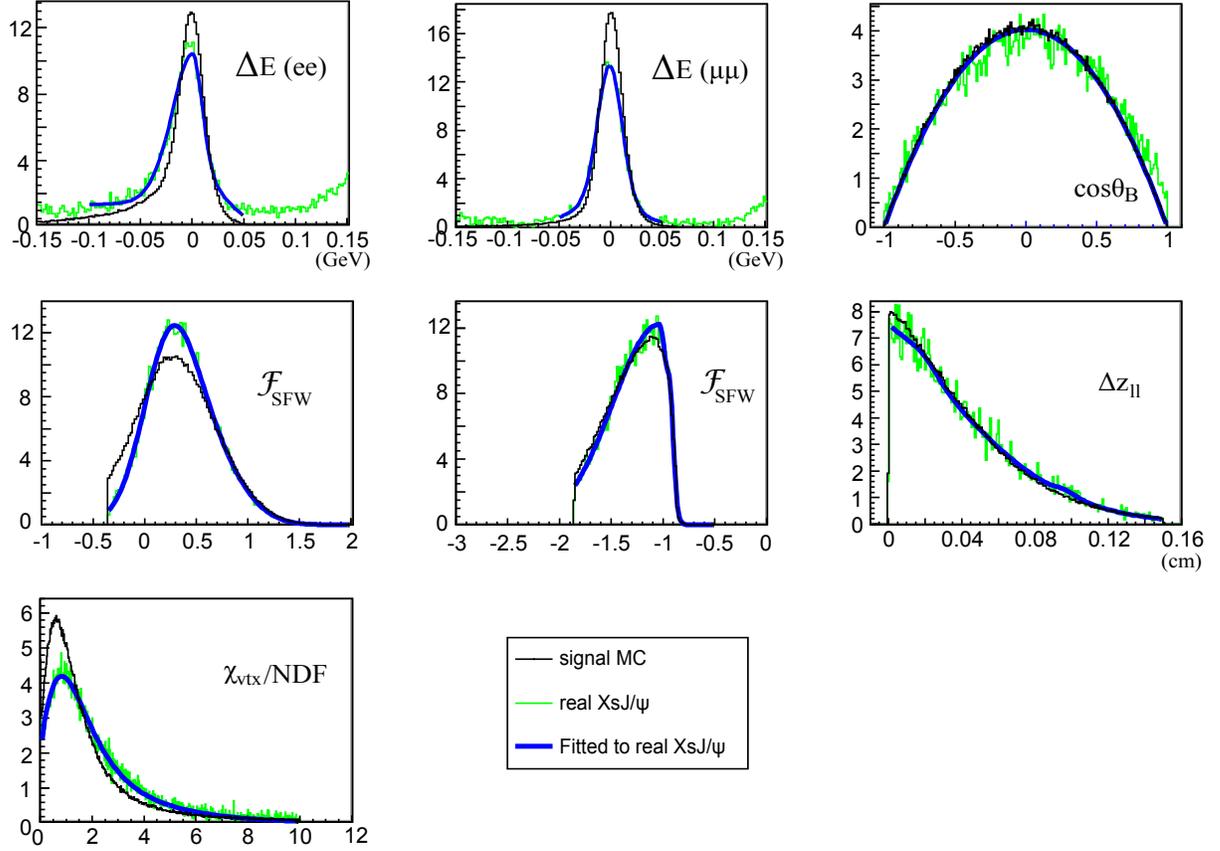


Figure 4.3: The distributions of discriminating variables: ΔE (e^+e^- modes), ΔE ($\mu^+\mu^-$ modes), $\cos\theta_B$, \mathcal{F}_{FW} , $\mathcal{F}_{\text{miss}}$, $|\Delta z_{\ell+\ell-}|$ and $\chi^2_{\text{vtx}}/\text{NDF}$, for the signal Monte Carlo sample (black) and the real $J/\psi X_s$ sample (green). The signal PDFs are determined by the fitting to the real $J/\psi X_s$ distribution. The obtained PDFs are also shown in thick black lines. Fitting to the ΔE distribution of $X_s e^+e^-$ modes and $X_s \mu^+\mu^-$ modes are separately done.

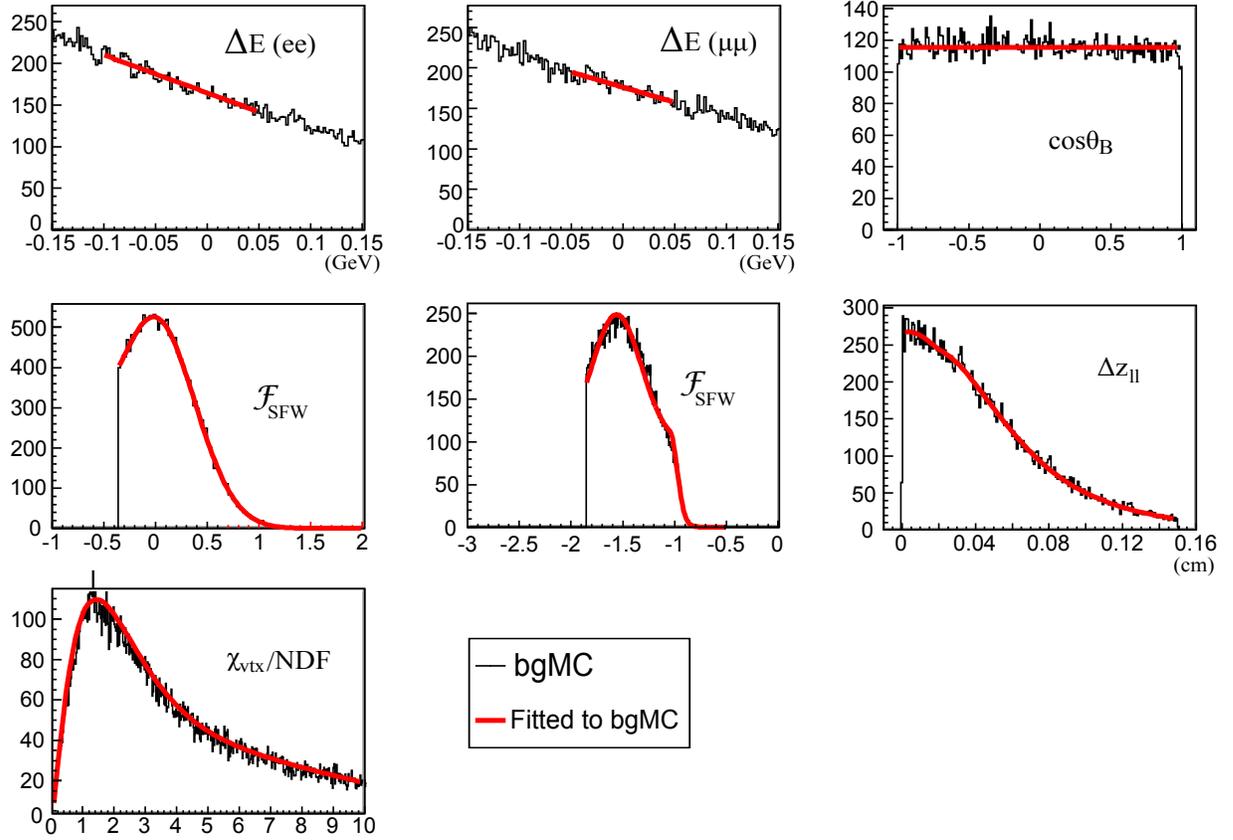


Figure 4.4: The distributions of discriminating variables: ΔE (e^+e^- modes), ΔE ($\mu^+\mu^-$ modes), $\cos\theta_B$, \mathcal{F}_{SFW} , $\mathcal{F}_{\text{miss}}$, $|\Delta z_{\ell+\ell-}|$ and $\chi^2_{\text{vtx}}/\text{NDF}$, for the background Monte Carlo events. The background PDFs are determined by the fitting to the background Monte Carlo distribution. The obtained PDFs are also shown in thick black lines. Fitting to the ΔE distribution of $X_s e^+e^-$ modes and $X_s \mu^+\mu^-$ modes of are separately done.

We then calculate the likelihoods $\mathcal{L}_{\text{sig,bkg}} = \prod_{i=1}^6 p_{\text{sig,bkg}}^i$ where p_{sig}^i and p_{bkg}^i are the PDFs of the discriminating variable i for the signal (sig) and the background (bkg), respectively. In each event, only the B candidate with the largest value of the likelihood ratio $\mathcal{R} = \mathcal{L}_{\text{sig}}/(\mathcal{L}_{\text{sig}} + \mathcal{L}_{\text{bkg}})$ is retained. According to the MC simulation, after selecting the B candidates with the largest value of the likelihood ratio, we find that the probability that all daughter particles from the $B \rightarrow X_s \ell^+ \ell^-$ decay are selected and are correctly assigned is 80% in the $B \rightarrow X_s \ell^+ \ell^-$ signal events.

As a check of the distributions of the likelihood ratio \mathcal{R} , we make a comparison between real data and MC data (Figure 4.5). For the signal, we compare the real $J/\psi X_s$ candidates with $B \rightarrow X_s \ell^+ \ell^-$ signal MC events. In the real charmonium-veto \mathcal{R} distribution, we subtract the background using the shapes obtained from the M_{bc} side-bands and the normalization from the number of background events in the M_{bc} signal region, in the same way as in the signal PDF determination. For the background, we compare the real $B \rightarrow X_s e^\pm \mu^\mp$ candidates with the generic background MC events. Here the real $B \rightarrow X_s e^\pm \mu^\mp$ candidates are reconstructed using the nominal selection criteria but requiring that the two leptons have different flavor ($e^\pm \mu^\mp$). We take the difference between distributions of data and MC as the systematic uncertainty by comparing the efficiencies of likelihood ratio cuts on data and MC.

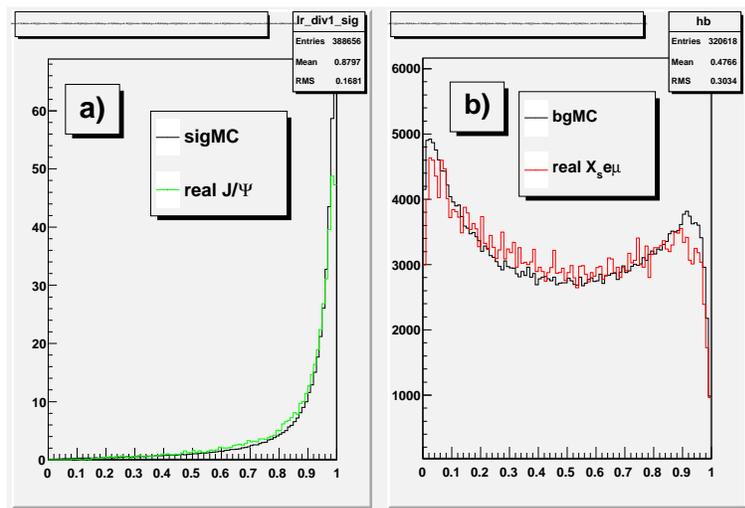


Figure 4.5: Likelihood ratio check for a) signal events and b) background events. For signal events, we compare signal MC (black) with real $J/\psi X_s$ events (green). For background events, we compare generic background MC (black) and real $B \rightarrow X_s e^\pm \mu^\mp$ events (red). The differences between distribution of data and MC are taken as the systematic uncertainty.

The final suppression of the combinatorial background is achieved with a cut on the likelihood ratio \mathcal{R} , optimized to maximize the statistical significance of the signal. This MC optimization is performed separately in the regions $M_{X_s} < 1.1 \text{ GeV}/c^2$ and $1.1 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, resulting in the cuts $\mathcal{R} > 0.77$ and $\mathcal{R} > 0.94$ respectively.

After applying all selection criteria, the remaining background events consist mostly of $B\bar{B}$ events (85% and 77% of the total background in the electron and muon channels, respectively), according to the background MC simulation. Using the signal MC simulation, the probability that all daughter particles from the $B \rightarrow X_s \ell^+ \ell^-$ decay are selected and are correctly assigned is estimated to be 88%.

All the $B \rightarrow X_s \ell^+ \ell^-$ selection cuts are summarized in Table 4.2.

Table 4.2: Summary of the $B \rightarrow X_s \ell^+ \ell^-$ event selection criteria.

Selection on the lepton pairs	
charmonium veto window	$-0.40 \text{ GeV}/c^2 < M_{ee(\gamma)} - M_{J/\psi} < 0.15 \text{ GeV}/c^2$
	$-0.25 \text{ GeV}/c^2 < M_{\mu\mu} - M_{J/\psi} < 0.10 \text{ GeV}/c^2$
	$-0.25 \text{ GeV}/c^2 < M_{ee(\gamma)} - M_{\psi(2S)} < 0.10 \text{ GeV}/c^2$
	$-0.15 \text{ GeV}/c^2 < M_{\mu\mu} - M_{\psi(2S)} < 0.10 \text{ GeV}/c^2$
γ -conv./ π_0 dalitz decay rejection	$M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$
Minimum z distance	$ \Delta z_{\ell^+\ell^-} < 150 \text{ mm}$
Selection on the reconstructed $B \rightarrow X_s \ell^+ \ell^-$ candidate	
B candidate signal box window	$5.27 < M_{bc} < 5.29 \text{ GeV}/c^2$
	$-0.10 < \Delta E < 0.05 \text{ GeV}$ for electron mode
	$-0.05 < \Delta E < 0.05 \text{ GeV}$ for muon mode
X_s mass cut	$M_{X_s} < 2.0 \text{ GeV}/c^2$
$X_s \ell^+ \ell^-$ vertex cut	$\chi^2_{\text{vtx}}/\text{NDF} < 10.0$
Continuum suppression	$\mathcal{F}_{\text{FW}} > -0.36$
$B\bar{B}$ suppression	$\mathcal{F}_{\text{miss}} > -1.856$
Likelihood-Ratio cut	$\mathcal{R} > 0.77$ for $M_{X_s} < 1.1 \text{ GeV}/c^2$
	$\mathcal{R} > 0.94$ for $M_{X_s} > 1.1 \text{ GeV}/c^2$

4.6 Data Samples

4.6.1 Real data sample

The Data sample used in this analysis comprises 657×10^6 B meson pairs, corresponding to an integrated luminosity of 605 fb^{-1} . All of the 140 fb^{-1} data samples used in the previous analysis are also included in this analysis.

4.6.2 Signal Monte Carlo sample and efficiency

For the $B \rightarrow X_s \ell^+ \ell^-$ event simulation, we use EVTGEN [40] as the event generator, JETSET [41] to hadronize the system consisting of a strange quark and a spectator quark, and GEANT [42] for the detector simulation. In the event generation, $B \rightarrow K \ell^+ \ell^-$, $B \rightarrow K^* \ell^+ \ell^-$, and non-resonant $B \rightarrow X_s \ell^+ \ell^-$ events are produced separately. Exclusive decays ($B \rightarrow K^{(*)} \ell^+ \ell^-$) are generated according to Refs [2, 43], where the relevant form factors are computed using light-cone QCD sum rules. Non-resonant $B \rightarrow X_s \ell^+ \ell^-$ generation is based on a model following Refs [2, 44] and the Fermi motion model of Ref. [45].

To normalize and mix these 3 different signal MC samples, we assumed the latest experimental branching fractions (See Table 2.2). For $B \rightarrow K \ell^+ \ell^-$ modes, we use the dilepton mode of HFAG world average, $\mathcal{B}(B \rightarrow K \ell^+ \ell^-) = 0.43 \times 10^{-6}$. For $B \rightarrow K^* \ell^+ \ell^-$ modes, we use the dimuon mode of HFAG world average, $\mathcal{B}(B \rightarrow K^* \ell^+ \ell^-) = 1.08 \times 10^{-6}$. For non-resonant $B \rightarrow X_s \ell^+ \ell^-$ mode, we assume the total $s \rightarrow \ell \ell$ branching fraction to be the dimuon mode of HFAG world average, $\mathcal{B}(B \rightarrow X_s \ell^+ \ell^-) = 4.3 \times 10^{-6}$. This means that the corresponding branching fraction for non-resonant $B \rightarrow X_s \ell^+ \ell^-$ is assumed to be $(4.3 - 0.43 - 1.08) \times 10^{-6} = 2.79 \times 10^{-6}$.

For $B \rightarrow K^* \ell^+ \ell^-$ and $B \rightarrow X_s \ell^+ \ell^-$ modes, we use the values for dimuon mode, since we assume that branching fraction of dielectron mode and dimuon mode are after $M_{\ell^+ \ell^-} > 0.2 \text{ GeV}/c^2$ cut in this analysis.

3.0(3.4) percent of $K^* e^+ e^-$ ($K^* \mu^+ \mu^-$) events exist in $M_{X_s} > 1.1 \text{ GeV}/c^2$ region. Those events overlap with non-resonant $B \rightarrow X_s \ell^+ \ell^-$ samples, but we just keep them in our analysis as did in Belle 140 fb^{-1} analysis [20], while BaBar 82 fb^{-1} analysis [21] didn't use them.

Table 4.3 summarizes the numbers of generated events N_{gen} for each sample and the scaling factors for 605 fb^{-1} . N_{rec} is the scaled number of reconstructed events after requiring all the analysis cuts. ϵ is the efficiency defined by $\epsilon = N_{\text{rec}}/N_{\text{gen}}$. The effect of missing modes is taken into account in the efficiency calculation.

Lepton identification efficiency for a Monte Carlo sample is known to be larger than that of the real data sample. The differences (MC < real data) are estimated to be 3.0% and 13.7% at 605 fb^{-1} , and 3.6% and 8.5% at 140 fb^{-1} , for $e^+ e^-$, $\mu^+ \mu^-$ and $\ell^+ \ell^-$ case, respectively. Hence, when we apply the efficiency in Table 4.3 to the real data, we should use the corrected smaller value.

4.6.3 Background Monte Carlo sample

For the background, I use the generic background Monte Carlo sample (charm, uds, charged, mixed) which corresponds to three times the amount of real data available.

Table 4.3: Summary of the signal Monte Carlo samples for 605 fb^{-1} . N_{gen} is the number of generated events satisfying $M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$. $M_{X_s} > 1.1 \text{ GeV}/c^2$ is also required for non-resonant $B \rightarrow X_s \ell^+ \ell^-$ sample. N_{gen} includes the missing modes such as K_L^0 modes or too many pion modes, etc. Each sample is scaled according to the assumed branching function and the luminosity. N_{rec} is the scaled number of reconstructed events after requiring all the analysis cuts. ϵ is the efficiency defined by $\epsilon = \frac{N_{\text{rec}}}{N_{\text{gen}}}$. The effect of missing modes is taken into account in the efficiency calculation. $\epsilon_{\text{corrected}}$ is the efficiency after the lepton identification correction.

Mode	Assumed branching fraction ($\times 10^{-6}$)	N_{gen}	N_{gen} (scaled)	N_{rec} (scaled)	ϵ	$\epsilon_{\text{corrected}}$
$B \rightarrow K e^+ e^-$	0.43	3930928	564.8	46.9	8.30 %	
$B \rightarrow K \mu^+ \mu^-$	0.43	3930928	564.8	58.1	10.29 %	
$B \rightarrow K^* e^+ e^-$	1.08	2914122	1418.5	53.4	3.76 %	
$B \rightarrow K^* \mu^+ \mu^-$	1.08	3930927	1418.5	66.5	4.69 %	
$B \rightarrow X_s e^+ e^-$ (non-resonance)	2.79	1508710	3664.5	49.0	1.34 %	
$B \rightarrow X_s \mu^+ \mu^-$ (non-resonance)	2.79	2210803	3664.5	49.8	1.35 %	
$b \rightarrow \text{see total}$	4.3	-	5647.8	149.2	2.64 %	2.56 %
$b \rightarrow s \mu^+ \mu^-$ total	4.3	-	5647.8	174.4	3.09 %	2.67 %
$b \rightarrow s ll$ total	4.3	-	11295.7	323.6	2.86 %	2.61 %

Normalization factor can be determined by two different methods. The first one is simple. Since we use three times the data, so use 0.33 as the factor. The second one is to determine the factor by equating the number of events in side-band region ($5.20 \text{ GeV}/c^2 < M_{bc} < 5.26 \text{ GeV}/c^2$) in the real $X_s \ell^+ \ell^-$ data sample and the background Monte Carlo sample, after applying all the event selections. For 605 fb^{-1} samples, we have 4569(18149) events for real(MC) sample, so the factor is 0.248. For 140 fb^{-1} samples, we have 1336(4899) events for real(MC) sample, so the factor is 0.273.

We take the factor determined with the second method for Fisher training and LR cut optimization.

Chapter 5

Maximum Likelihood Fit

5.1 Fitter method

We perform an extended unbinned maximum likelihood fit to the M_{bc} distribution in the region $M_{bc} > 5.20 \text{ GeV}/c^2$ for extracting the signal yield.

Fitting method is summarized in Table 5.1. We fit $B \rightarrow X_s \ell^+ \ell^-$ sample and $B \rightarrow X_s e^\pm \mu^\mp$ sample simultaneously. In the likelihood function, we consider four components for the $B \rightarrow X_s \ell^+ \ell^-$ sample: signal (sig), background (bkg), peaking background (pkg) and self cross-feed (scf). Those components are explained in the following subsections. For the $B \rightarrow X_s e^\pm \mu^\mp$ sample, we consider two components: background (bkg) and peaking background (pkg)¹.

We adopt a Gaussian function to model the signal PDF ($\mathcal{P}_{\text{sig}}^i$) and an ARGUS function [48]² to model the background PDF ($\mathcal{P}_{\text{bkg}}^i$). For the peaking background PDF ($\mathcal{P}_{\text{pkg}}^i$) and the self cross-feed PDF ($\mathcal{P}_{\text{scf}}^i$), we used the PDF constructed from the estimated M_{bc} distribution histogram shape using RooHistPdf method. For the peaking background PDF for higher ψ s ($\mathcal{P}_{\text{psi}}^i$), we adopt a Gaussian function.

For a dielectron and dimuon mode, we perform a simultaneous fit for $B \rightarrow X_s \ell^+ \ell^-$ and $B \rightarrow X_s e^\pm \mu^\mp$ samples, using a common ARGUS shape parameter for $\mathcal{P}_{\text{bkg}}^i$ and $\mathcal{P}_{\text{bkg},e\mu}^j$.

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

$$\begin{aligned} \mathcal{L} &= \frac{e^{-(N_{\text{sig}}+N_{\text{bkg}}+N_{\text{pkg}}+N_{\text{psi}}+N_{\text{scf}})}}{N!} \\ &\times \prod_{i=1}^N [N_{\text{sig}} \mathcal{P}_{\text{sig}}^i + N_{\text{bkg}} \mathcal{P}_{\text{bkg}}^i + N_{\text{pkg}} \mathcal{P}_{\text{pkg}}^i + N_{\text{psi}} \mathcal{P}_{\text{psi}}^i + N_{\text{scf}} \mathcal{P}_{\text{scf}}^i] \\ &\times \frac{e^{-(N_{\text{bkg},e\mu}+N_{\text{pkg},e\mu})}}{N_{e\mu}!} \times \prod_{j=1}^{N_{e\mu}} [N_{\text{bkg},e\mu} \mathcal{P}_{\text{bkg},e\mu}^j + N_{\text{pkg},e\mu} \mathcal{P}_{\text{pkg},e\mu}^j] \end{aligned}$$

²ARGUS function is defined as:

$$\mathcal{P}(x) = t \sqrt{1-t^2} \exp(\alpha(1-t^2)),$$

where x is M_{bc} , t is defined as $t = x/E_{\text{beam}}$, and α is the shape parameter of ARGUS function.

For a dilepton mode, we perform a combined fit for e^+e^- and $\mu^+\mu^-$ data sample: i.e. we fit $B \rightarrow X_s e^+ e^-$, $B \rightarrow X_s \mu^+ \mu^-$ and $B \rightarrow X_s e^\pm \mu^\mp$ samples simultaneously with three different likelihood functions corresponding to each sample, with a common ARGUS shape parameter.

In order to measure the differential branching fraction, we performed the fit on the partial data samples in M_{X_s} or q^2 regions. For M_{X_s} regions, we divide the dataset into 5 regions: [0.4, 0.6], [0.6, 0.8], [0.8, 1.0], [1.0, 1.4] and [1.4, 2.0] GeV/ c^2 . For q^2 regions, we divide the dataset into 4 regions: [0.04, 1.0], [1.0, 6.0], [6.0, 14.4] and [14.4, 25.0] (GeV/ c^2)².

Table 5.1: Summary of the fitting method. We fit real $B \rightarrow X_s \ell^+ \ell^-$ and $B \rightarrow X_s e^\pm \mu^\mp$ samples simultaneously, with a common ARGUS shape parameter. Floated parameters are N_{sig} , N_{bkg} , $N_{\text{bkg}, e\mu}$ and ARGUS shape parameter.

PDF component	Used function	fix or float
$X_s \ell^+ \ell^-$ fit		
Signal (sig)	Gaussian	N_{sig} :float, shape:fix
Background (bkg)	ARGUS	N_{bkg} :float, shape:float(common)
Peaking background (pkg)	histogram shape	N_{pkg} :fix
Peaking background from higher psi (psi)	Gaussian	N_{psi} :fix, shape:fix
Self cross-feed (scf)	histogram shape	$N_{\text{scf}}/N_{\text{sig}}$:fix
$X_s e\mu$ fit		
Background (bkg, $e\mu$)	ARGUS	$N_{\text{bkg}, e\mu}$:float, shape:float (common)
Peaking background (pkg, $e\mu$)	histogram shape	$N_{\text{pkg}, e\mu}$:fix

Compared to the previous Belle 140 fb⁻¹ analysis [20], we introduce following new fitting methods:

- We float the ARGUS shape parameter in the fit. In the previous analysis, it was fixed to the value obtained from the fit to the real $B \rightarrow X_s e^\pm \mu^\mp$ sample.
- We fit real $B \rightarrow X_s \ell^+ \ell^-$ and $B \rightarrow X_s e^\pm \mu^\mp$ samples simultaneously, with a common ARGUS shape parameter. In the previous analysis, a fit was made to the real $B \rightarrow X_s \ell^+ \ell^-$ sample only and the real $B \rightarrow X_s e^\pm \mu^\mp$ sample was used to determine the ARGUS shape parameter.
- We include two additional peaking background sources.

5.1.1 Signal component (sig)

The signal PDF $\mathcal{P}_{\text{sig}}^i$ is described by a Gaussian for the e^+e^- case as well as the $\mu^+\mu^-$ case, since the Bremsstrahlung recovery and selection procedure for the e^+e^- case lead to a negligible radiative tail in the M_{bc} distribution. The Gaussian shape

parameters, mean m_{sig} and resolution σ_{sig} , are fixed to the value determined by the ARGUS+Gaussian fit to the real $J/\psi X_s$ data sample. The obtained Gaussian shape parameters for the $\ell^+\ell^-$ sample are: $m_{\text{sig}} = 5279.396 \pm 0.015 \text{MeV}/c^2$ and $\sigma_{\text{sig}} = 2.669 \pm 0.013 \text{MeV}/c^2$. Figure 5.1 show the M_{bc} distributions for the real $J/\psi X_s$ sample. The signal yield N_{sig} is taken as a free parameter in the likelihood fit.

5.1.2 Background Component (bkg)

The background PDF $\mathcal{P}_{\text{bkg}}^i$ for continuum and $B\bar{B}$ combinatorial events is given by the ARGUS function [48]. The ARGUS cutoff is determined by the beam energy in the $\Upsilon(4S)$ rest frame, $E_{\text{beam}} = 5.290 \text{GeV}$. The ARGUS shape parameter is a free parameter in the likelihood fit ³. Note that we simultaneously fit the $B \rightarrow X_s \ell^+ \ell^-$ sample and $B \rightarrow X_s e^\pm \mu^\mp$ sample with the common shape parameter for background PDF. The yield N_{bkg} is also taken as a free parameter in the likelihood fit.

5.1.3 Peaking Background Component (pkg,psi)

We consider two sources for charmonium peaking background,

1. $B \rightarrow J/\psi X_s$ or $B \rightarrow \psi(2S) X_s$ decays with J/ψ or $\psi(2S) \rightarrow \ell^+ \ell^-$.
2. higher ψ resonances ($\psi(3770), \psi(4040), \psi(4160)$)

and two sources for hadronic peaking background.

3. $B \rightarrow X_s hh$ decays, mis-identify two hadrons(hh) as leptons
4. $B \rightarrow X_s h \ell \nu$ decays, mis-identify one hadron(h) as a lepton, missing ν energy

For 1, 3, and 4, we use the PDF $\mathcal{P}_{\text{pkg}}^i$ modeled by the shape of estimated M_{bc} histogram using the RooKeysPdf method ⁴. For 2, we use the same Gaussian PDF with signal component. $N_{\text{pkg}}, N_{\text{psi}}$ are fixed in the likelihood fit.

Charmonium Peaking Background

We have two sources for the charmonium peaking background.

The first source is $B \rightarrow J/\psi X_s$ or $B \rightarrow \psi(2S) X_s$ decays with J/ψ or $\psi(2S) \rightarrow \ell^+ \ell^-$. This background is efficiently removed with cuts on the dilepton mass $M_{\ell^+ \ell^-}$. To estimate this background, we use $B \rightarrow J/\psi X_s$ or $B \rightarrow \psi(2S) X_s$ Monte Carlo simulation. For the $X_s \ell^+ \ell^-$ modes including zero or one pion, we use exclusive Monte Carlo, including $J/\psi K, J/\psi K^*$, and various higher resonances. For the $X_s \ell^+ \ell^-$ modes

³In the previous analysis, the values of the ARGUS shape parameter was determined from the fit to the real $B \rightarrow X_s e^\pm \mu^\mp$ events selected using the nominal selection criteria but requiring that the two leptons have different flavor. The shape parameter obtained from 605fb^{-1} real $B \rightarrow X_s e^\pm \mu^\mp$ data is -17.3 ± 2.3 .

⁴Usually ‘‘peaking background’’ is a component around the signal peak ($5.28 \text{GeV}/c^2$) and modeled by the Gaussian which has the same shape parameter with the signal PDF. But in this analysis, we define the peaking background including both Gaussian component and combinatorial background from the peaking background.

including two or more pions, we use Monte Carlo where $B \rightarrow J/\psi X_s$ decays follow phase space. We compare the number of events in Monte Carlo and the real $J/\psi X_s$ sample which is vetoed by $M_{\ell^+\ell^-}$ charmonium cut and obtain the scaling factor between MC and real $J/\psi X_s$, for each of 36 $X_s \ell^+ \ell^-$ final states⁵. Then we apply these scaling factors to each $X_s \ell^+ \ell^-$ final state in the MC sample which remains after the $M_{\ell^+\ell^-}$ charmonium cut, and obtain the M_{bc} distribution of the charmonium peaking background. Figure 5.2 shows the M_{bc} distributions of the weighted MC charmonium peaking background events.

The second source is higher charmonium resonances, such as $\psi(3770)X_s$, $\psi(4040)X_s$ and $\psi(4160)X_s$. This is estimated to be 6.9 ± 4.3 events at 605 fb^{-1} (See Section A). For the dBR/dMX_s measurement, we assume the same M_{X_s} distribution for real $J/\psi X_s$ data. This source is taken into account for the first time in this 605 fb^{-1} analysis.

Hadronic Peaking Background

We have two sources for the hadronic peaking background.

The first source is the $B \rightarrow X_s hh$ decays, where $B \rightarrow D^{(*)}n\pi$ ($n > 0$), $D \rightarrow K\pi$. The two pions in the final state are misidentified as leptons. The normalization and shape of this peaking background is determined directly from the data. We repeat the same selection without the lepton identification requirements, picking up two hadrons (mostly pion in this case) instead of dileptons. We then weigh each event with the “fake rate”, which is the probability to mis-identify a hadron as an electron or muon as a function of hadron momentum and direction. Figure 5.3 shows the M_{bc} distributions of the weighted hadronic peaking background candidates.

The second source are $B \rightarrow X_s h\ell\nu$ decays, where $B \rightarrow D^{(*)}n\pi$, $D \rightarrow K\ell\nu$ or $B \rightarrow D^{(*)}\ell\nu$, $D \rightarrow K\pi$. In this case, one pion is misidentified as a lepton and the missing energy of the neutrino is compensated by picking up additional gammas (as π_0 daughters) from the other B meson, therefore these events remain in the M_{bc} signal region. To estimate this contribution, we repeat the same selection without the lepton identification requirements for one lepton (and keep them for the other lepton), and weigh each event with the fake rate. Note that the lepton-like particle we selected could be a pion with a probability of the fake rate. This means that the selected events are $\ell\pi + \pi\pi \times (\text{fakrate})$. Then we weigh each event with the fake rate of the other pion. Therefore the estimated M_{bc} distribution includes the full component of $B \rightarrow X_s hh$ events, not only $B \rightarrow X_s hh$ events. Figure A.1 shows the weighted M_{bc} distributions of the hadronic peaking background events (See Section A). This source is taken into account for the first time in this 605 fb^{-1} analysis.

5.1.4 Self cross-feed component (scf)

The last background, the self cross-feed PDF $\mathcal{P}_{\text{scf}}^i$ is also modeled by the histogram shape of the self cross-feed candidates in a signal MC samples. The yield N_{scf} should be proportional to N_{sig} , so the ratio $N_{\text{scf}}/N_{\text{sig}}$ is fixed to the value obtained from MC samples in the likelihood fit. Figure 5.4 shows the M_{bc} distributions of the MC self cross-feed events.

⁵We have 18 X_s final states and 2 dilepton states.

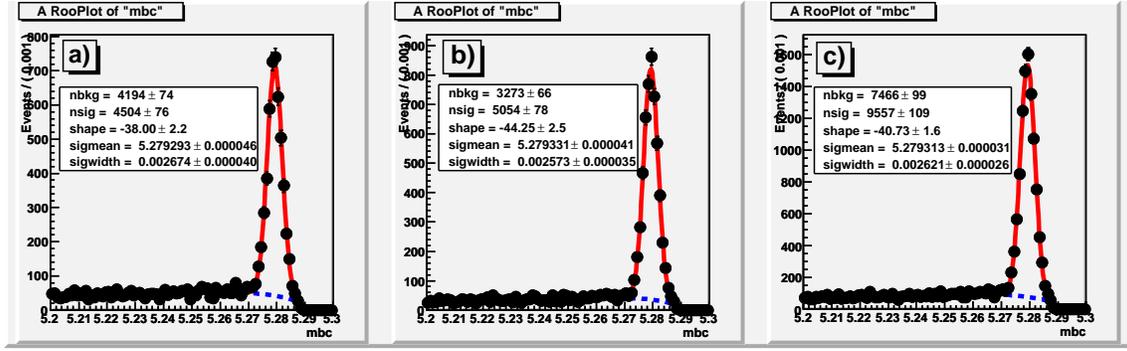


Figure 5.1: The M_{bc} distributions for the real $J/\psi X_s$ sample for a) e^+e^- , b) $\mu^+\mu^-$ and c) l^+l^- modes. The obtained fitted values are used for the $B \rightarrow X_s l^+l^-$ signal Gaussian fitting parameters, as described in the text. These histograms correspond to 605 fb^{-1} .

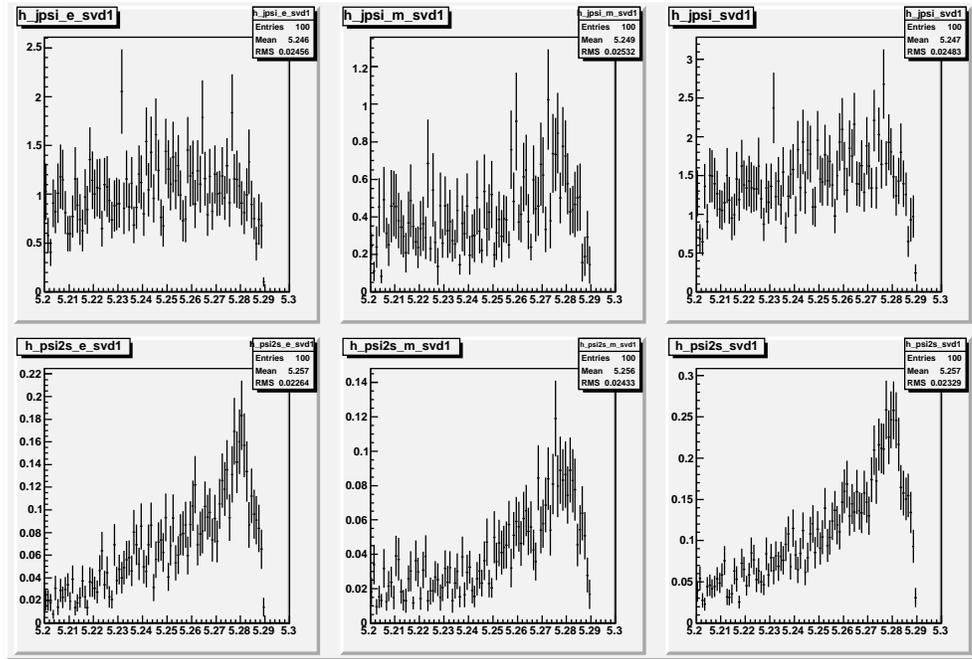


Figure 5.2: The estimated M_{bc} distributions for J/ψ (upper) and $\psi(2S)$ (lower) charmonium peaking background events, for dielectron (left), dimuon (center) and dilepton (right) cases. These histograms correspond to 605 fb^{-1} .

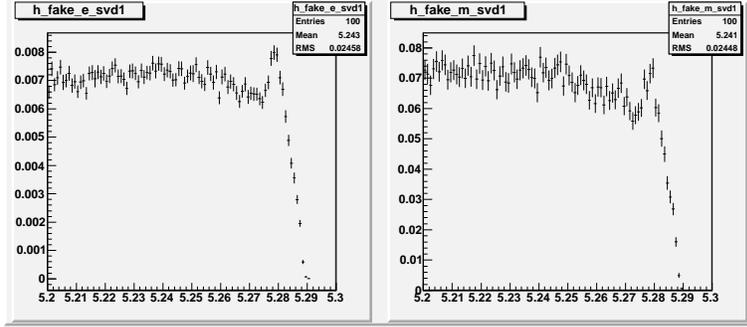


Figure 5.3: The estimated M_{bc} distributions for hadronic peaking background events, for dielectron(left) and dimuon(right) cases. These histograms include both of two hadronic peaking background events. These histograms correspond to 605 fb^{-1} .

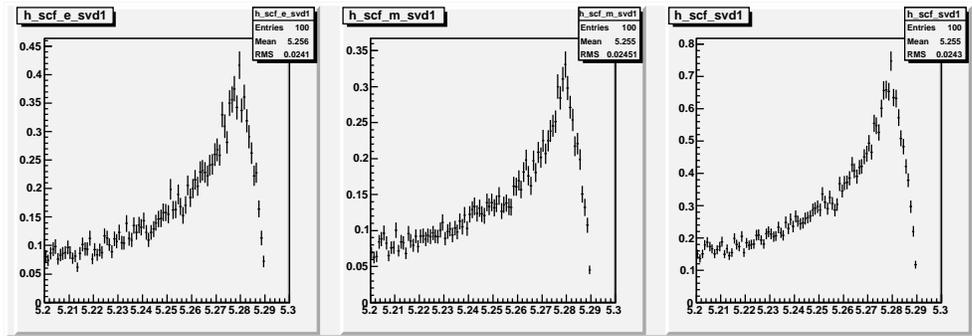


Figure 5.4: The estimated M_{bc} distributions for self cross-feed events, for dielectron(left), dimuon(center) and dilepton(right) cases. These histograms correspond to 605 fb^{-1} .

5.1.5 PDFs for each of M_{X_s} or $M_{\ell^+\ell^-}$ bins

The PDF distributions for each of M_{X_s} or $M_{\ell^+\ell^-}$ bins can be found in Section B.

5.2 Fitter validation test

To confirm that our fitter works properly, we performed two kinds of tests: a toy MC test and an ensemble test. For each case, we check the linearity between the input parameter and fitter output, and the pull distribution has a peak around 0 and its resolution is around 1. The tests are performed for a sample corresponding to 140 fb^{-1} .

5.2.1 Toy MC check

For the toy MC check, we generate the test event sample from the PDF for all the components (signal, background, peaking background and self cross feed). We checked the fitter response for the two input parameters: N_{sig} (Figure 5.5) and ARGUS shape parameter(Figure 5.6). Our fitter shows good response for each parameter.

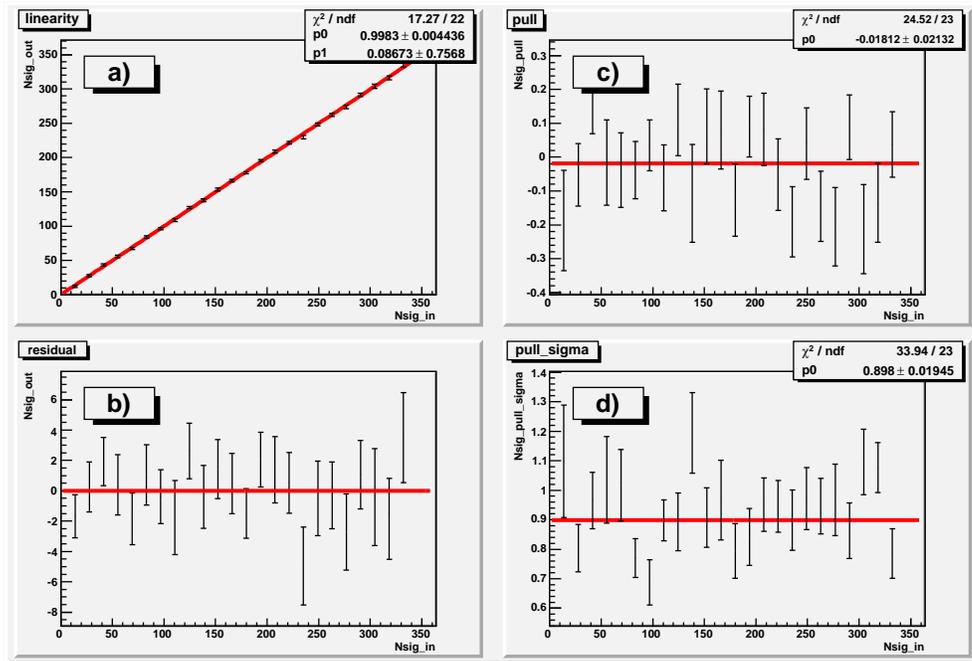


Figure 5.5: Toy MC test result for the N_{sig} . For 140 fb^{-1} , estimated number of signal is 69. a) is the fitter output as a function of the input, fitted with a linear function. b) is the residual from the fitted linear function. c) is the mean of the pull distribution. d) is the resolution of the pull distribution. We found no problem in the linearity and the pull distribution.

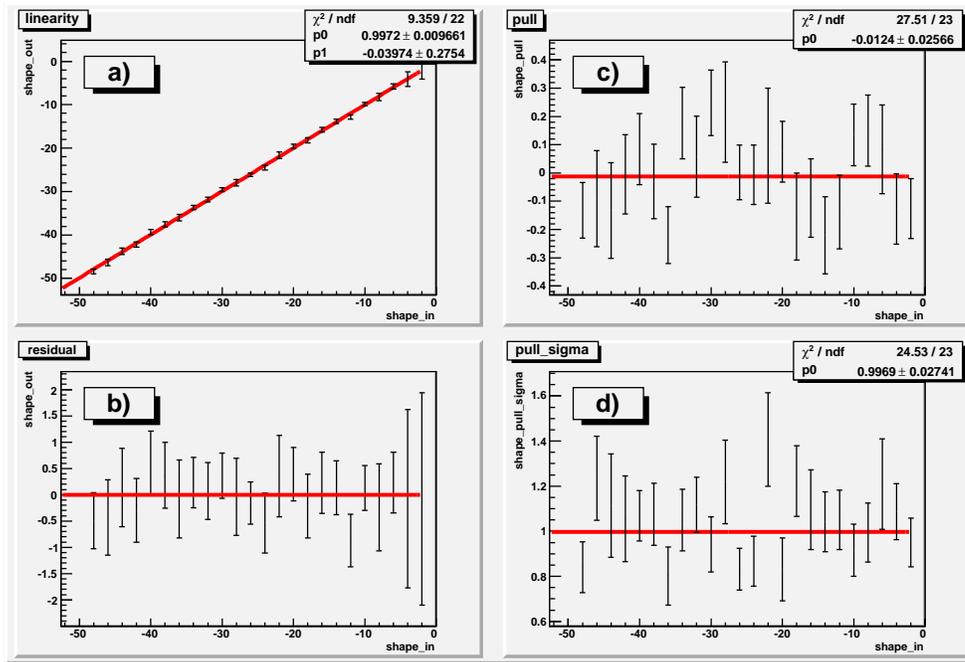


Figure 5.6: Toy MC test result for the ARGUS shape parameter. We found the linearity and the pull distribution is good. a) is the fitter output as a function of the input, fitted with linear function. b) is the residual from the fitted linear function. c) is the mean of the pull distribution. d) is the resolution of the pull distribution. We found no problem in the linearity and the pull distribution.

5.2.2 Ensemble check

For the ensemble check, signal, background and self cross feed components of the test event sample are randomly picked up from the corresponding MC sample. The peaking background component is generated from PDF. We checked the fitter response for N_{sig} and found an excess in the N_{sig} output.

The excess is independent from the N_{sig} input (see Figure 5.7 b)), but proportional to the N_{bkg} input (see Figure 5.8). We found that the peaking background events are included in the background MC sample, though we separately generate them from the PDF (the implemented branching fractions are different, though). Figure 5.9 shows that the peaking background events in the generic background MC are mainly from $B\bar{B}$ background MC, not from the continuum background MC. The number of the peaking background events is consistent with the excess in the N_{sig} output. So, we conclude that the reason for the excess is because we are double-counting the peaking background component, not because our fitter is biased. ⁶

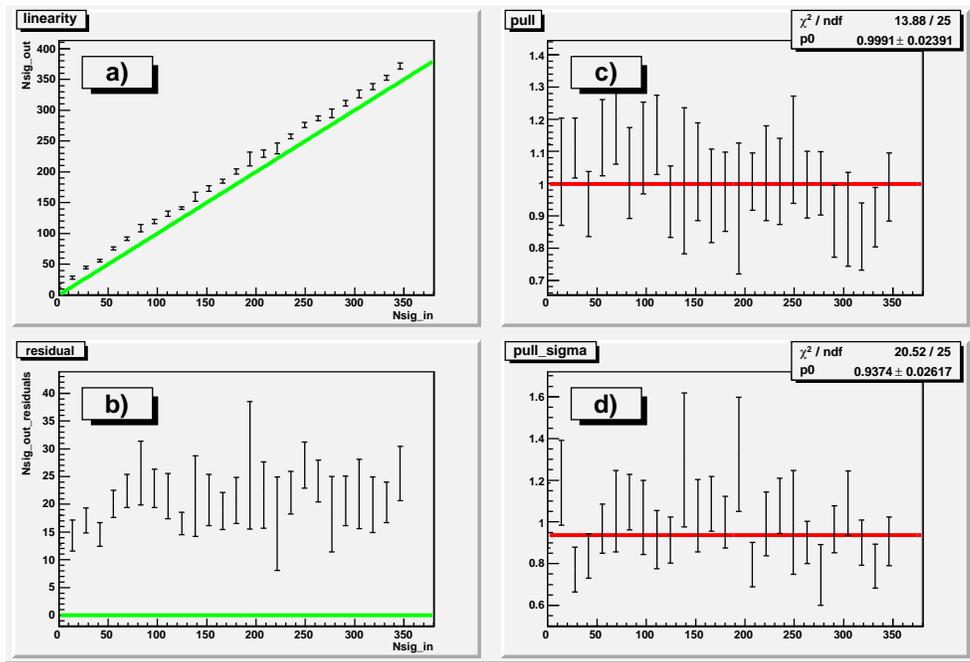


Figure 5.7: Ensemble test result for the fitter check. Signal, background and self cross feed components are randomly picked up from the corresponding MC sample. N_{sig} is estimated to be 69 for 140 fb^{-1} according to the MC. The peaking background component is generated from PDF. a) is the fitter output as a function of the input, fitted with a linear function. b) is the residual from the fitted linear function. c) is the mean of the pull distribution. d) is the resolution of the pull distribution. We found an excess in the N_{sig} output.

⁶At 605 fb^{-1} , the excess is ~ 75 in the ensemble test. This is consistent with the peaking component of the background Monte Carlo, 64 ± 12 (See Figure 5.9).

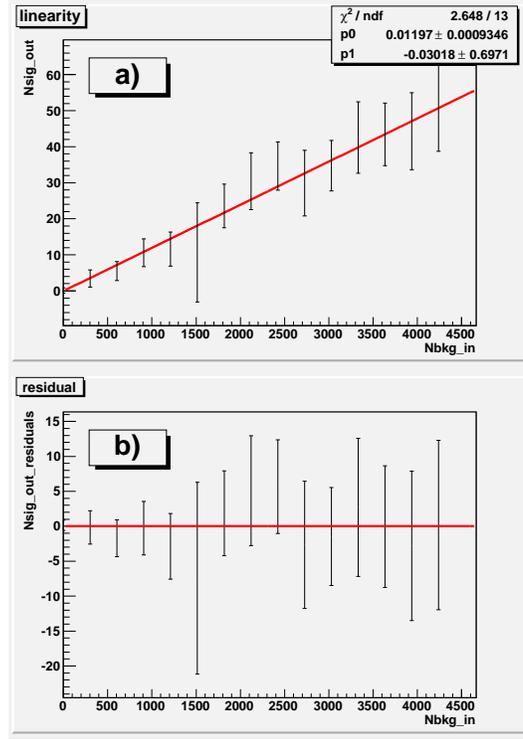


Figure 5.8: a) is N_{sig} excess as a function of the N_{bkg} input in the ensemble test, fitted with a linear function. b) is the residual from the fitted linear function. N_{bkg} is estimated to be ~ 1500 for 140 fb^{-1} according to the MC. We found the N_{sig} excess is proportional to the N_{bkg} input, which implies excess is purely from the background component, not from others. OB

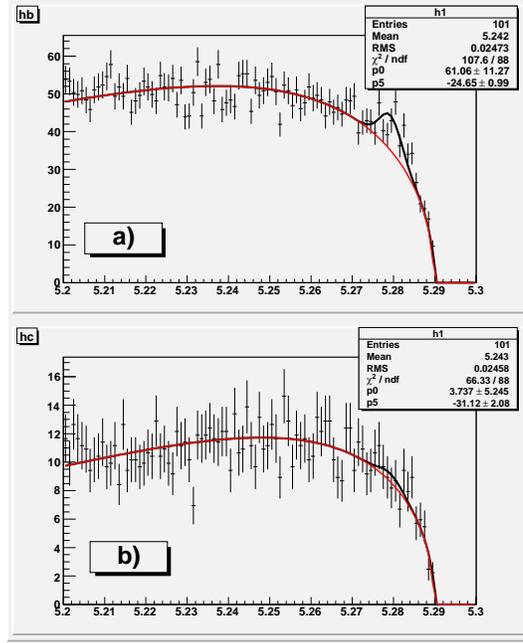


Figure 5.9: Fitting M_{bc} distribution of background MC with Gaussian + ARGUS function, for a) the $B\bar{B}$ background MC and b) the continuum background MC. Gaussian shape parameter is fixed to be the value determined from real $J/\psi X_s$ sample, in the same way of the data fitting. ARGUS shape parameter is floated.

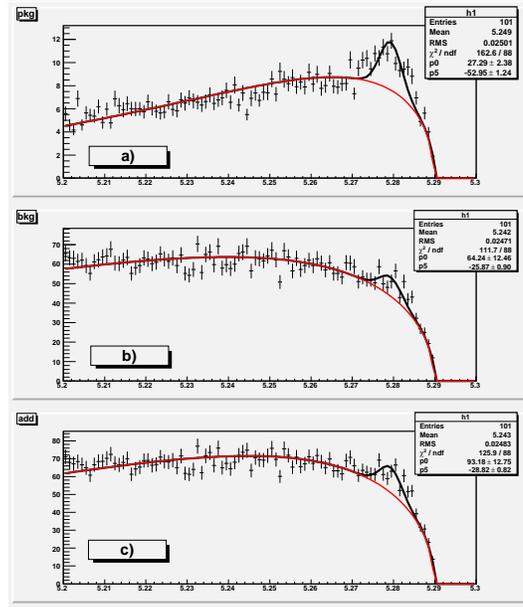


Figure 5.10: Fitting M_{bc} distribution of background MC with Gaussian + ARGUS function, for a) the estimated peaking background, b) the generic background MC and c) sum of a) and b). Gaussian shape parameter is fixed to be the same value as the real data fitting. ARGUS shape parameter is floated. The histogram for a) is the one estimated in Section 5.

Chapter 6

Results with full (605 fb^{-1}) data

After confirming the fitter validity in Section 5.2 and obtaining results consistent with the previous analysis in Chapter C.2, we finally opened the full box of 605 fb^{-1} data.

The statistical significance is defined by $\mathcal{S} = \sqrt{2 \ln(\mathcal{L}_{max}/\mathcal{L}_0)}$, where \mathcal{L}_{max} (\mathcal{L}_0) denotes the likelihood value at the maximum (with the signal yield fixed to zero).

The branching fraction \mathcal{B} for the signal is calculated as $\mathcal{B} = \frac{N_{\text{sig}}}{2N_{B\bar{B}} \epsilon}$, where $N_{B\bar{B}} = (656.7 \pm 8.9) \times 10^6$ is the number of $B\bar{B}$ pairs produced in 605 fb^{-1} and ϵ is the signal efficiency.

We also measured the differential branching fractions as a function of M_{X_s} and $q^2 \equiv M_{\ell^+\ell^-}^2$. To measure the differential branching fraction, we separately fit the partial dataset divided into each M_{X_s} or q^2 regions. For M_{X_s} regions, we divide the dataset into the following 5 regions: $[0.4, 0.6], [0.6, 0.8], [0.8, 1.0], [1.0, 1.4]$ and $[1.4, 2.0] \text{ GeV}/c^2$. For q^2 regions, we divide the dataset into the following 4 regions: $[0.04, 1.0], [1.0, 6.0], [6.0, 14.4]$ and $[14.4, 25.0] \text{ (GeV}/c^2)^2$.

The fit results are summarized in Table 6.1 and Table 6.2. The final branching fraction measurement results will be discussed in Chapter 8.

Table 6.1: Results of the simultaneous fit to the 605 fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample and 605 fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample: number of signal candidates in the M_{bc} signal region, fitted signal yield N_{sig} , and significance.

Mode	Candidates	ARGUS shape	N_{sig}	Significance
$B \rightarrow X_s e^+ e^-$	578	-14.8 ± 1.8	$123.6 \pm 19.5 \pm 2.0$	7.0
$B \rightarrow X_s \mu^+ \mu^-$	432	-15.4 ± 1.9	$118.3 \pm 17.3 \pm 1.5$	7.9
$B \rightarrow X_s \ell^+ \ell^-$	1010	-15.7 ± 1.7	$237.8 \pm 26.4 \pm 2.5$	10.0

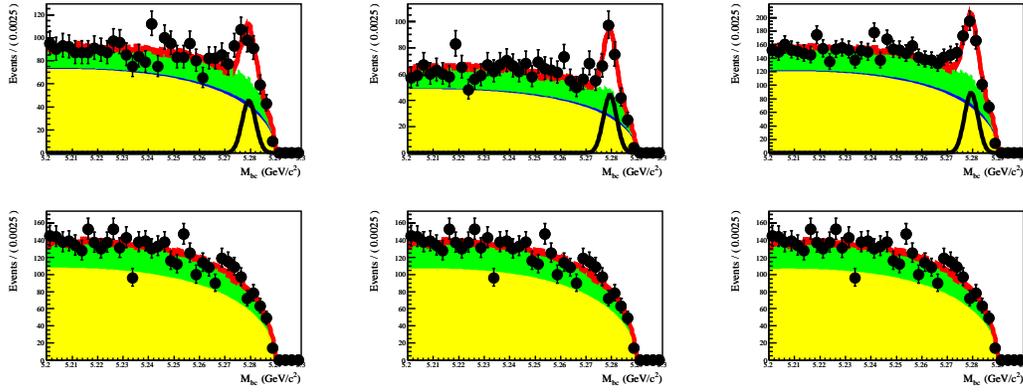


Figure 6.1: M_{bc} distribution of the 605 fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample (upper) and 605 fb^{-1} of $B \rightarrow X_s e^+ \mu^+ \mu^-$ sample (lower). Each column corresponds to $e^+ e^-$ (left), $\mu^+ \mu^-$ (center) and $\ell^+ \ell^-$ (right) cases. Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in the black line.

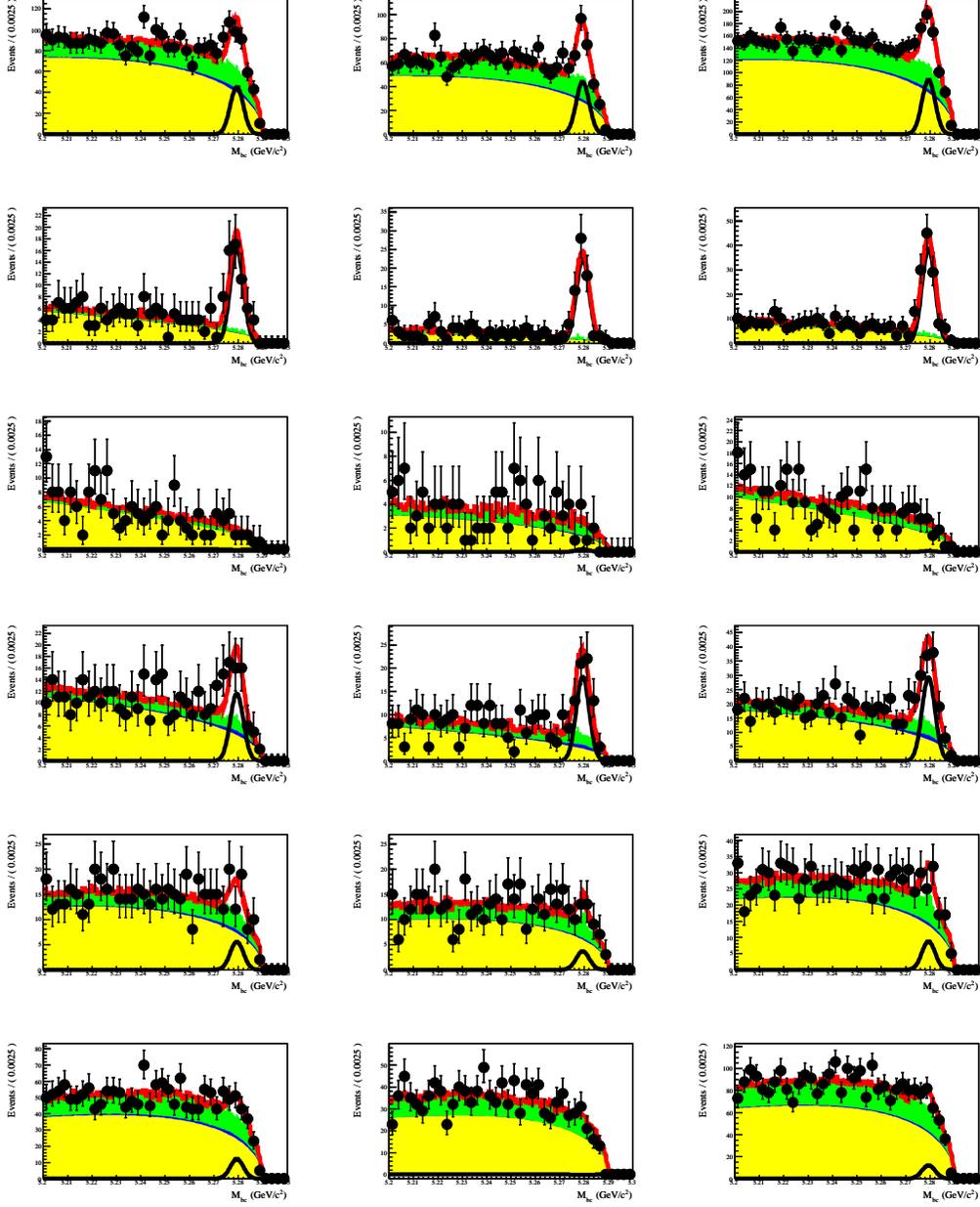


Figure 6.2: M_{bc} distribution of 605 fb^{-1} data for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- (left), $\mu^+\mu^-$ (center) and l^+l^- (right) cases. Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in the black line.

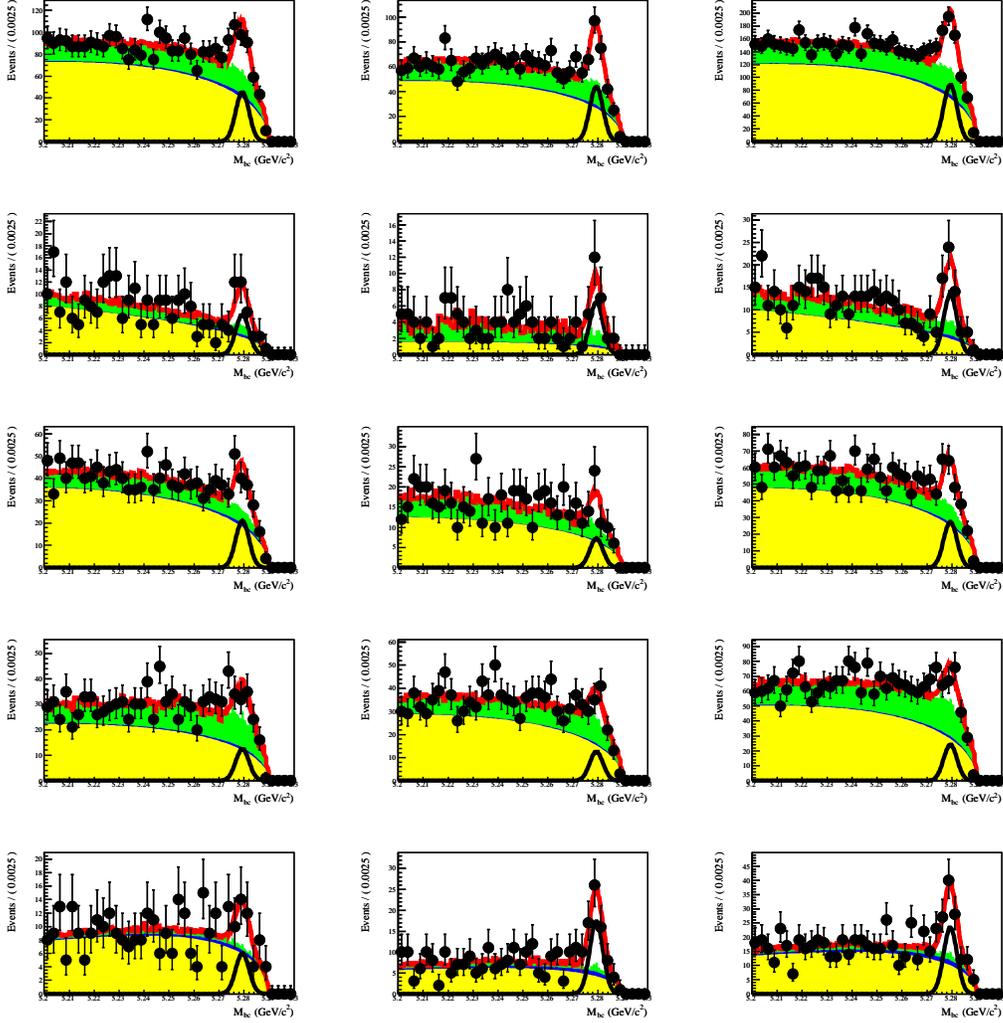


Figure 6.3: M_{bc} distribution of 605 fb^{-1} data for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- (left), $\mu^+\mu^-$ (center) and $\ell^+\ell^-$ (right) cases. Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in the black line.

Table 6.2: Results of the simultaneous fit to the 605 fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample and 605 fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample. Gaussian shape parameters (mean and width, in MeV/c^2) is fixed. N_{pkg} and $N_{\text{scf}}/N_{\text{sig}}$ are also fixed. N_{cand} is the number of candidates in M_{bc} signal region. Fitting results, N_{sig} , ARGUS shape parameter, significance (σ) are also shown.

	$B \rightarrow X_s e^+ e^-$	$B \rightarrow X_s \mu^+ \mu^-$	$B \rightarrow X_s \ell^+ \ell^-$
full sample	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 671.3$, $N_{\text{scf}}/N_{\text{sig}} = 0.40$ $N_{\text{cand}} = 578$, shape = -14.8 ± 1.8 $N_{\text{sig}} = 123.6 \pm 19.5 \pm 2.0$ $\sigma_{\text{stat}} = 7.1$, $\sigma = 7.0$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 559.9$, $N_{\text{scf}}/N_{\text{sig}} = 0.27$ $N_{\text{cand}} = 432$, shape = -15.4 ± 1.9 $N_{\text{sig}} = 118.3 \pm 17.3 \pm 1.5$ $\sigma_{\text{stat}} = 7.9$, $\sigma = 7.9$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 1231.2$, $N_{\text{scf}}/N_{\text{sig}} = 0.33$ $N_{\text{cand}} = 1010$, shape = -15.7 ± 1.7 $N_{\text{sig}} = 237.8 \pm 26.4 \pm 2.5$ $\sigma_{\text{stat}} = 10.1$, $\sigma = 10.0$
$M_{X_s} = [0.4, 0.6]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 26.0$, $N_{\text{scf}}/N_{\text{sig}} = 0.03$ $N_{\text{cand}} = 63$, shape = 3.7 ± 7.8 $N_{\text{sig}} = 44.2 \pm 7.8 \pm 0.9$ $\sigma_{\text{stat}} = 7.9$, $\sigma = 7.7$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 26.8$, $N_{\text{scf}}/N_{\text{sig}} = 0.02$ $N_{\text{cand}} = 72$, shape = 8.7 ± 8.8 $N_{\text{sig}} = 60.0 \pm 8.3 \pm 0.9$ $\sigma_{\text{stat}} = 11.8$, $\sigma = 11.3$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 52.9$, $N_{\text{scf}}/N_{\text{sig}} = 0.03$ $N_{\text{cand}} = 135$, shape = 3.4 ± 7.2 $N_{\text{sig}} = 103.7 \pm 11.5 \pm 1.8$ $\sigma_{\text{stat}} = 13.1$, $\sigma = 12.6$
$M_{X_s} = [0.6, 0.8]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 29.2$, $N_{\text{scf}}/N_{\text{sig}} = 3.37$ $N_{\text{cand}} = 22$, shape = 5.0 ± 7.7 $N_{\text{sig}} = -0.2 \pm 2.0 \pm 0.2$ $\sigma_{\text{stat}} = 0.1$, $\sigma = -5.5$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 34.4$, $N_{\text{scf}}/N_{\text{sig}} = 2.37$ $N_{\text{cand}} = 15$, shape = -5.7 ± 8.3 $N_{\text{sig}} = 0.6 \pm 2.2 \pm 0.2$ $\sigma_{\text{stat}} = 0.3$, $\sigma = 0.3$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 63.6$, $N_{\text{scf}}/N_{\text{sig}} = 2.82$ $N_{\text{cand}} = 37$, shape = 4.6 ± 7.1 $N_{\text{sig}} = 1.4 \pm 2.9 \pm 0.1$ $\sigma_{\text{stat}} = 0.5$, $\sigma = 0.5$
$M_{X_s} = [0.8, 1.0]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 109.3$, $N_{\text{scf}}/N_{\text{sig}} = 0.26$ $N_{\text{cand}} = 90$, shape = -4.8 ± 5.2 $N_{\text{sig}} = 31.6 \pm 7.9 \pm 0.4$ $\sigma_{\text{stat}} = 4.8$, $\sigma = 4.8$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 70.7$, $N_{\text{scf}}/N_{\text{sig}} = 0.18$ $N_{\text{cand}} = 89$, shape = -0.9 ± 5.6 $N_{\text{sig}} = 48.9 \pm 8.3 \pm 0.4$ $\sigma_{\text{stat}} = 8.1$, $\sigma = 8.0$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 131.1$, $N_{\text{scf}}/N_{\text{sig}} = 0.21$ $N_{\text{cand}} = 179$, shape = -6.5 ± 4.7 $N_{\text{sig}} = 78.3 \pm 11.6 \pm 0.4$ $\sigma_{\text{stat}} = 8.5$, $\sigma = 8.5$
$M_{X_s} = [1.0, 1.4]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 109.3$, $N_{\text{scf}}/N_{\text{sig}} = 0.86$ $N_{\text{cand}} = 98$, shape = -16.8 ± 4.3 $N_{\text{sig}} = 14.9 \pm 7.3 \pm 0.5$ $\sigma_{\text{stat}} = 2.2$, $\sigma = 2.2$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 111.2$, $N_{\text{scf}}/N_{\text{sig}} = 0.64$ $N_{\text{cand}} = 83$, shape = -18.4 ± 4.4 $N_{\text{sig}} = 9.9 \pm 6.8 \pm 0.5$ $\sigma_{\text{stat}} = 1.5$, $\sigma = 1.5$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 220.5$, $N_{\text{scf}}/N_{\text{sig}} = 0.74$ $N_{\text{cand}} = 181$, shape = -19.3 ± 3.9 $N_{\text{sig}} = 23.2 \pm 10.2 \pm 0.6$ $\sigma_{\text{stat}} = 2.4$, $\sigma = 2.4$
$M_{X_s} = [1.4, 2.0]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 446.5$, $N_{\text{scf}}/N_{\text{sig}} = 1.27$ $N_{\text{cand}} = 305$, shape = -20.4 ± 2.5 $N_{\text{sig}} = 32.5 \pm 12.7 \pm 1.4$ $\sigma_{\text{stat}} = 2.7$, $\sigma = 2.7$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 316.6$, $N_{\text{scf}}/N_{\text{sig}} = 0.95$ $N_{\text{cand}} = 173$, shape = -21.2 ± 2.6 $N_{\text{sig}} = -0.6 \pm 9.6 \pm 1.0$ $\sigma_{\text{stat}} = 0.1$, $\sigma = \text{nan}$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 763.1$, $N_{\text{scf}}/N_{\text{sig}} = 1.12$ $N_{\text{cand}} = 478$, shape = -20.8 ± 2.2 $N_{\text{sig}} = 29.9 \pm 16.1 \pm 1.4$ $\sigma_{\text{stat}} = 1.9$, $\sigma = 1.9$
$q^2 = [0.04, 1.0]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 56.4$, $N_{\text{scf}}/N_{\text{sig}} = 0.24$ $N_{\text{cand}} = 47$, shape = -0.6 ± 8.4 $N_{\text{sig}} = 18.0 \pm 6.0 \pm 0.1$ $\sigma_{\text{stat}} = 3.6$, $\sigma = 3.6$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 67.6$, $N_{\text{scf}}/N_{\text{sig}} = 0.13$ $N_{\text{cand}} = 33$, shape = -15.6 ± 11.9 $N_{\text{sig}} = 17.5 \pm 5.2 \pm 0.1$ $\sigma_{\text{stat}} = 4.5$, $\sigma = 4.5$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 124.0$, $N_{\text{scf}}/N_{\text{sig}} = 0.19$ $N_{\text{cand}} = 80$, shape = -0.9 ± 7.9 $N_{\text{sig}} = 36.4 \pm 7.9 \pm 0.1$ $\sigma_{\text{stat}} = 5.6$, $\sigma = 5.6$
$q^2 = [1.0, 6.0]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 260.2$, $N_{\text{scf}}/N_{\text{sig}} = 0.29$ $N_{\text{cand}} = 246$, shape = -12.6 ± 3.0 $N_{\text{sig}} = 57.4 \pm 13.0 \pm 0.9$ $\sigma_{\text{stat}} = 4.9$, $\sigma = 4.9$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 158.8$, $N_{\text{scf}}/N_{\text{sig}} = 0.17$ $N_{\text{cand}} = 92$, shape = -12.6 ± 3.6 $N_{\text{sig}} = 19.4 \pm 7.8 \pm 0.3$ $\sigma_{\text{stat}} = 2.8$, $\sigma = 2.8$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 419.0$, $N_{\text{scf}}/N_{\text{sig}} = 0.23$ $N_{\text{cand}} = 338$, shape = -13.7 ± 2.8 $N_{\text{sig}} = 74.4 \pm 15.3 \pm 0.8$ $\sigma_{\text{stat}} = 5.4$, $\sigma = 5.4$
$q^2 = [6.0, 14.4]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 331.3$, $N_{\text{scf}}/N_{\text{sig}} = 0.44$ $N_{\text{cand}} = 216$, shape = -14.3 ± 2.8 $N_{\text{sig}} = 34.1 \pm 11.5 \pm 1.4$ $\sigma_{\text{stat}} = 3.3$, $\sigma = 3.2$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 296.7$, $N_{\text{scf}}/N_{\text{sig}} = 0.28$ $N_{\text{cand}} = 214$, shape = -14.3 ± 2.6 $N_{\text{sig}} = 33.8 \pm 11.7 \pm 1.0$ $\sigma_{\text{stat}} = 3.2$, $\sigma = 3.1$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 628.1$, $N_{\text{scf}}/N_{\text{sig}} = 0.33$ $N_{\text{cand}} = 430$, shape = -15.6 ± 2.4 $N_{\text{sig}} = 64.6 \pm 16.6 \pm 0.8$ $\sigma_{\text{stat}} = 4.2$, $\sigma = 4.2$
$q^2 = [14.4, 25.0]$	mean = 5279.38, width = 2.74 $N_{\text{pkg}} = 23.4$, $N_{\text{scf}}/N_{\text{sig}} = 0.69$ $N_{\text{cand}} = 69$, shape = -26.0 ± 4.5 $N_{\text{sig}} = 16.3 \pm 6.8 \pm 1.1$ $\sigma_{\text{stat}} = 2.7$, $\sigma = 2.7$	mean = 5279.41, width = 2.64 $N_{\text{pkg}} = 36.8$, $N_{\text{scf}}/N_{\text{sig}} = 0.42$ $N_{\text{cand}} = 93$, shape = -25.2 ± 4.8 $N_{\text{sig}} = 46.7 \pm 8.5 \pm 1.1$ $\sigma_{\text{stat}} = 7.1$, $\sigma = 6.9$	mean = 5279.40, width = 2.69 $N_{\text{pkg}} = 60.1$, $N_{\text{scf}}/N_{\text{sig}} = 0.54$ $N_{\text{cand}} = 162$, shape = -25.6 ± 4.2 $N_{\text{sig}} = 63.2 \pm 11.0 \pm 2.2$ $\sigma_{\text{stat}} = 6.8$, $\sigma = 6.5$

Chapter 7

Systematic uncertainties

We have the following sources of systematic uncertainties: 1) signal yield uncertainty, 2) detector modeling uncertainty, 3) signal MC modeling uncertainty and 4) $N_{B\bar{B}}$ uncertainty.

The estimated systematic uncertainties are summarized in Table 7.2. The signal modeling uncertainty contributes mainly in the higher M_{X_s} region. The bin-by-bin breakdown for the signal modeling uncertainty is summarized in 7.3 and 7.4.

7.1 Signal yield uncertainty

Uncertainties affecting the extraction of the signal yield are evaluated as follows:

Signal Gaussian shape parameters

Signal shape parameters are taken from a fit to the real $J/\psi X_s$ sample. The mean and width of the signal Gaussian are randomly varied according to the Gaussian distribution whose width is the statistical error of the fit to the real $J/\psi X_s$ sample. We repeat the final fit to extract signal yield 100 times with varied factors and take the RMS of yield distribution as the systematic error.

Peaking backgrounds

The histogram shape of J/ψ or $\psi(2S)$ peaking background are obtained from the Monte Carlo events which remain after $M_{\ell^+\ell^-}$ veto, where MC events are weighted with a different scaling factor for each X_s mode, so as to reproduce the X_s topology of real $J/\psi X_s$ sample. Each scaling factor is randomly varied according to the Gaussian distribution whose width is the statistical error of the scaling factor. We repeat the final fit to extract signal yield 100 times with varied factors and take the RMS of yield distribution as the systematic error.

The number of higher charmonium peaking background events is varied for $\pm 1\sigma$, where σ is the estimated error.

Hadronic peaking backgrounds

The histogram shape of hadronic peaking background is estimated using the fake rates, which is the probability to mis-identify hadrons (pions) as electrons or muons, obtained from $K_S^0 \rightarrow \pi\pi$ events in the real data. We vary the fake rate for $\pm 1\sigma$, where σ is defined as the statistical error of the fake rate. The fake rate is a function of momentum and direction. When varying the fake rates, we assume full correlation between the fake rates at different regions of momentum/direction.

Cross-feed events

The histogram shape of cross-feed events are estimated from Signal Monte Carlo events. The entries in the bins are varied according to the Gaussian distribution whose width is the statistical error of the MC sample. We repeat the final fit to extract signal yield 100 times with varied factors and take the RMS of yield distribution as the systematic error.

7.2 Detector Modeling Uncertainty

The detector modeling uncertainty is sensitive to the tracking efficiency, lepton/kaon/pion identification efficiency, K_S^0/π^0 efficiency and likelihood ratio cut efficiency.

Tracking efficiency

The uncertainty in the charged track reconstruction efficiency is evaluated in the study by comparing Monte Carlo and real data in D^* decays [60].

Figure 7.1 shows the tracking efficiency errors for charged particles, as a function of momentum. In this analysis, we apply systematic uncertainty of 1.0% per track. We have 3.6 charged tracks on average in each event.

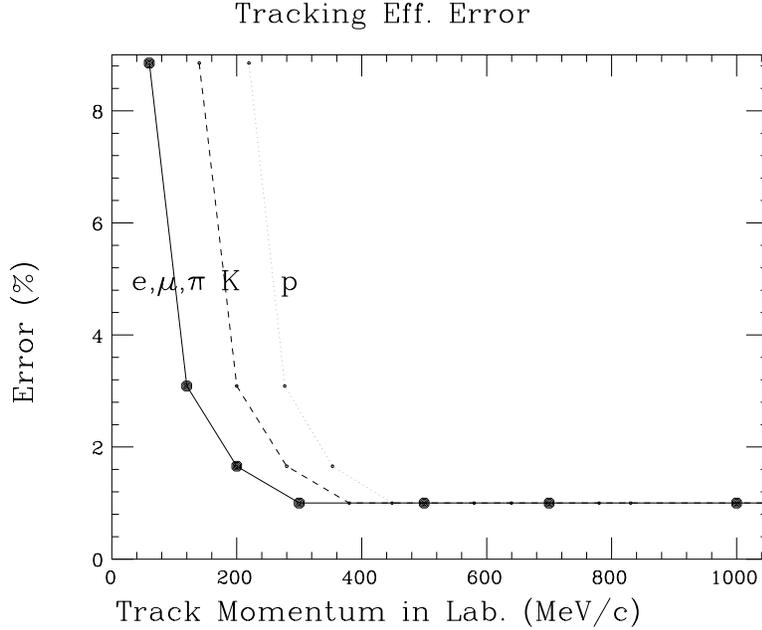


Figure 7.1: Tracking efficiency errors for several charged particles, as a function of momentum [60].

Lepton identification efficiency

The difference of lepton identification efficiency between Monte Carlo and real data is estimated using $\gamma\gamma \rightarrow \ell^+\ell^-$ process for MC/data comparison and inclusive $B \rightarrow XJ/\psi(\rightarrow \ell^+\ell^-)$ process for evaluating possible hadronic effects [61]. We obtain the correction factor C as a function of momentum and direction, where C is defined as $\epsilon_{data} = C \times \epsilon_{MC}$ and ϵ_{data} and ϵ_{MC} are the lepton identification efficiencies of the real data sample and Monte Carlo sample, respectively.

The estimated correction factor C for this analysis is 0.9850 ± 0.0210 per electron and 0.9435 ± 0.0219 per muon. Since $1 - C_{\ell^+}C_{\ell^-}$ is not small enough to assign it as

the systematic errors, we apply the efficiency correction for signal efficiency and take $\Delta(C_{\ell^+}C_{\ell^-})$ as the systematic uncertainty. The correction factor is 3.0%(13.7%) for the $e^+e^-(\mu^+\mu^-)$ mode. The systematic uncertainty from lepton identification efficiency is 2.1%(2.2%) for the $e^+e^-(\mu^+\mu^-)$ mode.

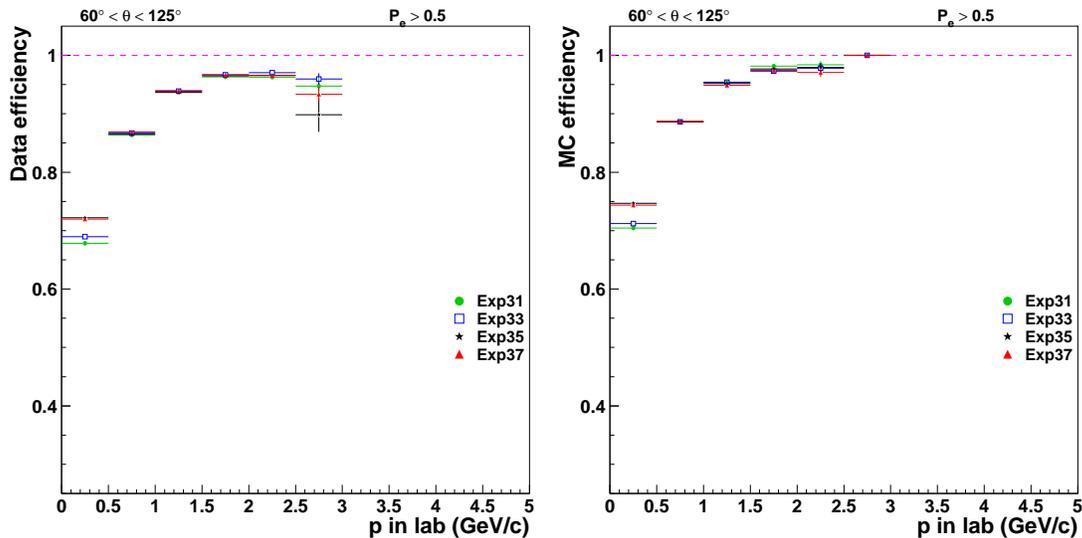


Figure 7.2: Typical electron identification efficiency for data (left) and MC (right), as a function of electron momentum [61].

Kaon identification efficiency

The uncertainty in the reconstruction efficiency of charged kaons is estimated using inclusive D^* decays [62]. The typical kaon identification efficiency as a function of kaon momentum is shown in Figure 7.4. In this analysis, the estimated correction factor C is consistent with one; 1.0011 ± 0.0092 (1.0016 ± 0.0093) for $e^+e^-(\mu^+\mu^-)$ mode. We take the errors of correction factor as the kaon identification systematic uncertainty, 0.9% for e^+e^- modes and 0.9% for the $\mu^+\mu^-$ mode.

Pion identification efficiency

The uncertainty in the reconstruction efficiency of the charged pion is estimated by the same study as the kaon case, using inclusive D^* decays [62]. The typical pion identification efficiency as a function of kaon momentum is shown in Figure 7.5. In this analysis, the estimated correction factor C is 0.9658 ± 0.0058 (0.9696 ± 0.0058) per pion for the $e^+e^-(\mu^+\mu^-)$ mode. We take $1 - (\prod_{i=1}^{n_\pi} C_i)$ as the systematic uncertainty, where n_π is number of charged pions in the final state. The estimated systematic uncertainty from pion identification is 3.4% for e^+e^- modes and 3.0% for the $\mu^+\mu^-$ mode.

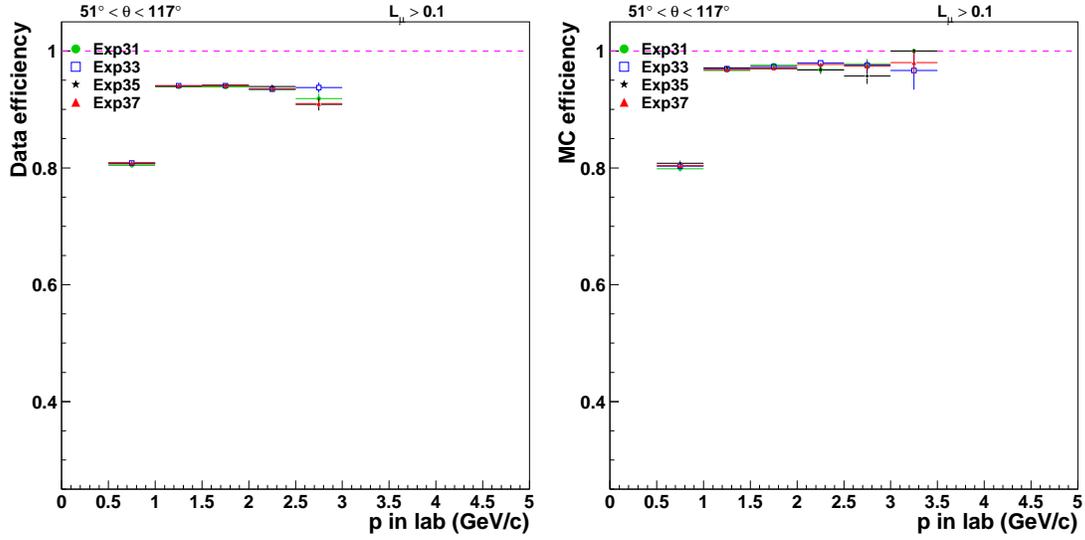


Figure 7.3: Typical muon identification efficiency for data (left) and MC (right), as a function of muon momentum [61].

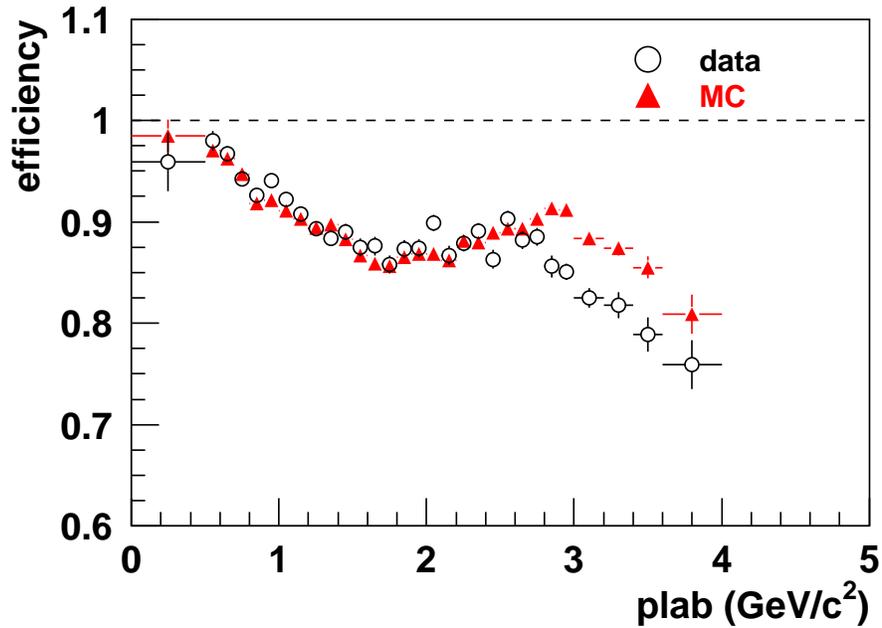


Figure 7.4: The typical kaon identification efficiency as a function of kaon momentum, for real data and Monte Carlo sample [62]. This figure is for $0.017 < \cos \theta < 0.029$.

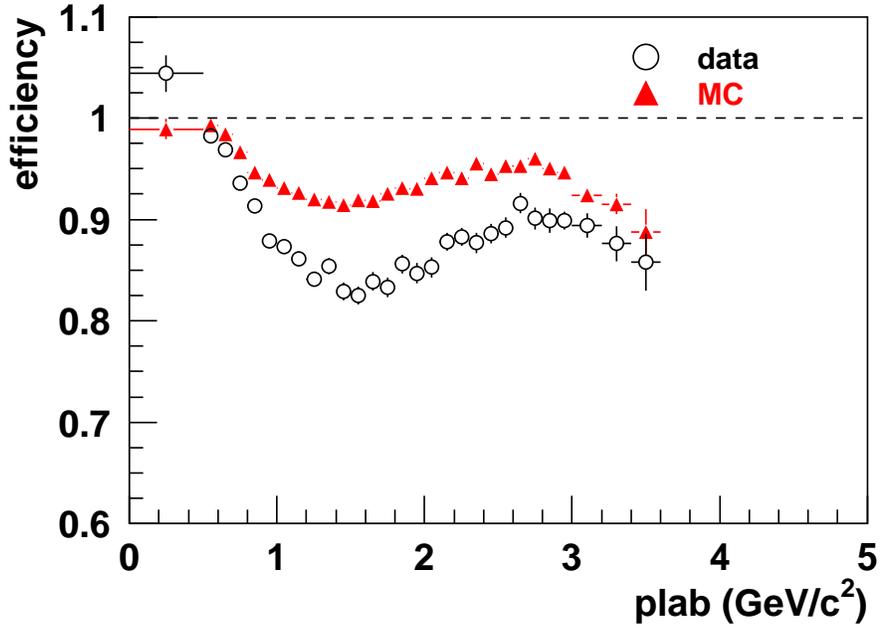


Figure 7.5: The typical pion identification efficiency as a function of pion momentum, for real data and Monte Carlo sample [62]. This figure is for $0.017 < \cos \theta < 0.029$.

K_S^0 and π^0 identification efficiency

The uncertainty in the reconstruction efficiency of K_S^0 is estimated using $D^+ \rightarrow K_S^0 \pi^+$ and $D^+ \rightarrow K^- \pi^+ \pi^+$ decays[63]. The estimated uncertainty is 4.5% per K_S^0 .

The uncertainty in the reconstruction efficiency of π^0 is estimated using η decays [64]. The estimated uncertainty is 3.8% per π^0 .

Likelihood ratio cut uncertainty

To suppress combinatorial background, we apply a likelihood ratio cut. The cut efficiency uncertainty is evaluated by comparing the real $J/\psi X_s$ sample and the signal Monte Carlo, and the efficiency discrepancy between them is taken as the systematic uncertainty. The estimated uncertainty is 5.3% (2.6%) for the $e^+e^-(\mu^+\mu^-)$ mode.

7.3 Signal Model Uncertainty

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 measurements of hadronic moments in semileptonic B decays [49] and the photon spectrum in inclusive $B \rightarrow X_s \gamma$ decays [50]. The varied range of the Fermi motion parameter is $200 \text{ MeV}/c < p_F < 480 \text{ MeV}/c$.

M_{X_s} and $M_{\ell^+\ell^-}$ distribution for varied Fermi motion parameter is shown in Figure 7.6.

This uncertainty is proportional to the fraction of non-resonant X_s .

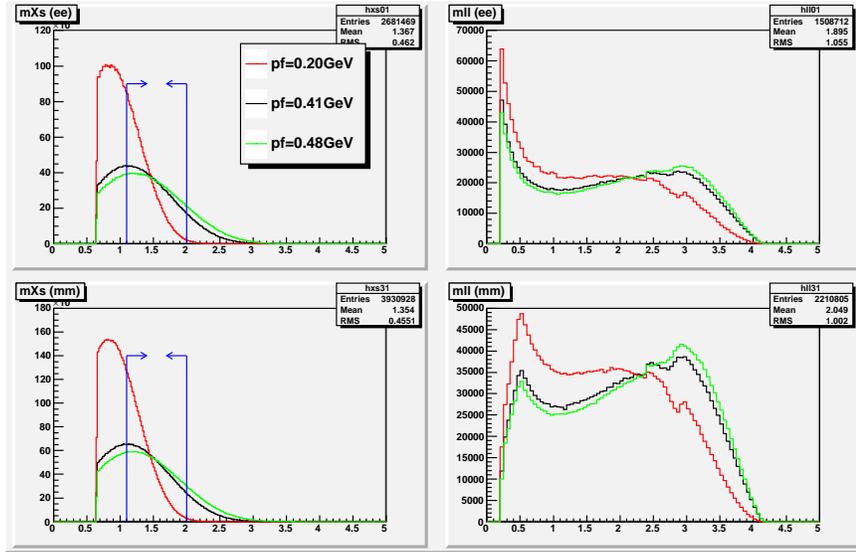


Figure 7.6: M_{X_s} (left) and $M_{\ell^+\ell^-}$ (right) distribution for the various Fermi motion parameter. Upper figures are for e^+e^- mode and lower figures are for $\mu^+\mu^-$ mode. The black, red and green line represents the distribution for $p_F = 410, 200$, and $480 \text{ MeV}/c$, respectively. M_{X_s} distributions are normalized to have the same entries in $1.1 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$.

Monte Carlo mixing ratio

The fractions of exclusive $B \rightarrow K\ell^+\ell^-$ and $B \rightarrow K^*\ell^+\ell^-$ decays are varied according to experimental (Table 2.2) and theoretical uncertainties [2], respectively ¹. This uncertainty is proportional to the fraction of non-resonant X_s .

¹If we take the total branching fraction as the sum of branching fractions in each $M_{\ell^+\ell^-}$ bin, this uncertainty will be much smaller. Instead, we suffer from low statistics by dividing the data samples in each bin.

X_s - K^* transition

In the efficiency calculation, we use $K^*\ell^+\ell^-$ MC sample for $M_{X_s} < 1.1 \text{ GeV}/c^2$ and non-resonant $X_s\ell^+\ell^-$ MC sample for $M_{X_s} > 1.1 \text{ GeV}/c^2$. The transition point in M_{X_s} between $K^*\ell^+\ell^-$ and non-resonant $X_s\ell^+\ell^-$ is varied by $\pm 0.1 \text{ GeV}/c^2$. This uncertainty is NOT proportional to the fraction of non-resonant X_s .

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. The signal efficiencies depend strongly on the X_s decay isotropy. To estimate this uncertainty, we vary the fraction of each X_s decay mode in signal MC so as to be the same fraction of a real $J/\psi X_s$ sample, and get the varied efficiency. The efficiency difference is taken as the systematic error. This uncertainty is proportional to the fraction of non-resonant X_s .

Missing modes

The 18 X_s final states reconstructed in this analysis only capture $\sim 62\%$ of the full inclusive rate. Approximately 21% are due to final states with a K_L^0 meson and their contribution can be determined from the K_S^0 modes. The remaining 17% are modes with too many pions or kaons (two extra kaons may be produced via $s\bar{s}$ popping), modes with η , η' , photons that do not originate from π^0 decays, and anything else.

For final states with $M_{X_s} > 1.1 \text{ GeV}/c^2$, we vary these fractions by $\pm 10\%$ per $N_{\pi^0} = 2$, $\pm 15\%$ per $N_{\pi^0} = 3$, $\pm 20\%$ per $N_{\pi^0} \geq 4$, $\pm 20\%$ for η , $\pm 30\%$ for $N_\pi > 5$, and $\pm 50\%$ for η' and others.

This uncertainty is proportional to the fraction of non-resonant X_s .

7.4 Uncertainty in number of B mesons

The number of B meson pairs in the real data used in this analysis is measured to be $(6.567 \pm 0.089) \times 10^8$. The relative uncertainty is 1.4%.

Table 7.1: Relative systematic uncertainties (in percent) on the $B \rightarrow X_s e^+ e^-$ and $B \rightarrow X_s \mu^+ \mu^-$ branching fraction measurements using simple method. The different contributions are combined in quadrature.

Source	$X_s e^+ e^-$	$X_s \mu^+ \mu^-$
Signal Gaussian shape	± 0.3	± 0.1
$J/\psi, \psi(2S)$ peaking background	± 1.2	± 0.9
Higher ψ peaking background	± 0.9	± 0.9
Hadronic peaking background	$+0.4$ -0.5	$+0.2$ -0.3
Self Cross-feed error	± 0.1	± 0.1
Signal yield subtotal	± 1.6	± 1.3
Tracking efficiency	± 3.6	± 3.6
Lepton identification efficiency	± 2.1	± 2.2
Kaon identification efficiency	± 0.4	± 1.0
π^\pm identification efficiency	± 3.4	± 3.0
K_S^0 efficiency	± 0.9	± 0.9
π^0 efficiency	± 0.5	± 0.5
\mathcal{R} cut efficiency	± 5.3	± 2.6
Detector model subtotal	± 7.6	± 6.0
Fermi motion model	-4.9 $+1.3$	-2.0 $+0.6$
$\mathcal{B}(B \rightarrow K \ell^+ \ell^-)$	± 6.0	± 6.8
$\mathcal{B}(B \rightarrow K^* \ell^+ \ell^-)$	± 6.8	± 6.8
K^*-X_s transition	-6.8 $+2.3$	-7.1 $+2.7$
Hadronization	± 5.8	± 5.5
Missing modes	± 1.7	± 1.7
Signal model subtotal	$+11.2$ -13.7	$+11.5$ -13.4
Monte Carlo statistics	< 0.1	< 0.1
$B\bar{B}$ counting	± 1.4	± 1.4
Total	$+13.6$ -15.8	$+13.1$ -14.8
$M_{X_s} > 2.0 \text{ GeV}/c^2$ extrapolation	$+6.5$ -10.9	$+3.0$ -5.0
Total(extrapolated)	$+15.1$ -19.2	$+13.4$ -15.6

Table 7.2: Relative systematic uncertainties (in percent) in case of sum-up method. Difference of uncertainties from signal yield and Detector model, between sum-up method and simple method is negligible. Therefore only uncertainties from signal model is shown. The different contributions are combined in quadrature.

Source	$X_s \ell^+ \ell^-$
Signal Gaussian shape	± 0.3
$J/\psi, \psi(2S)$ peaking background	± 1.2
Higher ψ peaking background	± 0.9
Hadronic peaking background	$^{+0.4}_{-0.5}$
Self Cross-feed error	± 0.1
Signal yield subtotal	± 1.6
Tracking efficiency	± 3.6
Lepton identification efficiency	± 2.2
Kaon identification efficiency	± 1.0
π^\pm identification efficiency	± 3.4
K_S^0 efficiency	± 0.9
π^0 efficiency	± 0.5
\mathcal{R} cut efficiency	± 5.3
Detector model subtotal	± 7.6
Fermi motion model	$^{+3.6}_{-0.5}$
$\mathcal{B}(B \rightarrow K \ell^+ \ell^-)$	0.0
$\mathcal{B}(B \rightarrow K^* \ell^+ \ell^-)$	0.0
$K^* - X_s$ transition	$^{+5.5}_{-0.0}$
Hadronization	± 6.0
Missing modes	± 1.7
Signal model subtotal	$^{+9.1}_{-6.2}$
Monte Carlo statistics	< 0.1
$B\bar{B}$ counting	± 1.4
Total	$^{+12.0}_{-9.9}$
$M_{X_s} > 2.0 \text{ GeV}/c^2$ extrapolation	$^{+5.6}_{-9.4}$
Total(extrapolated)	$^{+13.2}_{-13.7}$

Table 7.3: Relative systematic uncertainties (in percent) from signal model in each M_{X_s} bin for $B \rightarrow X_s \ell^+ \ell^-$ mode.

Source	M_{X_s} bin 1	M_{X_s} bin 2	M_{X_s} bin 3	M_{X_s} bin 4	M_{X_s} bin 5
Fermi motion	0	0	0	+1.13 -0.14	+9.21 -1.28
$X_s - K^*$ transition	0	0	0	+10.95 -0.00	0
Hadronization	0	0	0	± 7.87	± 13.85
Missing modes	0	0	0	± 0.28	± 0.07

Table 7.4: Relative systematic uncertainties (in percent) from signal model in each $M_{\ell^+ \ell^-}$ bin for $B \rightarrow X_s \ell^+ \ell^-$ mode..

Source	$M_{\ell^+ \ell^-}$ bin 1	$M_{\ell^+ \ell^-}$ bin 2	$M_{\ell^+ \ell^-}$ bin 3	$M_{\ell^+ \ell^-}$ bin 4
Fermi motion	+0.00 -0.01	+0.03 -0.15	+1.09 -0.13	+0.00 -0.01
$X_s - K^*$ transition	+1.50 -0.00	+2.34 -0.00	+3.36 -0.62	+2.35 -1.83
Hadronization	± 2.88	± 2.17	± 1.80	± 0.13
Missing modes	± 2.09	± 2.05	± 1.25	± 0.15

Chapter 8

Final results

In this chapter, we discuss the final results of the branching fraction measurements.

8.1 A simple method

Using the fit results in Chapter 6 and the systematic uncertainty study in Chapter 7, we obtain the branching fractions results shown in Table 8.1 and shown in Table 8.2.

Table 8.1: Signal yield (N_{sig}), significance, signal efficiency (ϵ) and branching fraction (\mathcal{B}) are shown in the table. Note that these branching fractions include only $M_{X_s} < 2.0\text{GeV}/c^2$

Mode	N_{sig}	Significance	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$B \rightarrow X_s e^+ e^-$	$123.6 \pm 19.5 \pm 2.0$	7.0	$2.56 \pm 0.01^{+0.34}_{-0.33}$	$3.67 \pm 0.58^{+0.49}_{-0.47}$
$B \rightarrow X_s \mu^+ \mu^-$	$118.3 \pm 17.3 \pm 1.5$	7.9	$2.67 \pm 0.01^{+0.36}_{-0.34}$	$3.38 \pm 0.49^{+0.46}_{-0.44}$
$B \rightarrow X_s \ell^+ \ell^-$	$237.8 \pm 26.4 \pm 2.5$	10.0	$2.62 \pm 0.00^{+0.35}_{-0.33}$	$3.46 \pm 0.38^{+0.47}_{-0.45}$

Table 8.2: Signal yields (N_{sig}), signal efficiency (ϵ) and Branching fractions (\mathcal{B}) for each bin of M_{X_s} and q^2 . Note that q^2 branching fractions are for $M_{X_s} < 2.0 \text{ GeV}$

Mode	N_{sig}	Significance	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$M_{X_s} (\text{GeV}/c^2)$				
= [0.4, 0.6]	$103.7 \pm 11.5 \pm 1.8$	12.6	$8.44 \pm 0.01^{+0.64}_{-0.64}$	$0.47 \pm 0.05^{+0.04}_{-0.04}$
= [0.6, 0.8]	$1.4 \pm 2.9 \pm 0.1$	0.5	$3.87 \pm 0.04^{+0.29}_{-0.29}$	$0.01 \pm 0.03^{+0.00}_{-0.00}$
= [0.8, 1.0]	$78.3 \pm 11.6 \pm 0.4$	8.5	$3.90 \pm 0.01^{+0.30}_{-0.30}$	$0.77 \pm 0.11^{+0.06}_{-0.06}$
= [1.0, 1.4]	$23.2 \pm 10.2 \pm 0.6$	2.4	$1.68 \pm 0.01^{+0.26}_{-0.18}$	$0.52 \pm 0.23^{+0.08}_{-0.06}$
= [1.4, 2.0]	$29.9 \pm 16.1 \pm 1.4$	1.9	$0.99 \pm 0.01^{+0.18}_{-0.16}$	$1.15 \pm 0.62^{+0.22}_{-0.19}$
$q^2 ((\text{GeV}/c^2)^2) \quad (M_{X_s} < 2.0 \text{ GeV}/c^2)$				
= [0.04, 1.0]	$36.4 \pm 7.9 \pm 0.1$	5.6	$1.97 \pm 0.01^{+0.17}_{-0.17}$	$0.70 \pm 0.15^{+0.06}_{-0.06}$
= [1.0, 6.0]	$74.4 \pm 15.3 \pm 0.8$	5.4	$2.81 \pm 0.01^{+0.24}_{-0.23}$	$1.01 \pm 0.21^{+0.09}_{-0.08}$
= [6.0, 14.4]	$64.6 \pm 16.6 \pm 0.8$	4.2	$1.85 \pm 0.01^{+0.16}_{-0.15}$	$1.33 \pm 0.34^{+0.12}_{-0.11}$
= [14.4, 25.0]	$63.2 \pm 11.0 \pm 2.2$	6.5	$7.17 \pm 0.02^{+0.57}_{-0.56}$	$0.34 \pm 0.06^{+0.03}_{-0.03}$

8.2 The sum-up method

The total branching fraction measurements shown in Table 8.1, however, suffer from a large systematic uncertainty from signal MC mixing ratio ($K\ell^+\ell^-$, $K^*\ell^+\ell^-$ and non-resonant $X_s\ell^+\ell^-$). In place of this, we will take the sum of the branching fraction in each M_{X_s} bin as the final result. In this case, the systematic uncertainty from MC mixing ratio is much smaller. As a drawback, we have larger statistical error since we divide the data sample for each M_{X_s} bins.

The branching fraction results by summing up M_{X_s} bins are shown in Table 8.3.

Table 8.3: Sum of branching fractions for M_{X_s} bins. Compared to the results in Table 8.1, we don't suffer from the efficiency uncertainty from branching fraction assumption when mixing 3 signal MC samples. Note that these branching fractions include only $M_{X_s} < 2.0 \text{ GeV}/c^2$.

Mode	$\mathcal{B} (\times 10^{-6})$
$B \rightarrow X_s e^+ e^-$	$4.04 \pm 0.98^{+0.47}_{-0.43}$
$B \rightarrow X_s \mu^+ \mu^-$	$1.84 \pm 0.87^{+0.13}_{-0.12}$
$B \rightarrow X_s \ell^+ \ell^-$	$2.92 \pm 0.67^{+0.24}_{-0.21}$

8.3 Extrapolation for entire M_{X_s} region (over $M_{X_s} > 2.0 \text{ GeV}/c^2$)

The results in Table 8.3 are for $M_{X_s} < 2.0 \text{ GeV}/c^2$. We extrapolate the results to entire M_{X_s} region, based on the signal MC. The extrapolation suffers from additional systematic uncertainty of Fermi motion model parameter.

According to our signal MC model, the extrapolation factor for $X_s\ell^+\ell^-$ branching fraction $\frac{\mathcal{B}(M_{X_s} > 2.0 \text{ GeV}/c^2)}{\mathcal{B}(1.1 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2)}$ is $0.185^{+0.106}_{-0.178}$ for dielectron mode and $0.173^{+0.100}_{-0.165}$ for dimuon mode. The errors come from the systematic uncertainty of Fermi motion parameters in the signal model. The extrapolated results are shown in Table 8.4,

Table 8.4: Total branching fractions extrapolated for the entire M_{X_s} region.

Mode	$\mathcal{B} (\times 10^{-6})$
$B \rightarrow X_s e^+ e^-$	$4.59 \pm 1.15^{+0.56}_{-0.51}$
$B \rightarrow X_s \mu^+ \mu^-$	$1.91 \pm 1.02^{+0.15}_{-0.14}$
$B \rightarrow X_s \ell^+ \ell^-$	$3.22 \pm 0.79^{+0.28}_{-0.25}$

8.4 MC normalization from measured branching fraction

For the final results for the differential branching fractions, we use the measured branching fractions to normalize the Monte Carlo and calculate the efficiencies.

As shown in Section 4.6, to normalize and mix 3 kinds of signal MC sample, we assume the HFAG world averages (See Table 2.2). We use the branching fraction of the dilepton mode for the $B \rightarrow K\ell^+\ell^-$ MC sample (0.43×10^{-6}), and the dimuon mode for the $B \rightarrow K^*\ell^+\ell^-$ (1.08×10^{-6}) and non-resonant $B \rightarrow X_s\ell^+\ell^-$ MC samples ($(4.3 - 0.43 - 1.08) \times 10^{-6} = 2.79 \times 10^{-6}$). The last one is mainly determined by the BELLE 140 fb^{-1} measurement.

To improve our signal MC model, we will use the branching fraction measured in this analysis with 605 fb^{-1} data for the final differential branching fraction measurement. We have measured the branching fraction of $\mathcal{B}(B \rightarrow X_s\ell^+\ell^-)$ to be $(3.00 \pm 0.68_{-0.21}^{+0.25}) \times 10^{-6}$ (see Table 8.3). Hence, for the branching fraction for the non-resonant $X_s\ell^+\ell^-$ sample, we will use $(3.00 - 0.43 - 1.08) \times 10^{-6} = 1.49 \times 10^{-6}$ from now on. For the $K\ell^+\ell^-$ and $K^*\ell^+\ell^-$ we keep the HFAG values.

The differential branching fractions for M_{X_s} or q^2 bins are shown in Figure 8.1 and 8.2.

Table 8.5: Differential branching fractions for q^2 bins. Note that the measured branching fractions are used for MC normalization. So the efficiencies and the branching fraction are different from those in Table 8.2. Note that these branching fractions include the entire M_{X_s} region (including $M_{X_s} > 2.0 \text{ GeV}/c^2$).

bin	N_{sig}	Significance	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$q^2((\text{GeV}/c^2)^2) \quad (M_{X_s} < 2.0 \text{ GeV}/c^2)$				
= [0.04, 1.0]	$36.4 \pm 7.9 \pm 0.1$	5.6	$2.41 \pm 0.01_{-0.20}^{+0.21}$	$0.58 \pm 0.13_{-0.05}^{+0.05}$
= [1.0, 6.0]	$74.4 \pm 15.3 \pm 0.8$	5.4	$3.48 \pm 0.01_{-0.28}^{+0.30}$	$0.81 \pm 0.17_{-0.07}^{+0.07}$
= [6.0, 14.4]	$64.6 \pm 16.6 \pm 0.8$	4.2	$2.19 \pm 0.01_{-0.17}^{+0.19}$	$1.12 \pm 0.29_{-0.09}^{+0.10}$
= [14.4, 25.0]	$63.2 \pm 11.0 \pm 2.2$	6.5	$7.47 \pm 0.02_{-0.58}^{+0.59}$	$0.32 \pm 0.06_{-0.03}^{+0.03}$
$q^2((\text{GeV}/c^2)^2) \quad (\text{entire } M_{X_s} \text{ region})$				
= [0.04, 1.0]				$0.77 \pm 0.17_{-0.07}^{+0.07}$
= [1.0, 6.0]				$0.99 \pm 0.20_{-0.08}^{+0.09}$
= [6.0, 14.4]				$1.16 \pm 0.30_{-0.09}^{+0.10}$
= [14.4, 25.0]				$0.32 \pm 0.06_{-0.03}^{+0.03}$

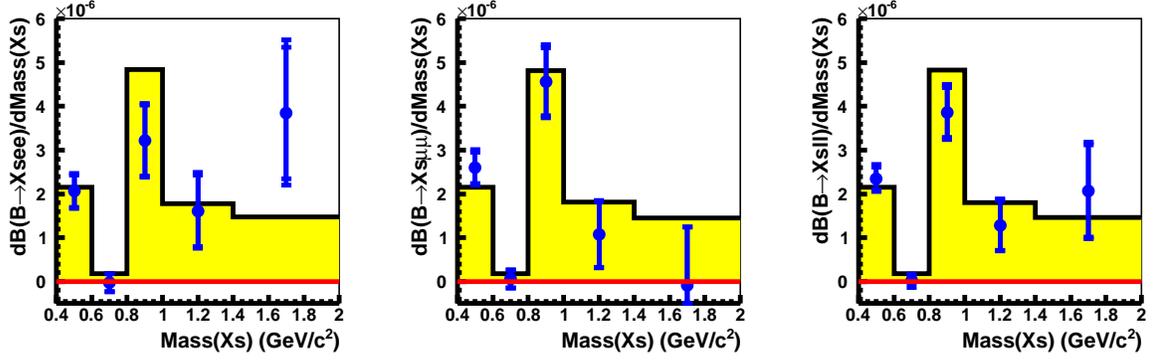


Figure 8.1: Differential branching fractions for M_{X_s} bins, by simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 605 fb^{-1} results and the histogram represents the signal MC simulation. The signal MC normalization for non-resonant $X_s \ell^+ \ell^-$ sample is based on the measurements in this 605 fb^{-1} analysis, not the HFAG values.

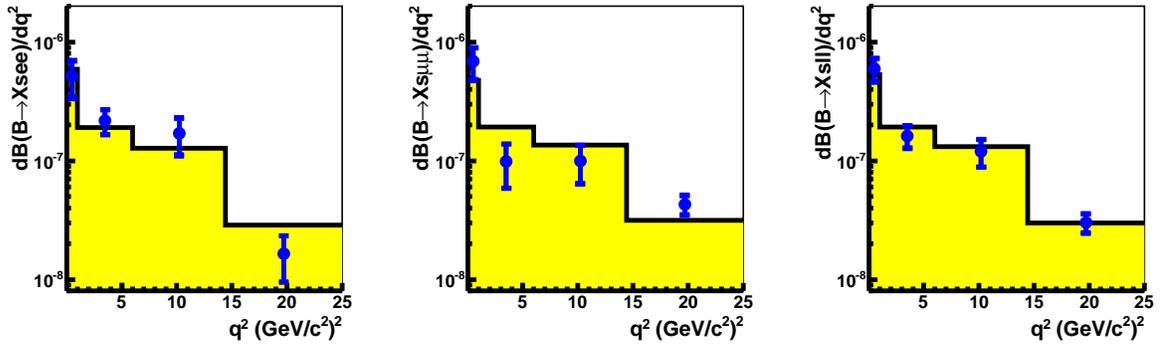


Figure 8.2: Differential branching fractions for q^2 bins, by simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 605 fb^{-1} results and the histogram represents the signal MC simulation. The signal MC normalization for non-resonant $X_s \ell^+ \ell^-$ sample is based on the measurements in this 605 fb^{-1} analysis, not the HFAG values.

Chapter 9

Conclusion

9.1 Summary of the results

We have measured an FCNC decay $B \rightarrow X_s \ell^+ \ell^-$ with a data sample of 605 fb^{-1} . The numbers of observed signal events N_{sig} are:

$$\begin{aligned} N_{\text{sig}}(B \rightarrow X_s e^+ e^-) &= 123.6 \pm 19.5 \pm 2.0 \text{ (significance} = 7.0 \sigma), \\ N_{\text{sig}}(B \rightarrow X_s \mu^+ \mu^-) &= 118.3 \pm 17.3 \pm 1.5 \text{ (significance} = 7.9 \sigma), \\ N_{\text{sig}}(B \rightarrow X_s \ell^+ \ell^-) &= 237.8 \pm 26.4 \pm 2.5 \text{ (significance} = 10.0 \sigma). \end{aligned}$$

Here the first and second errors are statistical and systematic errors, respectively.

The measured branching fractions for $M_{\ell^+ \ell^-} > 0.2 \text{ GeV}/c^2$ are

$$\begin{aligned} \mathcal{B}(B \rightarrow X_s e^+ e^-) &= \left(4.59 \pm 1.15(\text{stat.})_{-0.51}^{+0.56}(\text{syst.}) \right) \times 10^{-6}, \\ \mathcal{B}(B \rightarrow X_s \mu^+ \mu^-) &= \left(1.91 \pm 1.02(\text{stat.})_{-0.14}^{+0.15}(\text{syst.}) \right) \times 10^{-6}, \\ \mathcal{B}(B \rightarrow X_s \ell^+ \ell^-) &= \left(3.22 \pm 0.79(\text{stat.})_{-0.25}^{+0.28}(\text{syst.}) \right) \times 10^{-6} \end{aligned}$$

These results are consistent with the SM predictions within the errors. The fraction of total error is $\sim 26\%$ for $B \rightarrow X_s \ell^+ \ell^-$ mode, which is the current best measurements in the world. The latest HFAG world averages including these results are summarized in Figure 9.1.

We have also measured the differential branching fractions. The fraction of total error is 10–50% for M_{X_s} bins and $\sim 20\%$ for q^2 bins. We didn't observe any significant discrepancy from the SM predictions either.

9.2 Discussion

Systematic error reduction

We have performed the world-best measurements using about 4 times more data than the previous analysis. The results of this analysis are consistent with the SM predictions within the errors, and we don't observe any significant discrepancy from the SM predictions.

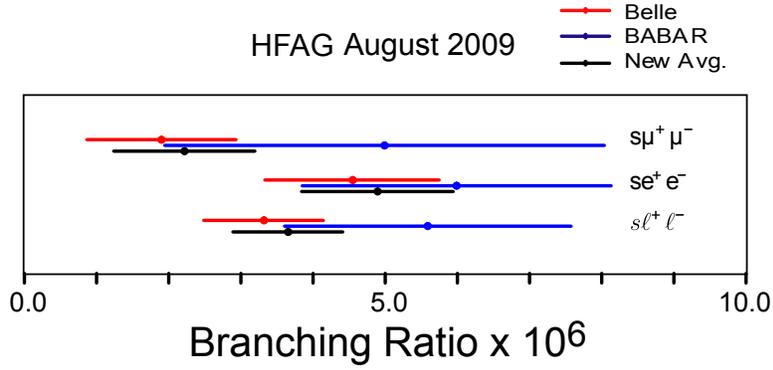


Figure 9.1: The latest HFAG averages for Electroweak radiative B decays (2009 summer).

In this analysis, we reduce the systematic uncertainty as follows:

- The theoretical uncertainty of the M_{X_s} shape is the largest systematic error source in the previous $B \rightarrow X_s \ell^+ \ell^-$ analysis. We reduce this uncertainty by dividing the selected event samples in M_{X_s} bins and applying the fit in each M_{X_s} bin. This new method allows us to perform more model-independent measurements. Instead, the statistical error will increase because we divide the event sample to the M_{X_s} bins.
- The uncertainty of the ARGUS shape parameter fixed in the fit is also large systematic error source in the previous $B \rightarrow X_s \ell^+ \ell^-$ analysis. We reduce this uncertainty by floating the ARGUS shape parameter in the fit. Instead, the statistical error of the fit will increase because we have wider parameter space in the fit.

As a result, we can greatly reduce the relative systematic error from 21% (previous 140 fb^{-1} analysis) to 13%. As a drawback, we cannot reduce the statistical error as we expected from the statistics increase. In total, the $B \rightarrow X_s \ell^+ \ell^-$ BR measurement accuracy becomes 26.1% in this thesis (28.9% in the previous analysis).

Updated constraints on Wilson coefficients: branching fraction

Table 9.1 shows the branching fractions for the previously measured, measured in this thesis, the SM prediction, and the prediction with the non-SM C_7 sign ($C_7 = -C_7^{SM}$), respectively. Comparing these branching fractions, both our results in $M_{\ell^+\ell^-} > 0.2 \text{ GeV}/c^2$ and $1 < M_{\ell^+\ell^-} < 6 (\text{GeV}/c^2)^2$ regions favor the SM predictions than the predictions with the non-SM C_7 sign. Here we should note that the recent measurements of $B \rightarrow K^* \ell^+ \ell^-$ forward-backward asymmetry [15] shows the opposite implication to favor the non SM-like C_7 sign, as shown in Section 2.4.

Table 9.1: The $B \rightarrow X_s \ell^+ \ell^-$ branching fractions (in $\times 10^{-6}$) measured in this thesis and predicted by the theoretical calculations. The new results shows that the non SM-like sign of C_7 is unlikely.

$M_{\ell^+ \ell^-}$ range	World average in 2005 [1]	Measured by this thesis	SM	$C_7 = -C_7^{SM}$
$M_{\ell^+ \ell^-} > 0.2 \text{ GeV}/c^2$	4.5 ± 1.1	$3.3 \pm 0.80^{+0.37}_{-0.44}$	4.4 ± 0.7	8.8 ± 0.7
$1 < M_{\ell^+ \ell^-}^2 < 6 (\text{GeV}/c^2)^2$	1.60 ± 0.5	$0.99 \pm 0.20^{+0.09}_{-0.08}$	1.57 ± 0.16	3.30 ± 0.25

Updated constraints on Wilson coefficients: C_9, C_{10} plane

Using the new branching fraction measurement results, the constraints on Wilson coefficients C_9, C_{10} are updated. Figure 9.2 shows the preliminary result of updated constraints on $C_9 - C_{10}$ plane, provided by H. Haisch in private communication (for previous constraints, see Figure 2.4). As you can see, the shape of allowed region is no longer donut-like but crescent. The updated results put the constraints on new physics parameters, for example, on the Extended MFV model (See Figure 9.3).

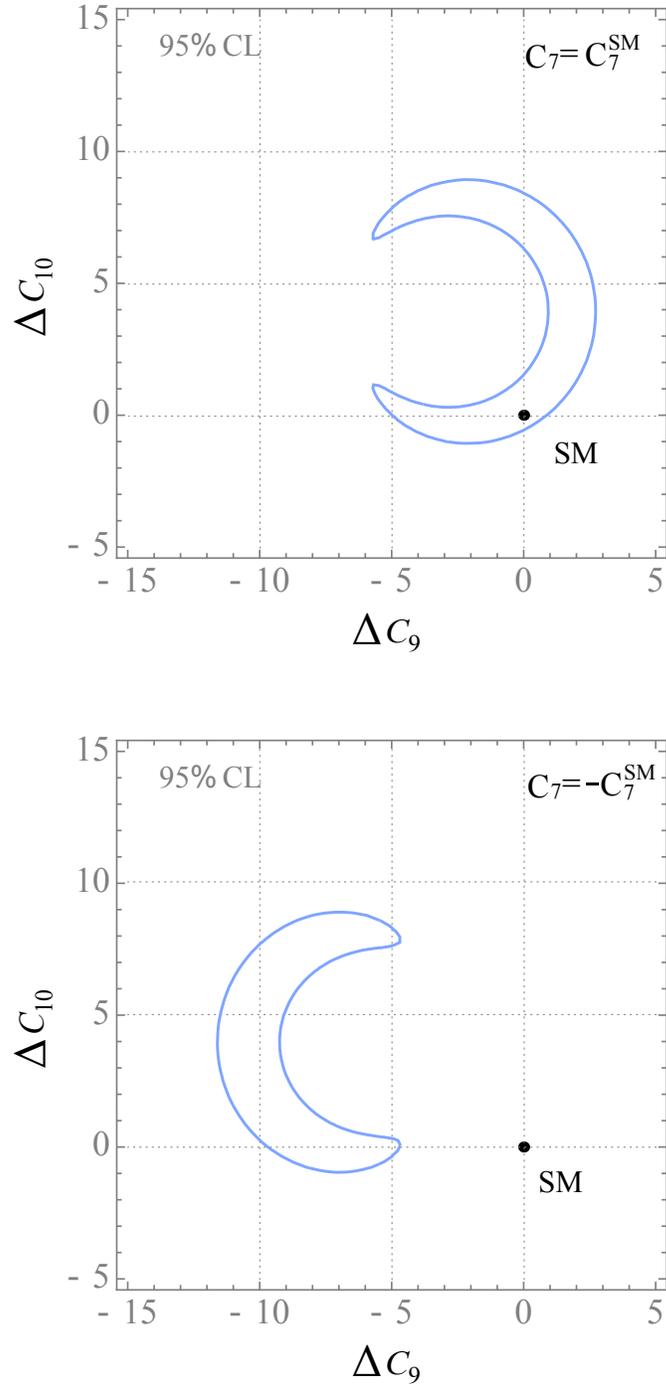


Figure 9.2: Model-independent constraints on additive new physics contributions to $C_{9,10}$ at 95% C.L for the SM-like (upper plot) and opposite (lower plot) sign of C_7 . The dot at $(C_9, C_{10}) = (0, 0)$ indicates the SM case for $C_{9,10}$.

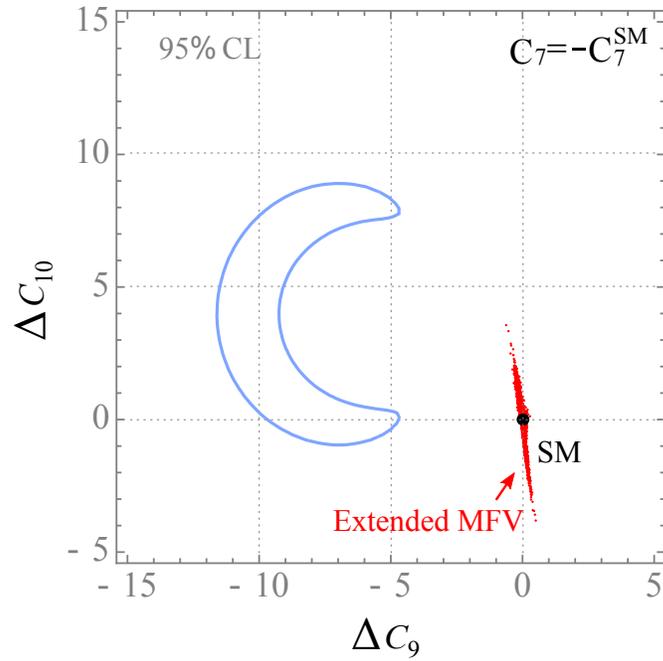
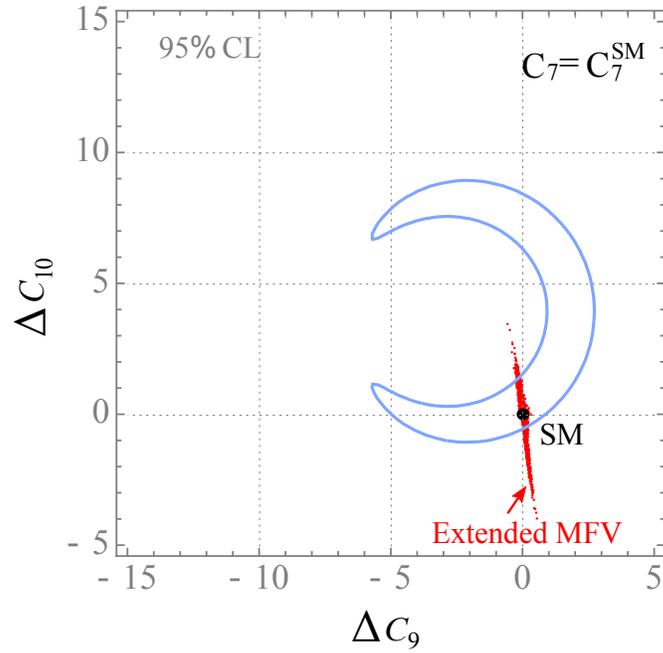


Figure 9.3: The constraints from this $B \rightarrow X_s \ell^+ \ell^-$ measurement on new physics parameters in the Extended MFV model, in case of SM-like (upper) and opposite (lower) sign of C_7 . Red dots represents the possible $C_{9,10}$ values in the Extended MFV model (shown in Figure 2.10).

Appendix A

The new peaking background sources

A.1 Charmonium higher resonances

We found that not only $J/\psi X_s$ or $\psi(2S)X_s$, but also the higher resonances like $\psi(3770)X_s$, $\psi(4040)X_s$ and $\psi(4160)X_s$ can be a peaking background events.

Unfortunately, we don't have any branching fraction measurements for such higher resonances, except for $\text{BR}(B \rightarrow \psi(3770)K^+) = (0.49 \pm 0.13) \times 10^{-3}$. Therefore we need many assumptions for estimating this peaking background.

First we assume $\text{BR}(B \rightarrow \psi(4040)K^+)$ and $\text{BR}(B \rightarrow \psi(4160)K^+)$ to be $(3.3 \pm 3.3) \times 10^{-3}$, which is $50\% \pm 50\%$ of the 90% CL upper limit of $\text{BR}(B \rightarrow \psi(3770)K^+)$. Second we assume isospin conservation, $\text{BR}(B^+ \rightarrow \psi K^+) = \text{BR}(B^0 \rightarrow \psi K^0) = 2 \text{BR}(B^0 \rightarrow \psi K_S^0)$. We also assume $\text{BR}(\psi \rightarrow \mu^+ \mu^-)$ to be same as $\text{BR}(\psi \rightarrow e^+ e^-)$ in PDG.

We have generated exclusive $B \rightarrow \psi K$ and obtained efficiencies for each mode. These assumption and efficiency calculation leads to 3.0 ± 1.8 event of $\text{BR}(B \rightarrow \psi K)$ at 605 fb^{-1} (see Table A.1).

Finally, we assume $\frac{N_{\text{rec}}(B \rightarrow \psi X_s)}{N_{\text{rec}}(B \rightarrow \psi K)} = 1.33$, which is estimated from the M_{X_s} distributions of signal MC sample ($4.00 \text{ GeV}/c^2 < M_{X_s} < 4.20 \text{ GeV}/c^2$).

A.2 Hadronic with neutrino

We found that not only $B \rightarrow X_s hh$, but also $B \rightarrow D^{(*)} n \pi, D \rightarrow K \ell \nu$ or $B \rightarrow D^{(*)} \ell \nu, D \rightarrow K \pi$ can be a peaking background events. In such events, one pion is misidentified as a lepton. The missing energy of the neutrino is compensated by picking up additional gammas (as pi0 daughters) from the other B meson, therefore these events remains in M_{bc} signal region.

We repeat the same selection without the lepton identification requirements for one lepton (and keep them for the other lepton), picking up one hadron (mostly pion in this case) and one lepton. Then we weight each event with the fake rate, which is also used to estimate hadronic peaking background. Figure A.1 shows the weighted M_{bc}

Table A.1: Peaking background from $B \rightarrow \psi X_s$ charmonium higher resonances. We have 6.9 ± 4.3 events at 605 fb^{-1} , where the correlation is taken into account for the error.

Mode	$\text{BR}(B \rightarrow \psi K)(\times 10^{-4})$	$\text{BR}(\psi \rightarrow \ell^+ \ell^-)(\times 10^{-5})$	efficiency (%)	$N_{\text{pkg}} @ 605 \text{ fb}^{-1}$
$\psi(3770)K^+$	4.4 ± 1.3	0.97 ± 0.07	4.9 ± 0.2	0.3 ± 0.1
$\psi(4040)K^+$	3.3 ± 3.3	1.07 ± 0.16	24.7 ± 0.5	1.1 ± 1.1
$\psi(4160)K^+$	3.3 ± 3.3	0.81 ± 0.09	25.3 ± 0.5	0.9 ± 0.9
$\psi(3770)K_S^0$	2.2 ± 0.7	0.97 ± 0.07	2.7 ± 0.2	0.1 ± 0.0
$\psi(4040)K_S^0$	1.7 ± 1.7	1.07 ± 0.16	13.8 ± 0.5	0.3 ± 0.3
$\psi(4160)K_S^0$	1.7 ± 1.7	0.81 ± 0.09	12.9 ± 0.5	0.2 ± 0.2
ψK				3.0 ± 1.8
ψX_s				3.8 ± 2.3

distributions of the hadronic with neutrino peaking background events.

By the way, we select one lepton-like particle in this estimation. That particle could be a pion with a probability of the fake rake. This means that the selected events are $\ell\pi + \pi\pi \times \text{rate}$. Then we weight each events with the fake rate of the other pion. Therefore the estimated M_{bc} distribution includes the full component of $B \rightarrow X_s hh$ events, not only $B \rightarrow X_s hh$ events.

This peaking background is included in the fitter using the histogram shape as its PDF.

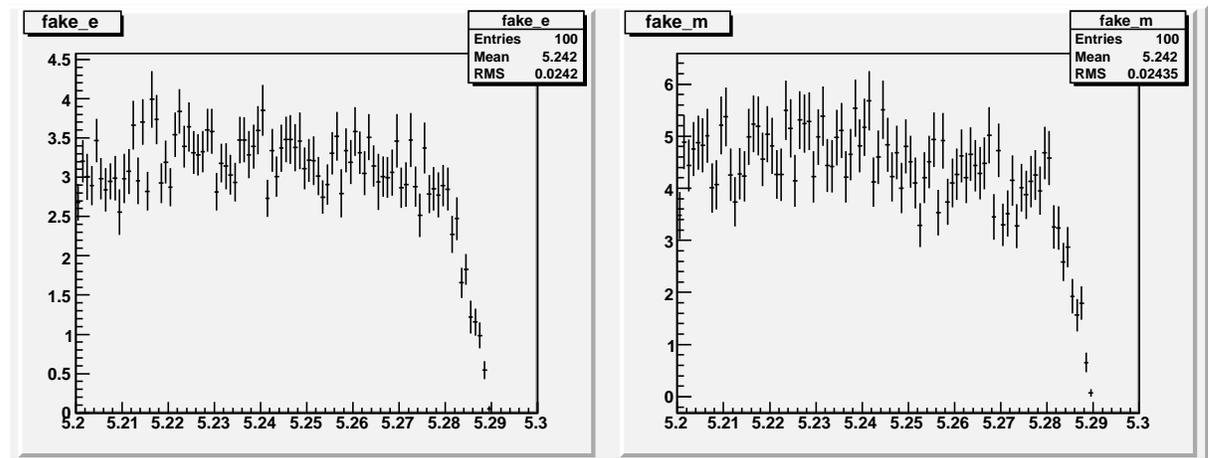


Figure A.1: M_{bc} distributions for hadronic peaking background events with neutrino, for electron(left) and muon(right) cases. These histograms correspond to 605 fb^{-1} .

A.3 Cross check with generic MC

We have compared our estimation with generic MC sample.

In 10 streams of generic MC, we observed 603 events of $J/\psi X_s$ ($S/N \sim 1/2$), 157 events of $\psi(2S)X_s$ ($S/N \sim 1/2$), 155 events of $\psi(4040)X_s$ ($S/N \gg 1$), 112 events of $\psi(4160)X_s$ ($S/N \gg 1$), 130 events of hadronic peaking background ($S/N \sim 3$), and 168 events of hadronic with neutrino peaking background ($S/N \sim 3$), in the M_{bc} signal region, after all the analysis cuts and requiring all X_s daughters except π^0 are coming from the same B meson.

If we fit the M_{bc} distributions of our peaking background estimation with Gaussian + ARGUS function with the same Gaussian shape parameter of the fit and floating ARGUS shape parameter, the Gaussian yield of each histogram is 21.1 ± 2.3 events for $J/\psi X_s$ and $\psi(2S)X_s$, 4.3 ± 0.1 events for hadronic peaking background, 16.6 ± 1.8 events for hadronic peaking background with neutrino. Taking into account the difference of fake rate for data and MC, our estimation and generic MC results are consistent within errors.

For charmonium higher resonances, ~ 6 times larger BR are used for generic MC and $\psi(3770)$ decays and $N_\pi > 2$ decays are not implemented. Comparing $\psi(4040)K$ and $\psi(4160)K$, our estimation and generic MC are consistent.

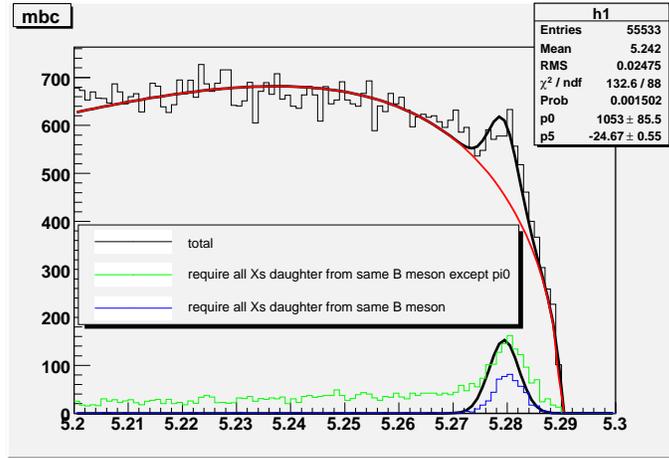


Figure A.2: M_{bc} distributions for generic MC sample (10 streams). The black histogram corresponds to the events after all analysis cut. The blue histogram corresponds to the events with an additional requirement that all X_s daughters are coming from the same B meson. The green histogram corresponds to the events with same requirements except for π^0 in X_s daughter.

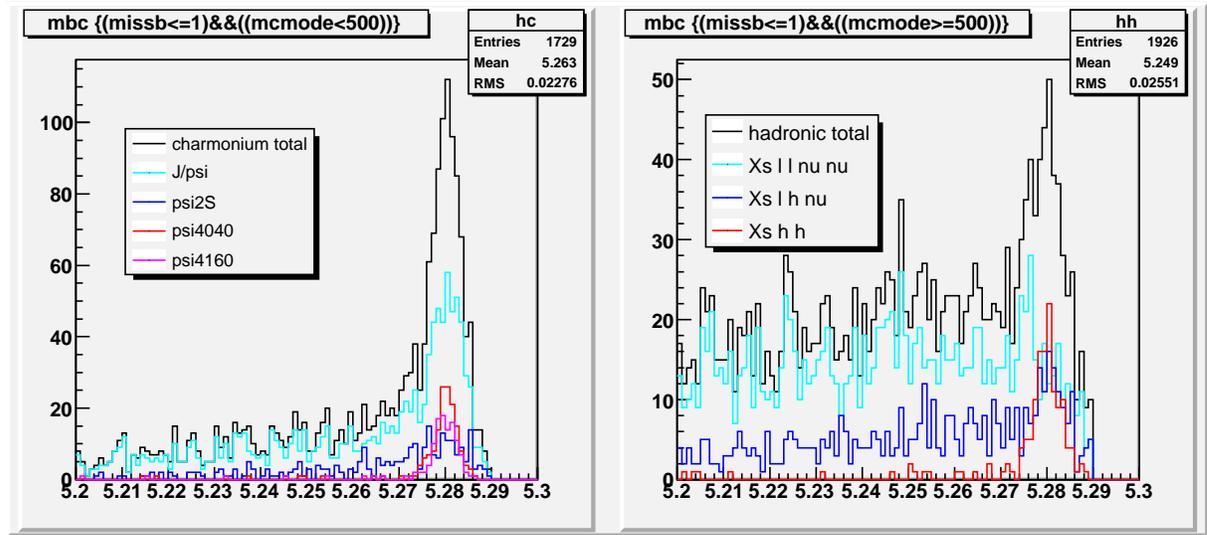


Figure A.3: M_{bc} distributions for generic MC sample (10 streams). The black histogram corresponds to the events after all analysis cut. The blue histogram corresponds to the events with an additional requirement that all X_s daughters are coming from the same B meson. The green histogram corresponds to the events with same requirements except for the π^0 in X_s daughter.

Appendix B

PDF for each of M_{X_s} or $M_{\ell^+\ell^-}$ bins

In this chapter, we show the M_{bc} distributions of each component used to form PDFs in the maximum likelihood fit, in each of M_{X_s} or $M_{\ell^+\ell^-}$ bins, for e^+e^- (left), $\mu^+\mu^-$ (center) and $\ell^+\ell^-$ (right) mode.

Figures B.1 and B.2 are M_{bc} distributions of real $J/\psi X_s$ samples, used to determine the mean and sigma of Gaussian used as the signal PDF $\mathcal{P}_{\text{sig}}^i$.

Figures B.3 and B.4 are M_{bc} distributions of total peaking background samples, used to determine the shape of the peaking background PDF $\mathcal{P}_{\text{pkg}}^i$. The higher charmonium resonances are not included.

Figures B.5 and B.6 are M_{bc} distributions of charmonium peaking background samples, except for the higher charmonium resonances.

Figures B.7 and B.8 are M_{bc} distributions of $B \rightarrow X_s hh$ component of hadronic peaking background samples.

Figures B.9 and B.10 are M_{bc} distributions of total hadronic peaking background samples, including both $B \rightarrow X_s hh$ component and $B \rightarrow X_s h\ell\nu$ component.

Figures B.11 and B.12 are M_{bc} distributions of charmonium peaking background samples in generic MC. Figures B.13 and B.14 are M_{bc} distributions of hadronic peaking background samples in generic MC.

Figures B.15 and B.16 are M_{bc} distributions of self cross-feed samples, used to determine the shape of the self cross-feed PDF $\mathcal{P}_{\text{scf}}^i$.

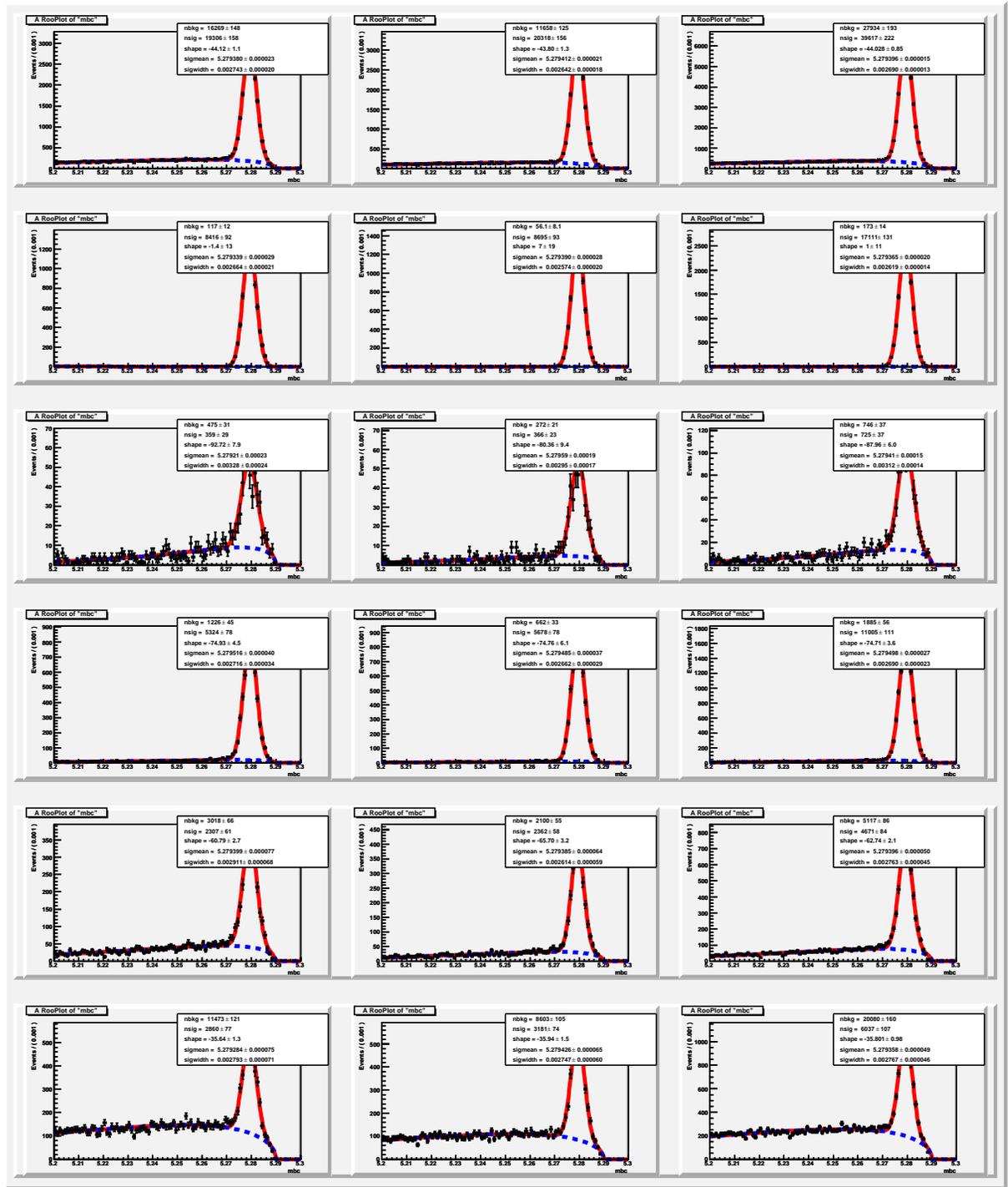


Figure B.1: M_{bc} distribution of real $J/\psi X_s$ sample for each M_{X_s} bin, fitted with Gaussian + ARGUS function. Obtained Gaussian shape parameters are used for signal PDF \mathcal{P}_{sig}^i . Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 605 fb^{-1} .

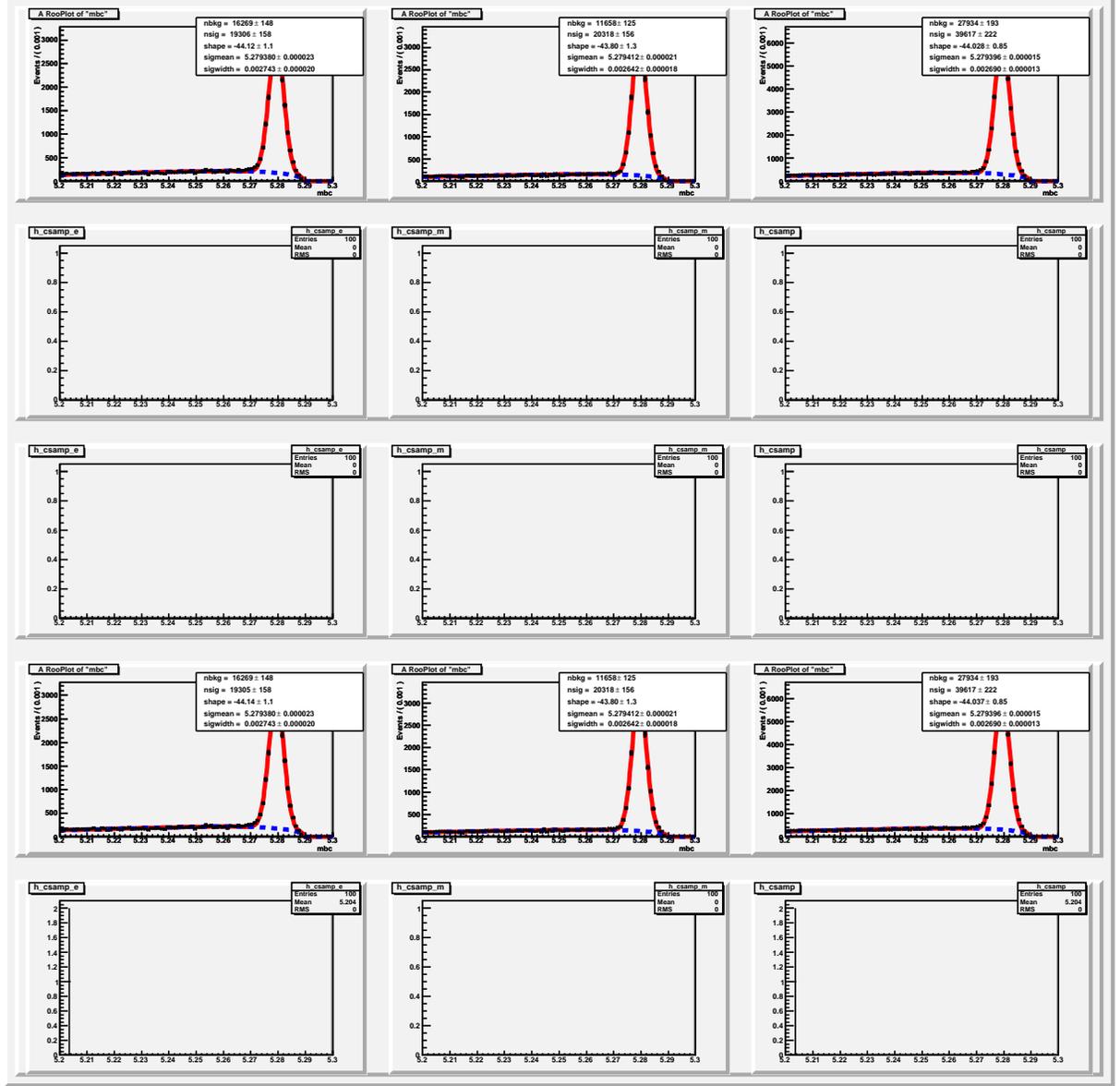


Figure B.2: M_{bc} distribution of real $J/\psi X_s$ sample for each M_{X_s} bin, fitted with Gaussian + ARGUS function. Each row corresponds to the full dataset, $0.04 \text{ (GeV}/c^2)^2 < q^2 < 1.0 \text{ (GeV}/c^2)^2$, $1.0 \text{ (GeV}/c^2)^2 < q^2 < 6.0 \text{ (GeV}/c^2)^2$, $6.0 \text{ (GeV}/c^2)^2 < q^2 < 14.4 \text{ (GeV}/c^2)^2$ and $14.4 \text{ (GeV}/c^2)^2 < q^2 < 25.0 \text{ (GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} . Because the real $J/\psi X_s$ sample has all entries in $q^2 = [6.0, 14.4]$, and no entry in other q^2 regions, we use the value obtained from the full sample for $\mathcal{P}_{\text{sig}}^i$ for each q^2 region.

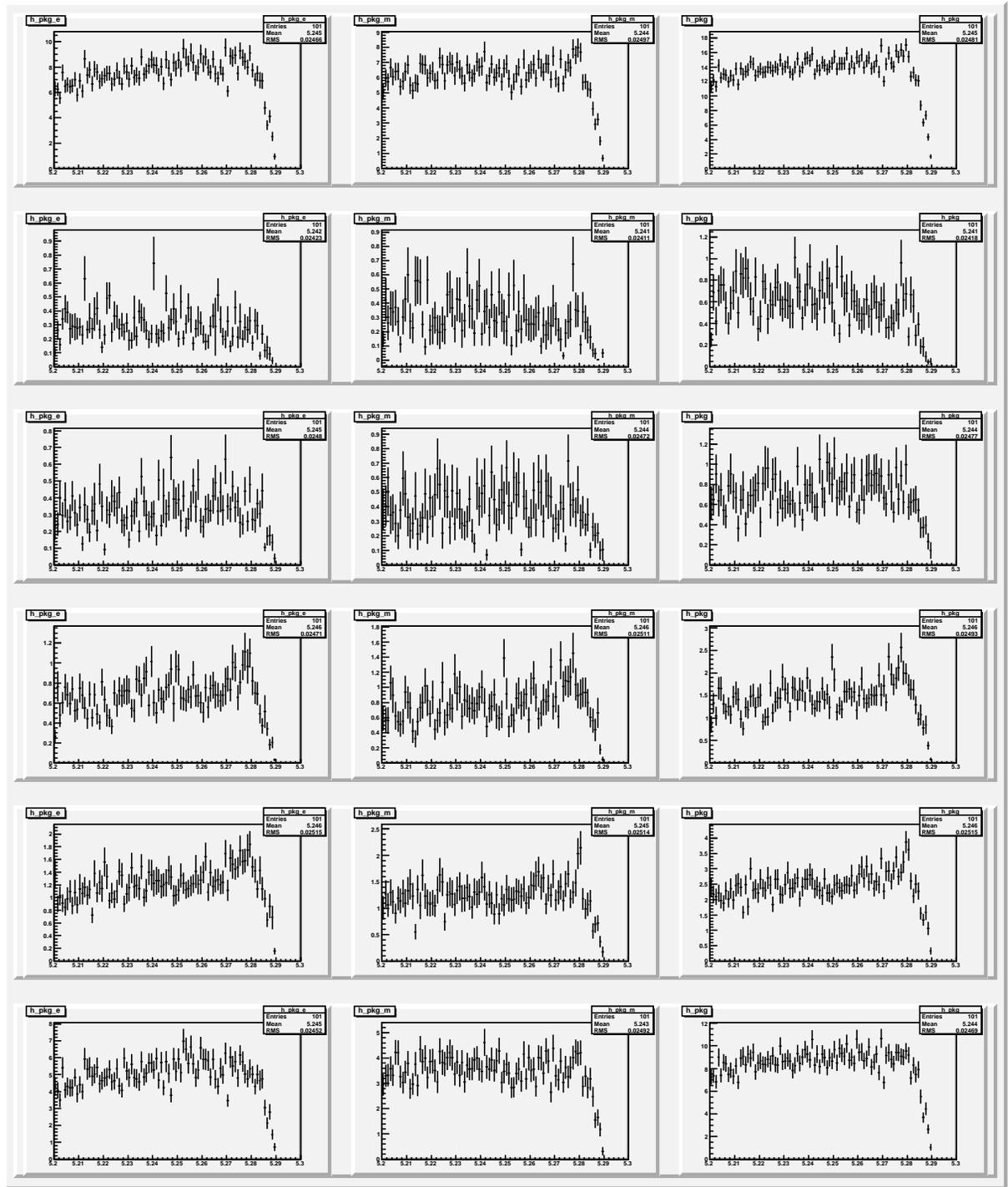


Figure B.3: $M_{b,c}$ distribution of peaking background for each M_{X_s} bin, estimated from the charmonium MC and real hadronic candidates. Histogram shapes are used for peaking background PDF $\mathcal{P}_{\text{pkg}}^i$. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

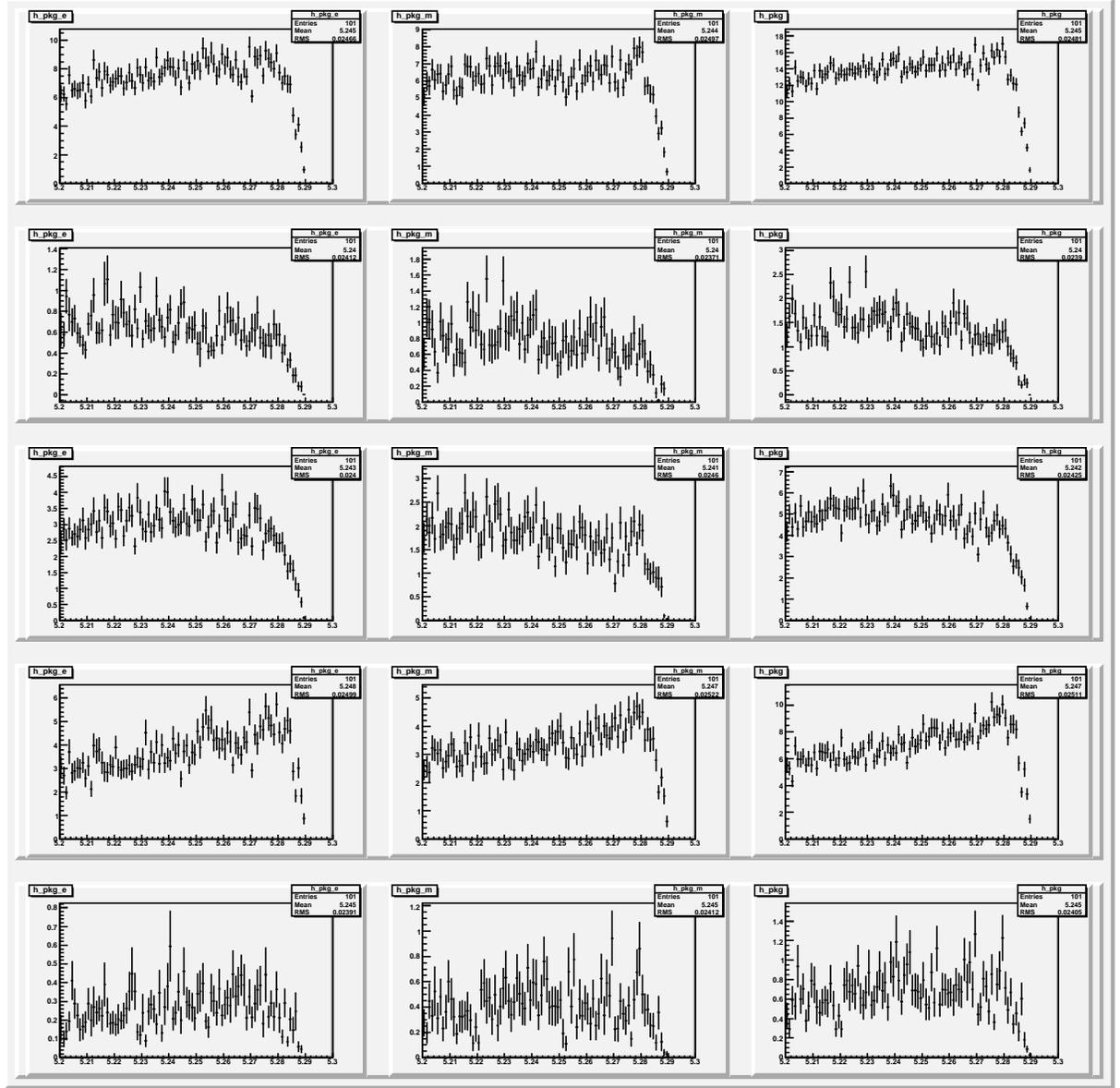


Figure B.4: M_{bc} distribution of peaking background for each q^2 bin, estimated from the charmonium MC and real hadronic candidates. Histogram shapes are used for peaking background PDF $\mathcal{P}_{\text{pkg}}^i$. Each row corresponds to the full dataset, $0.04 \text{ (GeV}/c^2)^2 < q^2 < 1.0 \text{ (GeV}/c^2)^2$, $1.0 \text{ (GeV}/c^2)^2 < q^2 < 6.0 \text{ (GeV}/c^2)^2$, $6.0 \text{ (GeV}/c^2)^2 < q^2 < 14.4 \text{ (GeV}/c^2)^2$ and $14.4 \text{ (GeV}/c^2)^2 < q^2 < 25.0 \text{ (GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 605 fb^{-1} .

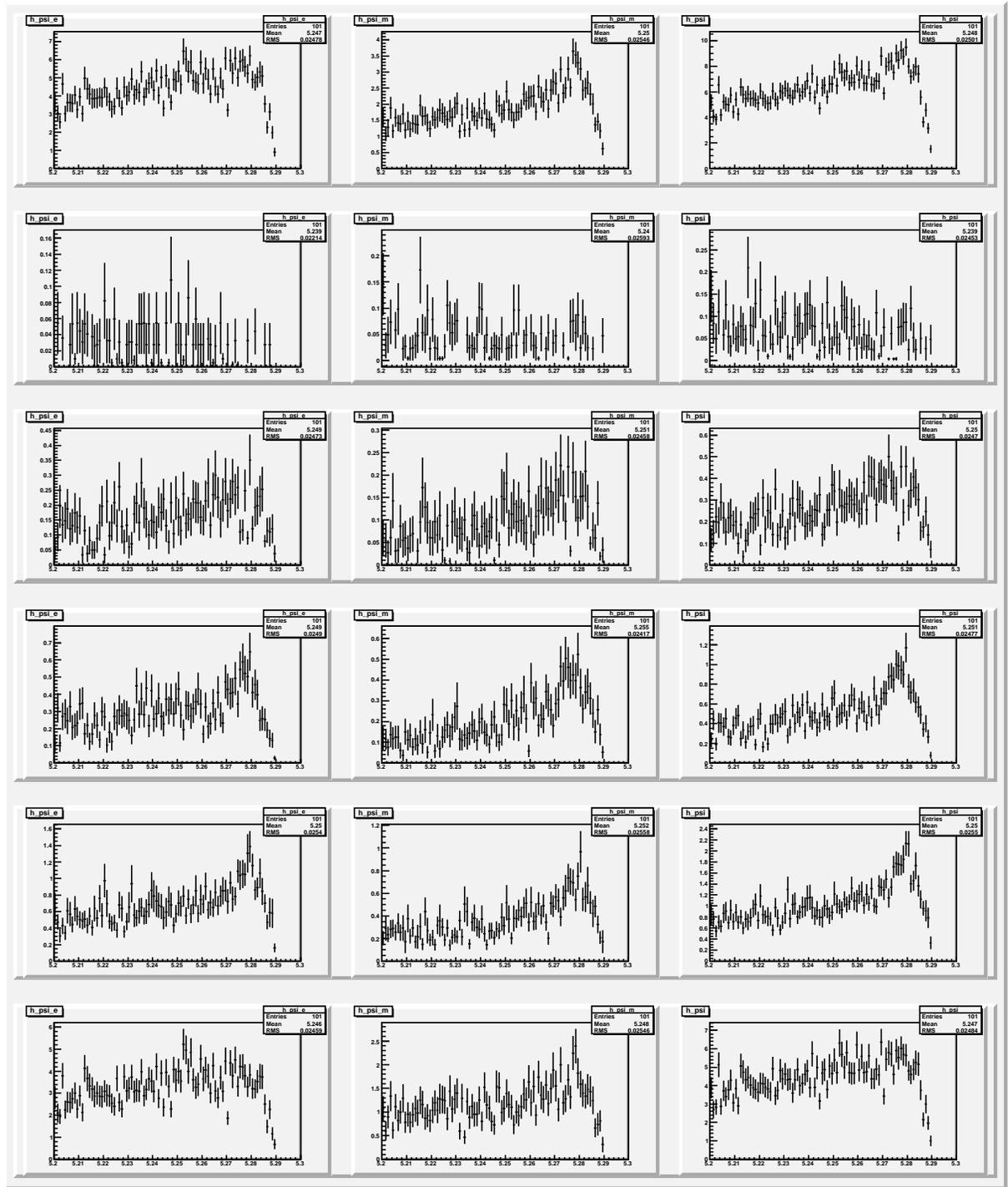


Figure B.5: M_{bc} distribution of $J/\psi X_s + \psi(2S)X_s$ peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

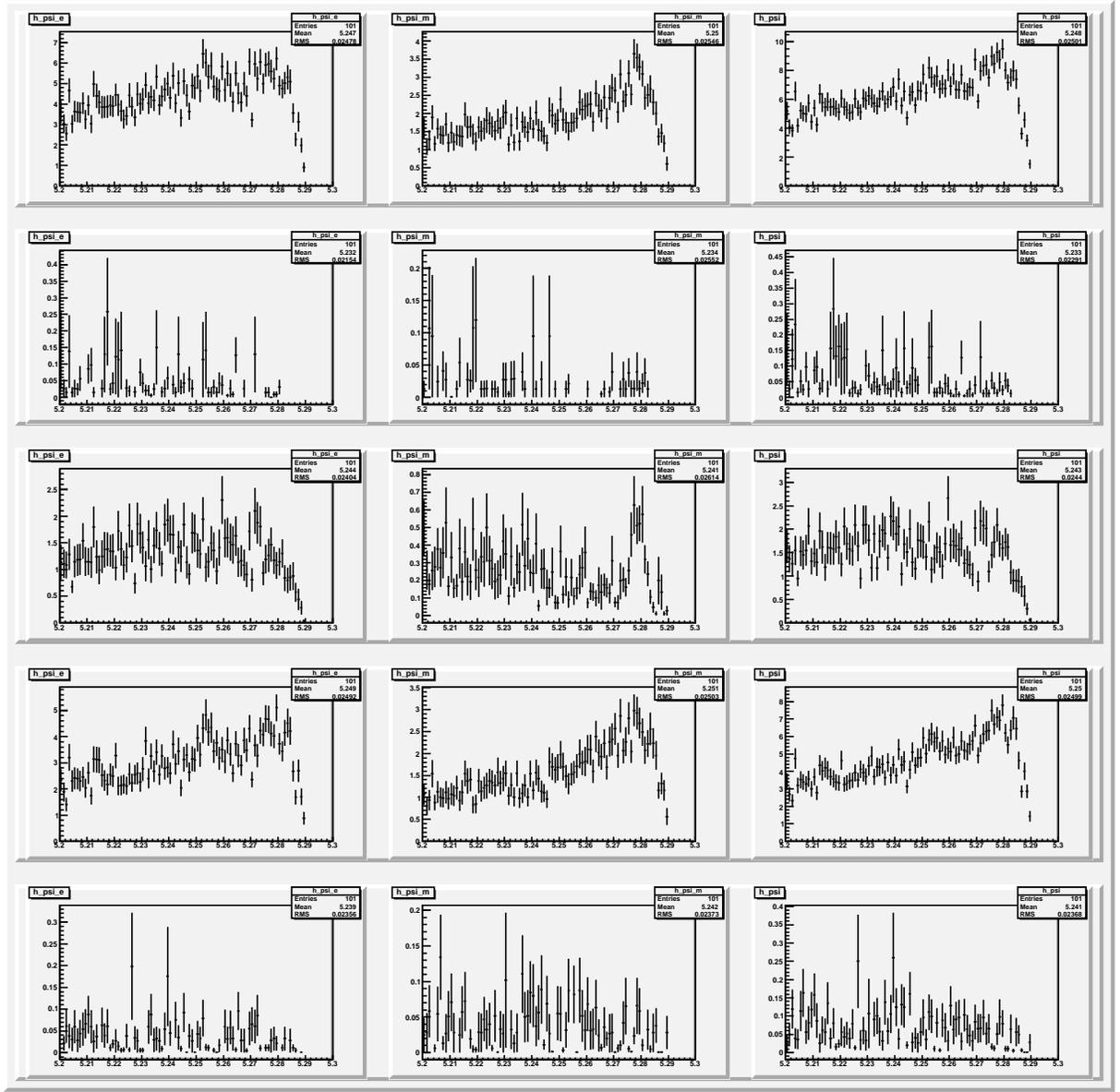


Figure B.6: M_{bc} distribution of $J/\psi X_s + \psi(2S) X_s$ peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 \text{ (GeV}/c^2)^2 < q^2 < 1.0 \text{ (GeV}/c^2)^2$, $1.0 \text{ (GeV}/c^2)^2 < q^2 < 6.0 \text{ (GeV}/c^2)^2$, $6.0 \text{ (GeV}/c^2)^2 < q^2 < 14.4 \text{ (GeV}/c^2)^2$ and $14.4 \text{ (GeV}/c^2)^2 < q^2 < 25.0 \text{ (GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

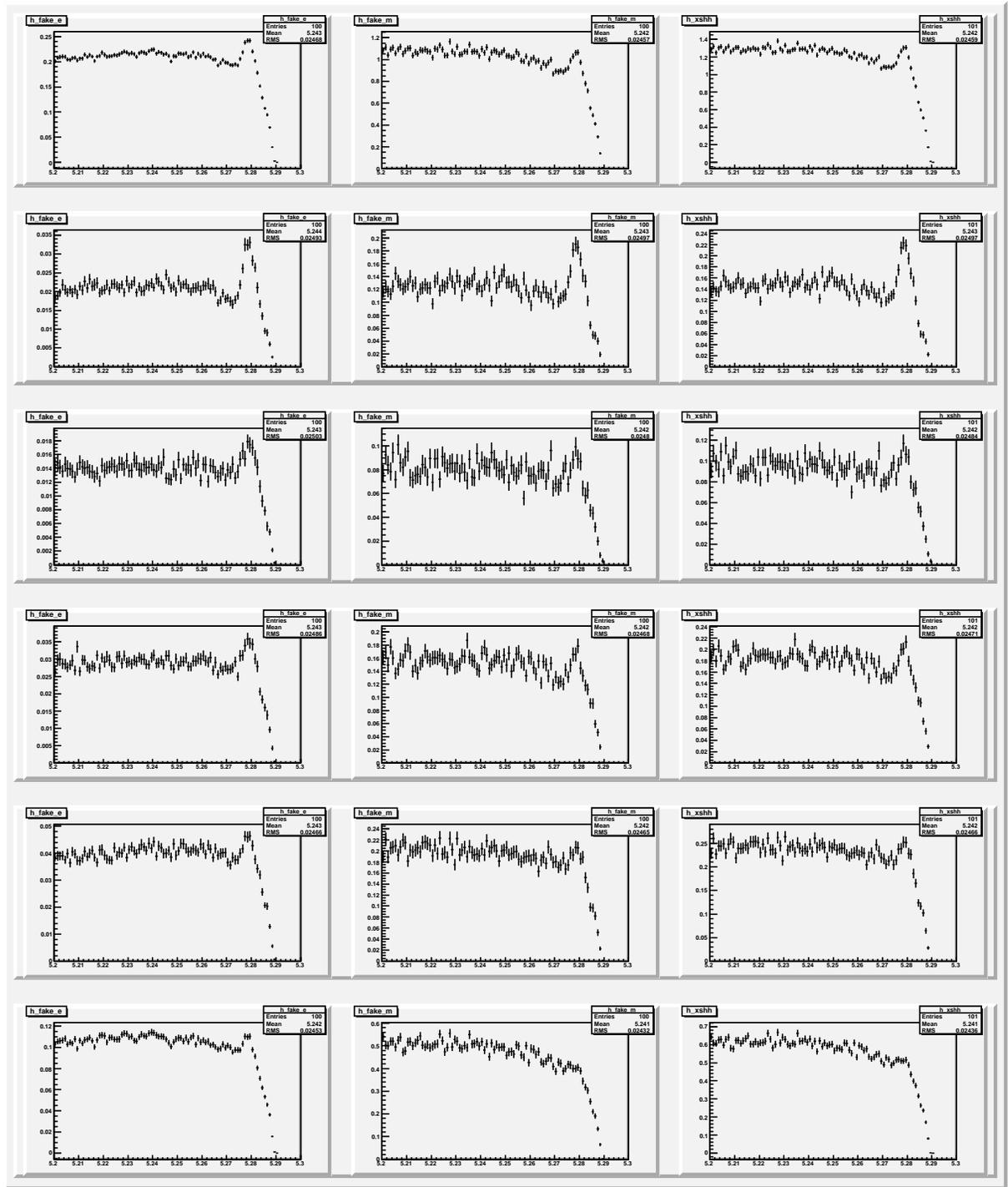


Figure B.7: M_{bc} distribution of $B \rightarrow X_s hh$ peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

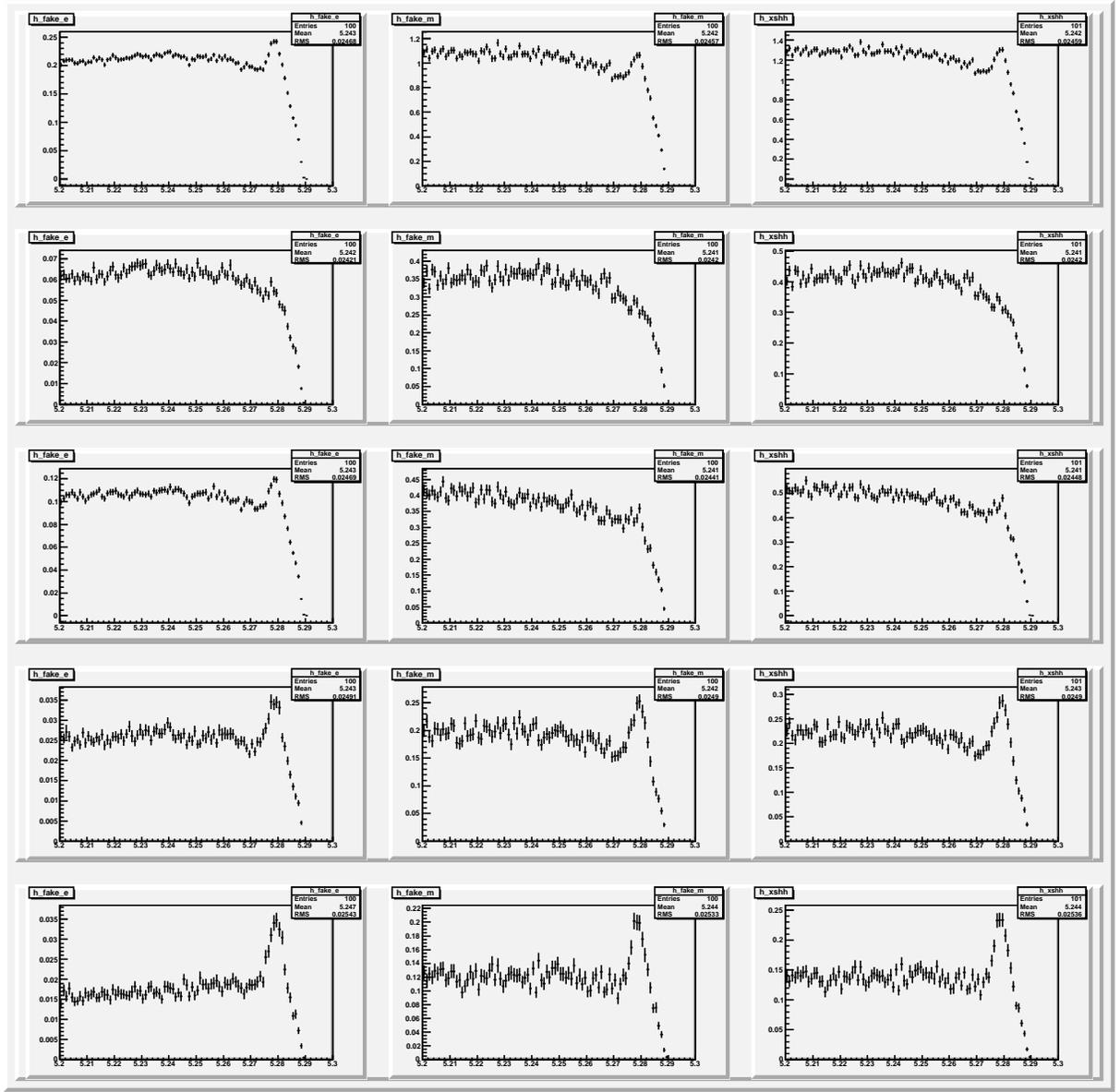


Figure B.8: M_{bc} distribution of $B \rightarrow X_s hh$ peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 \text{ (GeV}/c^2)^2 < q^2 < 1.0 \text{ (GeV}/c^2)^2$, $1.0 \text{ (GeV}/c^2)^2 < q^2 < 6.0 \text{ (GeV}/c^2)^2$, $6.0 \text{ (GeV}/c^2)^2 < q^2 < 14.4 \text{ (GeV}/c^2)^2$ and $14.4 \text{ (GeV}/c^2)^2 < q^2 < 25.0 \text{ (GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

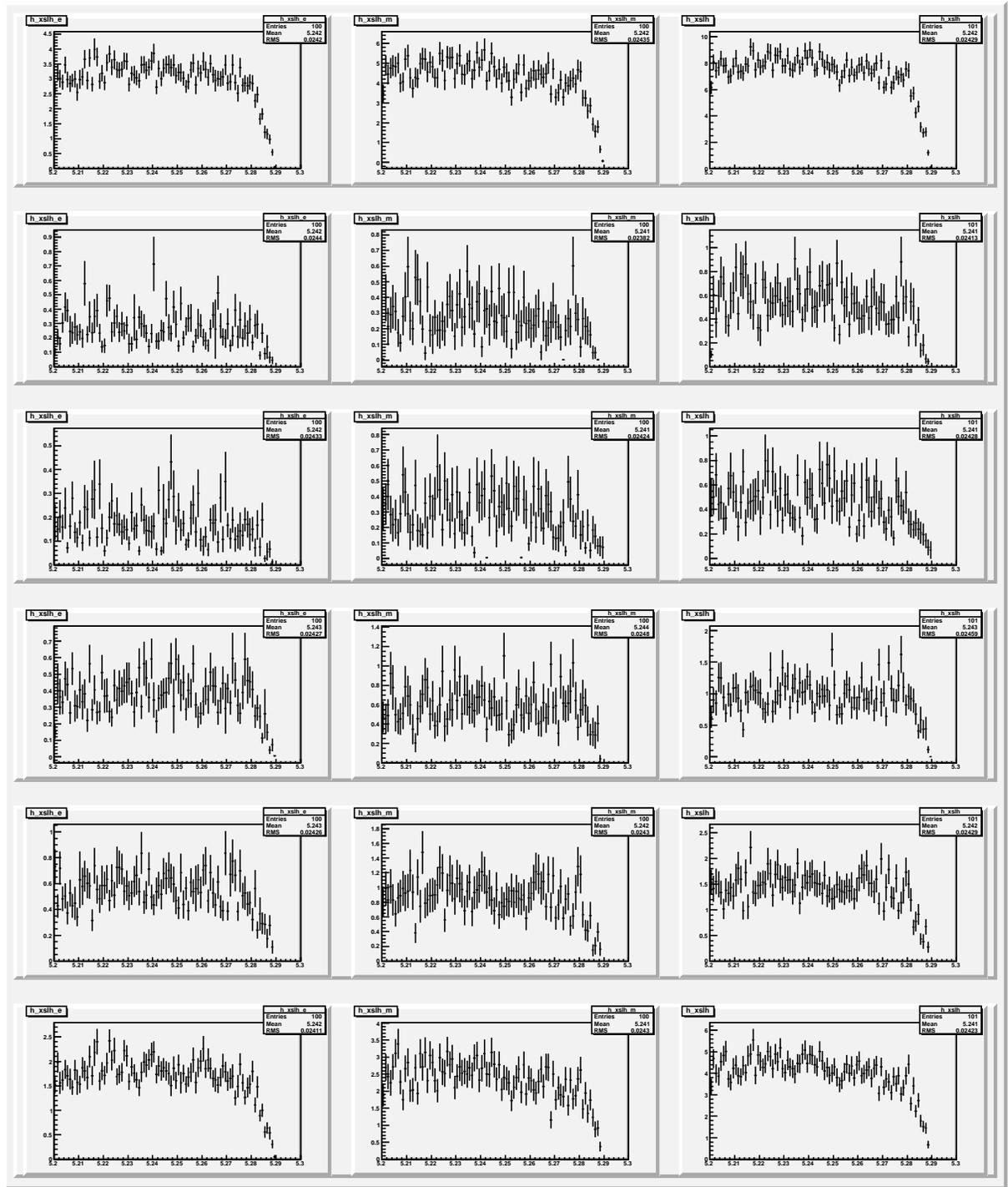


Figure B.9: M_{bc} distribution of $B \rightarrow X_s h l \nu + B \rightarrow X_s h h$ peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

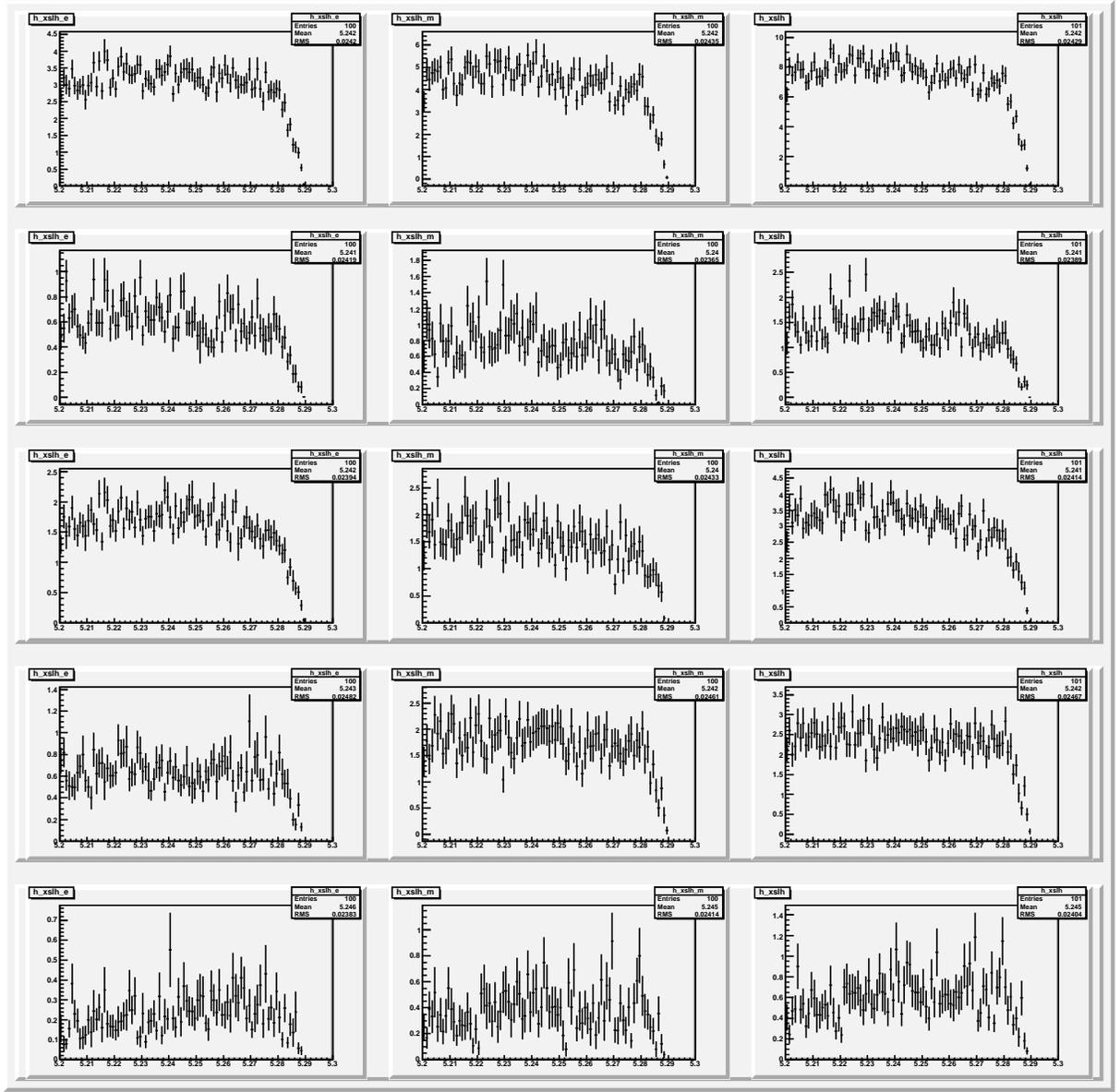


Figure B.10: M_{bc} distribution of $B \rightarrow X_s h l \nu + B \rightarrow X_s h h$ peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 \text{ (GeV}/c^2)^2 < q^2 < 1.0 \text{ (GeV}/c^2)^2$, $1.0 \text{ (GeV}/c^2)^2 < q^2 < 6.0 \text{ (GeV}/c^2)^2$, $6.0 \text{ (GeV}/c^2)^2 < q^2 < 14.4 \text{ (GeV}/c^2)^2$ and $14.4 \text{ (GeV}/c^2)^2 < q^2 < 25.0 \text{ (GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 605 fb^{-1} .

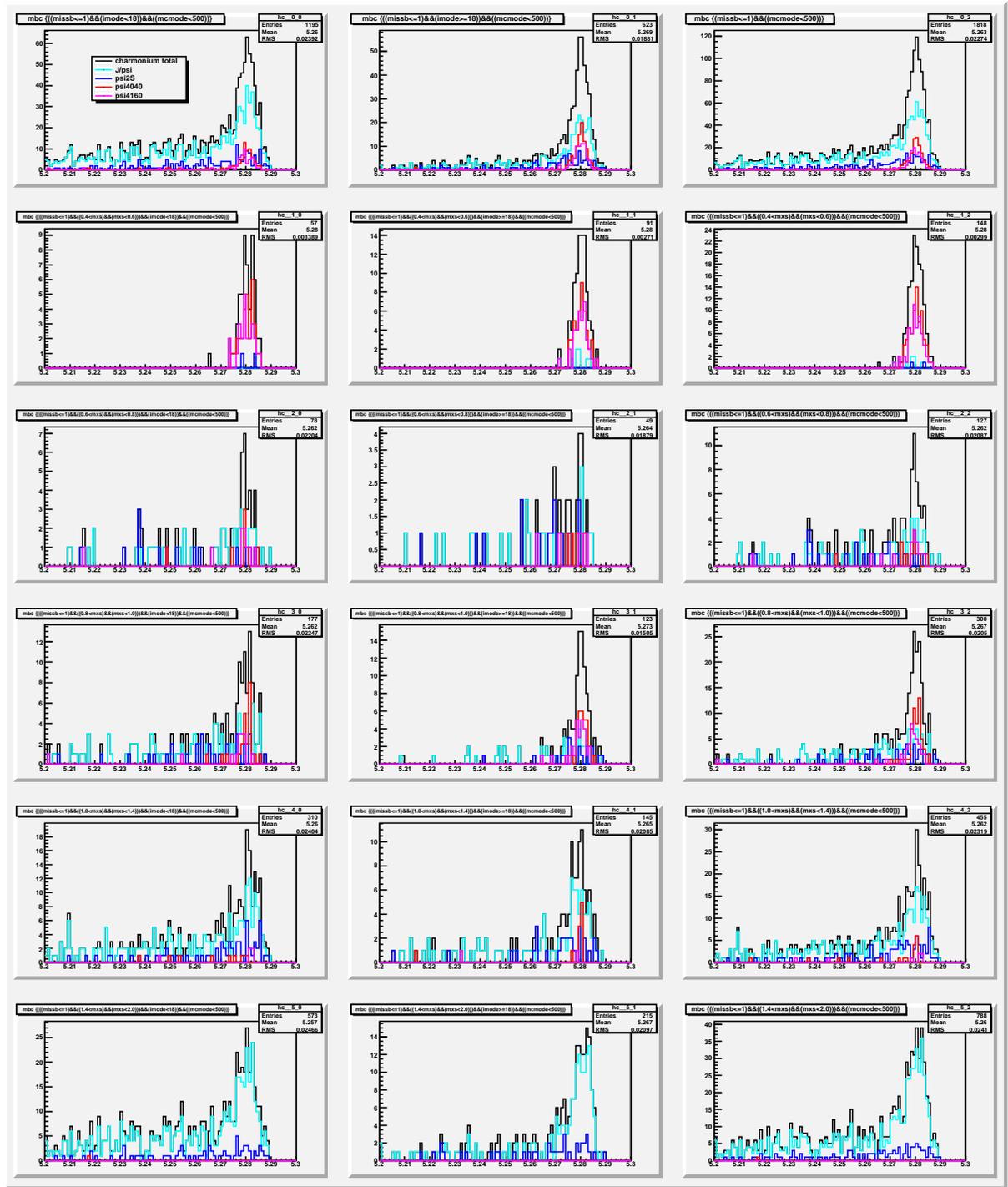


Figure B.11: M_{bc} distribution of generic MC charmonium peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson.

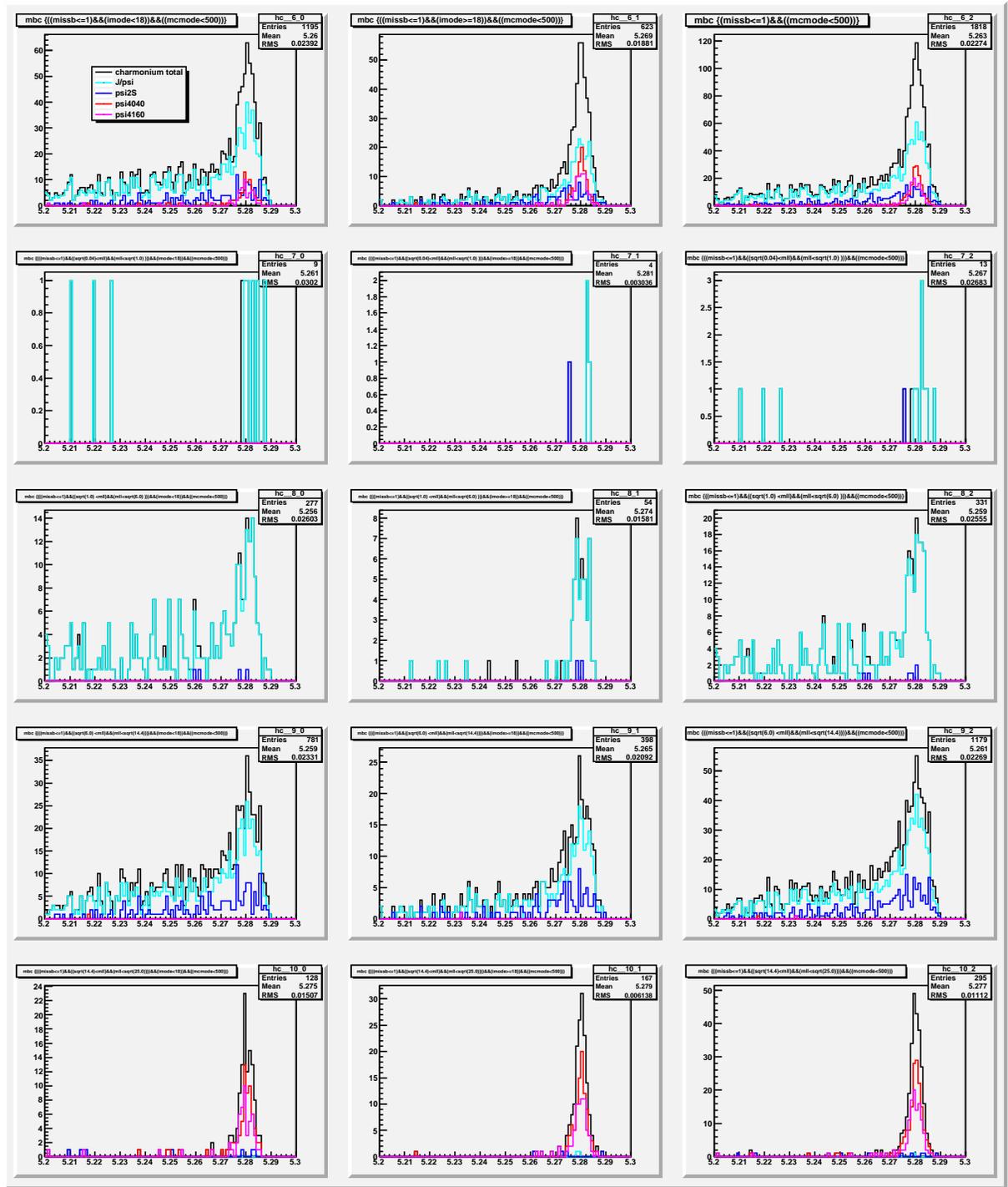


Figure B.12: M_{bc} distribution of generic MC charmonium peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson. 120

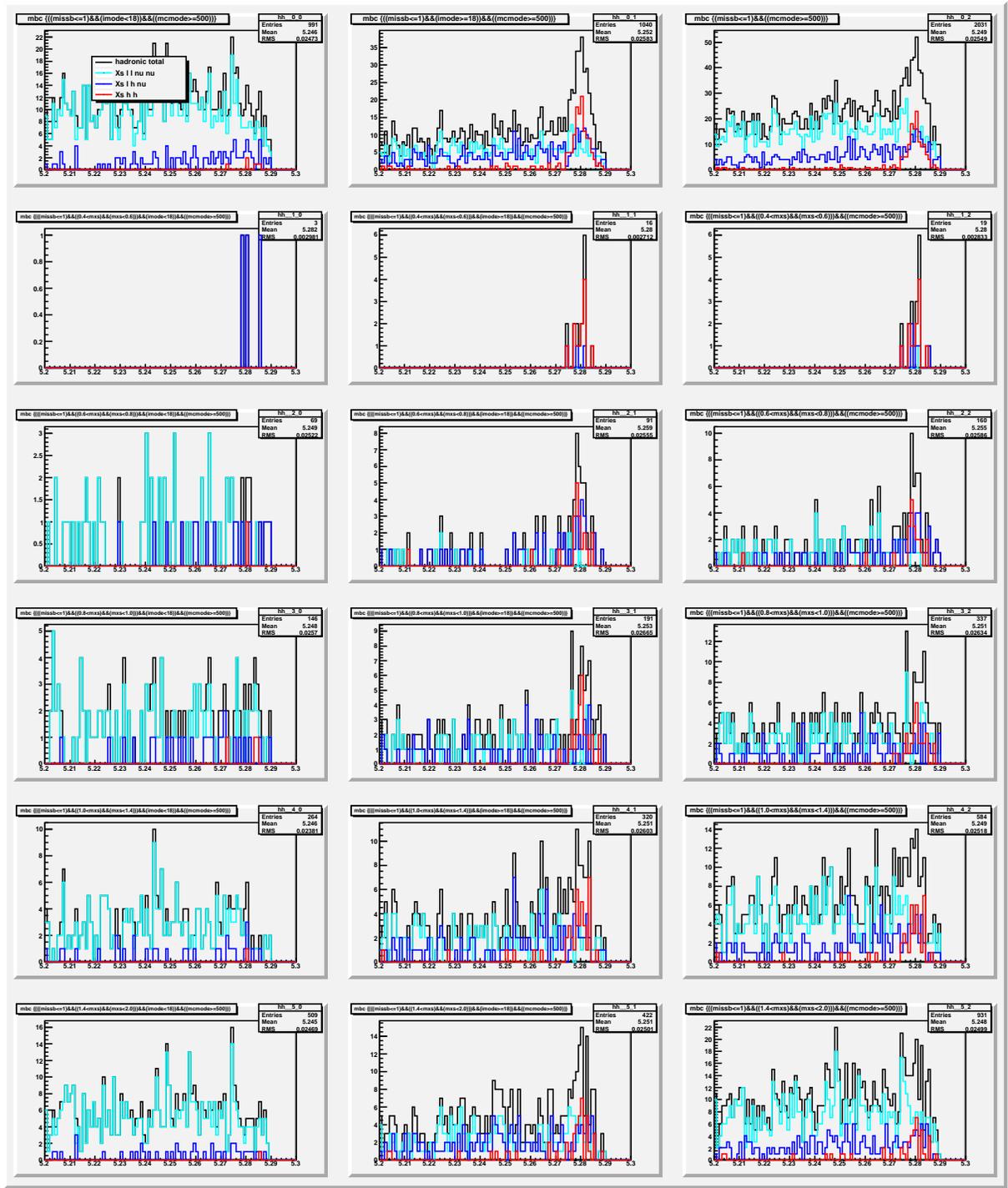


Figure B.13: M_{bc} distribution of generic MC hadronic peaking background for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson.

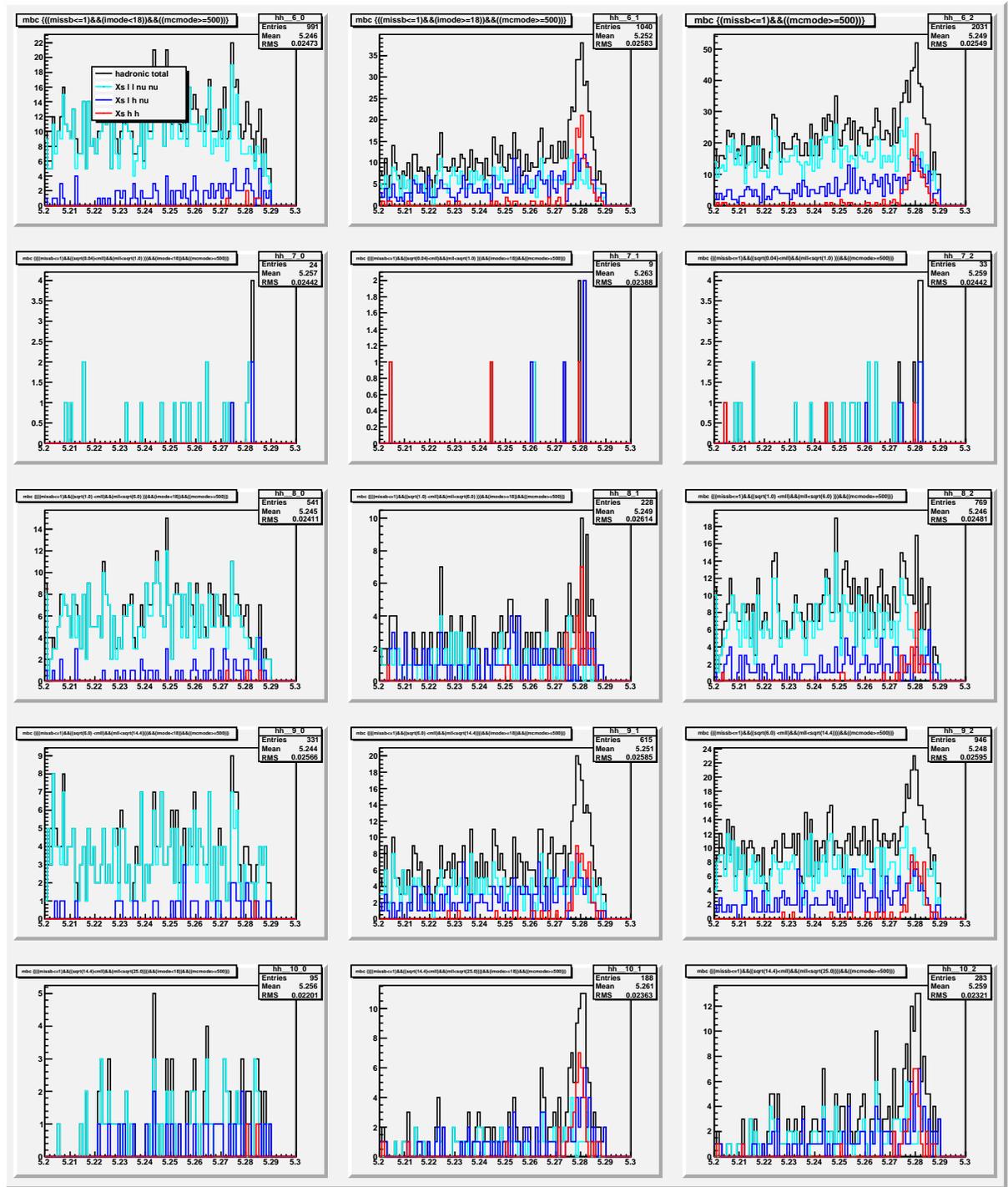


Figure B.14: M_{bc} distribution of generic MC hadronic peaking background for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and l^+l^- , from left to right. These distributions correspond to 10 streams of 605 fb^{-1} MC, requiring all analysis cuts and requiring all X_s children except π^0 is coming from the same B meson. 122

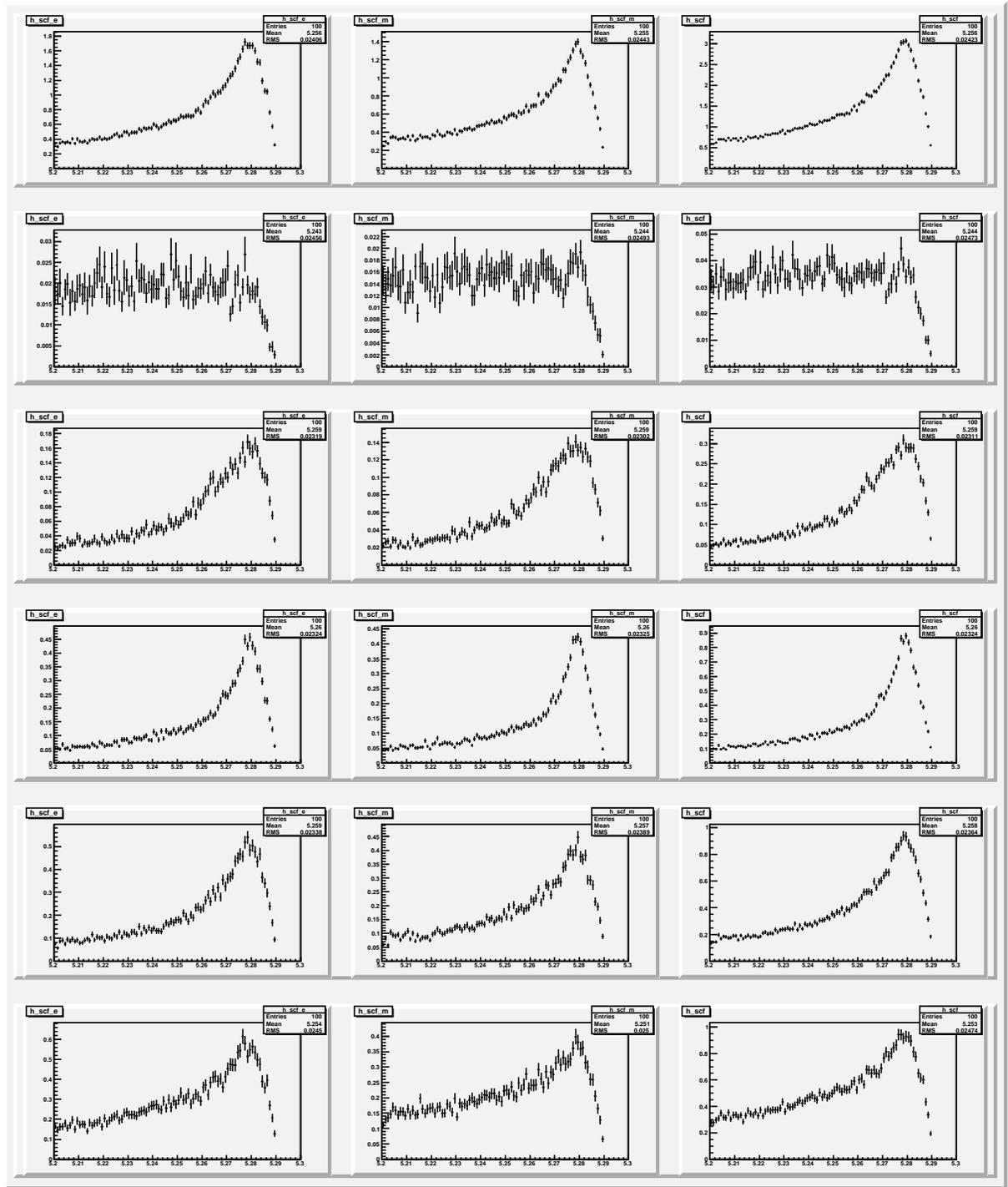


Figure B.15: M_{bc} distribution of self cross-feed candidates in signal MC for each M_{X_s} bin. Histogram shapes are used for self cross-feed PDF $\mathcal{P}_{\text{scf}}^i$. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

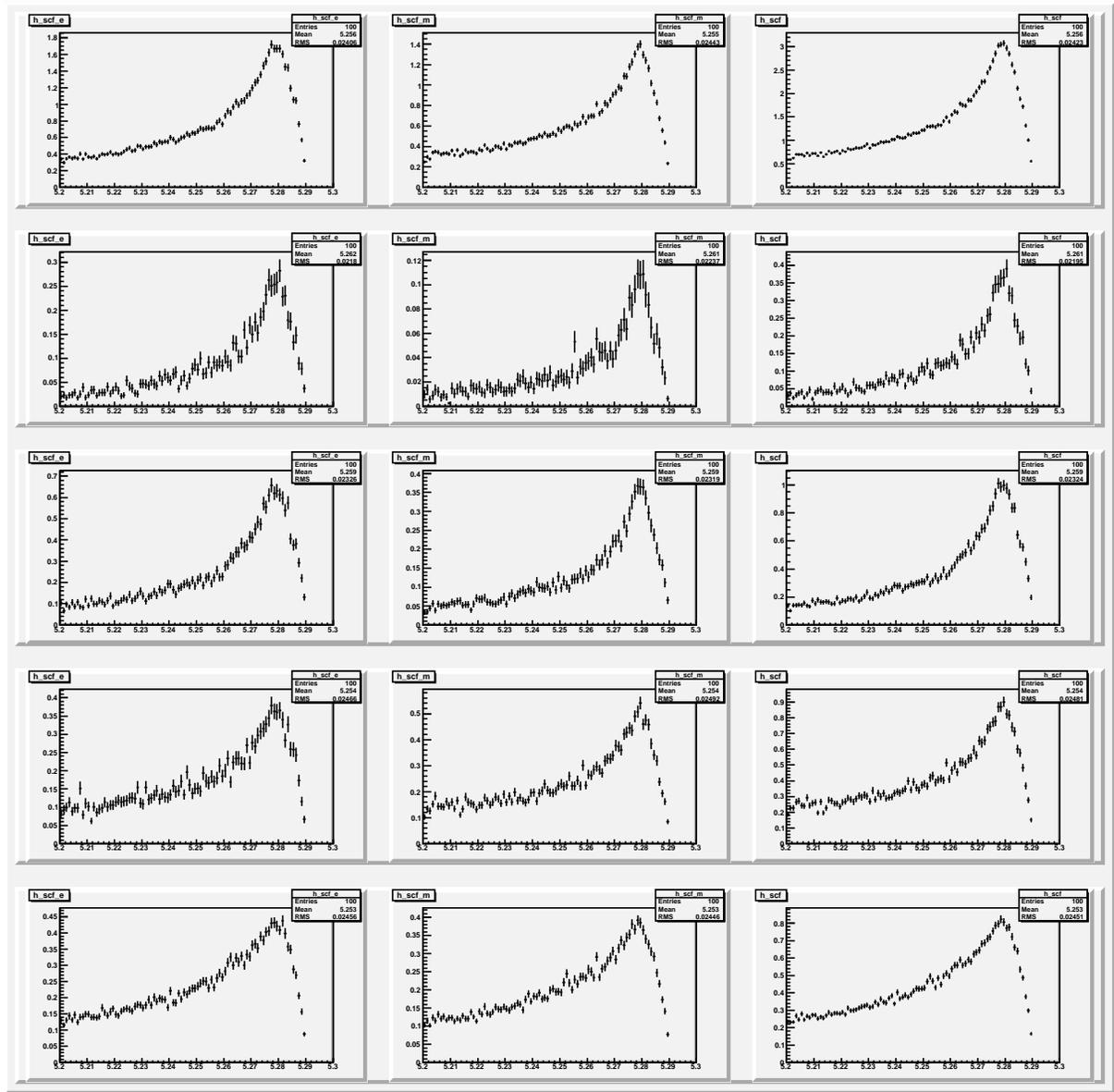


Figure B.16: M_{bc} distribution of self cross-feed candidates in signal MC for each M_{X_s} bin. Histogram shapes are used for self cross-feed PDF $\mathcal{P}_{\text{scf}}^i$. Each row corresponds to the full dataset, $0.04(\text{GeV}/c^2)^2 < q^2 < 1.0(\text{GeV}/c^2)^2$, $1.0(\text{GeV}/c^2)^2 < q^2 < 6.0(\text{GeV}/c^2)^2$, $6.0(\text{GeV}/c^2)^2 < q^2 < 14.4(\text{GeV}/c^2)^2$ and $14.4(\text{GeV}/c^2)^2 < q^2 < 25.0(\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. These distributions correspond to 605 fb^{-1} .

Appendix C

Cross check with the previous 140 fb⁻¹ analysis

As a cross-check with the previous analysis [20], we perform the measurement using only a part of real data sample (140 fb⁻¹), not a whole sample (605 fb⁻¹). The 140 fb⁻¹ data sample which is used in this chapter is the same one used in [20].

The study is done with several fitting method. In Section C.2, we show the results obtained by the same method we use for the 605 fb⁻¹ analysis. In Section C.3, we show the results with non-simultaneous method, which is described there.

C.1 Summary of the previous results

Following two tables are taken from the previous Belle analysis with 140 fb⁻¹ data[20].

They used the fixed ARGUS shape parameter, -13.6 ± 2.8 , obtained from the ARGUS fitting to the real $B \rightarrow X_s e^\pm \mu^\mp$ sample.

In the simultaneous fit method used in our analysis, we obtain smaller ARGUS shape parameter(-17.5). With smaller shape parameters, ARGUS distribution has less entries in M_{bc} sideband and more entries in M_{bc} signal region. Therefore we obtain smaller yields (48.4 ± 13.2) than the previous analysis and the measured branching fractions are also smaller.

Table C.1: Summary of results: signal yield (N_{sig}), significance, total signal efficiency ϵ and branching fraction (\mathcal{B}). The first and second errors quoted on N_{sig} and \mathcal{B} are statistical and systematic, respectively. The first error on ϵ corresponds to uncertainties in detector modeling, $B\bar{B}$ counting, and Monte Carlo statistics, and the second error on ϵ to the uncertainties in the signal model.

Mode	N_{sig}	Significance	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$X_s e^+e^-$	$31.8 \pm 10.2 \pm 3.1$	3.2	$2.59 \pm 0.20^{+0.45}_{-0.42}$	$4.04 \pm 1.30^{+0.87}_{-0.83}$
$X_s \mu^+\mu^-$	$36.3 \pm 9.3 \pm 2.1$	4.4	$2.89 \pm 0.24^{+0.52}_{-0.49}$	$4.13 \pm 1.05^{+0.85}_{-0.81}$
$X_s \ell^+\ell^-$	$68.4 \pm 13.8 \pm 5.0$	5.4	$2.74 \pm 0.22^{+0.48}_{-0.45}$	$4.11 \pm 0.83^{+0.85}_{-0.81}$

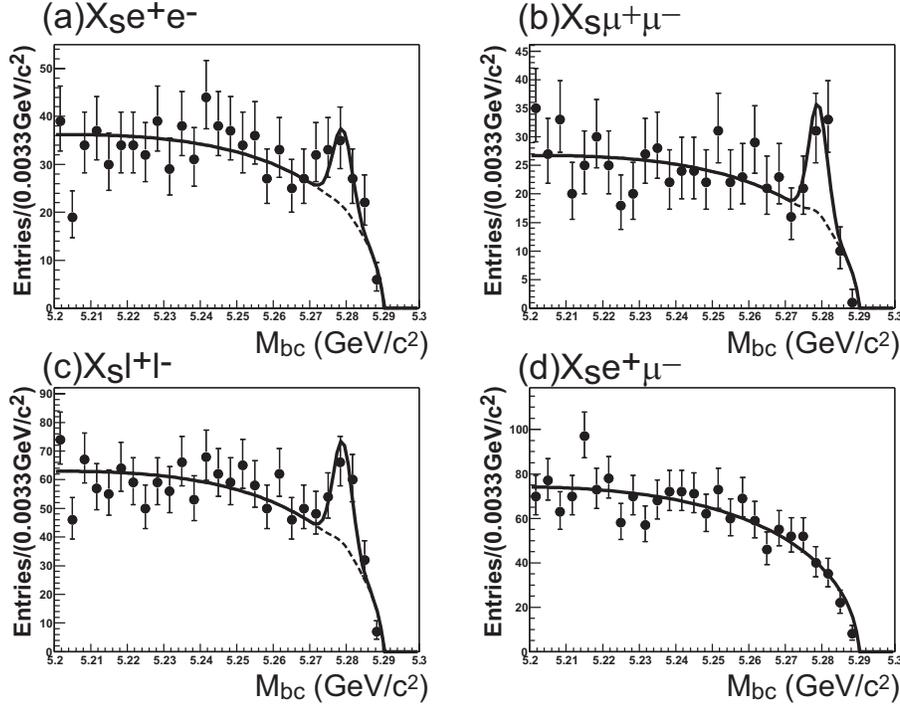


Figure C.1: M_{bc} distributions of selected (a) $B \rightarrow X_s e^+ e^-$, (b) $B \rightarrow X_s \mu^+ \mu^-$, (c) $B \rightarrow X_s \ell^+ \ell^-$ ($\ell = e, \mu$), and (d) $B \rightarrow X_s e^\pm \mu^\mp$ candidates. The solid lines represent the result of the fits, and dashed lines represent the sum of all background components under the signal peaks, respectively.

Table C.2: Signal yields (N_{sig}), signal efficiency (ϵ) and Branching fractions (\mathcal{B}) for each bin of M_{X_s} , $M_{\ell^+ \ell^-}$ and q^2 . The first and second errors for N_{sig} and \mathcal{B} are statistical and systematic, respectively.

bin	N_{sig}	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$M_{X_s} (\text{GeV}/c^2)$			
[0.4, 0.6]	$22.1 \pm 5.6 \pm 0.9$	$9.67 \pm 0.78^{+0.00}_{-0.00}$	$0.375 \pm 0.096^{+0.034}_{-0.034}$
[0.6, 0.8]	$1.1 \pm 2.7 \pm 0.2$	$4.98 \pm 0.57^{+0.00}_{-0.00}$	$0.036 \pm 0.088^{+0.008}_{-0.008}$
[0.8, 1.0]	$18.7 \pm 6.3 \pm 1.1$	$4.62 \pm 0.39^{+0.00}_{-0.00}$	$0.665 \pm 0.225^{+0.068}_{-0.068}$
[1.0, 1.4]	$8.8 \pm 5.8 \pm 0.9$	$1.38 \pm 0.13^{+0.18}_{-0.19}$	$1.050 \pm 0.690^{+0.201}_{-0.209}$
[1.4, 2.0]	$17.3 \pm 8.7 \pm 2.1$	$0.61 \pm 0.06^{+0.28}_{-0.11}$	$4.659 \pm 2.337^{+2.253}_{-1.097}$
$q^2 ((\text{GeV}/c^2)^2)$			
[0.04, 1.0]	$11.2 \pm 4.8 \pm 0.6$	$1.63 \pm 0.14^{+0.64}_{-0.34}$	$1.134 \pm 0.483^{+0.458}_{-0.268}$
[1.0, 6.0]	$25.1 \pm 8.5 \pm 2.0$	$2.77 \pm 0.23^{+0.66}_{-0.46}$	$1.493 \pm 0.504^{+0.395}_{-0.300}$
[6.0, 14.4]	$9.0 \pm 7.6 \pm 1.6$	$2.03 \pm 0.17^{+0.27}_{-0.31}$	$0.732 \pm 0.614^{+0.176}_{-0.184}$
[14.4, 25.0]	$21.1 \pm 5.9 \pm 1.0$	$8.32 \pm 0.69^{+0.79}_{-0.99}$	$0.418 \pm 0.117^{+0.056}_{-0.064}$

C.2 140 fb⁻¹ results by simultaneous fit

As a cross-check with the previous analysis [20], we performed the measurement using a part of real data sample (140 fb⁻¹), and not the whole sample (605 fb⁻¹). The 140 fb⁻¹ of data samples which is used in this chapter is the same one used in [20].

The statistical significance is defined by $\mathcal{S} = \sqrt{2 \ln(\mathcal{L}_{max}/\mathcal{L}_0)}$, where \mathcal{L}_{max} (\mathcal{L}_0) denotes the likelihood value at the maximum (with the signal yield fixed to zero).

The branching fraction \mathcal{B} for the signal is calculated as $\mathcal{B} = \frac{N_{sig}}{2N_{B\bar{B}}\epsilon}$, where $N_{B\bar{B}} = (152.0 \pm 1.2) \times 10^6$ is the number of $B\bar{B}$ pairs produced in 140 fb⁻¹ and ϵ is the signal efficiency.

We also measured the the differential branching fractions as a function of M_{X_s} and $q^2 \equiv M_{\ell^+\ell^-}^2$. To measure the differential branching fraction, we separately fit the partial dataset divided into each of M_{X_s} or q^2 regions. For M_{X_s} regions, we divide the dataset into following 5 regions: [0.4, 0.6],[0.6, 0.8],[0.8, 1.0],[1.0, 1.4] and [1.4, 2.0] GeV/c². For q^2 regions, we divide the dataset into following 4 regions: [0.04, 1.0],[1.0, 6.0],[6.0, 14.4] and [14.4, 25.0] (GeV/c²)².

Our results with 140 fb⁻¹ data are consistent with the previous analysis within the errors. The previous results are shown in Appendix C.1. Note that we have many differences from the previous analysis.

The fit results are summarized in Table C.3 and C.4. The branching fraction measurement results are summarized in Table C.5 and C.6.

Table C.3: Results of the simultaneous fit to the 140 fb⁻¹ of $B \rightarrow X_s \ell^+ \ell^-$ sample and 605 fb⁻¹ of $B \rightarrow X_s e^\pm \mu^\mp$ sample: number of signal candidates in the M_{bc} signal region, fitted signal yield N_{sig} , and significance.

Mode	Candidates	ARGUS shape	N_{sig}	Significance
$B \rightarrow X_s e^+ e^-$	161	-19.1 ± 2.1	$28.3 \pm 10.0 \pm 0.5$	3.1
$B \rightarrow X_s \mu^+ \mu^-$	122	-17.8 ± 2.1	$28.1 \pm 8.8 \pm 0.1$	3.6
$B \rightarrow X_s \ell^+ \ell^-$	283	-19.3 ± 2.0	$54.4 \pm 13.5 \pm 0.4$	4.5

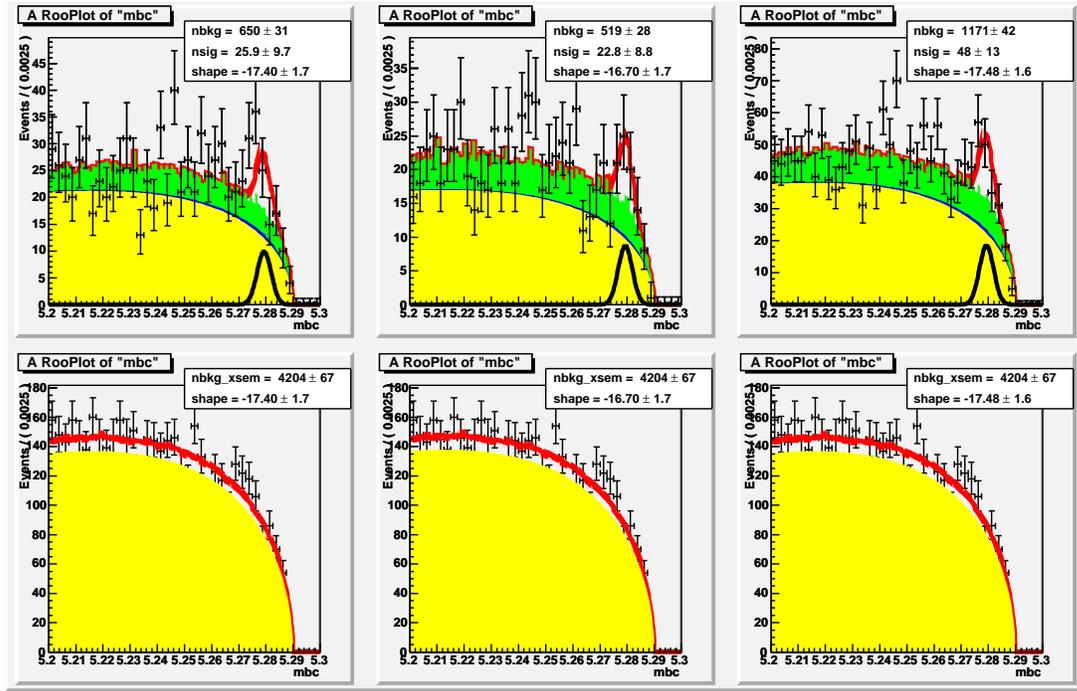


Figure C.2: M_{bc} fit to the 140 fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample (upper) and 605 fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample (lower). Each column corresponds to e^+e^- (left), $\mu^+\mu^-$ (center) and $\ell^+\ell^-$ case (right). Background, peaking background, self cross feed components are shown in yellow, green, blue areas, respectively. Signal component is shown in black line.

Table C.4: Summary of the simultaneous fit to the 140fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample and 605fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample. Gaussian shape parameters (mean and width, in MeV/c^2) is fixed. N_{pkg} and $N_{\text{scf}}/N_{\text{sig}}$ are also fixed. N_{cand} is the number of candidates in M_{bc} signal region. Fitting results, N_{sig} , ARGUS shape parameter, significance (σ) are also shown.

	$B \rightarrow X_s e^+ e^-$	$B \rightarrow X_s \mu^+ \mu^-$	$B \rightarrow X_s \ell^+ \ell^-$
full sample	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 97.3, N_{\text{scf}}/N_{\text{sig}} = 0.48$ $N_{\text{cand}} = 161, \text{shape} = -19.1 \pm 2.1$ $N_{\text{sig}} = 28.3 \pm 10.0 \pm 0.5$ $\sigma_{\text{stat}} = 3.1, \sigma = 3.1$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 40.3, N_{\text{scf}}/N_{\text{sig}} = 0.34$ $N_{\text{cand}} = 122, \text{shape} = -17.8 \pm 2.1$ $N_{\text{sig}} = 28.1 \pm 8.8 \pm 0.1$ $\sigma_{\text{stat}} = 3.6, \sigma = 3.6$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 137.6, N_{\text{scf}}/N_{\text{sig}} = 0.40$ $N_{\text{cand}} = 283, \text{shape} = -19.3 \pm 2.0$ $N_{\text{sig}} = 54.4 \pm 13.5 \pm 0.4$ $\sigma_{\text{stat}} = 4.5, \sigma = 4.5$
$M_{X_s} = [0.4, 0.6]$	mean = 5279.18, width = 2.58 $N_{\text{pkg}} = 0.6, N_{\text{scf}}/N_{\text{sig}} = 0.04$ $N_{\text{cand}} = 13, \text{shape} = 0.2 \pm 8.8$ $N_{\text{sig}} = 8.5 \pm 3.5 \pm 0.1$ $\sigma_{\text{stat}} = 3.4, \sigma = 3.4$	mean = 5279.24, width = 2.54 $N_{\text{pkg}} = 0.8, N_{\text{scf}}/N_{\text{sig}} = 0.02$ $N_{\text{cand}} = 18, \text{shape} = 2.6 \pm 9.1$ $N_{\text{sig}} = 14.8 \pm 4.1 \pm 0.0$ $\sigma_{\text{stat}} = 6.1, \sigma = 6.1$	mean = 5279.21, width = 2.56 $N_{\text{pkg}} = 1.4, N_{\text{scf}}/N_{\text{sig}} = 0.03$ $N_{\text{cand}} = 31, \text{shape} = -1.1 \pm 8.3$ $N_{\text{sig}} = 23.4 \pm 5.4 \pm 0.1$ $\sigma_{\text{stat}} = 6.5, \sigma = 6.4$
$M_{X_s} = [0.6, 0.8]$	mean = 5278.77, width = 2.44 $N_{\text{pkg}} = 3.2, N_{\text{scf}}/N_{\text{sig}} = 3.61$ $N_{\text{cand}} = 5, \text{shape} = -3.2 \pm 8.6$ $N_{\text{sig}} = -1.0 \pm 0.9 \pm 1.3$ $\sigma_{\text{stat}} = \text{nan}, \sigma = \text{nan}$	mean = 5279.58, width = 2.96 $N_{\text{pkg}} = 2.0, N_{\text{scf}}/N_{\text{sig}} = 2.67$ $N_{\text{cand}} = 2, \text{shape} = -20.0 \pm 8.3$ $N_{\text{sig}} = 0.9 \pm 3.3 \pm 1.3$ $\sigma_{\text{stat}} = 5.3, \sigma = 0.4$	mean = 5279.20, width = 2.81 $N_{\text{pkg}} = 5.1, N_{\text{scf}}/N_{\text{sig}} = 3.10$ $N_{\text{cand}} = 7, \text{shape} = -5.3 \pm 7.9$ $N_{\text{sig}} = -2.0 \pm 1.2 \pm 2.0$ $\sigma_{\text{stat}} = \text{nan}, \sigma = \text{nan}$
$M_{X_s} = [0.8, 1.0]$	mean = 5279.50, width = 2.71 $N_{\text{pkg}} = 5.9, N_{\text{scf}}/N_{\text{sig}} = 0.29$ $N_{\text{cand}} = 25, \text{shape} = -9.4 \pm 5.9$ $N_{\text{sig}} = 7.3 \pm 3.9 \pm 0.2$ $\sigma_{\text{stat}} = 2.3, \sigma = 2.3$	mean = 5279.36, width = 2.61 $N_{\text{pkg}} = 4.3, N_{\text{scf}}/N_{\text{sig}} = 0.21$ $N_{\text{cand}} = 22, \text{shape} = -3.7 \pm 6.0$ $N_{\text{sig}} = 12.2 \pm 4.0 \pm 0.1$ $\sigma_{\text{stat}} = 4.3, \sigma = 4.3$	mean = 5279.43, width = 2.66 $N_{\text{pkg}} = 10.2, N_{\text{scf}}/N_{\text{sig}} = 0.24$ $N_{\text{cand}} = 47, \text{shape} = -7.9 \pm 5.6$ $N_{\text{sig}} = 19.7 \pm 5.6 \pm 0.2$ $\sigma_{\text{stat}} = 4.5, \sigma = 4.5$
$M_{X_s} = [1.0, 1.4]$	mean = 5279.49, width = 2.70 $N_{\text{pkg}} = 15.0, N_{\text{scf}}/N_{\text{sig}} = 0.68$ $N_{\text{cand}} = 21, \text{shape} = -22.9 \pm 4.8$ $N_{\text{sig}} = 0.0 \pm 3.3 \pm 7.2$ $\sigma_{\text{stat}} = 0.0, \sigma = \text{nan}$	mean = 5279.44, width = 2.52 $N_{\text{pkg}} = 7.4, N_{\text{scf}}/N_{\text{sig}} = 0.54$ $N_{\text{cand}} = 25, \text{shape} = -22.3 \pm 4.7$ $N_{\text{sig}} = 0.4 \pm 3.4 \pm 0.9$ $\sigma_{\text{stat}} = 0.1, \sigma = \text{nan}$	mean = 5279.46, width = 2.61 $N_{\text{pkg}} = 22.3, N_{\text{scf}}/N_{\text{sig}} = 0.60$ $N_{\text{cand}} = 46, \text{shape} = -23.3 \pm 4.4$ $N_{\text{sig}} = 0.2 \pm 4.8 \pm 4.6$ $\sigma_{\text{stat}} = 0.0, \sigma = \text{nan}$
$M_{X_s} = [1.4, 2.0]$	mean = 5279.18, width = 2.82 $N_{\text{pkg}} = 72.7, N_{\text{scf}}/N_{\text{sig}} = 1.25$ $N_{\text{cand}} = 97, \text{shape} = -23.8 \pm 2.9$ $N_{\text{sig}} = 14.9 \pm 7.5 \pm 0.7$ $\sigma_{\text{stat}} = 2.1, \sigma = 2.1$	mean = 5279.41, width = 2.60 $N_{\text{pkg}} = 25.8, N_{\text{scf}}/N_{\text{sig}} = 0.95$ $N_{\text{cand}} = 55, \text{shape} = -22.8 \pm 2.8$ $N_{\text{sig}} = 1.2 \pm 5.7 \pm 0.1$ $\sigma_{\text{stat}} = 0.2, \sigma = 0.2$	mean = 5279.31, width = 2.70 $N_{\text{pkg}} = 98.5, N_{\text{scf}}/N_{\text{sig}} = 1.10$ $N_{\text{cand}} = 152, \text{shape} = -24.2 \pm 2.6$ $N_{\text{sig}} = 14.5 \pm 9.7 \pm 0.6$ $\sigma_{\text{stat}} = 1.6, \sigma = 1.6$
$q^2 = [0.04, 1.0]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 0.9, N_{\text{scf}}/N_{\text{sig}} = 0.30$ $N_{\text{cand}} = 14, \text{shape} = -12.0 \pm 7.2$ $N_{\text{sig}} = 4.0 \pm 3.0 \pm 0.0$ $\sigma_{\text{stat}} = 1.6, \sigma = 1.6$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 0.5, N_{\text{scf}}/N_{\text{sig}} = 0.17$ $N_{\text{cand}} = 11, \text{shape} = -15.3 \pm 7.4$ $N_{\text{sig}} = 6.3 \pm 3.0 \pm 0.0$ $\sigma_{\text{stat}} = 2.9, \sigma = 2.9$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 1.4, N_{\text{scf}}/N_{\text{sig}} = 0.25$ $N_{\text{cand}} = 25, \text{shape} = -12.7 \pm 6.7$ $N_{\text{sig}} = 10.4 \pm 4.2 \pm 0.0$ $\sigma_{\text{stat}} = 3.1, \sigma = 3.1$
$q^2 = [1.0, 6.0]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 30.1, N_{\text{scf}}/N_{\text{sig}} = 0.37$ $N_{\text{cand}} = 63, \text{shape} = -14.3 \pm 3.3$ $N_{\text{sig}} = 7.0 \pm 6.0 \pm 0.2$ $\sigma_{\text{stat}} = 1.3, \sigma = 1.3$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 5.6, N_{\text{scf}}/N_{\text{sig}} = 0.22$ $N_{\text{cand}} = 27, \text{shape} = -13.1 \pm 3.5$ $N_{\text{sig}} = 4.5 \pm 4.0 \pm 0.0$ $\sigma_{\text{stat}} = 1.3, \sigma = 1.3$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 35.6, N_{\text{scf}}/N_{\text{sig}} = 0.30$ $N_{\text{cand}} = 90, \text{shape} = -14.8 \pm 3.1$ $N_{\text{sig}} = 10.9 \pm 7.2 \pm 0.2$ $\sigma_{\text{stat}} = 1.6, \sigma = 1.6$
$q^2 = [6.0, 14.4]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 65.8, N_{\text{scf}}/N_{\text{sig}} = 0.57$ $N_{\text{cand}} = 67, \text{shape} = -20.5 \pm 3.6$ $N_{\text{sig}} = 14.4 \pm 6.5 \pm 0.3$ $\sigma_{\text{stat}} = 2.5, \sigma = 2.5$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 33.5, N_{\text{scf}}/N_{\text{sig}} = 0.38$ $N_{\text{cand}} = 64, \text{shape} = -18.7 \pm 3.3$ $N_{\text{sig}} = 8.0 \pm 6.2 \pm 0.1$ $\sigma_{\text{stat}} = 1.4, \sigma = 1.4$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 99.2, N_{\text{scf}}/N_{\text{sig}} = 0.44$ $N_{\text{cand}} = 131, \text{shape} = -21.1 \pm 3.2$ $N_{\text{sig}} = 20.0 \pm 9.1 \pm 0.2$ $\sigma_{\text{stat}} = 2.4, \sigma = 2.4$
$q^2 = [14.4, 25.0]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 0.6, N_{\text{scf}}/N_{\text{sig}} = 0.77$ $N_{\text{cand}} = 17, \text{shape} = -31.5 \pm 5.4$ $N_{\text{sig}} = 2.4 \pm 3.2 \pm 0.0$ $\sigma_{\text{stat}} = 0.8, \sigma = 0.8$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 0.8, N_{\text{scf}}/N_{\text{sig}} = 0.49$ $N_{\text{cand}} = 20, \text{shape} = -28.3 \pm 5.4$ $N_{\text{sig}} = 7.4 \pm 3.7 \pm 0.0$ $\sigma_{\text{stat}} = 2.5, \sigma = 2.5$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 1.4, N_{\text{scf}}/N_{\text{sig}} = 0.61$ $N_{\text{cand}} = 37, \text{shape} = -30.1 \pm 5.0$ $N_{\text{sig}} = 9.8 \pm 5.0 \pm 0.0$ $\sigma_{\text{stat}} = 2.3, \sigma = 2.3$

Table C.5: Results of the simultaneous fit to the 140fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample and 605fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample. signal yield (N_{sig}), significance, total signal efficiency ϵ (including the fraction of X_s states considered in this analysis) and branching fraction (\mathcal{B}).

Mode	N_{sig}	Significance	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$B \rightarrow X_s e^+ e^-$	$28.3 \pm 10.0 \pm 0.5$	3.1	$2.32 \pm 0.01^{+0.27}_{-0.31}$	$4.01 \pm 1.41^{+0.48}_{-0.54}$
$B \rightarrow X_s \mu^+ \mu^-$	$28.1 \pm 8.8 \pm 0.1$	3.6	$2.67 \pm 0.01^{+0.31}_{-0.36}$	$3.46 \pm 1.09^{+0.41}_{-0.46}$
$B \rightarrow X_s \ell^+ \ell^-$	$54.4 \pm 13.5 \pm 0.4$	4.5	$2.50 \pm 0.01^{+0.29}_{-0.33}$	$3.58 \pm 0.89^{+0.42}_{-0.48}$

Table C.6: Bin-by-Bin results of the simultaneous fit to the 140fb^{-1} of $B \rightarrow X_s \ell^+ \ell^-$ sample and 605fb^{-1} of $B \rightarrow X_s e^\pm \mu^\mp$ sample. Signal yields (N_{sig}), signal efficiency (ϵ) and Branching fractions (\mathcal{B}) for each bin of M_{X_s} , $M_{\ell^+ \ell^-}$ and q^2 .

bin	N_{sig}	ϵ (%)	$\mathcal{B} (\times 10^{-6})$
$M_{X_s} (\text{GeV}/c^2)$			
= [0.4, 0.6]	$23.4 \pm 5.4 \pm 0.1$	$8.07 \pm 0.02^{+0.94}_{-1.07}$	$0.48 \pm 0.11^{+0.06}_{-0.06}$
= [0.6, 0.8]	$-2.0 \pm 1.2 \pm 2.0$	$3.69 \pm 0.09^{+0.43}_{-0.49}$	$-0.09 \pm 0.05^{+0.09}_{-0.09}$
= [0.8, 1.0]	$19.7 \pm 5.6 \pm 0.2$	$3.73 \pm 0.02^{+0.44}_{-0.50}$	$0.87 \pm 0.25^{+0.10}_{-0.12}$
= [1.0, 1.4]	$0.2 \pm 4.8 \pm 4.6$	$1.61 \pm 0.02^{+0.19}_{-0.21}$	$0.02 \pm 0.49^{+0.47}_{-0.47}$
= [1.4, 2.0]	$14.5 \pm 9.7 \pm 0.6$	$0.94 \pm 0.01^{+0.11}_{-0.13}$	$2.53 \pm 1.69^{+0.32}_{-0.36}$
$q^2 ((\text{GeV}/c^2)^2)$			
= [0.04, 1.0]	$10.4 \pm 4.2 \pm 0.0$	$1.91 \pm 0.02^{+0.22}_{-0.25}$	$0.90 \pm 0.36^{+0.11}_{-0.12}$
= [1.0, 6.0]	$10.9 \pm 7.2 \pm 0.2$	$2.67 \pm 0.02^{+0.31}_{-0.36}$	$0.67 \pm 0.45^{+0.08}_{-0.09}$
= [6.0, 14.4]	$20.0 \pm 9.1 \pm 0.2$	$1.77 \pm 0.01^{+0.21}_{-0.24}$	$1.85 \pm 0.84^{+0.22}_{-0.25}$
= [14.4, 25.0]	$9.8 \pm 5.0 \pm 0.0$	$6.80 \pm 0.04^{+0.80}_{-0.91}$	$0.24 \pm 0.12^{+0.03}_{-0.03}$

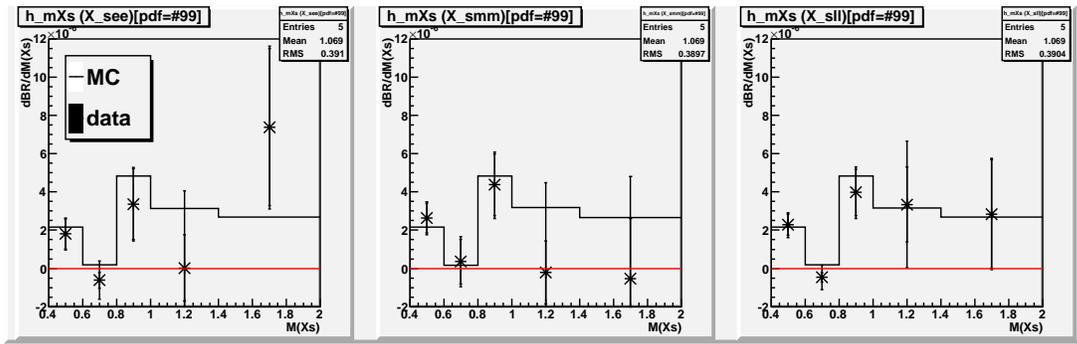


Figure C.3: Differential branching fraction in each M_{X_s} bin, by the simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 140 fb^{-1} results and the histogram represents the signal MC simulation.

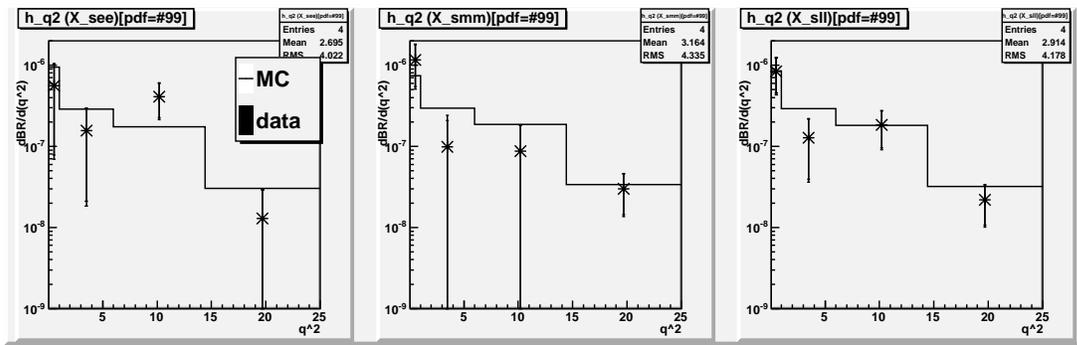


Figure C.4: Differential branching fraction in each q^2 bin, by the simultaneous fit, for e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right. Points represent the 140 fb^{-1} results and the histogram represents the signal MC simulation.

C.3 140 fb^{-1} results by non-simultaneous fit

In this section, we also show the 140 fb^{-1} results, but without using the simultaneous fit method. We fit the real $B \rightarrow X_s \ell^+ \ell^-$ sample only, without using the real $B \rightarrow X_s e^\pm \mu^\mp$ sample. The ARGUS shape parameter is floated in the fit.

With this simpler fitting method, we obtain smaller ARGUS shape parameters (-24.6) than the simultaneous fit (-17.5). With smaller shape parameters, ARGUS distribution has less entries in M_{bc} sideband and more entries in the M_{bc} signal region. Therefore we obtain smaller yields (44.9 ± 14.6) than the simultaneous fitting (48.4 ± 13.2) and the measured branching fractions are also smaller.

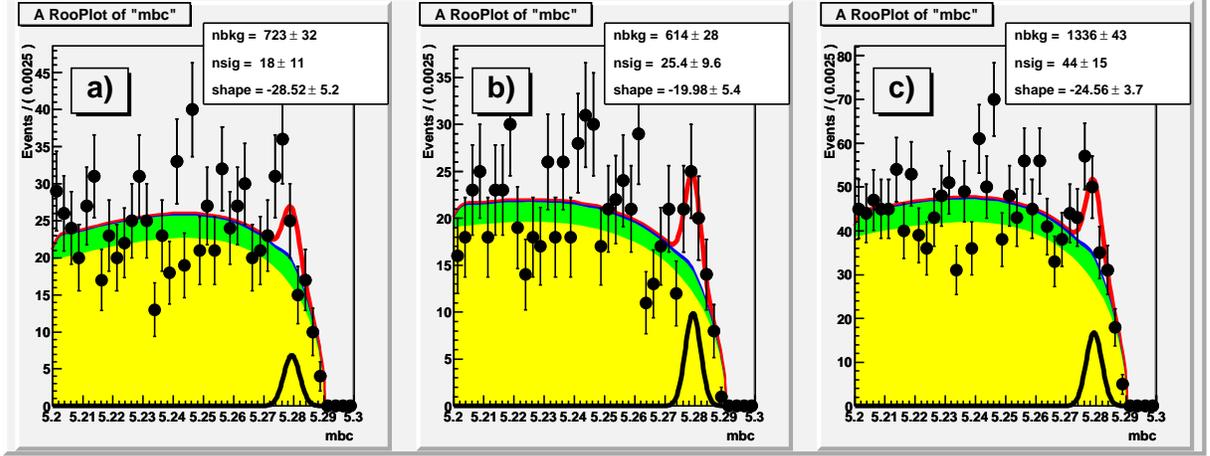


Figure C.5: M_{bc} fit with 140 fb^{-1} data, for a) $B \rightarrow X_s e^+ e^-$, b) $B \rightarrow X_s \mu^+ \mu^-$ and c) $B \rightarrow X_s \ell^+ \ell^-$. Background, peaking background, self cross feed components are shown in yellow, green, blue area, respectively. Signal component is shown in black line.

Table C.7: Results of the fit to the 140 fb^{-1} data : number of signal candidates in the M_{bc} signal region, fitted signal yield N_{sig} , and significance.

Mode	Candidates	ARGUS shape	N_{sig}	Significance
$B \rightarrow X_s e^+ e^-$	161	-28.5 ± 5.2	$18.4 \pm 11.0 \pm 0.6$	1.8
$B \rightarrow X_s \mu^+ \mu^-$	122	-20.2 ± 5.2	$26.3 \pm 9.6 \pm 0.1$	3.0
$B \rightarrow X_s \ell^+ \ell^-$	283	-24.6 ± 3.6	$44.9 \pm 14.6 \pm 0.5$	3.3

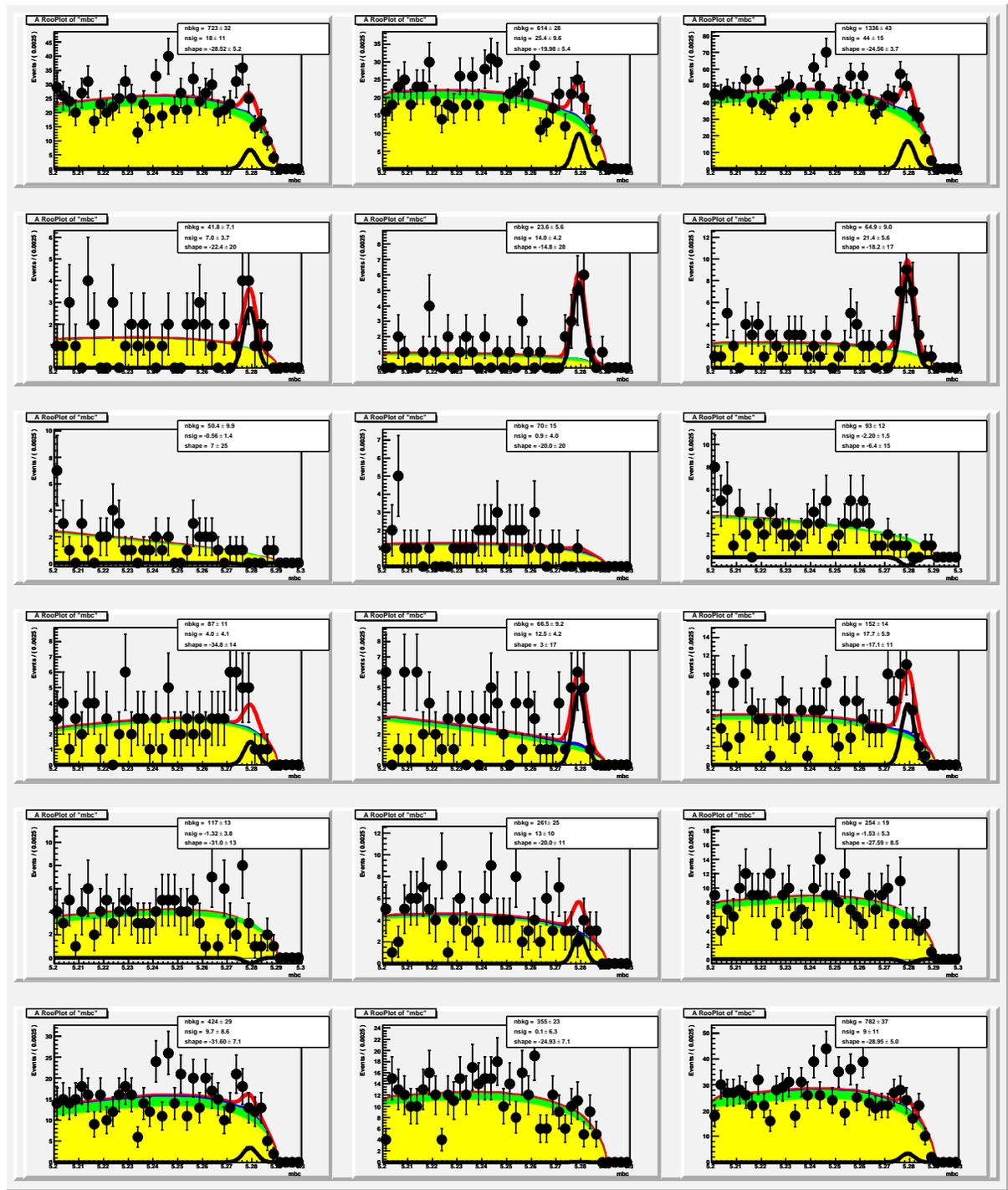


Figure C.6: M_{bc} fit with 140 fb^{-1} data for each M_{X_s} bin. Each row corresponds to the full dataset, $0.4 \text{ GeV}/c^2 < M_{X_s} < 0.6 \text{ GeV}/c^2$, $0.6 \text{ GeV}/c^2 < M_{X_s} < 0.8 \text{ GeV}/c^2$, $0.8 \text{ GeV}/c^2 < M_{X_s} < 1.0 \text{ GeV}/c^2$, $1.0 \text{ GeV}/c^2 < M_{X_s} < 1.4 \text{ GeV}/c^2$ and $1.4 \text{ GeV}/c^2 < M_{X_s} < 2.0 \text{ GeV}/c^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right.

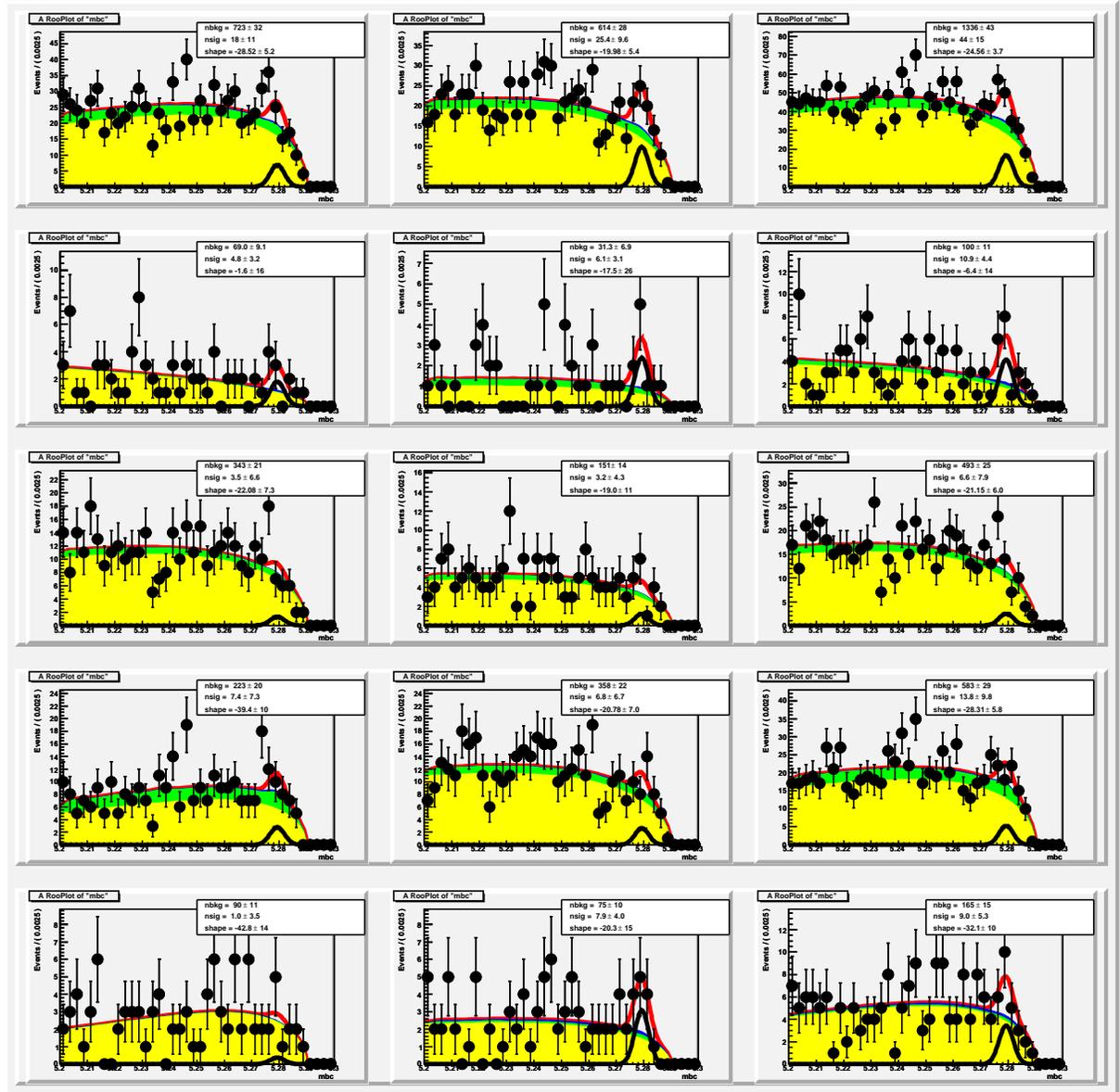


Figure C.7: M_{bc} fit with 140 fb^{-1} data for each q^2 bin. Each row corresponds to the full dataset, $0.04 (\text{GeV}/c^2)^2 < q^2 < 1.0 (\text{GeV}/c^2)^2$, $1.0 (\text{GeV}/c^2)^2 < q^2 < 6.0 (\text{GeV}/c^2)^2$, $6.0 (\text{GeV}/c^2)^2 < q^2 < 14.4 (\text{GeV}/c^2)^2$ and $14.4 (\text{GeV}/c^2)^2 < q^2 < 25.0 (\text{GeV}/c^2)^2$, from top to bottom. Each column corresponds to e^+e^- , $\mu^+\mu^-$ and $\ell^+\ell^-$, from left to right.

Table C.8: Summary of the fitter input parameters and the fitting results, with 140 fb^{-1} data. Gaussian shape parameters (mean and width, in MeV/c^2) is fixed. N_{pkg} and $N_{\text{scf}}/N_{\text{sig}}$ are also fixed. N_{cand} is the number of candidates in M_{bc} signal region. Fitting results, ARGUS shape parameter, N_{sig} , are also shown.

	$B \rightarrow X_s e^+ e^-$	$B \rightarrow X_s \mu^+ \mu^-$	$B \rightarrow X_s \ell^+ \ell^-$
full sample	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 97.3, N_{\text{scf}}/N_{\text{sig}} = 0.48$ $N_{\text{cand}} = 161, \text{shape} = -28.5 \pm 5.2$ $N_{\text{sig}} = 18.4 \pm 11.0 \pm 0.6$ $\sigma_{\text{stat}} = 1.8, \sigma = 1.8$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 40.3, N_{\text{scf}}/N_{\text{sig}} = 0.34$ $N_{\text{cand}} = 122, \text{shape} = -20.2 \pm 5.2$ $N_{\text{sig}} = 26.3 \pm 9.6 \pm 0.1$ $\sigma_{\text{stat}} = 3.0, \sigma = 3.0$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 137.6, N_{\text{scf}}/N_{\text{sig}} = 0.40$ $N_{\text{cand}} = 283, \text{shape} = -24.6 \pm 3.6$ $N_{\text{sig}} = 44.9 \pm 14.6 \pm 0.5$ $\sigma_{\text{stat}} = 3.3, \sigma = 3.3$
$M_{X_s} = [0.4, 0.6]$	mean = 5279.18, width = 2.58 $N_{\text{pkg}} = 0.6, N_{\text{scf}}/N_{\text{sig}} = 0.04$ $N_{\text{cand}} = 13, \text{shape} = -22.4 \pm 20.2$ $N_{\text{sig}} = 7.1 \pm 3.7 \pm 0.1$ $\sigma_{\text{stat}} = 2.3, \sigma = 2.3$	mean = 5279.24, width = 2.54 $N_{\text{pkg}} = 0.8, N_{\text{scf}}/N_{\text{sig}} = 0.02$ $N_{\text{cand}} = 18, \text{shape} = -14.9 \pm 25.2$ $N_{\text{sig}} = 14.2 \pm 4.2 \pm 0.1$ $\sigma_{\text{stat}} = 4.5, \sigma = 4.5$	mean = 5279.21, width = 2.56 $N_{\text{pkg}} = 1.4, N_{\text{scf}}/N_{\text{sig}} = 0.03$ $N_{\text{cand}} = 31, \text{shape} = -18.3 \pm 15.9$ $N_{\text{sig}} = 21.7 \pm 5.6 \pm 0.1$ $\sigma_{\text{stat}} = 4.9, \sigma = 4.9$
$M_{X_s} = [0.6, 0.8]$	mean = 5278.77, width = 2.44 $N_{\text{pkg}} = 3.2, N_{\text{scf}}/N_{\text{sig}} = 3.61$ $N_{\text{cand}} = 5, \text{shape} = 6.8 \pm 25.3$ $N_{\text{sig}} = -0.6 \pm 1.4 \pm 1.0$ $\sigma_{\text{stat}} = \text{nan}, \sigma = \text{nan}$	mean = 5279.58, width = 2.96 $N_{\text{pkg}} = 2.0, N_{\text{scf}}/N_{\text{sig}} = 2.67$ $N_{\text{cand}} = 2, \text{shape} = -20.0 \pm 19.5$ $N_{\text{sig}} = 0.9 \pm 3.9 \pm 1.5$ $\sigma_{\text{stat}} = 2.0, \sigma = \text{nan}$	mean = 5279.20, width = 2.81 $N_{\text{pkg}} = 5.1, N_{\text{scf}}/N_{\text{sig}} = 3.10$ $N_{\text{cand}} = 7, \text{shape} = -6.9 \pm 14.7$ $N_{\text{sig}} = -2.1 \pm 1.5 \pm 2.1$ $\sigma_{\text{stat}} = \text{nan}, \sigma = \text{nan}$
$M_{X_s} = [0.8, 1.0]$	mean = 5279.50, width = 2.71 $N_{\text{pkg}} = 5.9, N_{\text{scf}}/N_{\text{sig}} = 0.29$ $N_{\text{cand}} = 25, \text{shape} = -34.8 \pm 14.2$ $N_{\text{sig}} = 4.0 \pm 4.1 \pm 0.2$ $\sigma_{\text{stat}} = 1.0, \sigma = 1.0$	mean = 5279.36, width = 2.61 $N_{\text{pkg}} = 4.3, N_{\text{scf}}/N_{\text{sig}} = 0.21$ $N_{\text{cand}} = 22, \text{shape} = 1.2 \pm 15.9$ $N_{\text{sig}} = 12.6 \pm 4.2 \pm 0.1$ $\sigma_{\text{stat}} = 3.9, \sigma = 3.9$	mean = 5279.43, width = 2.66 $N_{\text{pkg}} = 10.2, N_{\text{scf}}/N_{\text{sig}} = 0.24$ $N_{\text{cand}} = 47, \text{shape} = -17.4 \pm 10.6$ $N_{\text{sig}} = 17.8 \pm 5.9 \pm 0.2$ $\sigma_{\text{stat}} = 3.5, \sigma = 3.5$
$M_{X_s} = [1.0, 1.4]$	mean = 5279.49, width = 2.70 $N_{\text{pkg}} = 15.0, N_{\text{scf}}/N_{\text{sig}} = 0.68$ $N_{\text{cand}} = 21, \text{shape} = -31.0 \pm 12.7$ $N_{\text{sig}} = -1.3 \pm 3.8 \pm 10.2$ $\sigma_{\text{stat}} = \text{nan}, \sigma = \text{nan}$	mean = 5279.44, width = 2.52 $N_{\text{pkg}} = 7.4, N_{\text{scf}}/N_{\text{sig}} = 0.54$ $N_{\text{cand}} = 25, \text{shape} = -20.0 \pm 10.6$ $N_{\text{sig}} = 13.1 \pm 10.0 \pm 9.4$ $\sigma_{\text{stat}} = 3.8, \sigma = \text{nan}$	mean = 5279.46, width = 2.61 $N_{\text{pkg}} = 22.3, N_{\text{scf}}/N_{\text{sig}} = 0.60$ $N_{\text{cand}} = 46, \text{shape} = -27.6 \pm 8.3$ $N_{\text{sig}} = -1.3 \pm 5.3 \pm 10.2$ $\sigma_{\text{stat}} = \text{nan}, \sigma = \text{nan}$
$M_{X_s} = [1.4, 2.0]$	mean = 5279.18, width = 2.82 $N_{\text{pkg}} = 72.7, N_{\text{scf}}/N_{\text{sig}} = 1.25$ $N_{\text{cand}} = 97, \text{shape} = -31.6 \pm 7.1$ $N_{\text{sig}} = 9.8 \pm 8.6 \pm 0.8$ $\sigma_{\text{stat}} = 1.2, \sigma = 1.2$	mean = 5279.41, width = 2.60 $N_{\text{pkg}} = 25.8, N_{\text{scf}}/N_{\text{sig}} = 0.95$ $N_{\text{cand}} = 55, \text{shape} = -24.8 \pm 6.8$ $N_{\text{sig}} = 0.3 \pm 6.3 \pm 1.2$ $\sigma_{\text{stat}} = 0.1, \sigma = \text{nan}$	mean = 5279.31, width = 2.70 $N_{\text{pkg}} = 98.5, N_{\text{scf}}/N_{\text{sig}} = 1.10$ $N_{\text{cand}} = 152, \text{shape} = -28.8 \pm 4.9$ $N_{\text{sig}} = 9.3 \pm 10.7 \pm 0.6$ $\sigma_{\text{stat}} = 0.9, \sigma = 0.9$
$q^2 = [0.04, 1.0]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 0.9, N_{\text{scf}}/N_{\text{sig}} = 0.30$ $N_{\text{cand}} = 14, \text{shape} = -1.9 \pm 16.1$ $N_{\text{sig}} = 4.8 \pm 3.2 \pm 0.0$ $\sigma_{\text{stat}} = 1.7, \sigma = 1.7$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 0.5, N_{\text{scf}}/N_{\text{sig}} = 0.17$ $N_{\text{cand}} = 11, \text{shape} = -18.9 \pm 19.8$ $N_{\text{sig}} = 6.1 \pm 3.1 \pm 0.3$ $\sigma_{\text{stat}} = 2.4, \sigma = 2.4$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 1.4, N_{\text{scf}}/N_{\text{sig}} = 0.25$ $N_{\text{cand}} = 25, \text{shape} = -8.0 \pm 12.5$ $N_{\text{sig}} = 11.0 \pm 4.4 \pm 0.0$ $\sigma_{\text{stat}} = 2.9, \sigma = 2.9$
$q^2 = [1.0, 6.0]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 30.1, N_{\text{scf}}/N_{\text{sig}} = 0.37$ $N_{\text{cand}} = 63, \text{shape} = -22.1 \pm 7.3$ $N_{\text{sig}} = 3.5 \pm 6.6 \pm 0.2$ $\sigma_{\text{stat}} = 0.5, \sigma = 0.5$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 5.6, N_{\text{scf}}/N_{\text{sig}} = 0.22$ $N_{\text{cand}} = 27, \text{shape} = -18.9 \pm 10.1$ $N_{\text{sig}} = 3.4 \pm 4.3 \pm 0.0$ $\sigma_{\text{stat}} = 0.8, \sigma = 0.8$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 35.6, N_{\text{scf}}/N_{\text{sig}} = 0.30$ $N_{\text{cand}} = 90, \text{shape} = -21.2 \pm 5.9$ $N_{\text{sig}} = 6.9 \pm 7.9 \pm 0.2$ $\sigma_{\text{stat}} = 0.9, \sigma = 0.9$
$q^2 = [6.0, 14.4]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 65.8, N_{\text{scf}}/N_{\text{sig}} = 0.57$ $N_{\text{cand}} = 67, \text{shape} = -39.4 \pm 10.0$ $N_{\text{sig}} = 7.5 \pm 7.3 \pm 0.3$ $\sigma_{\text{stat}} = 1.1, \sigma = 1.1$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 33.5, N_{\text{scf}}/N_{\text{sig}} = 0.38$ $N_{\text{cand}} = 64, \text{shape} = -20.9 \pm 6.9$ $N_{\text{sig}} = 7.0 \pm 6.7 \pm 0.1$ $\sigma_{\text{stat}} = 1.1, \sigma = 1.1$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 99.2, N_{\text{scf}}/N_{\text{sig}} = 0.44$ $N_{\text{cand}} = 131, \text{shape} = -28.3 \pm 5.7$ $N_{\text{sig}} = 14.1 \pm 9.8 \pm 0.2$ $\sigma_{\text{stat}} = 1.5, \sigma = 1.5$
$q^2 = [14.4, 25.0]$	mean = 5279.29, width = 2.67 $N_{\text{pkg}} = 0.6, N_{\text{scf}}/N_{\text{sig}} = 0.77$ $N_{\text{cand}} = 17, \text{shape} = -42.8 \pm 13.5$ $N_{\text{sig}} = 1.0 \pm 3.5 \pm 1.0$ $\sigma_{\text{stat}} = 0.3, \sigma = 0.5$	mean = 5279.33, width = 2.57 $N_{\text{pkg}} = 0.8, N_{\text{scf}}/N_{\text{sig}} = 0.49$ $N_{\text{cand}} = 20, \text{shape} = -21.0 \pm 14.7$ $N_{\text{sig}} = 8.2 \pm 4.0 \pm 0.0$ $\sigma_{\text{stat}} = 2.4, \sigma = 2.4$	mean = 5279.31, width = 2.62 $N_{\text{pkg}} = 1.4, N_{\text{scf}}/N_{\text{sig}} = 0.61$ $N_{\text{cand}} = 37, \text{shape} = -32.2 \pm 10.0$ $N_{\text{sig}} = 9.3 \pm 5.3 \pm 0.0$ $\sigma_{\text{stat}} = 1.9, \sigma = 1.9$

Bibliography

- [1] P. Gambino, U. Haisch and M. Misiak, Phys. Rev. Lett. **94**, 061803 (2005) [arXiv:hep-ph/0410155].
- [2] A. Ali, E. Lunghi, C. Greub, and G. Hiller, Phys. Rev. **D66**, 034002 (2002).
- [3] T. Hurth, hep-ph/0212304, SLAC-PUB-9604 (2003).
- [4] A. Ali, hep-ph/0210183, CERN-TH/2002-284 (2002).
- [5] E. Lunghi, A. Masiero, I. Scimemi and L. Silvestrini, Nucl. Phys. B **568**, 120 (2000) [arXiv:hep-ph/9906286].
- [6] G. Hiller and F. Kruger, Phys. Rev. D **69**, 074020 (2004) [arXiv:hep-ph/0310219].
- [7] Z. j. Xiao and L. x. Lu, Phys. Rev. D **74**, 034016 (2006)
- [8] T. Huber and S. Schilling, arXiv:hep-ph/0512067.
- [9] K. S. M. Lee, Z. Ligeti, I. W. Stewart and F. J. Tackmann, Phys. Rev. D **75**, 034016 (2007)
- [10] J. R. Ellis, M. K. Gaillard, D. V. Nanopoulos and S. Rudaz, Nucl. Phys. B **131**, 285 (1977) [Erratum-ibid. B **132**, 541 (1978)].
- [11] K. Abe *et al.* [BELLE Collaboration], Phys. Rev. Lett. **88**, 021801 (2002) [arXiv:hep-ex/0109026].
- [12] A. Ishikawa *et al.* [Belle Collaboration], Phys. Rev. Lett. **91**, 261601 (2003) [arXiv:hep-ex/0308044].
- [13] A. Ishikawa *et al.*, Phys. Rev. Lett. **96**, 251801 (2006) [arXiv:hep-ex/0603018].
- [14] I. Adachi *et al.* [Belle Collaboration], BELLE-CONF-0822 arXiv:0810.0335 [hep-ex].
- [15] J. T. Wei *et al.* [BELLE Collaboration], Phys. Rev. Lett. **103**, 171801 (2009) [arXiv:0904.0770 [hep-ex]].
- [16] B. Aubert *et al.* [BABAR Collaboration], SLAC-PUB-9323 (2002) [arXiv:hep-ex/0207082],

- [17] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. Lett. **91**, 221802 (2003) [arXiv:hep-ex/0308042].
- [18] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. D **73**, 092001 (2006) [arXiv:hep-ex/0604007].
- [19] J. Kaneko *et al.* [Belle Collaboration], Phys. Rev. Lett. **90**, 021801 (2003) [arXiv:hep-ex/0208029].
- [20] M. Iwasaki *et al.* [Belle Collaboration], Phys. Rev. D **72**, 092005 (2005) [arXiv:hep-ex/0503044].
- [21] B. Aubert *et al.* [BABAR Collaboration], Phys. Rev. Lett. **93**, 081802 (2004) [arXiv:hep-ex/0404006].
- [22] The semi-inclusive technique was introduced in the context of inclusive $B \rightarrow X_s \gamma$ decays; see M. S. Alam *et al.* [CLEO Collaboration], Phys. Rev. Lett. **74**, 2885 (1995).
- [23] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
For the reference of the CKM matrix itself, one also needs to cite N. Cabibbo, Phys. Rev. Lett. **8**, 214 (1964).
- [24] S. Kurokawa and E. Kikutani, Nucl. Instrum. Meth. A **499**, 1 (2003), and other papers included in this volume.
- [25] A. Abashian *et al.* [Belle Collaboration], Nucl. Instrum. Meth. A **479**, 117 (2002).
- [26] G. Alimonti *et al.* [BELLE Collaboration], Nucl. Instrum. Meth. A **453**, 71 (2000).
- [27] Z. Natkaniec *et al.*, “Status of the Belle silicon vertex detector,” Nucl. Instrum. Meth. A **560**, 1 (2006).
- [28] H. Hirano *et al.*, “A high resolution cylindrical drift chamber for the KEK B-factory,” Nucl. Instrum. Meth. A **455**, 294 (2000).
- [29] M. Akatsu *et al.*, “Cathode Image Readout In The Belle Central Drift Chamber,” Nucl. Instrum. Meth. A **454**, 322 (2000).
- [30] T. Iijima *et al.*, “Aerogel Cherenkov counter for the BELLE detector,” Nucl. Instrum. Meth. A **453**, 321 (2000).
- [31] T. Iijima *et al.*, “Study on fine-mesh PMTs for detection of aerogel Cherenkov light,” Nucl. Instrum. Meth. A **387**, 64 (1997).
- [32] H. Kichimi *et al.*, “The BELLE TOF system,” Nucl. Instrum. Meth. A **453** (2000) 315.
- [33] H. Ikeda *et al.*, “A detailed test of the CsI(Tl) calorimeter for BELLE with photon beams of energy between 20-MeV and 5.4-GeV,” Nucl. Instrum. Meth. A **441**, 401 (2000).

- [34] A. Abashian *et al.* [BELLE Collaboration], “The K(L) / Mu Detector Subsystem For The Belle Experiment At The Kek B Factory,” Nucl. Instrum. Meth. A **449**, 112 (2000).
- [35] R. Akhmetshin *et al.*, “Survey of the properties of BGO crystals for the extreme forward calorimeter at BELLE,” Nucl. Instrum. Meth. A **455**, 324 (2000).
- [36] Y. Ushiroda, A. Mohapatra, H. Sakamoto, Y. Sakai, M. Nakao, Q. An and Y. F. Wang, “Development of the central trigger system for the BELLE detector at the KEK B-factory,” Nucl. Instrum. Meth. A **438**, 460 (1999).
- [37] T. Ziegler *et al.*, “The Belle trigger system with the new silicon vertex detector SVD2,” IEEE Trans. Nucl. Sci. **51** (2004) 1852.
- [38] T. Higuchi *et al.*, “Modular Pipeline Readout Electronics For The Superbelle Drift Chamber,” IEEE Trans. Nucl. Sci. **52**, 1912 (2005).
- [39] T. Higuchi, R. Itoh, M. Nakao, S. Y. Suzuki and H. N. Nakayama, “Commissioning a Pipelined Data Acquisition System for the Belle Central Drift Chamber,” Presented at 15th IEEE Real Time Conference 2007 (RT 07), Batavia, Illinois, 29 Apr - 4 May 2007.
- [40] D. J. Lange, “The Evtgen Particle Decay Simulation Package,” Nucl. Instrum. and Meth. **A462**, 152 (2001).
- [41] T. Sjöstrand, Computer Physics Commun. **82**, 74 (1994).
- [42] R. Brun *et al.*, GEANT 3 Manual, CERN Program Library Long Writeup W5013, 1994.
- [43] A. Ali, P. Ball, L.T. Handoko, and G. Hiller, Phys. Rev. **D61**, 074024 (2000).
- [44] F. Krüger and L.M. Sehgal, Phys. Lett. **B380**, 199 (1996).
- [45] A. Ali and E. Pietarinen, Nucl. Phys. **B154**, 519 (1979); G. Altarelli *et al.*, Nucl. Phys. **B208**, 365 (1982).
- [46] R. A. Fisher, Annals Eugen. **7**, 179 (1936).
- [47] G.C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978).
- [48] H. Albrecht *et al.* [ARGUS Collaboration], Phys. Lett. **B192**, 245 (1987).
- [49] D. Cronin-Hennessy *et al.* [CLEO Collaboration], Phys. Rev. Lett. **87**, 251808 (2001).
- [50] S. Chen *et al.* [CLEO Collaboration], Phys. Rev. Lett. **87**, 251807 (2001).
- [51] “Flavor Tagging”: H. Kakuno *et al.*, Nucl. Instr. and Meth.A **533** 516 (2004).

- [52] “SFW”: The Fox-Wolfram moments were introduced in G. C. Fox and S. Wolfram, Phys. Rev. Lett. **41**, 1581 (1978). The Fisher discriminant used by Belle, based on modified Fox-Wolfram moments (SFW), is described in K. Abe *et al.* (Belle Collab.), Phys. Rev. Lett. **87**, 101801 (2001) and K. Abe *et al.* (Belle Collab.), Phys. Lett. B **511**, 151 (2001).
- [53] “KSFW”: S. H. Lee *et al.* (Belle Collab.), Phys. Rev. Lett. **91**, 261801 (2003).
- [54] “ $b \rightarrow s$ penguin TCPV”: K-F Chen *et al.* (Belle Collab.), Phys. Rev. Lett. **98** 031802 (2007).
- [55] “Vertexing & Resolution function”: H. Tajima *et al.*, Nucl. Instr. and Meth.A **533** 370 (2004).
- [56] “Feldman-Cousins”: G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998).
- [57] “POLE program”: The Highland-Cousins method for calculating upper limits with systematic errors is described in J. Conrad *et al.*, Phys. Rev. D **67**, 012002 (2003).
- [58] Y.-M. Yao *et al.* (Particle Data Group), J. Phys. G **33**, 1 (2006).
- [59] Heavy Flavor Averaging Group, <http://www.slac.stanford.edu/xorg/hfag>.
- [60] P.Koppenburg, Belle Note 621.
- [61] L.Hinz, Belle Note 954.
- [62] S.Nishida, Belle Note 779.
- [63] M.Nakao, Belle Note 513.
- [64] S.W.Lin, Belle Note 645.
- [65] M.Nakao, shown at 1st Open Meeting of the SuperKEKB Collaboration.
- [66] J. L. . Hewett *et al.*, arXiv:hep-ph/0503261.